Dissertation zur Erlangung des Doktorgrades an der
Fakultät Erziehungswissenschaft,
Psychologie und Bewegungswissenschaft

Fachbereich Psychologie
Universität Hamburg

Persönlichkeit als Prädiktor für Leistung in hoch automatisierten Mensch-Maschine-Teams der Luftfahrt

Vorgelegt von Dipl.-Psych. Solveig Caroline Sophie Eschen

Hamburg, 2013

Erstgutachter: Prof. Dr. Matthias Burisch
Zweitgutachterin: Prof. Dr. Monique Janneck

Tag der Disputation: 16.12.2013

Kurzfassung

Durch zunehmende Automatisierung im Luftfahrtbereich sind veränderte Anforderungen an Luftfahrtoperateure (Fluglotsen und Piloten) zu erwarten. So ist davon auszugehen, dass vermehrt enge Zusammenarbeit von Menschen mit einem hoch automatisierten System erforderlich sein wird. Dies kann neben veränderten Fähigkeitsanforderungen auch Anforderungsänderungen im Bereich der Persönlichkeit verursachen. Die vorliegende Arbeit beleuchtet den Einfluss von Aspekten der Persönlichkeit und Einstellung auf die Leistungsfähigkeit hinsichtlich heutiger sowie zukünftiger Anforderungen an Luftfahrtoperateure. Das Ergebnis ist eine Zusammenstellung von Skalen, die zukünftig relevante Persönlichkeits- und Einstellungsmerkmale erfassen. Einer Erprobung des konzipierten Forschungsfragebogens HTQ (Hybrid Team Questionnaire) folgen Studien zu Zusammenhängen mit Anforderungen an Luftfahrtoperateure. Es zeigen sich sowohl Effekte von Facetten der allgemeinen Persönlichkeit als auch von technikbezogenen Persönlichkeits- und Einstellungsskalen sowie Computernutzungsvariablen. Für drei Prädiktoren kann zusätzlich inkrementelle Validität nachgewiesen werden.

Schlagwörter: Mensch-Maschine-Interaktion, Mensch-Maschine-Team, Automation, Persönlichkeit, Leistung, zukünftige Anforderungen, Luftfahrt, Eignungsdiagnostik

Abstract

Generated by rising automation in aviation changing demands on aviation operators (air traffic controllers and pilots) are to be expected. Thus it can be assumed that close cooperation of people with a highly automated system will be increasingly required. In addition to changing skill requirements this can also cause changes on requirements in the area of personality. The present work examines the influence of aspects of personality and attitude on the performance in terms of today's and future requirements for aviation operators. The result is a set of scales that capture the relevant future personality and attitude features. An investigation of the developed research questionnaire HTQ (Hybrid Team Questionnaire) is followed by studies on correlations of the scales with current and future requirements for aviation operators. Shown are effects of the facets of general personality as well as technology-related personality and attitude scales plus computer usage variables. In addition, three predictors show incremental validity.

Keywords: Human-Machine-Interaction, Human-Machine-Team, Automation, Personality, Performance, Future Requirements, Aviation, Personnel Assessment

Inhaltsverzeichnis

Kurzfassung **3**

Abstract **3**

Inhaltsverzeichnis **5**

Vorwort **9**

1 Einleitung **11**

A. Theoretischer Teil **13**

2 Zentrale Oberbegriffe **14**

2.1 Persönlichkeit 14

2.2 Leistung 16

2.3 Mensch-Maschine-Interaktion 17

2.4 Automation 20

3 Persönlichkeit als Einflussfaktor **23**

3.1 Persönlichkeit und Leistung 23

3.2 Persönlichkeit in der Mensch-Maschine-Interaktion 25

3.3 Persönlichkeit und Automation 29

4 Forschungskontext: Eignungsdiagnostik für Fluglotsen und Piloten **31**

4.1 Eignungsdiagnostik heute 31

4.1.1 Auswahl von Nachwuchsfluglotsen im DLR 31

4.1.2 Auswahl von Nachwuchspiloten im DLR 34

4.2 Eignungsdiagnostik morgen 36

4.3 Das Forschungsprojekt HYBRID 39

B. Empirischer Teil **41**

5 Fragestellung **42**

6 Hypothesen **43**

7 Methode **44**

7.1 Entwicklung des Hybrid-Team-Questionnaire 45

7.1.1 Auswahl geeigneter Konstrukte und Skalen 46

7.1.2 Erprobung des HTQ-1 46

7.1.3 Erprobung des HTQ-2 48
7.1.4 Variablen aus der Eignungsdiagnostik 49
7.1.5 Korrelationsanalysen zum HTQ-2 50
7.1.6 Vom HTQ-2 zum HTQ-S 51
7.2 Entwicklung der ATM-Simulation HINT 52
7.2.1 Beschreibung des Verfahrens 52
7.2.2 Variablen aus der ATM-Simulation HINT 54
7.2.3 Korrelationsanalysen zu den HINT-Variablen 56
7.3 Kombination von HTQ-S und HINT 56
7.3.1 Stichprobe 57
7.3.2 Korrelationsanalysen zum HTQ-S 58
7.3.3 Gruppenvergleiche 58
7.3.4 Inkrementelle Validität 58
8 Ergebnisse 60
8.1 Der HTQ 60
8.1.1 Die Skalen des HTQ 60
8.1.2 Technische Umsetzbarkeit der HTQ-Skalen 81
8.1.3 Deskriptive Statistiken der HTQ-Skalen 81
8.1.4 Psychometrische Kennwerte der HTQ-Skalen 85
8.1.5 Endgültige Skalen des HTQ-S 89
8.2 Deskriptive Statistiken der HINT-Variablen 89
8.3 Korrelationsergebnisse 91
8.3.1 Korrelationen mit aktueller Eignungsdiagnostik 91
8.3.2 Korrelationen mit der der ATM-Simulation HINT 101
8.4 Ergebnisse der Gruppenvergleiche 107
8.5 Inkrementelle Validität des HTQ-S 114
8.6 Zusammenfassung der Ergebnisse 124

9 Diskussion 131
9.1 Zusammenhänge zu aktueller Eignungsdiagnostik – Diskussion zu Hypothese 1 ..131
9.2 Zusammenhänge mit zukünftigen Anforderungen – Diskussion zu Hypothese 2 ...135
9.3 Gruppenunterschiede – Diskussion zu Hypothese 3 139
9.4 Inkrementelle Validität – Diskussion zu Hypothese 4 144
9.5 Fazit und Ausblick 146

Literatur 153

Abbildungsverzeichnis 165

Tabellenverzeichnis 167

Abkürzungsverzeichnis 171

Anhang 173
Anhang A: Histogramme zum HTQ-S 174
Anhang B: Histogramme der HINT-Variablen 179

Vorwort

Im Zusammenhang mit der Fertigstellung dieser Arbeit möchte ich die Gelegenheit nutzen, einigen Wegbegleiterinnen und Wegbegleitern[1] meinen Dank auszusprechen.

Bedanken möchte ich mich zunächst bei den Mitarbeitern der Abteilung für Luft- und Raumfahrtpsychologie des DLR, insbesondere bei Dr. Hinnerk Eißfeldt und Dr. Yvonne Pecena, die mir stets mit wertvoller Unterstützung und fachlichen Ratschlägen zur Seite stehen. Namentlich möchte ich mich ebenfalls bei Dr. Peter Maschke sowie den Mitarbeiterinnen im Hybrid-Projekt Dr. Doris Keye, Dipl.-Psych. Katja Knappe, Hanna Willborn, Katharina Pommerening, Stefanie Nowitzke, Constanze Bohlmann und Anne Meusel bedanken. Aber auch allen anderen Mitarbeitern danke ich herzlich für die stets vorhandene Hilfsbereitschaft und die angenehme Zusammenarbeit!

Der Deutschen Flugsicherung GmbH und der Deutschen Lufthansa gilt mein Dank für ihre Kooperationsbereitschaft. Dabei insbesondere Herrn Dipl.-Psych. Alexander Heintz und Herrn Cpt. Joachim Graefe zu Baringdorf, die diese Zusammenarbeit ermöglichen.

Ganz besonders danke ich meinem Doktorvater Prof. Dr. Matthias Burisch, der immer ein offenes Ohr für mich hat und dessen fachliche und persönliche Unterstützung mir sehr wertvoll sind!

Meiner Familie danke ich für ihre verlässliche Unterstützung bei allem was ich tue. Allen voran meiner Mutter für ihren unermüdlichen Einsatz in der Kinderbetreuung und dafür, dass sie immer für mich da ist. Meiner Tochter danke ich dafür, dass sie mir immer wieder zeigt, was im Leben wirklich wichtig ist und meinem Freund für seinen unerschütterlichen Glauben an mich.

> "The more one does and sees and feels, the more one is able to do, and the more genuine may be one's appreciation of fundamental things like home, and love, and understanding companionship."
>
> *Amelia Earhart (1897-1937) US-amerikanische Flugpionierin und Frauenrechtlerin*

Solveig C.S. Eschen Hamburg, September 2013

[1] Ein Hinweis zur Sprachform der vorliegenden Arbeit: Da es für die meisten Personenbezeichnungen im Deutschen leider keine geschlechtsneutrale Form gibt, wähle ich in dieser Arbeit ab sofort die im allgemeinen Sprachgebrauch nach wie vor als neutrale Form verstandene männliche Personenbezeichnung, sofern beide Geschlechter angesprochen werden sollen. Wenn explizit von weiblichen Personen berichtet wird, gebrauche ich selbstverständlich die weibliche Schreibweise.

1 Einleitung

> „Die Rechenautomaten haben etwas von den Zauberern im Märchen. Sie geben einem wohl, was man sich wünscht, doch sagen sie einem nicht, was man sich wünschen soll."
>
> *Norbert Wiener (1894-1964)*
>
> *US-amerikanischer Mathematiker, Begründer der Kybernetik*

Dies galt für Wieners Rechenmaschinen in den 1940er Jahren und gilt auch aktuell noch für die modernen Computer heutiger Zeit. Dennoch ist die Kommunikation zwischen Mensch und Maschine intensiver geworden. Auch wenn Computernutzern nach wie vor nicht von der Maschine gesagt werden kann, was sie sich *wünschen* sollen, kommt es durch zunehmende Automation vor, dass Computer vorschlagen, was ein Mensch – in unserem Fall ein Operateur – *tun* soll. Die enge Zusammenarbeit von Mensch und Maschine kann, wie die Erläuterungen im theoretischen Teil dieser Arbeit zeigen, zur Teamarbeit werden.

Automation und Teamarbeit zwischen Mensch und Maschine erhalten in vielen Lebensbereichen Einzug, so auch in der Luftfahrt. Wie unter 2.4 näher erläutert werden wird, ist die Einführung neuer automatischer Funktionen bereits im Gange. Gleichzeitig ist die Möglichkeit der Automatisierung von Air Traffic Management begrenzt, so dass auch in Zukunft menschliche Akteure die Verantwortung für Entscheidungsprozesse tragen werden (Hansman, 2001). Nach Expertenmeinung wird es in Zukunft keine hundertprozentige Automation in der Luftfahrt und auch nicht in anderen Domänen geben. Der menschliche Akteur wird durch Automation unterstützt, nicht aber ersetzt werden.

Doch was macht einen zukünftigen Luftfahrtoperateur aus? Was sind angemessene Fähigkeiten und Kompetenzen? Mit der vorliegenden Arbeit sollen Erkenntnisse darüber, welchen Beitrag Persönlichkeit und Einstellung zur Eignung von Bewerbern als Luftfahrtoperateure der Zukunft leisten, voran gebracht werden. Der Fokus liegt dem Titel entsprechend auf Persönlichkeit als Prädiktor für Leistung in hoch automatisierten Mensch-Maschine-Teams der Luftfahrt.

Da sowohl Aspekte heutiger Eignungsdiagnostik auch in Zukunft noch relevant sein können, gleichzeitig aber von veränderten Anforderungen ausgegangen werden muss, wird ein Ansatz gewählt, der sowohl Bezüge zu heutigen Auswahlverfahren als auch zu zukünftigen Szenarien ermöglicht. Als Leistungskriterien stehen zum einen die Auswahlergebnisse der Fluglotsen- und Pilotenauswahl im Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR) zur Verfügung, zum anderen Leistungsvariablen aus einer eigens für die Fragestellung ent-

wickelten Simulation zukünftiger Anforderungen. Als Prädiktoren dienen diverse Persönlichkeits- und Einstellungsmerkmale. Bewusst wird hier ein sehr offenes und exploratives Vorgehen gewählt, um nicht von Vornherein Merkmale auszuschließen, die von Relevanz sein könnten. Da jedoch im Endergebnis ein ökonomisches Instrument für weitere Forschung und einen späteren Einsatz in der Auswahl von Luftfahrtoperateuren entstehen soll, ist auch das Aussortieren von Prädiktoren Ziel dieser Arbeit.

Die nun folgenden Kapitel sind der Übersicht halber in einen theoretischen und einen empirischen Teil gegliedert. Zunächst werden die zentralen Oberbegriffe der Arbeit erläutert, das Thema Persönlichkeit als Einflussfaktor näher beleuchtet sowie der Forschungskontext beschrieben. Im darauf folgenden empirischen Teil werden die Fragestellung spezifiziert, die zugehörigen Hypothesen dargestellt und im Anschluss Methode und Ergebnisse erläutert. Abschließend werden die Ergebnisse diskutiert. Ein Ausblick legt dar, welche weiterführende Beschäftigung mit dem Thema für sinnvoll erachtet wird.

A. Theoretischer Teil

2 Zentrale Oberbegriffe

In den folgenden Abschnitten werden die für diese Arbeit zentralen Oberbegriffe Persönlichkeit, Leistung, Automation und Mensch-Maschine-Interaktion beschrieben. Persönlichkeit (2.1) und Leistung (2.2) werden dabei als Grundbegriffe der Psychologie überblicksartig vorgestellt. Für detailliertere Informationen zu diesen jeweils sehr umfänglichen Forschungsbereichen sei folgende Grundlagenliteratur empfohlen, die auch in den einzelnen Abschnitten in Auszügen zitiert wird: Asendorpf (2007) gibt einen detaillierten Einblick in die verschiedenen Persönlichkeitstheorien. Zum Thema Persönlichkeitsdiagnostik seien Schmidt-Atzert und Amelang (2012) genannt. Letztere ebenfalls für das Thema Testleistung, während das Thema berufliche Leistung z.B. bei Marcus und Schuler (2006) näher ausgeführt wird.

Die Begriffe Mensch-Maschine-Interaktion (2.3) sowie Automation (2.4) werden aufgrund ihrer Spezifität etwas ausführlicher erläutert. Im Abschnitt Automation wird speziell auf die Themen Automation in der Luftfahrt sowie die Einführung von Automatisierung in anderen Domänen eingegangen. Als weiterführende Literatur zu den sich überschneidenden Themen Automation und Mensch-Maschine-Interaktion sind Wickens, Lee, Liu und Gordon Becker (2004) sowie Sheridan und Parasuraman (2005) und zudem alle weiteren in den Abschnitten 2.3 und 2.4 genannten Quellen von Parasuraman zu empfehlen.

2.1 Persönlichkeit

Eine gängige Definition von Persönlichkeit stammt von Gerrig und Zimbardo (2008), die selbige „als eine komplexe Menge von einzigartigen psychischen Eigenschaften, welche die für ein Individuum charakteristischen Verhaltensmuster in vielen Situationen und über einen längeren Zeitraum hinweg beeinflussen.“ (S. 504) beschreiben. Bezüglich der Frage, wie diese Verhaltensmuster strukturiert und kategorisiert werden können, gibt es verschiedenste Ansätze, die in unterschiedlichen Persönlichkeitstheorien ihren Ausdruck finden. Diese Persönlichkeitstheorien können als hypothetische Aussagen über die Struktur und die Funktionsweise individueller Persönlichkeiten beschrieben werden. Die verschiedenen Aussagen beinhalten dabei unter anderem Vorhersagen darüber, wie Menschen auf bestimmte Bedingungen reagieren und sich an diese anpassen. (Gerrig & Zimbardo, 2008).

Um Persönlichkeitseigenschaften zu erfassen, gibt es in der Psychologie verschiedene diagnostische Möglichkeiten. Schmidt-Atzert und Amelang (2012) nennen hierzu z.B. Persön-

lichkeitsfragebögen, nichtsprachliche und objektive Persönlichkeitstests, projektive Verfahren, Verhaltensbeobachtung und –beurteilung sowie diagnostische Interviews. Persönlichkeitsfragebögen gehören dabei zu den am häufigsten verwendeten persönlichkeitsdiagnostischen Verfahren und sind auch in dieser Arbeit die zentrale Methode. Die nach Asendorpf (2007) wohl am besten untersuchte Persönlichkeitseigenschaft überhaupt – Intelligenz – spielt hier eine Rolle in Form von kognitiven Leistungsvariablen, die als Kriterien dienen (siehe Abschnitte 7.1.4 und 7.2.2) und naturgemäß stark mit Intelligenz verknüpft sind. Die allgemeine Intelligenz wird in der Psychologie meist als die allen kognitiven Leistungstests gemeinsame Varianz verstanden. Damit erfasst jeder kognitive Leistungstest einen gewissen Teil der allgemeinen Intelligenz (Schmidt-Atzert, Deter, & Jaeckel, 2004).

Hinsichtlich der Entwicklung von Persönlichkeitsfragebögen ist eine gängige Methode die Persönlichkeitsklassifikation. Hierbei wird der zu klassifizierende Persönlichkeitsbereich eingegrenzt, eine möglichst umfassende Liste aller Eigenschaften in dem eingegrenzten Bereich erstellt und dann auf möglichst wenige voneinander unabhängige Dimensionen reduziert. Damit wird jede Dimension durch eine Skala gemessen, die eine Variable darstellt, in der sich Personen unterscheiden (Asendorpf, 2007).

Der mittlerweile wahrscheinlich bekannteste Klassifikationsansatz dieser Art ist das sogenannte Big-Five-Modell, welches die menschliche Persönlichkeit in fünf Hauptfaktoren beschreibt (Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit, Gewissenhaftigkeit). Cattell (1996) gibt einen guten Überblick über die Entstehungsgeschichte des Modells. Auch Abschnitt 7.1.1 geht mittels der Beschreibung des NEO-PI-R (Ostendorf & Angleitner, 2004) noch einmal näher auf die Big Five ein.

Zusätzlich zu Persönlichkeitsfragebögen kommt in dieser Arbeit auch ein objektiver Persönlichkeitstest zum Einsatz. Eine übliche Definition hierzu stammt von Schmidt (1975) und kennzeichnet diejenigen Verfahren als objektive Persönlichkeitstests, die das Verhalten von Individuen in standardisierten Situationen erfassen, ohne dass diese sich selbst beurteilen müssen. Abschnitt 8.1.1 beschreibt das verwendete Verfahren.

2.2 Leistung

Laut physikalischen Definitionen (z.B. nach Häcker & Stapf, 1998; Stichwort Leistung) handelt es sich bei Leistung um das Verhältnis von verrichteter Arbeit zu der dafür benötigten Zeitspanne, bzw. um die Fähigkeit, in einer bestimmten Zeiteinheit eine bestimmte Arbeit zu verrichten. Diese Definition trifft im Kern auch auf den psychologischen Begriff der Leistung zu, die hier in zwei ihrer Facetten relevant ist: zum einen als Testleistung und zum anderen als berufliche Leistung.

Testleistung ist nach Schmidt-Atzert und Amelang (2012) als Quantifizierung dessen zu verstehen, wie viele Aufgaben in einer feststehenden Bearbeitungszeit gelöst werden können oder wie viel Zeit zur Bearbeitung einzelner oder aller Aufgaben eines Tests gebraucht wird. Die Autoren erwähnen in diesem Kontext Aufmerksamkeits- und Konzentrationstests, spezielle Fähigkeitstests, Intelligenztests, Entwicklungstests und Schultests. Für diese Arbeit relevante Leistungstests gehören den beiden ersten Kategorien an.

Der Bereich der beruflichen Leistung überschneidet sich begrifflich mit dem Thema Berufserfolg. Letzterer kann sowohl mit neutralen Kennzahlen (z.B. Lohn und Arbeitsniveau, beruflicher Aufstieg oder Anzahl der Mitarbeiter) als auch mit Bezugsstandards (z.B. Vergleich mit Zielen und deren Erreichung, Vergleich mit anderen Personen oder Vergleich mit bestehenden Normen) gemessen werden (Dette, Abele, & Renner, 2004). Marcus und Schuler (2006) sprechen für die Beurteilung beruflicher Leistung von objektiven Quellen (z.B. Stückzahlen, Ausschussquoten, Anzahl an Kundenbeschwerden) und subjektiven Quellen (z.B. Vorgesetztenbeurteilungen, Beurteilungen durch Kollegen, Mitarbeiterurteile oder Selbstbeurteilungen).

Die Abschnitte 7.1.4 und 7.2.2 beschreiben die in dieser Arbeit relevanten Leistungsvariablen. In der verwendeten Simulation überschneiden sich Leistungstest und objektive Quelle für berufliche Leistung, da zukünftige berufliche Anforderungen simuliert und deren Erfüllung gemessen werden.

2.3 Mensch-Maschine-Interaktion

Der Begriff Mensch-Maschine-Interaktion ist nur einer von vielen möglichen Ausdrücken für ein weites Feld, das sich mit der Interaktion von Menschen mit in irgendeiner Form technischen, automatisierten oder computerisierten Systemen oder Maschinen befasst. Entsprechend viele Begriffe existieren in diesem Bereich und überschneiden sich in großen Teilen oder können sogar komplett synonym verwendet werden. Häufig verwendet werden die Begriffe „Human-Machine-Interaction“ (z.B. Hoc, 2000), „Human-Computer-Interaction“ (z.B. Wickens, Lee, Liu, & Gordon Becker, 2004), „Human-Agent-Interaction“ (z.B. Bradshaw, Feltovich, & Johnson, 2011) oder „Human Automation-Interaction“ (z.B. Sheridan & Parasuraman, 2005).

Eine einleuchtende Definition des Begriffs „Human-Automation-Interaction“, die so auch für die anderen genannten Begriffe dienen könnte, liefern Sheridan und Parasuraman (2005). Sie beschreiben den Begriff als „Umstand“, in dem Personen (a) der Automation (einem Computer) die Ziele und Einschränkungen (tue X aber vermeide Y) sowie die Kompromisse zwischen Zielen und Einschränkungen vorgeben; (b) die Aufgabenerfüllung der Automation starten, beenden oder modifizieren; und (c) von der Automation Informationen, Energie, physische Objekte oder Substanzen erhalten.

Parasuraman und Riley (1997) stellen fest, dass selbst sogenannte „unbemannte“ Systeme, wie z.B. Raumfahrzeuge, stets menschliche Operateure in der Rolle des Überwachers einbeziehen. Der Hintergrund für den Verzicht darauf, menschliche Operateure komplett zu ersetzen, sei die allgemeine Wahrnehmung des Menschen als flexibler, anpassungsfähiger, kreativer und damit geeigneter, sich auf veränderte oder unvorhergesehene Zustände einzustellen.

Der Charakter der Mensch-Maschine-Interaktion hat sich in den letzten Jahrzehnten gewandelt. Während es zu Beginn der Forschung in diesem Bereich noch darum ging, festzustellen, welche Dinge besser vom Menschen und welche besser von Maschinen ausgeführt werden sollten, ist heute vorrangiges Ziel, menschliche Kooperation mit automatisierten Systemen bestmöglich zu gestalten. Dem viel zitierten HABA-MABA-Konzept (humans are better at…/ machines are better at…) von Fitts (1951), z.B. beschrieben bei Bradshaw, Feltovich und Johnson (2011), steht heute der Wunsch gegenüber, Mensch-Maschine-*Interaktion* zu Mensch-Maschine-*Kooperation* weiterzuentwickeln (Hoc, 2000).

Wenn man in diesem Kontext Definitionen von Teamarbeit betrachtet, kann prinzipiell auch enge Mensch-Maschine-Kooperation als Teamarbeit verstanden werden. So werden Teams im Allgemeinen als soziale Einheiten von Mitgliedern mit hoher Aufgabeninterdependenz und geteilten Zielen charakterisiert (Dyer, 1984, zitiert nach Salas, Cooke & Rosen, 2008).

Auch bei der Zusammenarbeit von Mensch und Maschine, im Folgenden am Beispiel der Flugsicherung geschildert, sind diese Charakteristika häufig gegeben. Die technischen Flugsicherungssysteme sind unter anderem darauf ausgelegt, Sicherheitsabstände zwischen Flugzeugen aufrecht zu erhalten. Werden Sicherheitsabstände unterschritten, gibt das System visuelle Warnungen, die vom Menschen zu Befehlen an die jeweiligen Piloten weiterverarbeitet werden. Schon dieses simple Beispiel zeigt, dass auch hier Aufgabeninterdependenz (sowohl Mensch als auch System sind für die Sicherheitsabstände zuständig) und geteilte Ziele (sowohl Mensch als auch System haben das Ziel, Sicherheitsabstände einzuhalten) vorliegen.

Erste Ideen in diese Richtung wurden von Hollnagel und Woods (1983) entwickelt, die feststellten, durch die zunehmende Ausreifung von Computeranwendungen werde die Mensch-Maschine-Schnittstelle immer mehr zur Interaktion zweier kognitiver Systeme. Woods (1996) entwickelte diesen Ansatz dahingehend weiter, dass die Einführung automatischer Systeme zur Unterstützung menschlicher Operateure im Grunde bedeute, neue Teammitglieder einzugliedern. Beim Design neuer automatisierter und computerisierter Systeme sei man mit mehr als nur der Hardware oder Software befasst. Man designe zusätzlich ein Team aus Personen und Maschinen, welches seine Bewertungen und Aktivitäten koordinieren müsse, sobald eine Situation an Tempo, Schwierigkeit und Gefährlichkeit zunehme.

Nass, Fogg und Moon (1996) konnten zudem zeigen, dass Menschen sozial auf Technik und Computer reagieren und zwar ähnlich wie auf menschliche Kollaborateure. Probanden, denen für eine Aufgabe mitgeteilt wurde, sie seien dafür gemeinsam mit einem Computer zuständig, der zu ihrem Team gehöre, fühlten sich auch tatsächlich mit diesem in einem Team. Sie zeigten die gleichen gruppendynamischen Effekte von Teamarbeit wie bei Teamarbeit mit anderen Menschen: sie empfanden den Computer beispielsweise als ähnlicher zu sich selbst, waren offener für Beeinflussungen durch den Computer, fanden dessen Informationen qualitativ hochwertiger und freundlicher und hielten sich eher an seine Vorschläge als eine Kontrollgruppe dies tat. Letzterer wurde mitgeteilt, ihr Computer gehöre eigentlich zu einem anderen Team. Hier blieben die beschriebenen Effekte aus.

Für die Teamarbeit zwischen Mensch und Maschine finden sich in der Literatur neben „human-machine teams“ (z.B. Neerincx et al., 2008) auch Begriffe wie „human-agent teams“ (z.B. Deshmukh, McComb, & Wernz, 2008) oder „human-robot teams“ (z.B. Hoffman & Breazeal, 2004), deren jeweiliges Ziel ist, die enge Kooperation zwischen Mensch und Maschine zu charakterisieren. Für das Forschungsprojekt HYBRID (siehe Abschnitt 4.3) wurde von Eißfeldt (2008) der Begriff des „hybriden Teams“ geprägt, der die Verschiedenartigkeit

der kooperierenden Teampartner hervorhebt („hybrida“: lat. für „Mischling“). **Abbildung 1** skizziert die Teamarbeit zwischen Mensch und Maschine am Beispiel der Flugsicherung.

Abbildung 1: Das Mensch-Maschine-Team in der Flugsicherung
©DLR

Eng verbunden mit dem Thema der Mensch-Maschine-Interaktion ist der Bereich der Automation. Dieser letzte noch zu klärende Oberbegriff wird im folgenden Abschnitt 2.4 näher erläutert.

2.4 Automation

Unter Automation oder auch Automatisierung wird im Allgemeinen verstanden, dass Tätigkeiten, die vorher von Menschen durchgeführt wurden, nun von Computern bzw. Maschinen ausgeführt werden (Parasuraman & Riley, 1997; Wickens et al., 2004). In der Literatur werden hierbei verschiedene Automatisierungslevel unterschieden. **Abbildung 2** auf der folgenden Seite gibt einen Überblick über acht Automatisierungslevel wie sie z.B. bei Parasuraman, Sheridan und Wickens (2000) oder auch Sheridan und Parasuraman (2005) beschrieben werden.

Automatisierungslevel	
1.	Die Automation stellt keine Hilfe bereit, die komplette Kontrolle liegt beim Menschen.
2.	Die Automation schlägt Alternativen vor, filtert und hebt die beste Alternative hervor.
3.	Die Automation wählt eine Alternative aus und schlägt diese der Person vor.
4.	Die Automation führt die Handlung aus, wenn die Person dies akzeptiert.
5.	Die Automation gibt der Person beschränkte Zeit, um vor der Ausführung ein Veto gegen die Handlung einzulegen.
6.	Die Automation führt die Handlung aus und informiert die Person im Nachhinein
7.	Die Automation führt die Handlung aus und informiert die Person nur auf Nachfrage.
8.	Die Automation wählt eine Methode aus, führt die Aufgabe aus und ignoriert den Menschen (der Mensch hat kein Vetorecht und wird nicht informiert).

Abbildung 2: Automatisierungslevel
(in Anlehnung an Parasuraman et al., 2000 / Sheridan & Parasuraman, 2005)

Wickens et al. (2004) teilen die Gründe für die Automatisierung von Tätigkeiten in vier grobe Kategorien ein:

1. Eine Ausführung wäre für den Menschen unmöglich oder gefährlich (z.B. Arbeit mit gefährlichen Materialien oder in gefährlichen Umgebungen)
2. Eine Ausführung wäre für den Menschen schwierig oder unangenehm (z.B. Rechnen mit Taschenrechner oder automatische Warnung bei niedrigem Ölstand im Auto)
3. Erweiterung der menschlichen Fähigkeiten (z.B. Assistenzsysteme im Flugzeug)
4. Die technischen Möglichkeiten sind da und nicht teuer, bzw. sogar billiger als menschliche Arbeitskräfte (z.B. automatische Menüführung bei telefonischen „Hotlines“)

Wie bereits in der Einleitung erwähnt, schreitet die Automatisierung von Arbeitsabläufen auch in der Luftfahrt stetig fort. Das europäische Luftfahrtprogramm Single European Sky (SES) sieht die Einführung von neuen automatischen Funktionen vor (SESAR Consortium, 2007). Bibby, Margulies, Rijnsdorp und Withers (1975, zitiert nach Bainbridge, 1983) stellen zu diesem Thema jedoch fest, dass sogar hoch automatisierte Systeme (z.B. in Elektrizitätswerken) Menschen zur Überwachung, Regulierung, Wartung, Erweiterung und Verbesserung benötigen. Demnach könne nur die paradoxe Aussage getroffen werden, dass automatisierte Systeme letztendlich immer Mensch-Maschine-Systeme seien, für die damit sowohl technische als auch menschliche Faktoren wichtig sind. Diese Aussage aus den 1970er Jahren hat nach wie vor Gültigkeit.

Nach Eißfeldt (1991) wird so auch in der Flugsicherung niemals 100%ige Automation vorhanden sein. Stattdessen sei davon auszugehen, dass Menschen immer die zentralen Entscheider in Flugsicherungssystemen sein werden. Automatische Systeme würden leistungsfähiger werden und dem Menschen assistieren, ihn aber niemals ganz ersetzen. Die richtigen Menschen für solch einen fordernden Job auszuwählen, sei daher nach wie vor entscheidend. Auch das SESAR-Konsortium (2007) äußert sich ähnlich, betont jedoch zudem die Notwendigkeit höherer Automatisierungslevel. Menschen (mit angemessenen Fähigkeiten und Kompetenzen sowie ordnungsgemäß autorisiert) bilden nach Aussage des Konsortiums das Herzstück zukünftiger europäischer ATM(Air Traffic Management)-Systeme. Ein höheres Level an Automatisierung sei aber dennoch von Nöten.

Wann immer Automation in eine Arbeitsdomäne eingeführt wird, sollten nicht nur die positiven Effekte wie z.B. eine Ausweitung der technischen Möglichkeiten oder Effizienzsteigerung betrachtet werden, sondern ebenfalls mögliche negative psychologische Folgen für die Menschen, die mit der Automation arbeiten (Parasuraman & Riley, 1997). Automatisierung kann Operateure durch die Übernahme manueller oder kognitiver Funktionen entlasten, sie aber gleichzeitig durch neu hinzugekommene Anforderungen bei der Überwachung stark – und anders als gewohnt – beanspruchen (Parasuraman, Mouloua, & Molloy, 1996). Ein weiteres in der Literatur viel diskutiertes Thema ist Vertrauen in Automation (vgl. z.B. Lee & See, 2004). Problematisch kann dieses in zweierlei Hinsicht sein. Auf der einen Seite kann einer *unzuverlässigen* Automation zu *stark* vertraut werden (Parasuraman et al., 1996; Riley, 1996), auf der anderen Seite kann einer *zuverlässigen* Automation zu *wenig* vertraut werden (Lee & Moray, 1994; Parasuraman et al., 1996). Hoch sicherheitsrelevant sind dabei die Folgen von zu großem Vertrauen, sogenannte Complacency-Fehler (Parasuraman & Manzey, 2010). Diese lassen sich in zwei Kategorien einteilen: „errors of omission“ bei denen von Operateuren nicht auf abweichendes Systemverhalten reagiert wird, weil die Automation die Abweichung nicht erkannt oder angezeigt hat, und „errors of commission“ bei

denen Operateure in einer Situation nach falschen Vorgaben der Automation agieren und dabei die Daten einer zuverlässigeren Informationsquelle, anhand derer die falsche Vorgabe erkennbar wäre, ignorieren (Mosier & Skitka, 1996). Ein weiteres wichtiges Thema ist der mögliche Abbau der Fertigkeiten von Operateuren durch die längerfristige Nutzung von Automation (Hopkin, 1991). Nach Sarter und Woods (1991) kann das Übertragen von Entscheidungsfunktionen auf eine Automation z.B. zu einem Rückgang des Situationsbewusstseins führen.

Anhand dieser Ausführungen wird deutlich, dass Automatisierung die Tätigkeit von Operateuren verändert. Dies kann sich auch in einem veränderten Fähigkeits- und Persönlichkeitsprofil niederschlagen (Eschen, Knappe, & Eißfeldt, 2012). Abschnitt 4.2 geht näher auf dieses Thema ein.

3 Persönlichkeit als Einflussfaktor

Nachdem im vorigen Kapitel die zentralen Oberbegriffe Persönlichkeit, Leistung, Mensch-Maschine-Interaktion und Automation geklärt wurden, greift dieses Kapitel Zusammenhänge zwischen den genannten Aspekten auf. Beleuchtet wird der jeweilige Einfluss der menschlichen Persönlichkeit auf die anderen Gebiete.

3.1 Persönlichkeit und Leistung

Die Vorhersage von Leistung durch Persönlichkeit ist vor allem im Bereich der Eignungsdiagnostik von Relevanz. In vielen Auswahlverfahren wird Persönlichkeitsdiagnostik heute im Rahmen von Assessment Centern (AC) betrieben. Eine klare Mehrheit von 71% der Dax-100-Unternehmen und 73% aller in einer Studie vom Arbeitskreis AC e.V. befragten Firmen setzen mittlerweile auf die Methode AC oder DC (Development Center), (Domke, 2008). Wann immer es jedoch gelingt, mit in der Durchführung vergleichsweise ökonomischen Persönlichkeitsfragebögen beruflichen Erfolg valide zu prognostizieren, ist dies eine kostengünstige Alternative zu den teureren Assessment Centern. Diese bedeuten bei korrekter Anwendung aufwendige Vorbereitungen und ständige Standardisierung durch z.B. Beobachtertrainings, um ein ausreichendes Maß an Objektivität sicherzustellen (vgl. hierzu z.B. Eschen-Léguedé, 2007 oder auch Kleinmann, 2003). Der Forschungsbereich „Personality and Performance“ ruft seit mittlerweile Jahrzehnten Kritiker auf den Plan, die die prognostische Validität von Persönlichkeitsfragebögen für berufliche Leistung generell anzweifeln (Morgeson et al., 2007). Dennoch beschäftigen sich diverse Forschungsgruppen mit dem Thema und können immer wieder (Teil-) Erfolge vermelden.

So betrachten z.B. Barrick, Mount und Judge (2001) in einer viel zitierten Metaanalyse die Beziehung zwischen den Big Five und beruflicher Leistung. Aus 15 einbezogenen vorherigen Metaanalysen ergab sich der Faktor Gewissenhaftigkeit als valider Prädiktor über verschiedene Leistungsvariablen hinweg. Ebenso zeigte der Faktor Neurotizismus negative Zusammenhänge zu allgemeiner Arbeitsleistung. Die anderen drei Faktoren des Big-Five-Modells (Extraversion, Offenheit für Erfahrungen und Verträglichkeit) konnten den Erfolg in einzelnen Berufen, bzw. einzelnen Kriterien vorhersagen. So ist beispielsweise hohe Extraversion in Berufen mit großer interpersonaler Komponente vorteilhaft. Eine Vorhersage des allgemeinen Berufserfolges war jedoch nicht möglich. Kriterien der Metaanalyse waren unter anderem die allgemeine Arbeitsleistung (z.B. Produktivitätsdaten), Leistungsbewertungen von Vorgesetzten, objektive Leistungsindikatoren (z.B. Beförderungen und Lohnhö-

he), Leistung bei Trainings und Leistung in der Teamarbeit. Eine jüngere Metaanalyse von Ones, Viswesvaran, Judge und Dilchert (2007) bestätigt diese Ergebnisse.

Ein aktuelles Beispiel für den Zusammenhang von Persönlichkeit und Leistung liefert die Validierungsstudie zur Fluglotsenauswahl im Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR) von Eißfeldt, Keye, Conzelmann, Grasshoff, Pecena und Heintz (2013). Der in der Auswahl eingesetzte Persönlichkeitsfragebogen TSS (Temperament-Struktur-Skalen, siehe auch Abschnitt 4.1) wurde hier hinsichtlich seiner prädiktiven Validität für Leistung in der Fluglotsenausbildung geprüft. Drei Skalen zeigten dabei prädiktive Validität hinsichtlich Ausbildungsleistungen: Rigidität, Aggressivität und Leistungsmotivation. Höhere Rigidität ging mit besserem Ausbildungserfolg insgesamt, besserer Leistung in der theoretischen Abschlussprüfung und besserer durchschnittlicher Leistung in praktischen Prüfungen einher. Niedrigere Aggressivität zeigte Zusammenhänge mit größerem Erfolg in der theoretischen Ausbildung, besseren Noten in der theoretischen Ausbildungsprüfung und besserer durchschnittlicher Leistung in praktischen Prüfungen. Höhere Leistungsmotivation korrespondierte mit geringerer Wahrscheinlichkeit, eine Prüfung wiederholen zu müssen sowie wiederum besserer Leistung in der theoretischen Abschlussprüfung und besserer durchschnittlicher Leistung in praktischen Prüfungen.

Zusammenfassend lässt sich feststellen, dass alle Studien zum Thema Persönlichkeit und Leistung von generell eher kleinen bis maximal mittleren Effekten gekennzeichnet sind. Dennoch zeigen sich diese konsistent und sollten daher nicht unbeachtet bleiben.

Ein wichtiger Kritikpunkt am Einsatz persönlichkeitspsychologischer Diagnostik ist das Thema der sozialen Erwünschtheit, in englischsprachiger Literatur oft als „Impression Management“ (Li & Bagger, 2006) oder drastischer „Faking“ (Ones & Viswesvaran, 1998) bezeichnet. Kritiker mahnen hier an, dass die Ergebnisse von Persönlichkeitsfragebögen als Selbstbeschreibungen – gerade im Bewerbungskontext – anfällig für unehrliche Beantwortung seien. Dem gegenüber stehen jedoch Ergebnisse, die dieses Thema als unkritisch bewerten. So kommen Ones, Viswesvaran, Judge und Dilchert (2007) unter Einbeziehung verschiedenster Studien zu der Schlussfolgerung, dass Verfälschungen durch Berufsbewerber überbewertet werden. Sie drücken ihr Verständnis für die Skepsis hinsichtlich des Einsatzes von Selbstbeschreibungen in Personalauswahlsettings aus, geben jedoch zu bedenken, dass es Nachweise aus Bewerberstichproben gibt, die die Kriteriums- und Konstruktvalidität von Persönlichkeitsmessungen in entsprechenden Kontexten belegen. So geben z.B. Li und Bagger (2006) an, dass die Auspartialisierung von Impression Management und der Neigung zu Selbstbetrug nur geringe bis vernachlässigbare Effekte auf die Kriteriumsvalidität

von Persönlichkeitskonstrukten zeigt. Damit scheiden diese nach Ansicht der Autoren als Suppressoren von Korrelationen zwischen Persönlichkeit und Kriteriumsvariablen aus.

3.2 Persönlichkeit in der Mensch-Maschine-Interaktion

Auch im Gebiet der Mensch-Maschine-Interaktion ist das Thema Persönlichkeit hoch relevant. Bisher gibt es jedoch wenig Forschung, die sich konkret mit Persönlichkeit als Einflussfaktor beschäftigt. Kain und Nachtwei (2009) vermissen in der Forschung zu Human Factors (HF) Fragestellungen zum Thema Technikgestaltung, die sich statt des Einflusses von Leistung, Beanspruchung und Zufriedenheit des Menschen mit Persönlichkeitspsychologie beschäftigen. Sie schlagen vor, dass statt der isolierten Betrachtung von Umwelt und Persönlichkeit vielmehr die Interaktion beider Faktoren im Zentrum der Betrachtungen stehen sollte. Personen würden bisher meist nur nach Alter, Bildungsgrad oder Geschlecht, nicht aber nach Persönlichkeit differenziert. In einer Veröffentlichung zur Rolle von Kontrollvariablen in der HF-Forschung fassen sie Studien zur Vorhersage von HF-relevanten Außenkriterien zusammen. **Tabelle 1** gibt einen Überblick.

Tabelle 1: Persönlichkeit in der Vorhersage von HF-relevanten Außenkriterien
(nach Kain & Nachtwei, 2009)

Persönlichkeitsmerkmal	*Form des Zusammenhangs*	*vorhergesagtes Verhalten / Befinden / Zustand*	*Kontext der Untersuchung*
Neurotizismus	+	riskantes Fahren	Straßenverkehr (Kfz)
Gewissenhaftigkeit	+	Anstrengung	simulierte Flugsicherung
Risikoneigung	-	Situationsbewusstsein	simulierte Flugzeugführung
Intelligenz (g-Faktor)	+	Prozessführungsleistung	simulierte Wasseraufbereitung

Neben Persönlichkeitseigenschaften im engeren Sinne, wie z.B. den Big Five, sind zusätzliche Aspekte von Persönlichkeitsbeschreibungen für die Mensch-Maschine-Interaktion relevant. Huber, Kain und Nachtwei (2008) gehen genauer darauf ein und geben in Anlehnung an Corsini (1977) einen Überblick über wesentliche Begriffe der wissenschaftlichen Persönlichkeitsbeschreibung. **Tabelle 2** auf der Folgeseite gibt die Merkmale wider, die nach Recherchen der Autoren statistisch nachweisbaren Bezug zum Verhalten von Personen im Human Factors Bereich (Operateuren, Bedienern etc.) haben.

Tabelle 2: Überblick über Persönlichkeitsmerkmale und ihre Zuordnung zu wissenschaftlichen Begriffen der Persönlichkeitsbeschreibung (nach Huber et al., 2008)

Merkmal	**Wissenschaftlicher Begriff zur Persönlichkeitsbeschreibung**	**Beschreibung** (modifiziert nach Corsini, 1977)
Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit, Gewissenhaftigkeit	Eigenschaft („trait“)	Konstante, überdauernde und spezifische Art des Verhaltens
Risikoneigung		
Leistungsmotivation	Disposition	Neigung oder Bereitschaft in einer Person, auf eine gegebene Situation in einer charakteristischen Weise zu reagieren
Vertrauen (in Technik)	Einstellung	Erlernte Neigung, Klassen von Gegenständen oder Menschen in Abhängigkeit von eigenen Überzeugungen und Gefühlen zu bewerten
Intelligenz	Fähigkeit	Hier: Befähigung, neuen Problemen oder Situationen gerecht zu werden, ohne dass es dazu im wesentlichen Ausmaß früherer Lernerfahrung bedarf

Der von Huber et al. (2008) in der Tabelle aufgelistete Bereich der Einstellung umfasst neben dem erwähnten Vertrauen in Technik auch darüber hinausgehende Einstellungsaspekte. So teilt z.B. das Konzept von Richter, Naumann und Groeben (2001) computerbezogene Einstellungen mit dem Fragebogen FIDEC (Fragebogen zur inhaltlich differenzierten Erfassung von computerbezogenen Einstellungen; siehe auch unter 8.1.1) nach evaluativem Fokus (persönliche Erfahrung vs. gesellschaftliche Folgen), Nutzungsdomänen (Lernen und Arbeiten vs. Unterhaltung und Kommunikation) sowie Valenz (positive Einstellungskomponenten - Computer als nützliches Werk-zeug/nützliche Technologie vs. negativen Einstellungskomponenten - Computer als unbeeinflussbare Maschine/unbeeinflussbare Technik) ein.

Im Bereich der bereits erwähnten Big Five ließen sich in mehreren Studien negative Zusammenhänge zwischen der Computernutzungsvariable Internetnutzung und Verträglichkeit, Gewissenhaftigkeit sowie Extraversion feststellen. Ausgeprägtere Internetnutzung geht mit geringerer Verträglichkeit, geringerer Gewissenhaftigkeit und geringerer Extraversion einher (Landers & Lounsbury, 2006). Ganz aktuell werden das sogenannte computerbezo-

gene Selbstkonzept (Janneck, Vincent-Höper, & Ehrhardt, 2012) sowie computerbezogene Attributionsstile (Janneck & Guczka, 2013) tiefergehend beforscht. Bei dem computerbezogenen Selbstkonzept (CSK) handelt es sich um ein psychologisches Konstrukt, das als „Teilbereich des akademischen Selbstkonzepts verstanden werden kann und computerbezogene Erfahrungen, Interessen, Motivationen, Einstellungen und Kompetenzen beschreibt." (Janneck et al., 2012, S. 244). **Abbildung 3** zeigt die von den Autorinnen postulierten Bestandteile des CSK, die in ähnlicher Form auch Bestandteil der Erhebungen für diese Arbeit sind (siehe Kapitel 7).

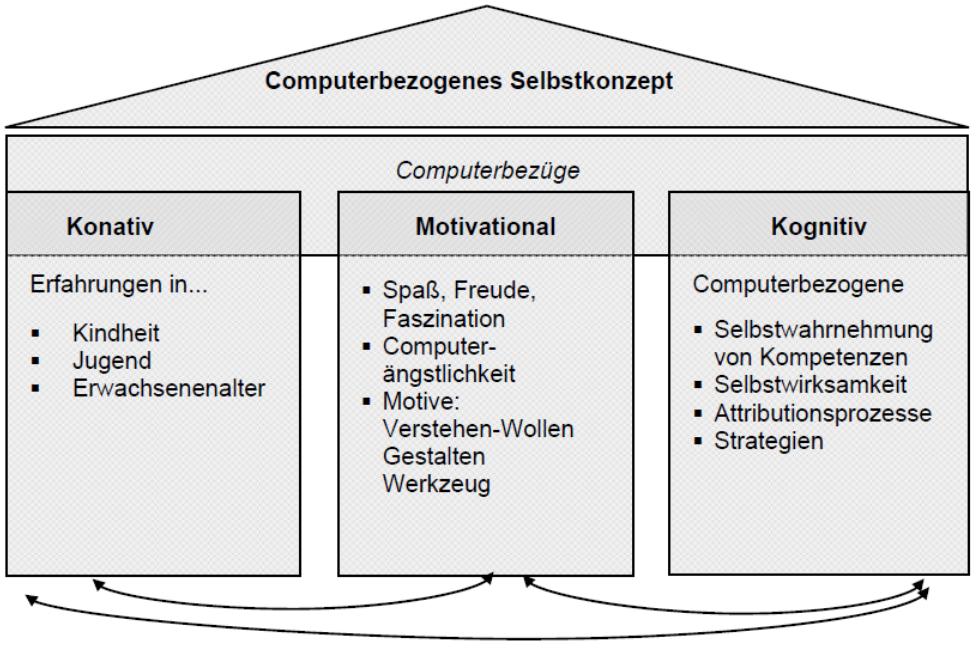


Abbildung 3: Das computerbezogene Selbstkonzept nach Janneck et al. (2012)

Im Rahmen des HYBRID-Projektes (siehe Abschnitt 4.3) wurde 2010 eine Diplomarbeit zum Thema „Messung von Persönlichkeitsaspekten mit Einfluss auf die Leistungsentfaltung in hochautomatisierten Systemen – Erprobung eines Forschungsfragebogens" (Knappe, 2010) angefertigt. Hierfür wurden aus verschiedenen Studien aktuell relevante Persönlichkeitsaspekte für Piloten und Fluglotsen zusammengestellt, deren Arbeitsplätze seit jeher stark von Mensch-Maschine-Interaktion geprägt und daher ein gutes Beispiel für die Relevanz von Persönlichkeit als Einflussfaktor sind. **Tabelle 3** auf den folgenden Seiten gibt einen Überblick (Anordnung der Persönlichkeitsaspekte in alphabetischer Reihenfolge).

Tabelle 3: Relevante Persönlichkeitsaspekte für Piloten und Fluglotsen
(modifiziert nach Knappe, 2010)

Persönlichkeitsaspekte	**Literaturangaben**
Aggressivität	Byrdorf (1998)
Arbeitsstil	Höft & Pecena (2005)
Commitment	Leonhardt (2005); Höft & Pecena (2005); Pecena (2002)
Disziplin	Höft & Pecena (2005)
Durchsetzungsfähigkeit	Goeters, Maschke & Eißfeldt (2004); Byrdorf (1998)
Emotionale Stabilität	Maschke (1987); Byrdorf (1998); Rathje, Golany & Eißfeldt (2004)
Entscheidungsfindung	Leonhardt (2005); Goeters, Maschke & Eißfeldt (2004); Höft & Pecena (2005); Pecena (2002); Maschke & Goeters (2003)
Entschlossenheit	Rathje, Golany & Eißfeldt (2004); Byrdorf (1998)
Flexibilität	Goeters, Maschke & Eißfeldt (2004); Höft & Pecena (2005); Pecena (2002); Maschke & Goeters (2003); Eißfeldt & Heintz (2002); Rathje, Golany & Eißfeldt (2004)
Führungsverhalten	Goeters, Maschke & Eißfeldt (2004); Eißfeldt & Heintz (2002); Byrdorf (1998); Flin (2005)
Gewissenhaftigkeit	Rathje, Golany & Eißfeldt (2004); Broach (2002)
Kommunikation	Broach (2002); Leonhardt (2005); Goeters, Maschke & Eißfeldt (2004); Maschke & Goeters (2003); Maschke (2005)
Kooperation	Goeters, Maschke & Eißfeldt (2004), Höft & Pecena (2005); Pecena (2002); Maschke & Goeters (2003); Byrdorf (1998); Rathje, Golany & Eißfeldt (2004); Flin (2005)
Motivation, Leistungsmotivation	Voller & Fowler (2005); Goeters, Maschke & Eißfeldt (2004); Maschke & Goeters (2003); Maschke (1987); Byrdorf (1998); Rathje, Golany & Eißfeldt (2004)
Resilienz	Goeters, Maschke & Eißfeldt (2004); Maschke & Goeters (2003); Eißfeldt & Heintz (2002); Rathje, Golany & Eißfeldt (2005)

Fortsetzung auf der Folgeseite

Persönlichkeitsaspekte	Literaturangaben
Fortsetzung von der vorigen Seite	
Resistenz gegenüber voreiligen Entscheidungen	Goeters, Maschke & Eißfeldt (2004); Eißfeldt & Heintz (2002)
Selbstbewusstsein	Goeters, Maschke & Eißfeldt (2004); Maschke & Goeters (2003); Rathje, Golany & Eißfeldt (2004)
Selbstkritik	Voller & Fowler (2005)
Selbstreflexion	Höft & Pecena (2005)
Social Sensitivity	Goeters, Maschke & Eißfeldt (2004); Eißfeldt & Heintz (2002); Rathje, Golany & Eißfeldt (2004)
Stresstoleranz, Stressresistenz	Leonhardt (2005); Goeters, Maschke & Eißfeldt (2004); Maschke & Goeters (2003); Höft & Pecena (2005); Pecena (2002); Byrdorf (1998)
Teamwork	Maschke (2005); Leonhardt (2005)
Verantwortungsbewusstsein	Leonhardt (2005)
Vertrauen	Voller & Fowler (2005)

3.3 Persönlichkeit und Automation

Im Gegensatz zu den in den beiden letzten Abschnitten behandelten Themen Persönlichkeit und Leistung, bzw. Persönlichkeit in der Mensch-Maschine-Interaktion, existieren zum Thema Persönlichkeit und Automation bisher mehr Annahmen als konkrete Forschungsergebnisse. Parasuraman, Mouloua und Molloy (1996) gehen davon aus, dass die Einführung und Erweiterung von (hoch)automatisierten Prozessen in das Arbeitsumfeld viele Veränderungen herbeiführen können. Wie bereits unter 2.4 angesprochen, können automatisierte Prozesse den Operateur durch Übernahme manueller oder kognitiver Aufgaben einerseits entlasten. Andererseits wird der Operateur durch neu entstandene Anforderungen wie die Überwachung der Automation erneut stark beansprucht. Aufgrund der in Zukunft stark automatisierten Umgebung rechnet Eißfeldt (1991) mit einem wachsenden Einfluss computerorientierter Kenntnisse, so dass es sinnvoll werde, die Einstellung der Bewerber bezüglich der Arbeit am Computer zu erfassen. Es sei wichtig, dass neue Technologien vom User akzeptiert werden. Diese Akzeptanz wiederum hängt von Persönlichkeitsmerkmalen der Benutzer ab (Beier, Spiekermann, & Rothensee, 2006).

Durch Automation verändern sich Kommunikationsmuster und Entscheidungsfindungsprozesse (Mosier & Skitka, 1996). Nach Broach (2005) kommt es zu einer Zunahme an Koor-

dination und geteilter Entscheidungsfindung. Im Projekt Consequences of Future ATM Systems for Air Traffic Controller Selection and Training (CAST) der Europäischen Union wird prognostiziert, dass sich im Allgemeinen die zukünftigen Anforderungen an die Persönlichkeit von Fluglotsen nicht stark von den derzeitigen unterscheiden werden. Allerdings werde das System ein Teil des Teams werden, und es werde Veränderungen hinsichtlich der persönlichen Beziehungen und des Teamworks geben. Faktoren wie Flexibilität, Teamwork, Denkstil, Anpassungsfähigkeit oder Gewissenhaftigkeit würden an Wichtigkeit zunehmen (CAST, 1999).

Hinsichtlich des Einflusses der Einstellung von Personen wurde vor allem das Thema Vertrauen (in Technik) beleuchtet. Nach Lee und See (2004) ist Vertrauen ein Beispiel des wichtigen Effektes von Affekten und Emotionen auf Mensch-Technik-Interaktion. Emotionale Reaktionen auf Technologien seien demnach nicht nur wichtig für die Akzeptanz einer Technologie. Auch könne damit fundamental zu Sicherheit und Leistung beigetragen werden. Vertrauenswürdige Technologien zu designen, werde daher ein kritischer Faktor für den Erfolg der nächsten Generationen von Automation und Computertechnologie sein.

Abgesehen von der aktuellen Forschung im HYBRID-Projekt (siehe Abschnitt 4.3) ist zum Verfassungszeitpunkt dieser Arbeit nur eine konkrete Studie zum Thema Persönlichkeit und Automation auffindbar. Szalma und Taylor (2011) untersuchten in einer ersten Studie zu individuellen Unterschieden in der Reaktion auf Automation den Zusammenhang der Big Five mit den abhängigen Variablen Leistung (Anteil korrekter Entscheidungen und Reaktionszeit), Übereinstimmung mit der Automation, empfundener Workload, Stress und Coping unter Verwendung eines simulierten Szenarios, in dem teil-, bzw. vollautomatisierte unbemannte Fahrzeuge zu überwachen und Bedrohungen zu identifizieren sind. Alle fünf Persönlichkeitsfaktoren zeigten mindestens indirekte Zusammenhänge mit den genannten Variablen und hingen mit wenigstens einem Workload- bzw. Stressparameter zusammen. Neurotizismus korrelierte mit höherem empfundenen Workload sowie Stress, die anderen Faktoren mit niedrigeren Werten in diesen Variablen. Höherer Neurotizismus hing zudem in manchen Bedingungen mit schlechterer Leistung zusammen. Andere Effekte variierten unter verschiedenen Bedingungen in ihrer Richtung.

Nachdem nun die Klärung der wichtigsten Oberbegriffe dieser Arbeit sowie die Erläuterung von Zusammenhängen von Persönlichkeit mit den relevanten Themengebieten erfolgt ist, wird im folgenden Kapitel 4 der Forschungskontext erläutert, in dem sich die aktuelle Arbeit bewegt.

4 Forschungskontext: Eignungsdiagnostik für Fluglotsen und Piloten

Um die Fragestellung dieser Arbeit (siehe Kapitel 5) einordnen zu können, wird in den folgenden Abschnitten sowohl die heutige Eignungsdiagnostik für Nachwuchsfluglotsen und –piloten sowie der Forschungsstand zu zukünftiger Eignungsdiagnostik beschrieben. Abgerundet wird dieser Themenbereich durch Beschreibungen zum Forschungsprojekt HYBRID, in dessen Rahmen die vorliegende Arbeit entstanden ist.

4.1 Eignungsdiagnostik heute

Die heutige Eignungsdiagnostik für die Personalauswahl von Nachwuchsfluglotsen für die Deutsche Flugsicherung GmbH (DFS) und Nachwuchspiloten für die Deutsche Lufthansa (DLH) im DLR wird im Folgenden überblicksartig dargestellt. Selbstverständlich werden weltweit auch von anderen Institutionen verschiedenste Auswahlverfahren für Flugsicherungen bzw. Airlines betrieben. Die Anforderungen an Luftfahrtoperateure sind jedoch vom Grundsatz her identisch. Die nach ISO 9001:2008 zertifizierten Auswahlprogramme im DLR werden daher exemplarisch für aktuelle Eignungsdiagnostik im Luftfahrtbereich präsentiert. Die beschriebenen Testverfahren sind ferner auch Grundlage für einen Teil der in dieser Arbeit verwendeten Leistungsvariablen.

Der Vollständigkeit halber wird in den Abschnitten 4.1.1 und 4.1.2 jeweils das komplette Auswahlverfahren präsentiert, obwohl nicht alle Einzelverfahren in spätere Berechnungen mit einbezogen werden (siehe 7.1.4). Die für die Analysen relevanten Auswahlbausteine sind jeweils gekennzeichnet.

4.1.1 Auswahl von Nachwuchsfluglotsen im DLR

Die Erläuterungen erfolgen in Anlehnung an Eschen-Léguedé (2007) und Nowitzke (2013). Wörtliche Zitate sind nicht separat gekennzeichnet, da es sich bei den Beschreibungen um allgemeine Informationen zum Ablauf des Verfahrens handelt, die nicht als geistiges Eigentum einzelner Autoren betrachtet werden können.

Die Abteilung für Luft- und Raumfahrtpsychologie des DLR führt die Auswahl von Auszubildenden für die Flugsicherung in Deutschland seit 1982 durch. Auftraggeberin war bis 1992 die Bundesanstalt für Flugsicherung, seit deren Privatisierung nunmehr die Deutsche Flugsicherung GmbH (DFS).

Die psychologische Eignungsuntersuchung soll die Prognose ermöglichen, ob ein Bewerber sowohl geeignet ist, die Ausbildung zum Fluglotsen zu bestehen, als auch die Voraussetzungen mitbringt, den stark fordernden Beruf des Fluglotsen über Jahre hinweg auszuüben. Ziel ist die Minimierung von Fällen des Leistungsversagens im Training, da diese zu deutlich erhöhten Ausbildungskosten führen und die Personalplanung der DFS erschweren.

In Anforderungsanalysen des DLR bezüglich der Ausübung des Berufs Fluglotse haben sich zahlreiche Anforderungsdimensionen herauskristallisiert. Als überdurchschnittlich wichtig zeigten sich kognitive Fähigkeiten, sensorische Fähigkeiten sowie interaktive und soziale Fähigkeiten (Eißfeldt, Goeters, Lorenz, Scholz, & Stelling, 1996). Um der Wichtigkeit letzterer gerecht zu werden, wurde 1996 erstmalig das DLR-Assessment-Center VerDi (Verhaltens-orientierte Diagnostik) für Fluglotsen-Bewerber eingesetzt. Ziel war und ist dabei „eignungsrelevante Aspekte der Persönlichkeit – grob gesprochen die Teamkompetenz – in konkreten Gruppensituationen zu erfassen.“ (Eißfeldt, Goeters, Lorenz, & Pecena, 1997).

Die psychologische Eignungsdiagnostik des DLR für Nachwuchsfluglotsen gliedert sich in zwei Abschnitte: eine Vorauswahl (VA) und eine Hauptuntersuchung (HU) der Bewerber. **Abbildung 4** auf der folgenden Seite zeigt eine Übersicht der zu absolvierenden Auswahlstufen.

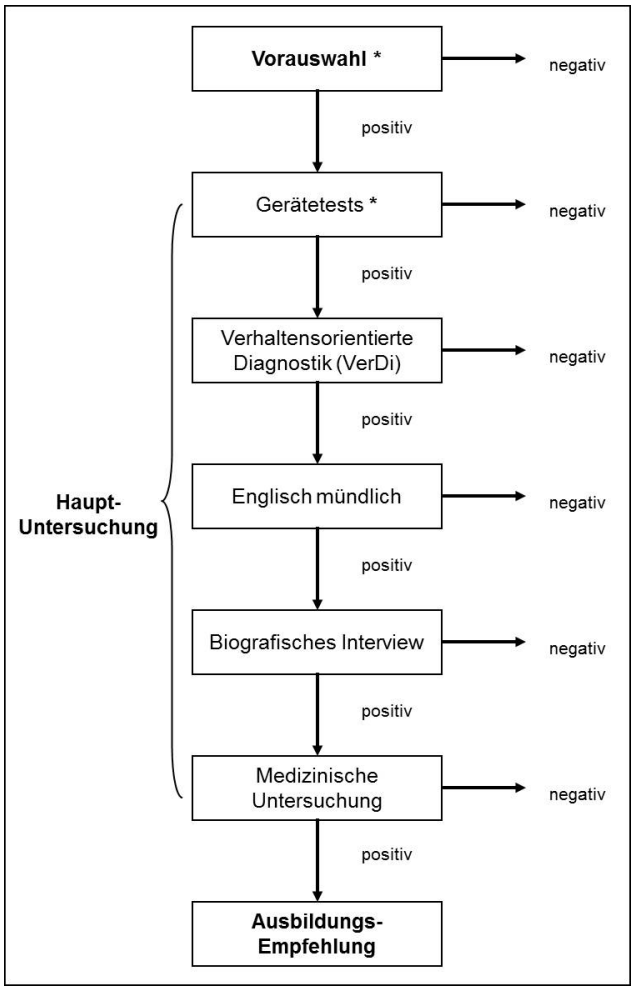


Abbildung 4: Stufen der Eignungsuntersuchung für Nachwuchsfluglotsen
(mit * gekennzeichnete Auswahlbausteine sind für die Analysen unter 7.1.5 relevant)

Beide Auswahlabschnitte enthalten mehrere Testverfahren bzw. Eignungsfeststellungen. Diese finden aufeinander folgend in zwei Teilen statt. Nach Durchführung der Vorauswahl und der Gerätetests reisen die Bewerber zunächst ab und im Falle positiver Ergebnisse zur Durchführung der restlichen Testverfahren erneut an. Nach jeder Stufe der Eignungsuntersuchung werden Bewerber, deren Leistungen nicht den Anforderungen entsprechen (negativ), vom weiteren Verfahren ausgeschlossen. Alle übrigen Probanden (positiv) nehmen an der nächsten Auswahlstufe teil (sequentielle Auswahlstrategie), sofern nicht die dafür vorgesehene Maximalzahl an Teilnehmern überschritten wird. In diesem Fall setzen überzählige Bewerber das Auswahlverfahren zu einem späteren Zeitpunkt fort.

Die Vorauswahl der Bewerber nimmt den kompletten ersten Tag der Auswahluntersuchung in Anspruch. Die dafür eingesetzte Testbatterie erfasst die Merkmalsbereiche Rechnen, Raumvorstellung, Aufmerksamkeit / Konzentration, visuelle und akustische Merkfähigkeit, Englisch, eine Arbeitsprobe in Form einer simulierten Anflugkontrolle, sowie einen Persönlichkeitstest. Alle Tests sind computergestützt und werden in Gruppen von maximal 50 Be-

werbern gleichzeitig bearbeitet. Eine ausführliche Beschreibung der einzelnen Testverfahren der Vorauswahl inklusive Literaturangaben findet sich z.B. bei Nowitzke (2013).

Die Hauptuntersuchung des Verfahrens beinhaltet, wie in **Abbildung 4** ersichtlich, Gerätetests, das Assessment Center VerDi, eine mündliche Englischprüfung, ein biografisches Interview und eine medizinische Untersuchung.

Die Gerätetests DAC (Dynamischer Airtraffic Control Test; Eißfeldt, 2002) und SDM (Strip Display Management Test; Grasshoff, 2001), prüfen die Bewerber als Arbeitsproben auf Mehrfacharbeitsfähigkeit, Problemlösefähigkeit und Entscheidungsverhalten in komplexen Situationen. Bestehen die Bewerber diese Auswahlstufe, werden sie in Gruppen von bis zu 12 Kandidaten zum Assessment-Center VerDi (Verhaltensorientierte Persönlichkeitsdiagnostik; Pecena, 2002) eingeladen. Eine Auswahlkommission, bestehend aus zwei Psychologen des DLR und zwei Fluglotsen der DFS beurteilt die Bewerber anhand von zwei AC-Übungen auf Verhaltensdimensionen wie z.B. Kooperation oder Engagement. Eine ausführlichere Beschreibung des Assessment Centers VerDi findet sich z.B. bei Eschen-Léguedé (2007).

Nach Bestehen des Assessment-Centers nehmen die Bewerber an einer mündlichen Englischprüfung und bei positiver Wertung an einem biografischen Interview teil. Im Anschluss an das Interview jeden Bewerbers berät sich die Auswahlkommission und gibt eine Laufbahnprognose für die Fluglotsenausbildung ab. Besteht der Bewerber diese letzte Auswahlstufe im DLR, steht noch eine externe flugmedizinische Untersuchung aus. Bei einem positiven Ergebnis erhält der Bewerber einen Ausbildungsvertrag.

4.1.2 Auswahl von Nachwuchspiloten im DLR

Die folgenden Erläuterungen erfolgen in Anlehnung an Nowitzke (2013) und Kalinina (2012). Wörtliche Zitate sind wiederum nicht separat gekennzeichnet, da es sich wie unter 4.1.1 auch hier um allgemeine Informationen zum Ablauf des Verfahrens handelt, die nicht als geistiges Eigentum einzelner Autoren betrachtet werden können.

Das Auswahlverfahren für Nachwuchspiloten wird in der Abteilung für Luft- und Raumfahrtpsychologie des DLR seit 1955 durchgeführt. Hauptauftraggeber ist die DLH. Auch hier soll die psychologische Eignungsuntersuchung die Prognose ermöglichen, ob ein Bewerber sowohl geeignet ist, die Ausbildung zum Piloten zu bestehen, als auch die Voraussetzungen mitbringt, den stark fordernden Beruf des Piloten über Jahre hinaus auszuüben.

In Anforderungsanalysen des DLR haben sich auch für den Pilotenberuf zahlreiche Anforderungsdimensionen herauskristallisiert. Als überdurchschnittlich wichtig zeigten sich kog-

nitive Fähigkeiten, psychomotorische Fähigkeiten, sensorische Fähigkeiten, interaktive und soziale Fähigkeiten sowie Kenntnisse und Fertigkeiten (z.B. das Lesen von Karten und Plänen), (Maschke & Goeters, 1999).

Die psychologische Eignungsdiagnostik des DLR für Nachwuchspiloten gliedert sich in zwei Abschnitte. Zunächst müssen die Bewerber die Berufsgrunduntersuchung (BU) absolvieren. Wird diese positiv bestanden, werden die Bewerber zur Firmenqualifikation (FQ) eingeladen. **Abbildung 5** zeigt eine Übersicht der zu absolvierenden Auswahlstufen.

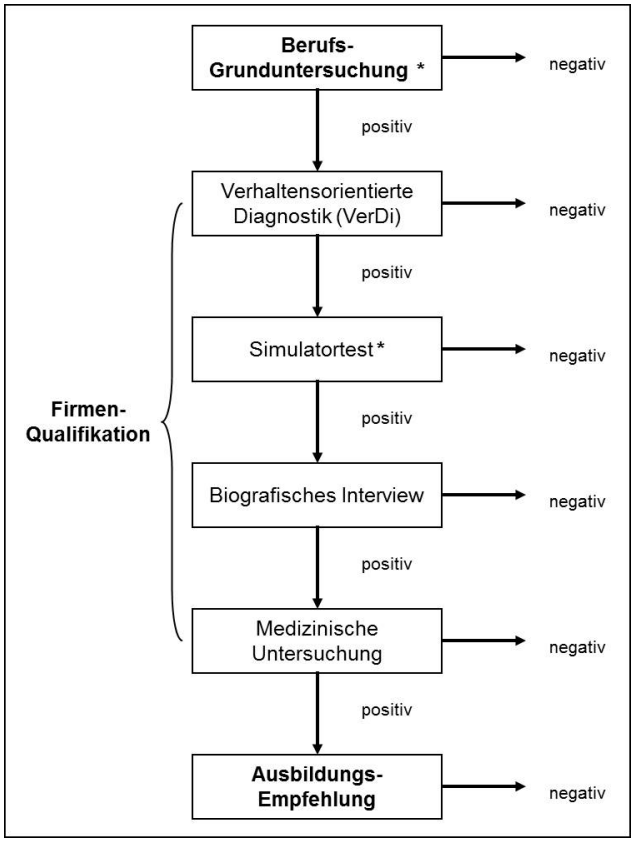


Abbildung 5: Stufen der Eignungsuntersuchung für Nachwuchspiloten
(mit * gekennzeichnete Auswahlbausteine sind für die Analysen unter 7.1.5 relevant)

Analog zur Fluglotsenauswahl enthalten beide Untersuchungsabschnitte mehrere Testverfahren bzw. Eignungsfeststellungen. Diese finden aufeinander folgend in zwei Teilen statt. Nach Durchführung der Berufsgrunduntersuchung reisen die Bewerber zunächst ab und im Falle positiver Ergebnisse zur Durchführung der restlichen Testverfahren erneut an. Auch in der Pilotenauswahl wird eine sequentielle Auswahlstrategie verfolgt: nach jeder Stufe der Eignungsuntersuchung werden Bewerber, deren Leistungen nicht den Anforderungen entsprechen (negativ), vom weiteren Verfahren ausgeschlossen. Alle übrigen Probanden (positiv) nehmen an der nächsten Auswahlstufe teil.

Die Berufsgrunduntersuchung (BU) dauert einen vollen Tag und wird mit bis zu 48 Bewerbern computergestützt durchgeführt. Erfasst werden die Merkmalsbereiche Englisch, Mathematik- und Technikkenntnisse, visuelle und akustische Merkfähigkeit, räumliche Orientierung, Aufmerksamkeit, Konzentration sowie Mehrfacharbeit und Psychomotorik. Zusätzlich werden berufsrelevante Persönlichkeitsmerkmale erhoben.

Die Firmenqualifikation für Nachwuchspiloten findet bei einem positiven BU-Ergebnis einige Wochen später in Bewerbergruppen von acht bis zehn Personen statt. Die erste Stufe der Firmenqualifikation stellt das Assessment-Center VerDi dar. Analog zur Fluglotsenauswahl wird auch bei der Pilotenauswahl die Leistung der Bewerber von einer Auswahlkommission beurteilt. Diese besteht in diesem Fall aus zwei Psychologen des DLR sowie einem Flugkapitän der DLH. Das Assessment-Center für die Nachwuchspilotenauswahl besteht aus mehreren Gruppenübungen, mit denen u.a. die soziale Kompetenz, die Fähigkeit zur Teamarbeit und die Fähigkeit zur gemeinsamen Handlungsplanung beurteilt werden. Bestehen die Bewerber diese Auswahlstufe, bearbeiten sie am folgenden Tag zunächst den Simulatortest (Soll, 2012). In dieser Arbeitsprobe müssen vorgegebene Flugrouten nachgeflogen werden. Beurteilt werden die Bewerber von speziell ausgebildeten Testleitern mit Pilotenlizenz. Fällt das Ergebnis des Simulatortests positiv aus, nehmen die Bewerber im Anschluss an einem biografischen Interview teil. Wie auch bei der Fluglotsenauswahl berät sich die Auswahlkommission im Anschluss an das Interview jeden Bewerbers und gibt eine Laufbahnprognose für die Ausbildung ab. Besteht ein Bewerber auch diese letzte Auswahlstufe im DLR, steht noch eine externe flugmedizinische Untersuchung aus. Bei einem positiven Ergebnis erhält der Bewerber einen Ausbildungsvertrag.

Wie bereits erwähnt, ist anzunehmen, dass eine Veränderung der Berufsbilder Pilot und Fluglotse auch mit veränderten Anforderungen an die Operateure einhergeht. Der nächste Abschnitt 4.2 geht daher, wie angekündigt, auf den aktuellen Forschungsstand zu zukünftiger Eignungsdiagnostik ein.

4.2 Eignungsdiagnostik morgen

Eißfeldt (2008) weist darauf hin, dass bisher nur wenige Studien die Frage nach zukünftigen Auswahlkriterien in Luftverkehrssystemen behandeln. Veröffentlichungen aus den neunziger Jahren gehen teilweise noch davon aus, dass keine veränderten Anforderungen an Luftfahrt-operateure zu erwarten seien. Manning und Broach (1992) begründen dies damit, dass auch in zukünftigen Luftverkehrssystemen bei Fehlfunktionen automatischer Routinen nach wie vor eine volle Verantwortungsübernahme durch den Menschen vorliegt. Im Projekt CAST wird jedoch darauf hingewiesen, dass angesichts von stärker passiv überwachenden

Rollen zukünftiger Operateure zentrale Fertigkeiten weniger gut geübt sein werden und daher sogar höhere Anforderungen an zugrunde liegende Fähigkeiten absehbar seien. Das Lösen von Problemen werde seltener nötig, aber *wenn,* dann umso schwieriger sein. Wie schon unter 3.3 erwähnt, wird zudem prognostiziert, dass der Bereich der Persönlichkeit für die Arbeit mit zukünftigen ATM-Systemen an Relevanz gewinnt. In den CAST-Studien stellten sich als Anforderungsdimensionen solche Skalen als relevant für die Zukunft heraus, die Charakteristiken wie Flexibilität, Teamarbeit, Denkstil, Anpassungsfähigkeit und Gewissenhaftigkeit betreffen (CAST, 1999).

Im DLR-Projekt Aviator 2030 wurden zum Thema zukünftiger Anforderungen an Luftfahrtoperateure unter anderem Workshops mit aktuell tätigen Fluglotsen und Piloten durchgeführt. Diese bewerteten im Rahmen einer Anforderungsanalyse heutige und antizipierte zukünftige Anforderungen an ihre Berufsgruppen mittels des Fleishman Job Analysis Survey (F-JAS, Fleishman, 1995). Hinsichtlich kognitiver Anforderungen zeigten sich einzelne, aber keine größeren Veränderungen. Im Bereich sozialer / interaktiver Fähigkeiten jedoch wurde von beiden Berufsgruppen ein genereller Anstieg prognostiziert. 11 von 18 bewerteten Fähigkeiten stiegen in ihrer Relevanz an, so z.B. Entscheidungsfindung, Situationsbewusstsein, Kommunikation, Kooperation und Selbstsicherheit (Eißfeldt et al., 2009). In einer anschließenden experimentellen Zukunftssimulationsstudie mit aktuellen Berufsinhabern ergaben sich bei folgenden F-JAS-Skalen signifikante Unterschiede hinsichtlich heutiger und zukünftiger Anforderungen (**Tabelle 4**, siehe Folgeseite).

Tabelle 4: Heutige (Baseline) vs. zukünftige (Future) Anforderungen an Fluglotsen und Piloten (nach Eißfeldt et al., 2009)

F-JAS Scales	ATC			Pilots		
	Baseline	Future	p	Baseline	Future	p
Oral Comprehension	4.8	4.1	0.08			
Written Expression	3.0	3.8	0.04			
Fluency of Ideas				3.2	3.9	0.10
Originality	4.6	4.0	0.07	2.5	3.8	0.00
Inductive Reasoning				3.2	4.0	0.01
Information Ordering				4.0	4.6	0.01
Speed of Closure	6.2	5.1	0.10			
Flexibility of Closure				3.5	4.4	0.01
Selective Attention				3.9	4.5	0.01
Situation Awareness				5.0	5.6	0.01
Impulse Control				3.4	4.1	0.03

Wie Eignungsdiagnostik „morgen“ konkret ablaufen wird, kann heute nur vermutet werden. Viele der heutigen Anforderungen werden den genannten Forschungsergebnissen nach weiterhin relevant sein (Eißfeldt et al., 2009), Überwachungsleistung wird mit größter Wahrscheinlichkeit einen höheren Stellenwert als heute einnehmen (Parasuraman, Mouloua, & Molloy, 1996; CAST, 1999) und wird z.B. als Operational Monitoring in dem DLR-Projekt Aviator II (Hasse, Grasshoff, & Bruder, 2012; Bruder, Eißfeldt, Maschke, & Hasse, 2013) aktuell weiter beforscht. Erkenntnisse darüber, welchen Beitrag Persönlichkeit und Einstellung zur Eignung von Bewerbern als Luftfahrtoperateure der Zukunft leisten, sollen mit dieser Arbeit voran gebracht werden, die im Rahmen des Projekts HYBRID angefertigt wird. Der folgende Abschnitt 4.3 gibt daher einen Kurzüberblick über das Projekt.

4.3 Das Forschungsprojekt HYBRID

Das DLR-Projekt "HYBRID: Entscheidungsverhalten in hybriden Teams" befasst sich seit April 2009 mit zukünftiger Teamarbeit zwischen menschlichen und maschinellen Teampartnern (siehe dazu auch 2.3).

Abbildung 6: Logo des HYBRID-Projekts
©DLR

Hintergrund ist der Wunsch nach rechtzeitiger Berücksichtigung von sich ändernden Fähigkeitsanforderungen an Luftfahrtoperateure, die durch zunehmende Automation zu erwarten sind (siehe 2.4). So ist davon auszugehen, dass vermehrt enge Zusammenarbeit von Menschen mit einem hoch automatisierten Partner erforderlich sein wird (siehe 2.3). Nach heutigen Annahmen könnte die Zukunft der Luftfahrt daher weg von klassischer Teamarbeit, hin zur Arbeit im hybriden Team führen. Durch solch eine Verschiebung der Tätigkeit ist für zukünftige Luftfahrtoperateure im Vergleich zu heutigen gegebenenfalls ein verändertes Anforderungsprofil gefragt. Dies kann neben Fähigkeitsanforderungen auch den Bereich der Persönlichkeit betreffen. In dem über fünf Jahre angelegten HYBRID-Projekt wird daher insbesondere der Einfluss von Aspekten der Persönlichkeit und der Einstellung auf die Leistungsfähigkeit in hybriden Teams erforscht. Wenn sich bestätigt, dass bestimmte Facetten der Persönlichkeit und/oder Einstellungen in der Lage sind, die Leistungsfähigkeit bei der Arbeit in einem hoch automatisierten System vorherzusagen, sollten diese Aspekte in die Eignungsdiagnostik der Zukunft eingebunden werden. Das Projekt HYBRID „berührt damit verschiedene Stränge aktueller luftfahrtpsychologischer Forschung wie z.B. Teamfor-

schung, Persönlichkeitspsychologie, kognitive Automationsforschung und Differentielle Psychologie. Die Verankerung auf der Schnittstelle dieser Teildisziplinen unterstreicht den innovativen Charakter des Projekts. Vergleichbare Vorläufer des Vorhabens lassen sich nicht identifizieren.“ (Eißfeldt, 2008, S. 4). Das Projekt wird im Rahmen der Spitzenclusterförderung des Luftfahrtclusters der Metropolregion Hamburg vom Bundesministerium für Bildung und Forschung gefördert und mit Unterstützung der DFS Deutsche Flugsicherung GmbH und der DLH Deutsche Lufthansa AG durchgeführt.

Nachdem mittels theoretischer Einbettung nun der Rahmen der vorliegenden Arbeit gegeben ist, beschäftigt sich der jetzt anschließende empirische Teil, wie angekündigt, mit der Fragestellung (Kapitel 5), den Hypothesen (Kapitel 6), der Methode (Kapitel 7) und den Ergebnissen (Kapitel 8) der Arbeit, die zu guter Letzt noch diskutiert werden (Kapitel 9).

B. Empirischer Teil

5 Fragestellung

Wie im Titel verdeutlicht, beschäftigt sich die vorliegende Arbeit mit Persönlichkeit als Prädiktor für Leistung in hoch automatisierten Mensch-Maschine-Teams der Luftfahrt. Wie bereits in den vorangegangenen Kapiteln beschrieben, muss langfristig abgeklärt werden, ob durch die Zunahme von Arbeit in Mensch-Maschine-Teams im Bereich der Luftfahrt, aber auch in anderen Domänen, Tätigkeiten soweit verändert werden, dass im Vergleich zu heute auch veränderte Fähigkeits- und Persönlichkeitsprofile gefordert sind. Aufgabe der Forschung ist, herauszufinden, ob in der Persönlichkeit begründete Faktoren die Leistungsentfaltung in Mensch-Maschine-Teams beeinflussen. Kann z.B. ein in der Persönlichkeit angelegtes Bedürfnis nach sozialer Interaktion bei der Arbeit, d.h. eine ausgeprägte Teamorientierung hinsichtlich menschlicher Teampartner, eventuell hinderlich für die Bewältigung von Aufgaben im Mensch-Maschine-Team sein?

Bisherige Routinen in der Luftfahrt sehen stets Teamarbeit und ein Vier-Augen-Prinzip vor (z.B. Pilot & Co-Pilot oder Radar- & Koordinationslotse). In die Eignungsdiagnostik für diese Berufe werden dahingehende Kompetenzen einbezogen und können ebenso wie mentale Fähigkeiten ausschlaggebend für die Entscheidung für oder gegen einen Kandidaten sein (Eißfeldt, Goeters, Lorenz, & Pecena, 1997; Pecena, 2002). Können daher Kandidaten, die heute als Luftfahrtoperateure unter Einbeziehung von Teamkompetenzen ausgewählt werden, auch ohne menschlichen Teampartner die nötige Leistung erbringen? Es ist anzunehmen, dass die Arbeit im Mensch-Maschine-Team voraussetzt, selbstverständlich und ohne Scheu mit der vorhanden Technik und Automation – eben mit dem maschinellen Teampartner – umzugehen und auch darauf zu vertrauen, dass man selbst in dieser Zusammenarbeit etwas bewirken kann. Sind also z.B. Computerkompetenz, Technikaffinität oder auch Kontrollüberzeugungen im Umgang mit Technik valide Prädiktoren für erfolgreiche Arbeit im Mensch-Maschine-Team?

Ziel dieser Arbeit ist die Entwicklung eines Forschungsfragebogens zu Persönlichkeits- und Einstellungsaspekten, die für die Arbeit im Mensch-Maschine-Team eine Rolle spielen. Für den Fragebogen sollen geeignete Skalen gefunden, erprobt und validiert werden. Damit entsteht erstmalig ein spezifischer Persönlichkeits- und Einstellungsfragebogen, der sich mit relevanten Aspekten für die Arbeit mit hoch automatisierten Systemen befasst. Voraussichtliche Einsatzmöglichkeiten bestehen in der Human-Factors- und Automationsforschung sowie auch ganz konkret in zukünftiger Eignungsdiagnostik für Luftfahrtoperateure.

6 Hypothesen

Die im Folgenden aufgestellten Hypothesen nehmen Bezug auf den Forschungsfragebogen Hybrid-Team-Questionnaire (HTQ), der in den anschließenden Kapiteln ausführlich beschrieben wird. Der Konzeption des Fragebogens liegt die Annahme zugrunde, dass die ausgewählten Skalen in der Lage sind, Leistungsaspekte vorherzusagen. Es wird daher davon ausgegangen, dass bereits bei heutigen Verfahren der Eignungsdiagnostik für Fluglotsen und Piloten Zusammenhänge mit den HTQ-Skalen vorhanden sind (Hypothese 1). Vor allem aber werden Zusammenhänge von den Skalen des HTQ mit der Leistung im Mensch-Maschine-Team postuliert (Hypothese 2). Weiterhin wird davon ausgegangen, dass der HTQ zwischen besseren und schlechteren Operateuren differenzieren kann (Hypothese 3). Die Relevanz des Fragebogens ergibt sich aus der Annahme, dass Persönlichkeit und Einstellung über kognitive Leistung hinaus einen inkrementellen Beitrag zur Eignungsfeststellung bei zukünftigen Luftfahrtoperateuren leisten (Hypothese 4).

Hypothese 1: *Es bestehen Zusammenhänge zwischen den Skalen zur Persönlichkeit und Einstellung des im Rahmen der Dissertation konzipierten Forschungsfragebogens HTQ und den Leistungen von Bewerbern in der heutigen Eignungsdiagnostik für Fluglotsen und Piloten:*

$H_1: \rho \neq 0$ $H_0: \rho = 0$

Hypothese 2: *Es bestehen Zusammenhänge zwischen den Skalen zur Persönlichkeit und Einstellung des im Rahmen der Dissertation konzipierten Forschungsfragebogens HTQ und den Leistungen menschlicher Operateure im Mensch-Maschine-Team:*

$H_1: \rho \neq 0$ $H_0: \rho = 0$

Hypothese 3*: Probanden mit besseren Leistungen im Mensch-Maschine-Team unterscheiden sich in den Skalen des HTQ signifikant von Probanden mit schlechteren Leistungen im Mensch-Maschine-Team:*

$H_1: \mu_1 \neq \mu_2$ $H_0: \mu_1 = \mu_2$

Hypothese 4*: Die Skalen des HTQ klären gegenüber kognitiven Variablen inkrementelle Validität zur Vorhersage der Leistung im Mensch-Maschine-Team auf:*

$H_1: R^2_{kog_HTQ} > R^2_{kog}$ $H_0: R^2_{kog_HTQ} = R^2_{kog}$

7 Methode

Um die oben beschriebenen Hypothesen zu prüfen, waren mehrere aufeinander aufbauende Schritte und Studien notwendig. Nachdem geeignete Konstrukte und Skalen zu deren Messung ausgewählt waren (7.1.1) wurden diese in Form des HTQ-1 einer ersten Erprobung unterzogen (7.1.2). Deren Ergebnisse flossen in die Konzeption des HTQ-2 ein, der ebenfalls einer Erprobung unterzogen wurde (7.1.3). Die Erprobungsergebnisse zum HTQ-2 erlaubten, den Fragebogen für die Prüfung von Hypothese 2 zu Zusammenhängen mit aktueller Eignungsdiagnostik einzusetzen. Dafür erfolgte zunächst die Bildung entsprechender Variablen aus den Auswahlverfahren (7.1.4). Im Anschluss wurden korrelative Zusammenhänge zwischen HTQ-2 und den Variablen des Auswahlverfahrens für Nachwuchsfluglotsen und –piloten analysiert (7.1.5). Im Laufe des Forschungsprojekts wurde zudem der Persönlichkeitsbereich Neigung zu risikoreichem Entscheidungsverhalten als möglicherweise zusätzlich wichtiger Prädiktor für Leistung in Mensch-Maschine-Teams erachtet. Dieser wird in Ergänzung zu den Selbstbeschreibungsskalen des HTQ-2 mittels objektiver Persönlichkeitstests erhoben und komplettiert den Fragebogen zum HTQ-S (7.1.6).

Um Zusammenhänge zwischen HTQ-S und Leistung im Mensch-Maschine-Team zu erheben, wurde die Entwicklung eines Verfahrens notwendig, das zukünftige Anforderungen an Luftfahrtoperateure simuliert. Ergebnis ist die ATM-Simulation HINT (Hybrid Interaction Simulation, 7.2), die zunächst beschrieben wird (7.2.1), um im Anschluss daran Variablen für die geplanten Analysen abzuleiten (7.2.2).

In einer letzten Studie (7.3) wurden sowohl HTQ-S als auch HINT-Daten erhoben, mittels derer die verbleibenden Hypothesen geprüft werden können: korrelative Zusammenhänge entsprechend Hypothese 1 (7.3.2), Gruppenvergleiche hinsichtlich signifikanter Unterschiede im HTQ-S entsprechend Hypothese 3 (7.3.3), inkrementelle Validität des Fragebogens entsprechend Hypothese 4 (7.3.4). Alle statistischen Auswertungen erfolgten mit der Software SPSS™ 18. **Abbildung 7** auf der nächsten Seite liefert eine Übersicht über die drei vorhandenen Kategorien an Variablen und die in diesem Kapitel beschriebenen Analysen.

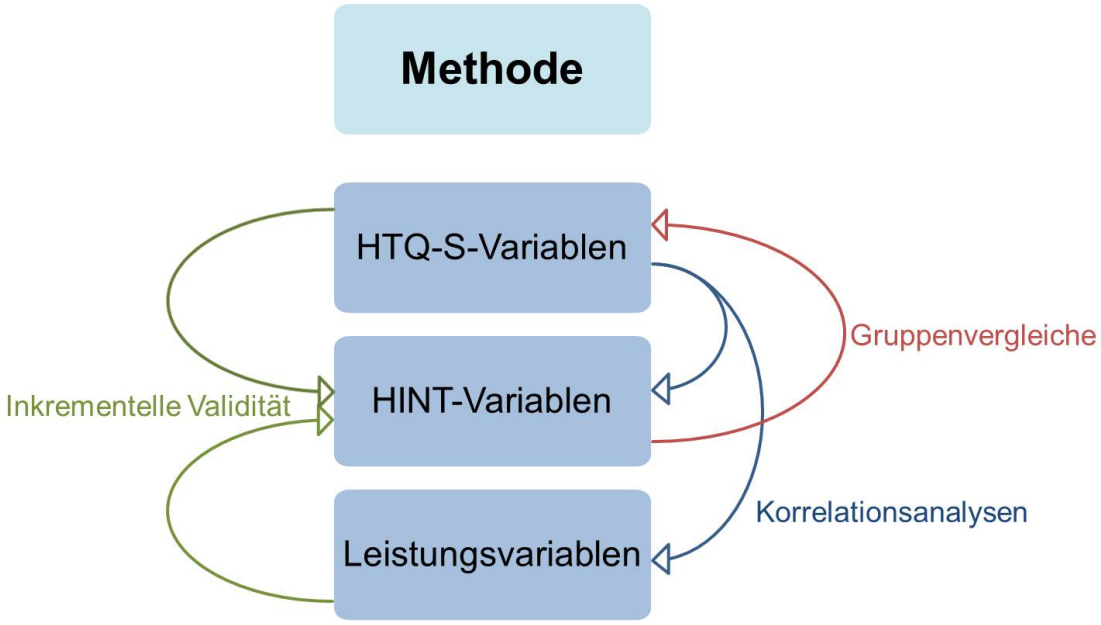


Abbildung 7: Übersicht über die angewandte Methode

7.1 Entwicklung des Hybrid-Team-Questionnaire

Die Entwicklung des Fragebogens Hybrid-Team-Questionnaire (HTQ) erfolgte in vier Schritten. Zunächst wurden geeignete Konstrukte und passende Skalen zu deren Messung ausgewählt (siehe Abschnitt 7.1.1). Die identifizierten Skalen wurden in einer ersten Version des HTQ (im Folgenden als HTQ-1 bezeichnet) einer Erprobung unterzogen, um die psychometrische Qualität sowie die Handhabbarkeit der Skalen zu prüfen. Abschnitt 7.1.2 beschreibt das Vorgehen. Anhand der Ergebnisse kann zudem festgestellt werden, ob die Skalen des HTQ in der betrachteten Population von Luftfahrtbewerbern genauso funktionieren wie in der Allgemeinbevölkerung. Da die Stichprobe ausschließlich aus Bewerbern mit Interesse am Beruf eines Piloten oder Fluglotsen besteht, ist eine Selektion der Stichprobe in Bezug auf einige Eigenschaften möglich. Diese Tatsache kann mit einer Varianzeinschränkung und Ungeeignetheit von Tests einhergehen. Ausgehend von den Erprobungsergebnissen wurde der Fragebogen modifiziert und in seiner zweiten Version (im Folgenden als HTQ-2 bezeichnet) erneut einer Erprobung unterzogen, die in Abschnitt 7.1.3 beschrieben wird. Der letzte Schritt beinhaltete die Ergänzung des Fragebogens mit objektiven Persönlichkeitstests zum HTQ-S (HTQ-Supplemented). Abschnitt 7.1.6 geht genauer darauf ein. Alle Versionen des HTQ wurden computerbasiert durchgeführt. Die vorläufigen Versionen des Fragebogens HTQ-1 und HTQ-2 wurden über das im DLR eingesetzte Testsystem CAT

(Computer Assisted Testing), die endgültige Version des HTQ-S auf der Testplattform Inquisit ™ vorgegeben.

7.1.1 Auswahl geeigneter Konstrukte und Skalen

Mittels Literaturrecherche wurden verschiedenste Aspekte der Persönlichkeit und Einstellung betrachtet, um Items bzw. Skalen für den Fragebogen HTQ zusammenzustellen. Hierbei wurden sowohl etablierte Persönlichkeitstests, wie z.B. NEO-PI-R (Ostendorf & Angleitner, 2004) herangezogen, als auch spezifischere Skalen, die eine Rolle hinsichtlich der Eignung für Mensch-Maschine-Teams spielen könnten. Grundlage hierfür sind die im Kapitel 3 beschriebenen Forschungsergebnisse zu Persönlichkeitsvariablen als Einflussfaktoren auf Leistung, Mensch-Maschine-Interaktion und Automation.

7.1.2 Erprobung des HTQ-1

Die erste Erprobung des HTQ (HTQ-1) wird bereits in der im Rahmen des HYBRID-Projekts entstandenen Diplomarbeit „Messung von Persönlichkeitsaspekten mit Einfluss auf die Leistungsentfaltung in hochautomatisierten Systemen – Erprobung eines Forschungsfragebogens“ von (Knappe, 2010) beschrieben. An dieser Stelle wird ein Überblick über das methodische Vorgehen gegeben, wie er ähnlich in der genannten Diplomarbeit zu finden ist. Wörtliche Zitate sind in diesem Fall nicht gesondert gekennzeichnet.

An der Erprobung des HTQ-1 nahmen 198 Bewerber für den Pilotenberuf bei der DLH und 205 Bewerber für den Fluglotsenberuf bei der DFS teil. Die Gesamtstichprobe beinhaltet somit 403 Probanden. 33% der Probanden sind weiblich. Der Altersdurchschnitt der Probanden liegt bei 20,23 Jahren (SD = 1.99). Alle weiteren erhobenen biografischen Merkmale der Teilnehmer (z.B. Nationalität, Muttersprache, Familienstand…) sind von sehr geringer Variabilität gekennzeichnet, so dass sie in den weiteren Analysen nicht berücksichtigt werden.

Die Studien fanden im Rahmen von Eignungsuntersuchungen des Zeitraums Mai bis Juli 2010 in der Abteilung für Luft- und Raumfahrtpsychologie des DLR in Hamburg statt. Die Bewerber wurden gebeten, sich nach der ersten Stufe der Eignungsuntersuchung an der Erprobung von Forschungsfragebögen zum Thema „Anforderungen an zukünftige Luftfahrtoperateure“ zu beteiligen. Die Teilnahme war freiwillig. Im Erhebungszeitraum stimmten bis auf einen Bewerber alle zu. Die Testung erfolgte in Gruppen von maximal 45 Personen und dauerte zwischen 90 und 120 Minuten. Die Erprobung wurde im DLR-Testsaal durchgeführt, in dem zuvor auch die Eignungsuntersuchung stattfand. Die Eingaben erfolgten über

Touchscreen. Die Computervorgabe erlaubte keine Auslassung von Fragen, so dass alle Teilnehmer auf jede Frage antworteten.

Hintergrund für die Überprüfung der Skalen ist, dass sich Luftfahrtoperateure hinsichtlich einiger Persönlichkeitsfaktoren von anderen Stichproben unterscheiden (Maschke, 2005). Da die Stichprobe des HTQ-1 ausschließlich aus Bewerbern für Luftfahrtberufe besteht, ist es möglich, dass diese in Bezug auf einige Eigenschaften vorselektiert sind (z.B. Technikaffinität). Dies kann mit einer Varianzeinschränkung und Ungeeignetheit von Tests einhergehen. Zur Kontrolle wurden für alle Skalen die psychometrischen Eigenschaften der Items sowie die Reliabilität untersucht.

Zur statistischen Auswertung des HTQ-1 wurde eine Itemanalyse durchgeführt, um zu prüfen, welche Items zwischen Teilnehmern differenzieren können und welche ggf. nicht. Berechnet wurden Standardabweichung, Range, Itemschwierigkeit, Trennschärfe sowie als Schätzer der Reliabilität die interne Konsistenz (Cronbachs α) der Komplettskalen sowie die interne Konsistenz wenn ein betrachtetes Item weggelassen wird.

Für die verwendeten Persönlichkeitsskalen wird wie bei den meisten Erhebungsinstrumenten Intervallskalenniveau angenommen (Bortz, 2005). Dies ermöglicht den Einsatz von statistischen Verfahren für intervallskalierte Daten wie z.B. die Berechnung von Mittelwerten.

Nicht komplett überprüft wurde das eingesetzte Verfahren NEO-PI-R (siehe 8.1.1). Da das Inventar das derzeit international am häufigsten eingesetzte Verfahren zur Messung des Fünf-Faktoren-Modells der Persönlichkeit ist und die Struktur und Validität der deutschen Version in mehreren Studien mit über 10.000 Probanden nachgewiesen werden konnte (Ostendorf & Angleitner, 2004), fand die Überprüfung der psychometrischen Eigenschaften der Items in der vorliegenden Untersuchung keine Berücksichtigung. Es wird davon ausgegangen, dass sich die aktuelle Stichprobe hinsichtlich der Variablen des NEO-PI-R von der Normalbevölkerung nicht unterscheidet. Zum Vergleich mit den von den Testautoren berichteten Werten werden jedoch die internen Konsistenzen dargestellt (**Tabelle 9**).

Gemäß Bühner (2006) wird zur Bestimmung der Itemschwierigkeit von Ratingskalen der Itemmittelwert berechnet. Items mit hohen Mittelwerten besitzen eine niedrige Schwierigkeit, da diese in Leistungstests „leicht“ richtig zu beantworten sind bzw. ihnen in Persönlichkeitsfragebögen „leicht“ zugestimmt wird. Im Umkehrschluss besitzen Items mit niedrigen Mittelwerten eine hohe Schwierigkeit. Besonders wünschenswert sind Skalen mit Items mittlerer Schwierigkeit, da für diese die Wahrscheinlichkeit einer hohen Trennschärfe und somit einer hohen Reliabilität ansteigt. Um auch im Randbereich differenzieren zu können, sind aber auch Items mit extremen Werten von Nutzen (Bühner, 2006).

Als weiterer Bestandteil der Itemanalyse wurde die Trennschärfe mit Part-Whole-Korrektur bestimmt. Diese erlaubt Aussagen darüber, wie gut ein Item inhaltlich alle anderen Items der Skala widerspiegelt (Bühner, 2006). Schwierigkeits-Trennschärfe-Diagramme stellen die Verteilung der Schwierigkeits- und Trennschärfeindizes grafisch dar und werden bei Knappe (2010) berichtet.

Des Weiteren erfolgte die Berechnung von Streuung und Spannweite. Diese Maße geben Aufschluss darüber, wie sehr sich die Probanden in der Beantwortung der Items unterscheiden und inwieweit die Antwortkategorien ausgenutzt wurden. Hohe Itemstreuungen begünstigen hohe Trennschärfen (Bühner, 2006).

Die interne Konsistenz (berechnet mittels Cronbachs α) wurde für jede Skala betrachtet. Bei der Interpretation der Werte ist zu beachten, dass einige Skalen lediglich aus wenigen Items bestehen und somit die Wahrscheinlichkeit hoch ist, dass die Reliabilität hier geringer ausfällt als bei Skalen mit hoher Itemanzahl. Ebenfalls berechnet wurde Cronbachs α, wenn das zu betrachtende Item weggelassen wird. Dies ermöglicht eine Aussage darüber, ob die interne Konsistenz der Skala steigt, wenn das Item entfernt wird.

7.1.3 Erprobung des HTQ-2

Die Erprobung des HTQ-2 wird bereits anteilig in der im Rahmen des HYBRID-Projektes entstandenen Diplomarbeit „Einflüsse von technikbezogenen Persönlichkeitskonstrukten und genereller Selbstwirksamkeitserwartung auf die Leistungsfähigkeit in computerbasierten Testverfahren“ von Nowitzke (2013) beschrieben. An dieser Stelle wird ein Überblick über das methodische Vorgehen gegeben, wie er ähnlich in der genannten Diplomarbeit zu finden ist. Auch hier sind wörtliche Zitate nicht gesondert gekennzeichnet.

An der Erprobung des HTQ-2 nahmen 196 Bewerber für den Pilotenberuf bei der DLH und 214 Bewerber für den Fluglotsenberuf bei der DFS teil. Eine Person wurde aufgrund fehlender Werte wegen verspäteter Ankunft von den Analysen ausgeschlossen. Die Gesamtstichprobe beinhaltet somit 410 Probanden. 23,9 % der Probanden sind weiblich. Der Altersdurchschnitt der Probanden liegt bei 19,60 Jahren (SD = 1.83). Erneut werden alle weiteren erhobenen biografischen Merkmale der Teilnehmer (z.B. Nationalität, Muttersprache, Familienstand…) in den weiteren Analysen nicht berücksichtigt, da sie von zu geringer Variabilität gekennzeichnet sind.

Die Untersuchungen fanden im Rahmen von Eignungsuntersuchungen des Zeitraums Februar bis März 2011 in der Abteilung Luft- und Raumfahrtpsychologie des DLR in Hamburg statt. Erneut wurden die Bewerber gebeten, sich nach der ersten Stufe der Eignungsuntersu-

chung an der Erprobung von Forschungsfragebögen zum Thema „Anforderungen an zukünftige Luftfahrtoperateure“ zu beteiligen. Die Teilnahme war freiwillig, alle Bewerber stimmten einer Beteiligung zu. Die Testung erfolgte in Gruppen von maximal 45 Personen und dauerte im Vergleich zum HTQ-1 aufgrund einer geringeren Itemzahl nur ca. 45 Minuten. Wie gehabt wurde die Erprobung im DLR-Testsaal durchgeführt, in dem zuvor auch die Eignungsuntersuchung stattfand. Die Eingaben erfolgten über Touchscreen. Die Computervorgabe erlaubte keine Auslassung von Fragen, so dass alle Teilnehmer auf jede Frage antworteten.

Der Datensatz des HTQ-2 wurde entsprechend der im vorangegangenen Abschnitt 7.1.2) geschilderten Analysen hinsichtlich der psychometrischen Eigenschaften der Items sowie der Reliabilität der Skalen untersucht.

Wie bereits unter 7 erwähnt, erlaubten die Erprobungsergebnisse zum HTQ-2 (siehe 8.1.4), den Fragebogen für die Prüfung von Hypothese 2 zu Zusammenhängen mit aktueller Eignungsdiagnostik einzusetzen. Dafür wurde die Bildung entsprechender Variablen aus den Auswahlverfahren nötig, die im folgenden Abschnitt geschildert wird.

7.1.4 Variablen aus der Eignungsdiagnostik

Wie unter 4.1 beschrieben, bestehen die eignungsdiagnostischen Untersuchungen für Nachwuchsfluglotsen und -piloten aus jeweils mehreren Stufen. Die erste Stufe (VA bzw. BU) enthält dabei Testverfahren zu verschiedenen Merkmalsbereichen. Da eine Auswertung der Daten auf der Ebene aller einzelnen Leistungstests zu detailliert wäre, werden Variablen zu den erhobenen Merkmalsbereichen gebildet. Zusätzlich zu den Basisleistungen aus VA und BU stehen die Ergebnisse aus den Gerätetests bzw. dem Simulatortest als Variablen zur Verfügung. Die daran anschließenden Auswahlstufen werden nicht berücksichtigt, da diese keine Komponenten der Mensch-Maschine-Interaktion enthalten. **Tabelle 5** auf der nächsten Seite gibt einen Überblick über die Variablen und ihre Bedeutung. Für alle Variablen gilt, dass ein höherer Wert bessere Leistung darstellt.

Tabelle 5: Variablen aus der Eignungsdiagnostik und ihre Bedeutung

Variable	*Bedeutung*
Mittlere Testleistung (VA)	Mittelwert aus allen Leistungstests der VA
Teilbereich Konzentration (VA)	Mittelwert aus allen Konzentrationstests der VA
Teilbereich Aufmerksamkeit (VA)	Mittelwert aus allen Aufmerksamkeitstests der VA
Teilbereich Merkfähigkeit (VA)	Mittelwert aus allen Merkfähigkeitstests der VA
Teilbereich Raumorientierung (VA)	Mittelwert aus allen Raumorientierungstests der VA
Mittlere Testleistung (BU)	Mittelwert aus allen Leistungstests der BU
Teilbereich Technikwissen (BU)	Mittelwert aus allen Technikwissenstests der BU
Teilbereich Mathematik (BU)	Mittelwert aus allen Mathematiktests der BU
Teilbereich Merkfähigkeit (BU)	Mittelwert aus allen Merkfähigkeitstests der BU
Teilbereich Konzentration (BU)	Mittelwert aus allen Konzentrationstests der BU
Teilbereich Raumorientierung (BU)	Mittelwert aus allen Raumorientierungstests der BU
DAC	Dynamischer Airtraffic Control Test (Radarsimulation)
SDM	Strip Display Management Test (Flugstreifensimulation)
MIC	Multi Instrument Coordination Test (Fluginstrumentensimulation)
SIM	Simulatortest (Flugsimulation)

7.1.5 Korrelationsanalysen zum HTQ-2

Um Zusammenhänge zwischen dem HTQ-2 und der Leistung in den Auswahlverfahren zu prüfen, werden Pearson-Korrelationen berechnet. Voraussetzung für die Anwendung der Pearson-Korrelation sind intervallskalierte Daten und Normalverteilung (Ziegler & Bühner, 2009). Intervallskalenniveau wird entsprechend der üblichen Forschungspraxis (Bortz, 2005) für die verwendeten Skalen angenommen. Die Normalverteilung der Variablen wird mittels Sichtung der Histogramme geprüft, diese sind im Anhang aufgeführt.

In der Arbeit von Nowitzke (2013) ergab sich, dass einzelne Korrelationen zwischen HTQ-2-Skalen und Auswahlvariablen lediglich auf die Variable Geschlecht zurückzuführen waren. Auch andere Forschungsergebnisse zeigen, dass es z.B. bei technikbezogenen Persönlichkeitsfragebögen Geschlechterunterschiede gibt. So fand z.B. Beier (2004) bei seinen Untersuchungen zum Fragebogen „Kontrollüberzeugungen im Umgang mit Technik“ Geschlechtseffekte bei der Beantwortung. Männer wiesen höhere Werte auf als Frauen (Beier,

2004), gehen also davon aus, „die Technik eher im Griff zu haben als Frauen“ (Beier, 2004, S. 106). Zu ähnlichen Ergebnissen kamen Karrer, Glaser, Clemens und Bruder (2009), die zeigen konnten, dass weibliche Personen ihre Kompetenz und Begeisterung im Umgang mit elektronischen Geräten geringer einschätzen, als männliche Personen. Um reine Geschlechtseffekte von tatsächlichen Korrelationen des HTQ-2 mit den abhängigen Variablen zu trennen, werden Partialkorrelationen für die Variable Geschlecht berechnet. Die Ergebnisse der Korrelationsanalysen werden in Abschnitt 8.3 erläutert.

7.1.6 Vom HTQ-2 zum HTQ-S

Im Laufe des Projekts wurde zusätzlich zu den Selbstbeschreibungsskalen des HTQ-2 der Persönlichkeitsbereich Neigung zu risikoreichem Entscheidungsverhalten als möglicherweise wichtiger Prädiktor für Leistung in Mensch-Maschine-Teams erachtet. Da dieser in der Literatur als nur schwer erfassbar mit Persönlichkeitsfragebögen gilt (Lejuez et al., 2002), wurde das Kurzexperiment Balloon Analogue Risk Task (BART) (Lejuez et al., 2002) ausgewählt, um den Persönlichkeitsaspekt in objektiver Form zu erfassen (zum Thema objektive Persönlichkeitstests siehe auch unter 2.1). Die BART komplettiert den Fragebogen zum HTQ-S (=HTQ-Supplemented). **Abbildung 8** auf der Folgeseite zeigt die Durchführung der Balloon Analogue Risk Task.

Abbildung 8: Durchführung der Balloon Analogue Risk Task (BART)
©DLR

7.2 Entwicklung der ATM-Simulation HINT

Auch wenn es schwer ist, bereits heute zu antizipieren, wie die Automatisierung der Zukunft genau aussehen wird und welche Aufgaben den menschlichen Akteuren dabei konkret zukommen, sollte dieses zumindest versucht werden, um angemessen darauf reagieren zu können (Manning & Broach, 1992). Um Mensch-Maschine-Teamarbeit im Luftfahrtkontext zu simulieren, entstand daher im Rahmen des HYBRID-Projekts die Air-Traffic-Management-Simulation HINT (Hybrid Interaction Scenario). Im folgenden Abschnitt wird diese zunächst beschrieben, um im Anschluss daran Variablen für statistische Analysen abzuleiten (siehe unter 7.2.2). Im Rahmen dieser Arbeit wird die Simulation als Validierungsinstrument verwendet. Die Entwicklung des Verfahrens erfolgte innerhalb des HYBRID-Projektes.

7.2.1 Beschreibung des Verfahrens

HINT ist eine stark vereinfachte Flugverkehrssimulation von zwei zunächst unabhängigen Sektoren Alpha und Bravo. Die Sektoren sind getrennte Arbeitsplätze, die entweder manuell von jeweils einer Person oder voll automatisiert bearbeitet werden können. Die Hauptaufgabe der Operateure ist das zwischen den Sektoren ausbalancierte Aufrechterhalten von Flugverkehr im Luftraum. Zusätzlich muss eine Nebenaufgabe bearbeitet werden, die so genannte Trajectory-Control-Aufgabe, bei der die Flughöhe und/oder die Geschwindigkeit von Flugzeugen (Critical Aircraft, CAC) korrigiert werden müssen. Im Vordergrund steht nicht die detailgetreue Nachbildung antizipierter zukünftiger Realität, sondern eine möglichst vereinfachte Umsetzung zentraler Anforderungen bei der Arbeit mit hoch automatisierten Systemen, z.B. Monitoring (Systemüberwachung) und Exception Handling (Ausnahmebehandlung). **Abbildung 9** zeigt einen Screenshot der HINT-Simulation.

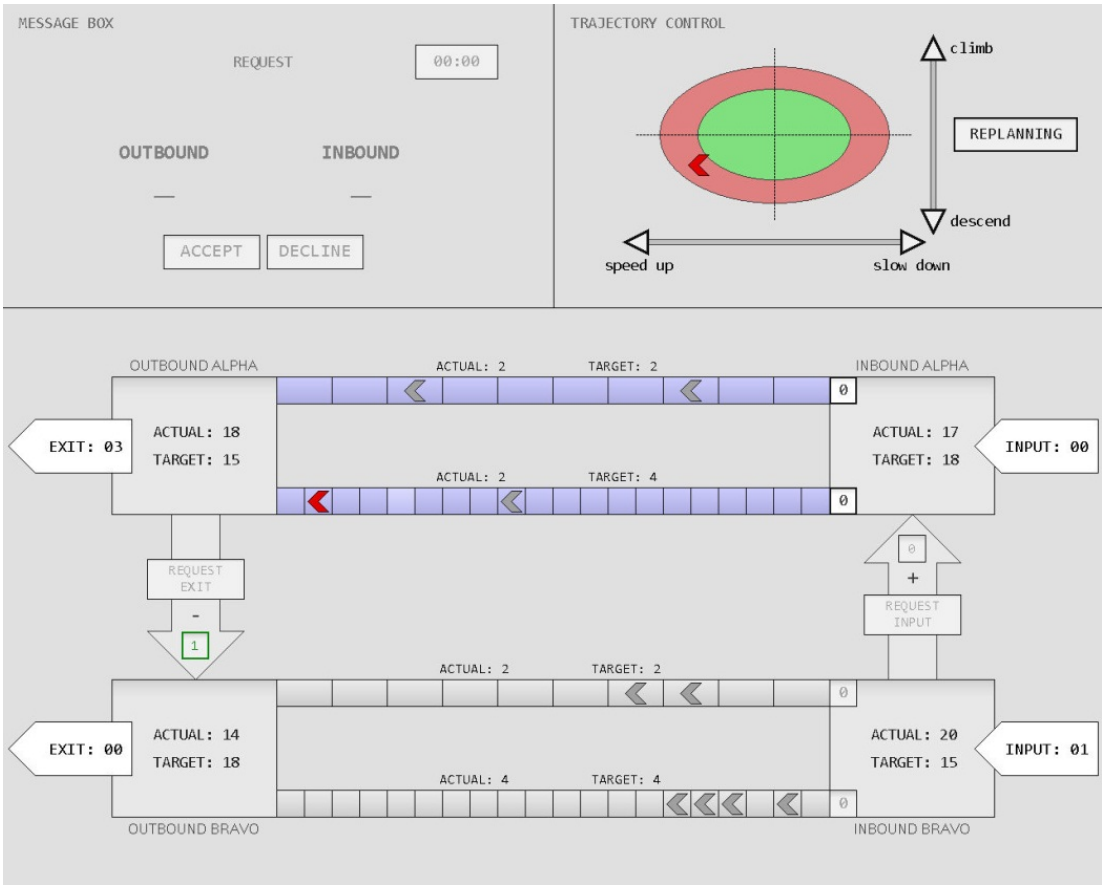


Abbildung 9: HINT (Hybrid Interaction Scenario)
©DLR

Die Probanden übernehmen die Rolle des Luftraumoperateurs im Sektor Alpha. Sektor Bravo wird dagegen voll automatisiert gesteuert. Beide Sektoren sollen vorgegebene Zielgrößen (Target-Werte) auf den vorhandenen Luftstraßen sowie Einflugbereichen (Inbound) und Ausflugbereichen (Outbound) so gut wie möglich erfüllen, damit ein bestmöglicher Verkehrsfluss erreicht wird. Hierbei sind die meisten Funktionen des Systems automatisiert, der Operateur muss jedoch „im Loop“ bleiben, um angemessen auf Anfragen zum Transfer von Luftfahrzeugen zwischen den Sektoren (Requests/RQ) des Nachbarsektors Bravo zu reagieren und/oder eigene Requests zu stellen, um die Target-Werte der Simulation zu optimieren. Wichtig hervorzuheben ist hierbei, dass die Probanden instruiert sind, eine bestmögliche Gesamtsituation hinsichtlich ihrer eigenen Target-Werte sowie der Target-Werte im automatisierten Sektor Bravo zu erreichen. Es genügt also nicht, nur die eigenen Target-Werte im Sektor Alpha zu optimieren. Die Kommunikation mit dem maschinellen Teampartner erfolgt ausschließlich „schriftlich-elektronisch“ (ähnlich wie Data Link im modernen Luftverkehr, siehe z.B. Kerns, 1991) durch Request-Stellung bzw. – Bearbeitung über eine Message Box.

Hinsichtlich der in Abschnitt 2.4 dargestellten Automatisierungslevel ist HINT vom Grundprinzip auf Level 8 einzuordnen, da die Automation selbstständig Aufgaben ausführt und den Menschen nicht einbezieht (Management des Flugverkehrs und Bearbeitung der CAC in Sektor Bravo). Bezogen auf den Austausch zwischen Mensch und Maschine mittels Requests ist Level 4 die passende Zuordnung, da die Automation einen Transfer nur ausführt, wenn die Person den entsprechenden Request akzeptiert. Gleiches gilt jedoch auch für die Requests des menschlichen Operateurs: auch sie werden nur ausgeführt, wenn die Automation zustimmt. Ob sie dies tut, ist abhängig von der Sinnhaftigkeit des Requests. Eine Programmschleife zur Request-Prüfung testet jede Anfrage des Operateurs darauf, ob diese der Annäherung an die Target-Werte des Gesamtsystems dienlich ist oder nicht.

In einer Vorstudie zu HINT nahmen 51 Fluglotsenbewerber der DFS gegen eine Aufwandsentschädigung von 20€ an der insgesamt 1,5-stündigen Untersuchung (inklusive Instruktion) teil. Nach ausführlicher Instruktion sowie einer Übung, in der durch die anwesende Testleiterin sichergestellt wurde, dass die Funktionsweise des Programms verstanden wurde, bearbeiteten alle Bewerber 8 je fünfminütige Szenarien mit unterschiedlicher Schwierigkeitskonfiguration. Jedes Szenario enthielt 151 Messzeitpunkte zu denen jeweils 12 Leistungsparameter erfasst wurden. Insgesamt entstanden so nahezu 14.500 Datenpunkte pro Bewerber. Durch die Schwierigkeitsmanipulation wurde die Leistung der Probanden in Abhängigkeit von der Konfiguration erhoben. Anhand dieser ersten Erhebung konnte die Funktionalität der Leistungsparameter festgestellt werden. Zusätzlich konnte bestimmt werden, welche Konfigurationen am ehesten in der Lage sind, zwischen Probanden zu differenzieren, d.h. „gute“ und „schlechte“ Leistung in der Zusammenarbeit mit einem hochautomatisierten System zu messen. Ergänzend zu der Bearbeitung von HINT füllten die Bewerber zu jedem Szenario einen Fragebogen aus. Diese Nachbefragung enthielt unter anderem Fragen zur subjektiv empfundenen Schwierigkeit der verschiedenen Konfigurationen, zu einzelnen Parametern der Szenarien, zur Bedienbarkeit von HINT und zur Beurteilung der Zusammenarbeit mit dem hochautomatisierten System. Die Ergebnisse der Vorstudie sowie detailliertere Beschreibungen der Simulation werden von Keye und Eschen (2013) berichtet.

7.2.2 Variablen aus der ATM-Simulation HINT

Wie im vorigen Abschnitt beschrieben, wird durch HINT eine Fülle an Datenpunkten erzeugt. Alle zwei Sekunden wird der Ist-Zustand von je 12 Leistungsparametern erfasst. In einem ersten Schritt wurden Mittelwerte der Leistungsparameter über die gesamte Bearbeitungszeit hinweg gebildet. Diese wurden in einem zweiten Schritt anhand inhaltlicher Überlegungen weiter zusammengefasst, so dass nun 7 endgültige Variablen mit Aussagekraft

hinsichtlich der Leistung bzw. des Verhaltens des menschlichen Operateurs Alpha zur Verfügung stehen. Im Rahmen des HYBRID-Projekts waren ausführliche Analysen zur zweckmäßigsten Zusammensetzung der Variablen vorangegangen. Neben der Zusammenfassung auf inhaltlicher Ebene wurden auch Auswertungen auf Ebene der einzelnen Parameter sowie eine faktorenanalytische Variablenbildung geprüft. Am geeignetsten erwies sich jedoch die inhaltliche Aggregation. Die daraus resultierenden Variablen werden in **Tabelle 6** beschrieben. Sofern verschiedene Ursprungsvariablen zu einer Variablen zusammengefasst wurden, wurden diese zuvor z-standardisiert, um sie miteinander verrechnen zu können. Bezüglich der Variablen RQ-Ablehnung und RQ-Annahme durch Alpha sei darauf hingewiesen, dass diese in der Summe ungleich der Gesamtanzahl der Requests sind, da ebenfalls die Möglichkeit des zeitlichen Verfalls von Requests besteht.

Tabelle 6: Variablen aus der HINT-Simulation und ihre Bedeutung

Variable	*Bedeutung*	*Polung*
Korrekte CAC-Bearbeitung	Relative Anzahl korrekt bearbeiteter CAC	hohe Werte = viele korrekt bearbeitete CAC
Reaktionszeit	Wert aus gemittelter Zeit bis ein CAC angeklickt und korrekt bearbeitet ist sowie relativer Anzahl verfallener, also nicht bearbeiteter CAC	hohe Werte = langsames Arbeiten
RQ Alpha an Bravo	Anzahl gestellter RQ von Alpha an Bravo	hohe Werte = hohe Anzahl gestellter RQ
RQ-Ablehnung durch Alpha	Relative Anzahl von Alpha abgelehnter RQ	hohe Werte = hohe Anzahl abgelehnter RQ
RQ-Annahme durch Alpha	Relative Anzahl von Alpha abgelehnter RQ	hohe Werte = hohe Anzahl angenommener RQ
Target-Abweichungen Alpha	Gemittelte Abweichung aller Target-Werte im Inbound, Outbound und auf den Luftstraßen von Alpha	hohe Werte = hohe Abweichungen
Target-Abweichungen Bravo	Gemittelte Abweichung der Target-Werte im Inbound und Outbound von Bravo (nur diese sind von Alpha mittels RQ beider Seiten beeinflussbar)	hohe Werte = hohe Abweichungen

Inhaltlich ist darauf hinzuweisen, dass im Gegensatz zu den übrigen aufgelisteten Variablen für die Request-Variablen nicht eindeutig definierbar ist, ob eine große Anzahl gestellter, bzw. abgelehnter oder angenommener Requests für gute oder schlechte Leistung steht. Dieses hängt jeweils sehr vom individuellen Kontext eines Requests ab. Jedoch sind die Request-Variablen Indikatoren für die Bereitschaft von Probanden, mit ihrem automatisierten Partner zusammen zu arbeiten, d.h. für ihre Kooperationsleistung.

7.2.3 Korrelationsanalysen zu den HINT-Variablen

Um die Interkorrelationen der HINT-Variablen zu prüfen, werden Pearson-Korrelationen berechnet. Hiermit soll festgestellt werden, ob z.B. Zusammenhänge zwischen der Reaktionszeit in der Simulation und den zusätzlichen Leistungsparametern vorliegen. Auf die Voraussetzungen geht bereits Abschnitt 6.1.5 ein. Die Ergebnisse der Korrelationsanalysen werden in Abschnitt 8.2 erläutert.

7.3 Kombination von HTQ-S und HINT

Die Kombinationsstudie zu HTQ-S und HINT beinhaltete zusätzlich zu den bereits beschriebenen Verfahren eine Nachbefragung zu HINT, eine Erhebung der gefühlten Beanspruchung durch HINT mittels NASA Task Load Index (NASA-TLX, Hart & Staveland, 1988) sowie Items zur Befindlichkeit der Probanden zu mehreren Zeitpunkten des Experiments. Außerdem wurde neben der BART noch ein weiteres Experiment zur Neigung zu risikoreichem Entscheidungsverhalten (die Iowa Gambling Task, beschrieben z.B. bei Bechara, Damasio, Damasio, & Anderson, 1994) eingesetzt. Die Auswertungen zu diesen zusätzlichen Verfahren werden in der vorliegenden Arbeit jedoch nicht behandelt, sondern sind Gegenstand weiterführender Analysen. Allerdings werden vier Items aus der Nachbefragung in die Analysen mit einbezogen. Diese waren ursprünglich Teil des HTQ-2, wurden jedoch für die Kombinationsstudie in die Nachbefragung ausgelagert. Sie besitzen konkreten Bezug zu der Automation, mit der gearbeitet wird, und waren daher sinnvoller im direkten Anschluss an HINT zu beantworten (siehe auch unter EgMMT in Abschnitt 8.1.1).

Die Durchführung der Studie dauerte ca. 3,5 Stunden pro Durchlauf. Maximal vier Bewerber konnten gleichzeitig teilnehmen. Die Probanden wurden zu verschiedenen Zeitpunkten im Rahmen ihrer Auswahluntersuchung für die DFS bzw. DLH angesprochen und gebeten, sich an einer Studie zum Thema „Anforderungen an zukünftige Luftfahrtoperateure“ zu beteiligen. Die Teilnahme war freiwillig. Für die Teilnahme erhielten die Probanden ein Versuchspersonenentgelt von 50€ bis 60€ als Aufwandsentschädigung. Aufgrund der langen Dauer des Experiments war die Teilnahme für die Bewerber aufwendig, was die Rekrutierung erschwerte.

Die Untersuchungen fanden während des Zeitraums März bis November 2012 in der Abteilung für Luft- und Raumfahrtpsychologie des DLR in Hamburg statt und wurden von Testleiterinnen des DLR betreut. **Abbildung 10** auf der Folgeseite zeigt den Versuchsaufbau in dem dafür genutzten Experimentalraum.

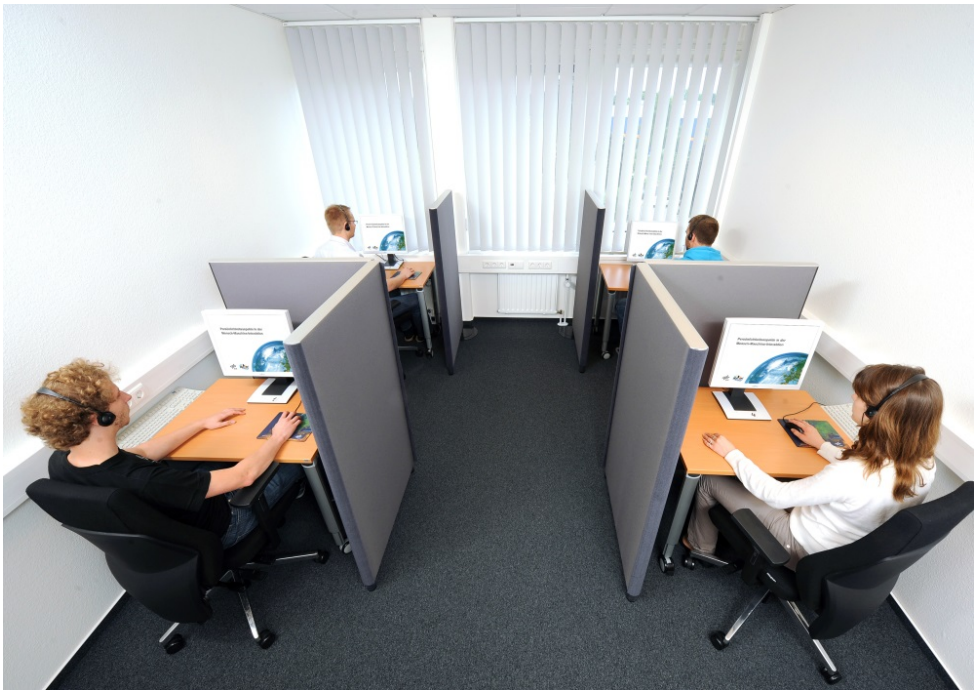


Abbildung 10: Durchführung der Kombinationsstudie zu HTQ-S und HINT
©DLR

7.3.1 Stichprobe

Insgesamt bearbeiteten N=156 Bewerber sowohl die ATM-Simulation HINT als auch den HTQ-S: 101 DFS-Bewerber sowie 55 DLH-Bewerber. Die unterschiedliche Größe der Teilstichproben resultiert daraus, dass eine Teilnahme für DLH-Bewerber schwieriger zu realisieren war. Hintergrund ist, dass die Struktur des DLH-Auswahlverfahrens weniger ausreichend große Zeitfenster als die des DFS-Auswahlverfahrens enthält. Weiterhin stehen für die verschiedenen Variablen aus HINT und HTQ-S unterschiedlich große Stichprobengrößen zur Verfügung. Alle 156 Probanden bearbeiteten HINT, jedoch konnten aufgrund eines Versionswechsels bei der Simulation während des Erhebungszeitraumes nur für 69 Personen alle Variablen erfasst werden. Die HINT-Variable „RQ Alpha an Bravo“ sowie beide Variablen zu Target-Abweichungen liegen daher für 69 Probanden vor. Alle anderen Variablen wurden grundsätzlich für die volle Stichprobe erhoben, fehlende Werte gibt es jedoch für sechs Probanden in der Nachbefragung sowie für zwei Probanden bei den Variablen zu RQ-Annahme bzw. –Ablehnung durch Alpha. Probanden mit fehlenden Werten wurden nicht komplett aus der Untersuchung ausgeschlossen, sondern wo Daten vorhanden waren, in die Analysen einbezogen. Zu allen berichteten Ergebnissen in Kapitel 8 ist die jeweilige Stichprobengröße angegeben.

27,6% der Probanden aus der Gesamtstichprobe sind weiblich. Der Altersdurchschnitt aller Probanden liegt bei 20,02 (SD = 2.08). Wie auch in den Untersuchungen zum HTQ-1 und

HTQ-2 werden alle weiteren erhobenen biografischen Merkmale der Teilnehmer (z.B. Nationalität, Muttersprache, Familienstand…) in den weiteren Analysen nicht berücksichtigt, da sie von zu geringer Variabilität gekennzeichnet sind. Geschlechts- und Bewerbergruppenunterschiede im HTQ-S werden in Abschnitt 8.1.3 berichtet.

7.3.2 Korrelationsanalysen zum HTQ-S

Um Zusammenhänge zwischen dem HTQ-S, aktueller Eignungsdiagnostik und der ATM-Simulation HINT zu prüfen, werden Pearson-Korrelationen berechnet. Auf die Voraussetzungen geht bereits Abschnitt 7.1.5 ein. Wie auch für die dort beschriebenen Korrelationsanalysen zur Eignungsdiagnostik werden die Zusammenhänge von HTQ-S und HINT ebenfalls mittels Partialkorrelationen auf Geschlechtseffekte hin untersucht. Die Ergebnisse der Korrelationsanalysen werden in Abschnitt 8.3 erläutert.

7.3.3 Gruppenvergleiche

Ergänzend zu korrelativen Zusammenhängen wird überprüft, ob eine Einteilung der Probanden in Gruppen anhand besserer und schlechterer Leistung in der HINT-Simulation unterschiedliche Beantwortungen der HTQ-S-Skalen sichtbar macht. Hintergrund ist, dass bei einem eventuellen Einsatz des Fragebogens in zukünftiger Eignungsdiagnostik Grenzwerte gesetzt werden müssen. Mit der Einteilung in zwei Gruppen soll eine erste Annäherung an eventuelle spätere Grenzwerte geschaffen werden. Für jede HINT-Variable werden daher anhand eines Median-Splits eine bessere und eine schlechtere Gruppe von Probanden gebildet. Für die Request-Variablen wird zunächst neutral eine Gruppe mit niedrigeren bzw. höheren Werten gebildet. Bei diesen rein quantitativen Variablen kann nicht allein aufgrund der Anzahl gestellter / angenommener / abgelehnter Requests von besserer oder schlechterer Leistung gesprochen werden. Ob sich die HTQ-S-Mittelwerte zwischen den Gruppen statistisch signifikant unterscheiden, wird durch t-Tests überprüft. Zusätzlich werden mittels Cohens d die jeweiligen Effektgrößen geschätzt, um zu untersuchen, ob die ermittelten signifikanten Unterschiede auch praktisch bedeutsam sind. In Abschnitt 8.4 werden die Ergebnisse dargelegt.

7.3.4 Inkrementelle Validität

Die inkrementelle Validität des HTQ-S wird mittels multipler Korrelationsstatistik geschätzt. Für jede HINT-Variable als Kriterium wird in einem ersten Schritt einzeln geprüft, welche Leistungsvariablen aus der Vorauswahl bzw. Berufsgrunduntersuchung der Bewer-

ber signifikant zur Kriteriumsvarianz R^2 beitragen. Im zweiten Schritt werden die Variablen des HTQ-S als zusätzliche Prädiktoren in die Analysen aufgenommen. Auf diese Art und Weise wird geprüft, ob sich durch die Einbeziehung der Variablen eine höhere Korrelation ergibt und somit ein größerer Anteil der Kriteriumsvarianz R^2 vorhersagbar wird. Hierzu werden Regressionsanalysen durchgeführt. Die Eingabe der Variablen in die Regressionsanalyse erfolgt in diesem Fall schrittweise, um zu gewährleisten, dass weitere Variablen nur dann in die Regressionsgleichung aufgenommen werden, wenn sie zusätzlich einen bedeutsamen Anteil zur Varianzaufklärung beitragen.

Für die multiple Regression müssen diverse Voraussetzungen erfüllt sein (beschrieben z.B. bei Bühner & Ziegler, 2009). Die einbezogenen Variablen sollten normalverteilt sein, dies wird mit Histogrammen überprüft. Die Varianz der Verteilung der abhängigen Variablen muss für alle Werte der unabhängigen Variablen konstant sein, d.h. mit steigenden Werten der unabhängigen Variablen sollten die Werte der abhängigen Variablen nicht weiter streuen. Die Annahme dieser sogenannten Homoskedastizität wird mit Streudiagrammen überprüft. Zudem sollten die Fehler der Residuen zufällig auftreten und normalverteilt sein. Dies wird mit Hilfe von Histogrammen der standardisierten Residuen geprüft. Ebenfalls wird verlangt, dass keine Kollinearität der Variablen vorliegt, dies lässt sich mittels Korrelationsanalysen testen.

Da multiple Korrelationen nach Bortz (2005) insbesondere bei kleinem Stichprobenumfang und vielen Prädiktoren dazu neigen, den wahren Korrelationskoeffizienten zu überschätzen, wird zu jeder der durchgeführten Berechnungen zusätzlich eine Schrumpfungskorrektur nach Olkin und Pratt (1958) durchgeführt. Da es verschiedene Formeln der Schrumpfungskorrektur gibt und diese nicht in dem verwendeten Auswertungsprogramm SPSS™ 18 zur Verfügung stehen, wird die verwendete Formel für *N Personen* und *K Prädiktoren* an dieser Stelle dargestellt:

Formel 1: $$\hat{R}^2 = 1 - \left(\frac{N-3}{N-K-2}\right) \cdot \left((1-R^2) + \left(\frac{2}{N-K}\right) \cdot (1-R^2)^2\right)$$

8 Ergebnisse

Dieses Kapitel stellt schrittweise die Ergebnisse der verschiedenen durchgeführten Studien vor. Begonnen wird mit der Darstellung des HTQ, dessen Konzeption ein großer Baustein der vorliegenden Arbeit ist. Nachdem die verwendeten Skalen und die dahinter stehenden Konstrukte inklusive ihrer Relevanz für die Fragestellung beschrieben sind (8.1.1), werden die Ergebnisse zur technischen Umsetzbarkeit der Skalen präsentiert (8.1.2), die deskriptiven Statistiken (8.1.3) und psychometrischen Kennwerte (8.1.4) dargestellt sowie im Folgenden die endgültige Skalenzusammenstellung des HTQ-S erläutert (8.1.5). Abschnitt 8.2 beschäftigt sich mit den deskriptiven Statistiken der HINT-Variablen, so dass nun für die darauf folgenden Korrelationsanalysen (8.3) alle verwendeten Variablen vorgestellt sind. Den Korrelationen des HTQ-S mit aktueller Eignungsdiagnostik (8.3.1) folgen die Korrelationsergebnisse mit der ATM-Simulation HINT (8.3.2). Die Gruppenvergleiche zum HTQ-S werden unter 8.4 vorgestellt, während die inkrementelle Validität des Fragebogens Thema des Abschnitts 8.5 ist. Im letzten Teil dieses Kapitels werden die zahlreichen Ergebnisse zusammengefasst (8.6).

8.1 Der HTQ

8.1.1 Die Skalen des HTQ

Im Folgenden werden alle in den verschiedenen Versionen des HTQ (HTQ-1, HTQ-2 und HTQ-S) verwendeten Skalen dargestellt. **Tabelle 7** auf der nächsten Seite zeigt eine Übersicht, im Anschluss werden die Skalen näher beschrieben sowie die jeweilige Relevanz der gemessenen Konstrukte für die Fragestellung der Arbeit erläutert. Die Darstellung erfolgt in alphabetischer Reihenfolge.

Tabelle 7: Übersicht über die HTQ-Skalen

Kürzel	*Testname*	*Testautoren*	*Verwendete Skalen & Itemanzahl*
BART	Balloon Analogue Risk Task	Lejuez, C.W., Read, J.P., Kahler, C.W., Richards, J.B., Ramsey, S.E., Stuart, G.L., Strong, D.R., & Brown, R.A. (2002)	Objektiver Persönlichkeitstest (30 Durchgänge)
BFI-10	Big Five Inventory	Rammstedt & John (2007)	N, E, O, G, V (10)
BIO	Biografische Angaben	HYBRID-Projekt	Biografische Angaben (9) Items zur Computernutzung (20)
BIP	Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung	Hossiep, R. & Paschen, M. (1998)	Flexibilität (14) Teamorientierung (13)
BnT	Bedürfnis nach Teamarbeit	HYBRID-Projekt	BnT (5)
CaP	Complacency als Potential	Bahner (2008)	Risikotoleranz (4)
CFQ	Cognitive Failures Questionnaire	Klumb, P.L. (1995)	18 ausgewählte Items des CFQ von Klumb
CNFB	Fragebogen zur Computernutzung	Schroeders, U., & Wilhelm, O. (2010)	Für das Projekt angepasste Version des CNFB (34)
EgA	Einstellung gegenüber Automation	HYBRID-Projekt	EgA (17)
EgMMT	Einstellung gegenüber Arbeit im Mensch-Maschine-Team	HYBRID-Projekt	EgMMT (4)
FEcA	Fragebogen zur Erfassung computerspezifischer Attributionen	Dickhäuser, O. & Stiensmeier-Pelster, J. (2000)	Lokalität, Kontrolle, Stabilität, Globalität für Erfolg und Misserfolg (10 Situationen à 8 Items)
FIT	Fragebogen zur Erfassung individueller Einstellungen gegenüber Teamarbeit	Mohiyeddini, C. (2001)	Bereitschaft zur Teamarbeit (4) Vorbehalte gegenüber Teamarbeit (4)
INCOBI-R	Inventar zur Computerbildung, revidierte Form	Richter, T., Nauman, J., & Horz, H. (2010)	FIDEC (77) COMA (8)
KUT	Kontrollüberzeugungen im Umgang mit Technik	Beier, G. (2004)	KUT (8)
NEO-PI-R	NEO-Persönlichkeitsinventar, revidierte Fassung	Ostendorf, F. & Angleitner, A. (2004)	Neurotizismus (48) Extraversion (48) Offenheit für Erfahrungen (48) Verträglichkeit (48) Gewissenhaftigkeit (48)
SWE	Allgemeine Selbstwirksamkeitserwartung	Schwarzer, R. & Jerusalem, M. (1999)	SWE (10)
TA-EG	Fragebogen zur Messung von Technikaffinität	Karrer, K., Glaser, C., Clemens, C. & Bruder, C. (2009)	Begeisterung für Technik (5) Kompetenz im Umgang mit Technik (4) Positive Folgen (5) Negative Folgen (5)

Balloon Analogue Risk Task (BART)

Die Balloon Analogue Risk Task (BART) wurde von Lejuez et al. (2002) entwickelt und ist ein objektiver Persönlichkeitstest. Mit der computerbasierten Messmethode wird risikoreiches Entscheidungsverhalten erfasst. Vergleichbar zu realen Situationen wird wagemutiges Entscheidungsverhalten bis zu dem Punkt belohnt, bei dem weitere Risiken zu schlechteren Resultaten (Platzen eines Luftballons) führen.

Hintergrund: Die Messung von risikoreichem Entscheidungsverhalten wurde bis zur Entwicklung von objektiven Persönlichkeitstests wie der BART jeweils durch Selbstberichte in Fragebögen zu z.B. Sensation Seeking oder Impulsivität vorgenommen. Laut den Testautoren der BART gibt es zwischen den genannten Konstrukten und der Zieldimension zwar Überschneidungen, diese seien jedoch nicht in der Lage, die breite und multidimensionale Natur der Persönlichkeitseigenschaft zu erfassen (Lejuez et al., 2002). Mit der BART sollen diese Restriktionen überwunden werden.

Aufbau: Die BART wurde in einer für das Projekt adaptierten Version und auf Deutsch vorgegeben. Auf dem Bildschirm ist dabei ein roter Luftballon mit einem zugehörigen Button „Aufpumpen“ zu sehen. Insgesamt sind 30 Ballons aufzupumpen, pro Aufpumpen erhält der Proband 50 Gewinnpunkte. Zusätzlich zu diesen beiden Informationen werden auf dem Bildschirm die Nummer des aktuellen Ballons, die aktuelle Anzahl der Pumpvorgänge und der Gesamtgewinn angezeigt. Über einen Button „Gewinn einsammeln“ können die Probanden entscheiden, das Aufpumpen eines Ballons zu beenden, bevor dieser platzt. In diesem Fall addieren sich die Gewinnpunkte des aktuellen Ballons zum bisherigen Gesamtgewinn. Platzt ein Ballon, werden die Gewinnpunkte für diesen Ballon auf Null gesetzt. Pro 100 Gewinnpunkte erhalten die Probanden am Ende des Gesamtversuchs 1 Cent zusätzlich zu ihrer allgemeinen Aufwandsentschädigung (maximal entstehen aus der BART 6€ Zusatzgewinn).

Anwendung: Seit ihrer Einführung wird die BART in verschiedensten Labor- sowie Feldstudien genutzt. So wurden z.B. Zusammenhänge zu Substanzmissbrauch und risikoreichem Sexualverhalten (Aklin, Lejuez, Zvolensky, Kahler, & Gwadz, 2005), zu Ängstlichkeit (Maner et al., 2007) oder auch zu Machtmotivation (Maner, Gailliot, Butz, & Peruche, 2007) näher untersucht.

Gütekriterien: In einer Studie von White, Lejuez und de Wit (2008) zeigte die BART mit einem Abstand von mehreren Tagen eine Retest-Reliabilität von .77. Zur Überprüfung der Kriteriumsvalidität wurden von Lejuez et al. (2002) unter anderem Korrelationen mit Impulsivitätsskalen, Sensation Seeking sowie risikoreichem Verhalten innerhalb des letzten Jahres erhoben. Zu Impulsivität ergab sich eine signifikante Korrelation von .28, zu Sensa-

tion Seeking von .35. Zusätzlich zeigten sich signifikante Korrelationen zu allen erhobenen Kriterien für selbstberichtetes risikoreiches Verhalten innerhalb des letzten Jahres, z.B. Alkohol-, Zigaretten- und Drogenkonsum (.28 -.36), Glücksspiel (.44), unangeschnalltes Autofahren (.25), ungeschützter Sexualverkehr mit verschiedenen Partnern (.25), Anzahl gestohlener Gegenstände aus Geschäften (.25).

Relevanz für die Fragestellung: Von Operateuren in sicherheitsrelevanten Arbeitsbereichen wie der Luftfahrt wird grundsätzlich risikoarmes Entscheidungsverhalten verlangt. Dies wird auch in Zukunft für die menschlichen Operateure in Mensch-Maschine-Teams der Luftfahrt gelten. Feuerberg, Bahner und Manzey (2005) konnten zeigen, dass eine höhere Risikotoleranz mit unzureichender Überwachungsleistung einhergeht. Geprüft werden soll daher, ob hinsichtlich der ATM-Simulation HINT negative Auswirkungen von risikoreichem Entscheidungsverhalten auf die erbrachte Leistung feststellbar sind.

Big Five Inventory (BFI-10)

Das Big-Five-Inventory (BFI-10) wurde von Rammstedt und John (2007) zur Erfassung der Big Five in kürzestmöglicher Zeit entwickelt. Für weiterführende Informationen zu den Persönlichkeitsfaktoren des Fünf-Faktoren-Modells sei an dieser Stelle auf den Abschnitt zum NEO-PI-R verwiesen, der die Faktoren in ausführlicher Weise misst.

Hintergrund: Grundlage des BFI-10 ist das BFI-44 nach John, Donahue und Kentle (1991). Ziel der Autoren des BFI-10 ist, mit der Kurzskala ein Messinstrument zur Verfügung zu stellen, das die Big Five in Kontexten messen kann, in denen die Zeit für das Ausfüllen von Fragebögen stark begrenzt ist. Unter Bezug auf Burisch (1997) weisen sie darauf hin, dass auch sehr kurze Skalen respektable psychometrische Charakteristiken zeigen können.

Aufbau: Im BFI-10 werden die Big Five mit je zwei Items gemessen. Für die Dimension Verträglichkeit gibt es zusätzlich ein drittes optionales Item, mit dem sich die Skala Verträglichkeit als prädiktiver erwies (Rammstedt & John, 2007). Für den HTQ-S wird die BFI-10 Version mit 11 Items und wie im Original mit einem fünfstufigen Antwortformat verwendet.

Anwendung: Die Anwendung des BFI-10 wird von den Testautoren im Forschungsbereich gesehen.

Gütekriterien: Die Testautoren berichten von substantiellen part-whole Korrelationen des BFI-10 zum BFI-44. Im Mittel zeigt sich eine Korrelation von .83 zwischen den Kurz- und den Vollskalen. Das BFI-10 kann 70% der Varianz der Vollversion erklären. Zur Bestim-

mung der konvergenten Validität wurde der NEO-PI-R als weiteres Messinstrument für die Big Five herangezogen. Die Korrelationen mit den Primärfaktoren des NEO-PI-R liegen im Mittel bei .67 (Rammstedt & John, 2007). Die Retest-Reliabilität des BFI-10 nach 6 bis 8 Wochen liegt bei durchschnittlich .75.

Normierung: Basierend auf einer bevölkerungsrepräsentativen deutschen Stichprobe liegen alters-, geschlechts- und bildungsspezifische Normwerte (t-Werte) vor (Rammstedt, 2007).

Relevanz für die Fragestellung: Die Verwendung des BFI-10 soll ermöglichen, Zusammenhänge der allgemeinen Persönlichkeit mit der Leistung in hoch automatisierten Mensch-Maschine-Teams der Luftfahrt zu erfassen. So soll geprüft werden, ob die unter 3.1 besprochenen Zusammenhänge zwischen den Big Five und beruflicher Leistung auch für zukünftige Luftfahrtoperateure replizierbar sind. Das BFI-10 wird zusätzlich zum NEO-PI-R eingesetzt, um zu prüfen, ob dieser durch seine Kürze sehr ökonomische Fragebogen mit 11 Items, die wesentlich aufwändigere Messung der Big Five mit den 241 Items des NEO-PI-R ersetzen kann.

Biografische Angaben (BIO)

Zusätzlich zu den Persönlichkeitsverfahren werden auch biografische Angaben der Probanden erhoben. Dabei handelt es sich neben Standardangaben wie Geschlecht, Alter, Bildungsstand etc. auch um Gewohnheiten bei der Computernutzung, z.B. hinsichtlich des Spielens am Computer in Wochenstunden. Die Items des Fragebogens werden nicht zu einer Skala zusammengefügt, sondern einzeln ausgewertet.

Relevanz für die Fragestellung: Der Fragebogen BIO wird eingesetzt, um den Einfluss biografischer Merkmale zu erheben. So ist z.B. anzunehmen, dass Personen mit der Angewohnheit, Computer sehr intensiv zu nutzen, Vorteile bei der Arbeit im Mensch-Maschine-Team haben.

Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)

Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP) wurde von Hossiep und Paschen (1998) zur Erfassung von im Berufsleben relevanten Persönlichkeitsfacetten entwickelt. Auf 14 Dimensionen werden persönliche Eignungsvoraussetzungen wie Arbeitsverhalten, berufliche Orientierung, soziale Kompetenzen und psychische Konstitution abgefragt. Seit 2003 existiert eine zweite vollständig überarbeitete Auflage des BIP, die

inzwischen 17 Dimensionen beinhaltet. Die folgenden Angaben beziehen sich auf die in dieser Arbeit verwendete Version von 1998.

Hintergrund: Bei der Personalauswahl ist es laut den Testautoren notwendig, neben Wissen und Erfahrung auch überfachliche Kompetenzen von Bewerbern zu berücksichtigen. An diesem Punkt setzt das BIP mit der Erfassung von im Berufsleben relevanten Persönlichkeitsfacetten an. Die Testentwicklung des BIP orientierte sich in erster Linie an den Erfordernissen der diagnostischen Praxis. Die Auswahl der Dimensionen erfolgte neben den Praxisanforderungen auf der Basis von Befunden zur Validität von Persönlichkeitsstrukturen sowie der Befragung von Personalexperten zu relevanten Persönlichkeitsattributen.

Aufbau: Das BIP besteht aus 14 Dimensionen, die sich vier Bereichen zuordnen lassen. Jeder Dimension sind zwölf bis 16 Items zugeordnet. Insgesamt werden über die 14 Dimensionen 196 Items erfasst. Im Original werden die Items auf einem sechsstufigen Format beantwortet. Für den Einsatz im HTQ-S wird wiederum aus Gründen der Konformität ein fünfstufiges Antwortformat verwendet. Die Bearbeitungszeit der Komplettversion beträgt rund 45 Minuten. Für den HTQ-S wurden die Skalen Teamorientierung und Flexibilität ausgewählt. Die übrigen Skalen wurden wegen Überschneidungen zu anderen verwendeten Skalen nicht eingesetzt. Bei Teamorientierung handelt es sich nach Angabe der Testautoren um „Hohe Wertschätzung von Teamarbeit und Kooperation; Bereitschaft zur aktiven Unterstützung von Teamprozessen; breitwillige Zurücknahme eigener Profilierungsmöglichkeiten zugunsten der Arbeitsgruppe“. Flexibilität wird als „Hohe Bereitschaft und Fähigkeit, sich auf neue einzustellen und Ungewissheit zu tolerieren; Offenheit für neue Perspektiven und Methoden; hohe Veränderungsbereitschaft“ verstanden (Hossiep & Paschen, 1998).

Anwendung: Das BIP ist für Anwendungen in den Bereichen Eignungsdiagnostik, Berufs- und Karriereberatung, Training und Coaching sowie der anwendungsbezogenen Forschung geeignet.

Gütekriterien: Die interne Reliabilität der Skalen liegt zwischen .75 und .92, die Retest-Reliabilität nach acht bis zehn Wochen zwischen .77 und .89. Zur Überprüfung der Kriteriumsvalidität wurden unter anderem Korrelationen mit Leistungsmaßen wie Schul- und Ausbildungsnoten berechnet. Es zeigten sich geringe, nicht signifikante Zusammenhänge. Diese Befunde deuten darauf hin, dass es kaum Überschneidungen mit kognitiven Leistungstests gibt (Hossiep & Paschen, 1998). Signifikante Zusammenhänge zeigten sich z.B. mit Variablen wie Entscheidungssicherheit für den nach dem Studienabschluss angestrebten Tätigkeitsbereich, gesellschaftliches Engagement, berufliches Entgelt und eigene Arbeitszufriedenheit. Zur Überprüfung der Fragebogenstruktur wurden Hauptkomponentenanalysen berechnet. Die gefunden Faktoren waren weitgehend deckungsgleich zu den 14 Dimensionen.

Normierung: Es liegen Stanine- und Stanten-Werte in Normierungstabellen vor. Zudem existieren verschiedene Vergleichsgruppen nach Alter, Geschlecht, Position, Studienbereichen und Status.

Relevanz für die Fragestellung: Die aus dem BIP ausgewählten Skalen Teamorientierung und Flexibilität könnten aus zwei verschiedenen Gründen relevant für die Leistung im Mensch-Maschine-Team sein. Teamorientierung könnte einerseits förderlich sein, sofern der maschinelle Part tatsächlich als Teampartner empfunden wird, allerdings auch hinderlich, wenn Personen besser mit menschlichen als mit maschinellen Teampartnern zusammenarbeiten können. Flexibilität könnte ein Indikator dafür sein, wie Probanden mit der für sie neuen Situation im Mensch-Maschine-Team zurechtkommen.

Bedürfnis nach Teamarbeit (BnT)

Der Fragebogen zum Bedürfnis nach Teamarbeit (BnT) wurde im Rahmen des HYBRID-Projektes entwickelt. Die aus fünf Items bestehende Kurzskala enthält Fragen zur Wichtigkeit der Zusammenarbeit und Kommunikation sowie des persönlichen Kontakts mit anderen Menschen am Arbeitsplatz. Sie wurde entwickelt, um zwischen Personen differenzieren zu können, denen es ein Bedürfnis ist, beruflich im Team mit anderen Menschen zu arbeiten versus Personen, die weniger oder keinen Wert auf diesen Kontakt legen. Wie auch für die anderen Skalen des HTQ-S wurde ein fünfstufiges Antwortformat verwendet. In den Abschnitten 8.1.3 und 8.1.4 werden deskriptive Statistiken sowie psychometrische Kennwerte der Skala berichtet. Weiterführende Analysen zu Gütekriterien, Einsatzmöglichkeiten etc. müssen in Zukunft erfolgen.

Relevanz für die Fragestellung: Der Fragebogen BnT wurde eigens für die Fragestellung dieser Arbeit entwickelt. Es ist anzunehmen, dass Personen mit einem ausgeprägten Bedürfnis nach Teamarbeit gegebenenfalls weniger leistungsfähig im Mensch-Maschine-Team agieren, da ihnen der Kontakt und Austausch mit menschlichen Kollegen fehlt.

Complacency als Potential (CaP)

Mit dem Konstrukt Complacency wird das menschliche Phänomen beschrieben, durch übersteigertes Vertrauen in die Funktions- und Leistungsfähigkeit eines automatisierten Systems unzureichende Überwachungsleistung zu zeigen (Bahner, 2008). Das Konstrukt ist sehr breit angelegt und umfasst Einstellungen gegenüber Technik sowie verschiedene Persönlichkeitsmerkmale, die in dem Fragebogen Complacency als Potential (CaP) abgefragt wer-

den. Da in der vorliegenden Arbeit ausschließlich der sehr kleine Complacency-Teilbereich der Risikotoleranz verwendet wird, wird auf eine detaillierte Beschreibung der weiteren Teilbereiche verzichtet. Informationen hierzu finden sich bei Feuerberg et al. (2005).

Hintergrund und Anwendung: Das Konstrukt Complacency soll einen Beitrag zur Aufklärung von Problemen in der Mensch-Maschine-Interaktion leisten. Ziel ist, mögliche Ansatzpunkte für ein verbessertes Zusammenspiel von Menschen und automatisierten Systemen zu liefern (Bahner, 2008).

Aufbau: Die Skala Risikotoleranz beinhaltet vier Items, die inhaltlich dem Konstrukt Flexibilität (z.B. aus dem BIP, siehe dort) ähneln. Die Skala wird im HTQ-S entsprechend dem Original auf einem fünfstufigen Antwortformat beantwortet. (Zum generellen Aufbau des CaP siehe Feuerberg et al., 2005)

Gütekriterien: Als Reliabilitätsmaß steht die interne Konsistenz der Skalen zur Verfügung, die von den Testautoren für die Skala Risikotoleranz mit .70 angegeben wird (Feuerberg, Bahner, & Manzey, 2005). In einer Studie von Mahlfeld (2011) ergab sich ein Cronbach's α von .63. Zur Validitätsprüfung wurde ein Extremgruppenvergleich zwischen Personen mit einer starken und einer geringen Verhaltensausprägung von Complacency durchgeführt. Die Gruppen unterschieden sich unter anderem auch signifikant hinsichtlich ihrer Risikotoleranz. Ausgeprägteres Complacency-Verhalten ging mit größerer Risikotoleranz einher (Feuerberg et al., 2005).

Relevanz für die Fragestellung: Die Gründe für den Einsatz der Skala Risikotoleranz entsprechen denen des Einsatzes der BART zu Messung der Neigung zu risikoreichem Entscheidungsverhalten: Von Operateuren in sicherheitsrelevanten Arbeitsbereichen wie der Luftfahrt wird grundsätzlich risikoarmes Entscheidungsverhalten verlangt. Dies wird auch in Zukunft für die menschlichen Operateure in Mensch-Maschine-Teams der Luftfahrt gelten. Wie bereits erwähnt, konnten Feuerberg et al. (2005) zeigen, dass eine höhere Risikotoleranz mit unzureichender Überwachungsleistung einhergeht. Geprüft werden soll daher, ob hinsichtlich der ATM-Simulation HINT negative Auswirkungen von hoher Risikotoleranz auf die erbrachte Leistung feststellbar sind.

Cognitive Failures Questionnaire (CFQ)

Der von Broadbent, Cooper, FitzGerald und Parkes (1982) entwickelte Cognitive Failures Questionnaire (CFQ) erfasst die Häufigkeit begangener Alltagsfehler innerhalb eines bestimmten Zeitraumes in den Bereichen Wahrnehmung, Gedächtnis und Handlungsregulati-

on. Die hier beschriebenen Skalen entstammen der deutschsprachigen Version des CFQ von Klumb (1995).

Hintergrund: Viele Unfälle werden menschlichen Fehlern zugeschrieben. Im Allgemeinen können Fehler nicht ausschließlich Personeneigenschaften zugeschrieben werden, sondern sind als Konsequenz der mangelhaften Passung zwischen Aufgabe, technischem und menschlichem oder sozialem System zu betrachten. Fehlerprävention und –management kann sich demnach an das technische, organisatorische oder soziale System richten (Klumb, 1995). Broadbent et al. (1982) fassen Fehler der Wahrnehmung, des Gedächtnisses und der Handlung unter dem Term „cognitive failures" zusammen und gehen in ihrer Originalversion mit 25 Items von einer eindimensionalen Struktur des Konstrukts aus. Spätere Studien bestätigen diese eindimensionale Struktur jedoch nicht. Allerdings kommen sie zu keiner einheitlichen Faktorenlösung, sondern zu sehr unterschiedlichen (Volz-Sidiropoulou et al., 2007).

Aufbau: Die deutsche Version des CFQ von Klumb (1995) besteht aus 32 Items mit einem fünfstufigen Antwortformat. Die Beantwortung der Items bezieht sich auf Vorkommnisse der letzten zwölf Monate. Für den HTQ-S werden 18 Items verwendet, von denen vermutet wird, dass sie im Kontext von Mensch-Maschine-Teams relevant sein könnten.

Anwendung: Der CFQ hat eine gute Vorhersagevalidität im Bereich Verkehrsdiagnostik. Menschen, die viele Aufmerksamkeitsfehler im CFQ angaben, hatten auch ein erhöhtes Unfallrisiko (Larson & Merritt, 1991). Des Weiteren gibt es Befunde, die auf eine Beziehung zwischen CFQ-Scores und Wohlbefinden hindeuten (Smith, Peck, & Clatworthy, 1989).

Gütekriterien: In Hinblick auf die Kriteriumsvalidität zeigen bisherige Studien keine klare Beziehung zwischen CFQ und tatsächlicher Performance in Bezug auf die Anfälligkeit für kognitive Fehler. Klumb (1995) fand einen Zusammenhang zwischen CFQ-Scores und alltäglichen Problemen wie z.B. ein ausgeliehenes Buch zurückzugeben. Die Korrelationen waren signifikant und fielen moderat aus. Zur Untersuchung der Konstruktvalidität wurden Korrelationen zu Angst, Copingstilen und Action Control berechnet. Die Zusammenhänge fielen signifikant und in die erwartete Richtung aus. Des Weiteren wurde die Dimensionalität des CFQ betrachtet. Eine Maximum-Likelihood-Faktorenanalyse mit Varimax-Rotation legt eine zweifaktorielle Lösung nahe mit den Faktoren Ablenkbarkeit und Sensomotorische Koordination.

Relevanz für die Fragestellung: Die oben beschriebene prädiktive Validität des CFQ im Bereich Verkehrsdiagnostik lässt vermuten, dass auch für Mensch-Maschine-Teams der Luftfahrt Zusammenhänge zu finden sind.

Fragebogen zur Computernutzung (CNFB)

Der Fragebogen zur Computernutzung (CNFB) wurde von Schroeders und Wilhelm (2011) entwickelt. Die Originalform enthält 18 Items zur Messung der Häufigkeit computerbezogener Aktivitäten und Softwarenutzung. Für das HYBRID-Projekt und diese Arbeit wird eine adaptierte Form mit 34 Items verwendet, die zusätzlich zur Originalskala ergänzende Nutzungsbeispiele enthält.

Hintergrund: Die Computernutzung kann andere Konzepte wie computerbezogene Einstellungen, Emotionen und situationale Faktoren beeinflussen. Die Konzeptualisierung der Computernutzung hängt eng mit dem Konzept der Computererfahrung zusammen. Ein Mangel an Erfahrungen, verursacht durch Mangel an Computernutzung, kann als Mediator zur Erklärung von unterschiedlicher Performance verstanden werden. Die Autoren gehen von einem Model zur Computernutzung aus, das aus einem Generalfaktor und drei verschachtelten inhaltlichen Faktoren - Office, Internet und Games besteht (Schroeders & Wilhelm, 2011).

Aufbau: Der verwendete Fragebogen besteht aus 34 Items mit einem fünfstufigen Antwortformat.

Anwendung: Die Autoren untersuchten Geschlechtsunterschiede mithilfe einer konfirmatorischen Faktorenanalyse und fanden, dass das Geschlecht einen Einfluss auf das Konstrukt der Computernutzung hat.

Gütekriterien: Schroeders und Wilhelm (2011) konnten in ihrer 18-Item-Version eine Struktur mit einem Generalfaktor der Computernutzung und drei verschachtelten inhaltlichen Faktoren (Office, Internet oder Games) nachweisen. Alle Items lassen sich dem Generalfaktor und mit einer Ausnahme ebenfalls einem untergeordneten Faktor zuordnen. Zur Überprüfung der Validität wurden Zusammenhänge zwischen Computernutzungs-Scores und anderen Konstrukten betrachtet. Es gibt positive signifikante Zusammenhänge zur Dauer der Nutzung pro Woche sowie zur Computererfahrung (seit wie vielen Jahren wird der Computer benutzt), zum Zugang zu Computern, zum Selbstkonzept in Bezug auf Computerskills und Bewertungen von Computern. Negative Zusammenhänge gibt es zur fluiden und kristallinen Intelligenz.

Relevanz für die Fragestellung: Beim (CNFB) handelt es sich in Abgrenzung zu den anderen ausgewählten Verfahren nicht um einen Persönlichkeits- oder Einstellungsfragebogen. Die Skala wird dennoch mit einbezogen, da die Computernutzung andere Konzepte wie computerbezogene Einstellungen, Emotionen und situationale Faktoren beeinflussen kann.

Aufgrund der stark computer- und technikbezogenen Arbeit ist ein Zusammenhang zwischen Computernutzung und Leistung im Mensch-Maschine-Team anzunehmen.

Einstellung gegenüber Automation (EgA)

Der Fragebogen zur Einstellung gegenüber Automation (EgA) wurde im Rahmen des HYBRID-Projektes entwickelt. Er setzt sich aus verschiedenen Komponenten zusammen und enthält insgesamt 13 Items. Einbezogen wurden zehn adaptierte und ins Deutsche übersetzte Items der „Complacency Potential Rating Scale“ CPRS von Singh, Molloy und Parasuraman (1993) sowie drei zusätzlich formulierte Items. Inhaltlich befassen sich die Fragen mit verschiedenen Aussagen zu alltäglichen Lebensbereichen, die heutzutage von Automatisierung geprägt sind. Mittels fünfstufigen Antwortformats können Probanden diese Aussagen ablehnen oder ihnen zustimmen. Wie auch zum Fragebogen BnT werden in den Abschnitten 8.1.3 und 8.1.4 deskriptive Statistiken sowie psychometrische Kennwerte der Skala berichtet. Weiterführende Analysen zu Gütekriterien, Einsatzmöglichkeiten etc. müssen in Zukunft erfolgen.

Relevanz für die Fragestellung: Der Fragebogen EgA wurde eigens für die Fragestellung dieser Arbeit zusammengestellt. Es ist zu vermuten, dass Personen mit positiver Einstellung gegenüber einem maschinellen Teampartner auch die Zusammenarbeit mit selbigem besser gelingt.

Einstellung gegenüber Arbeit im Mensch-Maschine-Team (EgMMT)

Auch der Fragebogen zur Einstellung gegenüber Arbeit im Mensch-Maschine-Team (EgMMT) wurde im Rahmen des HYBRID-Projektes entwickelt. Er enthält vier Items mit Aussagen zur Zusammenarbeit mit Computern versus Zusammenarbeit mit Menschen und wurde im Rahmen der Nachbefragung zu HINT eingesetzt (siehe auch unter 7.3). Erneut können Probanden diese Aussagen mittels fünfstufigen Antwortformats ablehnen oder ihnen zustimmen. In den Abschnitten 8.1.3 und 8.1.4 werden die Ergebnisse zum EgMMT berichtet. Weiterführende Analysen zu Gütekriterien, Einsatzmöglichkeiten etc. müssen in Zukunft erfolgen.

Relevanz für die Fragestellung: Der Fragebogen EgMMT wurde eigens für die Fragestellung dieser Arbeit zusammengestellt. Es ist zu vermuten, dass Personen mit positiver Einstellung gegenüber Arbeit im Mensch-Maschine-Team auch bessere Leistungen bei der tat-

sächlichen Arbeit mit einem maschinellen Teampartner erzielen, bzw. gewillter sind, mit diesem zu kooperieren.

Fragebogen zur Erfassung computerspezifischer Attributionen (FEcA)

Dickhäuser und Stiensmeier-Pelster (2000) entwickelten den Fragebogen zur Erfassung computerspezifischer Attributionen (FEcA) zwecks Erhebung computerspezifischer Ursachenerklärungen. Das Instrument misst computerspezifische Erfolgs- und Misserfolgsattributionen auf den Dimensionen Lokalität, Kontrolle, Stabilität, Globalität.

Unterschiede in computerspezifischen Attributionsstilen können Unterschiede im computerbezogenen Verhalten nach sich ziehen (Dickhaeuser & Stiensmeier-Pelster, 2000). Faber (1992) weist darauf hin, dass mit zunehmender Spezifität der Selbstinformation diese in stärkerem Maße handlungsbedeutsam wird. Demnach hängen bereichsspezifische Selbstüberzeugungen enger mit entsprechenden Leistungsmerkmalen zusammen als allgemeine. Die Erfassung computerbezogener Attributionen sollte damit eher Möglichkeiten zur Vorhersage computerbezogenen Verhaltens eröffnen als die Erfassung genereller Attributionstendenzen.

Hintergrund: Der Fragebogen unterscheidet die Dimensionen Lokalität, Kontrolle und Stabilität sowie das Konstrukt der erlernten Hilflosigkeit aus dem die vierte Dimension der Globalität abgeleitet wurde. Nach Dickhäuser und Stiensmeier-Pelster (2000) unterscheiden sich Personen in ihrer dimensionalen Perzeption ein und desselben Ursachenfaktors. Daraus folgt, dass allein die Angabe des Grundes für das Zustandekommen eines Ereignisses zur Erschließung der Attributionsdimensionen nicht ausreicht. Es sollte daher explizit nach der dimensionalen Wahrnehmung von Ursachen gefragt werden. Die Entwicklung der in dem Fragebogen verwendeten Beispielsituationen erfolgte durch die Befragung von 175 Studenten nach Erfolgen und Misserfolgen am Computer.

Aufbau: Der FEcA besteht aus zehn computerbezogenen Situationen (fünf Erfolge, fünf Misserfolge) für dessen Zustandekommen der Hauptgrund notiert werden soll. Anschließend wird der Hauptgrund auf acht Skalen (für jede Dimension wurden zwei Items entwickelt) mit einem fünfstufigen Antwortformat bewertet. Die Auswertung erfolgt für acht Skalen (2 [Erfolg/Mißerfolg] x 4 [Lokalität, Kontrolle, Stabilität, Globalität]).

Anwendung: Der FEcA kann laut Autoren als Forschungsinstrument und zur Untersuchung von Gruppenunterschieden (z.B. Geschlecht) hinsichtlich computerspezifischer Attribution eingesetzt werden.

Gütekriterien: Die Reliabilitätskoeffizienten der acht Skalen liegen zwischen .67 und .83 und werden von den Autoren als gut bis sehr gut interpretiert. Eine von ihnen durchgeführte Faktorenanalyse legt eine einfaktorielle Struktur der Skalen nahe. Korrelationsberechnung mit dem computerspezifischen Selbstkonzept eigener Begabung und Ursachenerklärungen in einer spezifischen Situation lassen auf die kriteriumsbezogene Validität des Instrumentes schließen (Dickhäuser & Stiensmeier-Pelster, 2000).

Relevanz für die Fragestellung: Da wie oben beschrieben, Unterschiede in computerspezifischen Attributionsstilen Unterschiede im computerbezogenen Verhalten nach sich ziehen können, ist anzunehmen, dass die Leistung im Mensch-Maschine-Team mit diesen zusammenhängt.

Fragebogen zur Erfassung individueller Einstellungen gegenüber Teamarbeit (FIT)

Mohiyeddini (2001) konstruierte den Fragebogen zur Erfassung individueller Einstellungen gegenüber Teamarbeit (FIT). Gemessen wird die Einstellung von Personen in Bezug auf „Bereitschaft zur Teamarbeit“ und „Vorbehalte gegenüber Teamarbeit“.

Hintergrund: Teamarbeit ist eine wichtige Arbeitsform zur Erhaltung und Steigerung von Arbeitsleistung und Wettbewerbsfähigkeit. Einstellung und Teamfähigkeit der Mitarbeiter spielen dabei eine wichtige Rolle und können Prädiktoren für das Verhalten sein (Mohiyeddini, 2001).

Aufbau: Erfasst werden die Dimensionen „Bereitschaft zur Teamarbeit“ und „Vorbehalte gegenüber Teamarbeit“. Die Items werden im Original auf einem sechsstufigen Antwortformat beantwortet. Für den Einsatz im HTQ-S wird aus Gründen der Konformität mit den anderen eingesetzten Skalen ein fünfstufiges Antwortformat verwendet. Die Bearbeitungsdauer liegt bei vier bis sieben Minuten.

Anwendung: Der FIT kann laut Testautor im Rahmen der Analyse von Teamproblematiken eigesetzt werden, seine Ergebnisse können dabei Informationen zur Analyse von problematischen Themen und Ableitung von Trainingsmaßnahmen liefern. Des Weiteren kann er bei Platzierungs- und Selektionsentscheidungen helfen und im Forschungsbereich verwendet werden.

Gütekriterien: Die interne Konsistenz der Skala BT liegt bei .84, der Skala VT bei .74. Zur Untersuchung der Validität wurde der Zusammenhang zwischen Selbst- und Fremdrating sowie zwischen der Selbstauskunft vom FIT und der Arbeitszufriedenheit berechnet. Alle Korrelationen fielen signifikant und zwar in die vorher postulierte Richtung aus. Positive

Zusammenhänge ließen sich zudem zwischen BT und der Einschätzung des Stationsklimas in einem Krankenhaus sowie der Zusammenarbeit mit Vorgesetzten finden.

Relevanz für die Fragestellung: Die individuelle Einstellungen gegenüber Teamarbeit bezieht sich selbstverständlich auf die Arbeit mit menschlichen Teampartnern. Trotzdem ist es interessant zu prüfen, ob sich Zusammenhänge mit der Leistung im Mensch-Maschine-Team zeigen.

Inventar zur Computerbildung, revidierte Form (INCOBI-R)

Die revidierte Fassung des Inventars zur Computerbildung (INCOBI-R) von Richter, Naumann und Horz (2010) dient der Erfassung von Computer Literacy (computerbezogene Kompetenzen) und inhaltlich differenzierten computerbezogenen Einstellungen. Es werden drei Aspekte von Computer Literacy erhoben: praktisches und theoretisches Computerwissen sowie Sicherheit im Umgang mit Computern. Unter computerbezogenen Einstellungen werden individuelle Bewertungen von Computern verstanden.

Hintergrund: Computeranwendungen spielen inzwischen in vielen Berufsfeldern, im Bildungs- und privaten Bereich eine wichtige Rolle. Daraus leitet sich ab, dass der Umgang mit dem Computer eine Schlüsselqualifikation in verschiedenen Lebensbereichen darstellt. Es wird davon ausgegangen, dass neben Computer Literacy auch computerbezogene Einstellungen relevant sind, da durch sie Ausmaß und Art der Nutzung von Informationstechnologien mitbestimmt werden können (Richter, Naumann, & Horz, 2010). Computerängstlichkeit, bzw. positiv gepolt Sicherheit im Umgang mit Computern, wird von den Autoren als zeitlich überdauernde Persönlichkeitsdisposition (Trait) mit kognitiven als auch affektiven Komponenten verstanden.

Aufbau: Der INCOBI-R besteht in seiner Komplettversion aus den folgenden Skalen:

- Fragebogen zur inhaltlich differenzierten Erfassung von computerbezogenen Einstellungen (**FIDEC**), 77 Items, fünfstufiges Antwortformat
- Fragebogen zur Sicherheit im Umgang mit Computeranwendungen (**COMA**), 8 Items, fünfstufiges Antwortformat
- Fragebogen zu theoretischem Computerwissen (**TECOWI**), 20 Mehrfachwahlaufgaben, 5 Antwortalternativen
- Fragebogen zu praktischem Computerwissen (**PRACOWI**), 20 Mehrfachwahlaufgaben, 5 Antwortalternativen

- Fragebogen zu soziodemographische Angaben, 10 Fragen

Die computerbezogenen Einstellungen werden über acht Einstellungsskalen operationalisiert, die sich aus der Kombination von drei dichotomen Unterscheidungen ergeben. Es wird inhaltlich unterschieden nach

- evaluativem Fokus (persönliche Erfahrung vs. gesellschaftliche Folgen)
- Nutzungsdomänen (Lernen und Arbeiten vs. Unterhaltung und Kommunikation)
- Valenz (positive Einstellungskomponenten - Computer als nützliches Werkzeug/nützliche Technologie vs. negativen Einstellungskomponenten - Computer als unbeeinflussbare Maschine/unbeeinflussbare Technik).

Für die aktuelle Arbeit wurden die vier der acht Unterskalen des FIDEC (computerbezogene Einstellungen), die sich auf persönliche Erfahrungen beziehen, sowie der COMA (Sicherheit im Umgang mit Computeranwendungen) ausgewählt. Die Skalen werden im Folgenden beschrieben:

- Die Skala FIDEC B kombiniert die Merkmale positive persönliche Erfahrungen mit Computern als Unterhaltungs- und Kommunikationsmittel mit positiver Valenz als nützliches Werkzeug.
- Die Skala FIDEC F kombiniert die Merkmale positive persönliche Erfahrungen mit Computern als Lern- und Arbeitsmittel mit positiver Valenz als nützliches Werkzeug.
- Die Skala FIDEC G kombiniert die Merkmale negative persönliche Erfahrungen mit Computern als Unterhaltungs- und Kommunikationsmittel mit negativer Valenz als unbeeinflussbare Maschine
- Die Skala FIDEC G kombiniert die Merkmale negative persönliche Erfahrungen mit Computern als Lern- und Arbeitsmittel mit negativer Valenz als unbeeinflussbare Maschine

Die übrigen Skalen werden wegen Überschneidungen mit dem CNFB bzw. aufgrund der fehlenden Relevanz reiner Abfragen von Computerwissen (TECOWI und PRACOWI) nicht verwendet.

Anwendung: Das Instrument wurde bereits in Untersuchungen zum computergestützten Lernen eingesetzt, zur computervermittelten Kommunikation, in populationsbeschreibende Untersuchungen, z.B. für nationale Ergänzungsstudien zu PISA sowie zur Untersuchung von Geschlechtsunterschieden.

Gütekriterien: Die interne Konsistenz der Skalen liegt zwischen .76 und .88. Zur Überprüfung der Konstruktvalidität wurden Korrelationen zwischen den Skalen berechnet. Es zeigten sich erwartungskonform höhere Zusammenhänge zwischen Skalen, die konzeptuell zusammenhängende Konstrukte erfassen. Zur Überprüfung der Kriteriumsvalidität wurden Korrelationen mit Maßen der tatsächlichen Computernutzung berechnet. Diese bestätigen die Validität des Instruments. Die inhaltlich differenzierte Struktur der computerbezogenen Einstellungen wurde mithilfe einer konfirmatorischen Faktorenanalyse belegt (Richter et al., 2010).

Relevanz für die Fragestellung: Aufgrund der stark computer- und technikbezogenen Arbeit im Mensch-Maschine-Team ist anzunehmen, dass sowohl computerbezogene Einstellungen als auch die Sicherheit im Umgang mit Computeranwendungen die Leistung beeinflussen können.

Kontrollüberzeugungen im Umgang mit Technik (KUT)

Der Fragebogen Kontrollüberzeugungen im Umgang mit Technik (KUT) von Beier (2004) erfasst ein „Persönlichkeitsmerkmal mit Einfluss auf die Mensch-Technik-Interaktion“ (S. 151). Das Konzept der Kontrollüberzeugungen (Locus of Conrol of Reinforcement) entwickelte Rotter (1954) im Rahmen seiner sozialen Lerntheorie der Persönlichkeit. Kontrollüberzeugungen können definiert werden als „die generalisierten Erwartungen einer Person darüber, in welchem Maße wichtige Ergebnisse und Verstärker vom eigenen Handeln oder eigenen Charakteristika abhängen“ (Beier, 2004, S. 48). Rotter (1966) geht davon aus, dass eine Person generalisierte Kontrollüberzeugungen in ihrer individuellen Lebensgeschichte, also in einem Lernprozess, ausbildet. Der KUT-Fragebogen misst die Kontrollüberzeugungen in einem bereichsspezifischen Kontext – im Umgang mit Technik. Erfasst wird also der Grad, in dem eine Person glaubt, selbst über die Ereignisse und Verstärker in der Interaktion mit Technik bestimmen zu können.

Hintergrund: Nach Beier (2004) ist der Erklärungswert generalisierter Persönlichkeitsmerkmale für menschliches Verhalten in Bezug auf technische Systeme eher gering. Deshalb erfolgte die Entwicklung des Fragebogens zur Messung des bereichsspezifischen Persönlichkeitsmerkmals der Kontrollüberzeugung im Bereich Mensch-Technik-Interaktion.

Aufbau: Der Fragebogen besteht in seiner 2004 veröffentlichten Version aus acht Items, die eine eindimensionale Skala bilden. Für den HTQ-S wurde jedoch eine um vier Items erweiterte Version des KUT verwendet, die eine bessere Differenzierung im oberen Ausprägungsbereich bewirken soll. Diese neue Version wurde direkt vom Testautoren zur Verfü-

gung gestellt. Wie vom Autor empfohlen, wird auch in dieser Untersuchung ein fünfstufiges Antwortformat verwendet.

Anwendung: Einsatz fand der KUT bisher beispielsweise im Bereich Internetrecherche oder Akzeptanz von Assistenzsystemen in der Automobilbranche. Forschungsergebnisse können unter anderem für die Entwicklung und den Vertrieb von technischen System sowie für eignungsdiagnostische Fragestellungen genutzt werden (Beier, 2004).

Gütekriterien: Das ursprünglich dreidimensionale Messinstrument konnte auf eine Dimension gekürzt werden. Diese einfaktorielle Lösung klärt 57,1% der Varianz auf (Beier, 2004). Die Reliabilität des KUT (interne Konsistenz und Retest-Reliabilität) ist mit jeweils rund .90 als sehr gut zu bewerten. Zur Validierung wurden Untersuchungen durchgeführt, die theoretisch vorhergesagte Zusammenhänge zwischen Kontrollüberzeugungen im Umgang mit Technik und unter anderem Variablen der Handlungssteuerung wie Fähigkeit im Umgang mit technischen Geräten, emotionales Erleben und Attribution von Misserfolg belegen (Beier, 2004).

Normierung: Da die KUT-Werte stark abhängig vom Setting sind, ist eine Normierung schwierig. Zur Vergleichbarkeit der Testwerte wurden Prozentskalen erstellt. Beim Vergleich von Teilstichproben zeigten sich höher KUT-Werte für Männer als für Frauen; und niedrigere KUT-Werte für jüngere Menschen als für ältere (Beier, 2004).

Relevanz für die Fragestellung: Laut Beier (2004) stehen Menschen mit hohen technischen Kontrollüberzeugungen neuen Systemen offener gegenüber und wollen mit dem System in Interaktion bleiben. Sie akzeptieren die informierende Assistenz stärker als Personen mit niedrigen Werten im KUT. Diese akzeptieren eine Assistenz dann am stärksten, wenn sie automatisch funktioniert. Es ist daher anzunehmen, dass auch die Leistung in einem Mensch-Maschine-Team, wie es durch HINT simuliert wird, von technischen Kontrollüberzeugungen beeinflusst werden kann.

NEO-Persönlichkeitsinventar, revidierte Form (NEO-PI-R)

Die revidierte Form des NEO-Persönlichkeitsinventars (NEO-PI-R) wurde von Costa und McCrae (1992) publiziert, die deutsche Adaption stammt von Ostendorf und Angleitner (2004). Das NEO-PI-R ist ein Persönlichkeitsstrukturtest zur Erfassung der Big-Five-Dimensionen auf fünf Hauptskalen (mit jeweils sechs Subskalen)

Hintergrund: Das NEO-PI-R beruht auf Annahmen eines Fünf-Faktoren-Modells. Dieses Modell der Big Five Persönlichkeitsfaktoren geht auf Tupes und Christal (1961) zurück. Berth und Goldschmidt (2006) berichten, dass die Ableitung der Faktoren/Facetten auf de-

skriptiven theoretischen Ableitungen der Autoren basiert. Ausgegangen wird von einem hierarchischen Persönlichkeitsmodell, mit fünf Hauptfaktoren der Persönlichkeit mit jeweils sechs Facetten. Dies ermöglicht eine umfassende und differenzierte Persönlichkeitsbeschreibung. Bei den fünf Hauptfaktoren handelt es sich um Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit und Gewissenhaftigkeit, welche im nächsten Abschnitt kurz beschrieben werden. Das Inventar ist derzeit international der „wohl am häufigste eingesetzte Fragebogen zur Messung des Fünf-Faktoren-Modells der Persönlichkeit" (Ostendorf, 2001, S. 409). Die Struktur und Validität der deutschen Version konnte in mehreren Studien mit über 10.000 Probanden nachgewiesen werden.

Aufbau: Die fünf Merkmalsbereiche des NEO-PI-R mit ihren jeweils sechs Facetten sind in **Tabelle 9** auf Seite 86 aufgelistet. In Anlehnung an Berth und Goldschmidt (2006) lassen sich die fünf Hauptfaktoren inhaltlich folgendermaßen beschreiben:

- *Neurotizismus*: Emotionale Empfindsamkeit bzw. Ansprechbarkeit vs. emotionale Robustheit
- *Extraversion*: Geselligkeit, Freundlichkeit, Unternehmungsfreude und Aktivität, Mögen von Menschen und Menschenansammlungen, Liebe von Aufregung und Anregung, Gesprächigkeit, Dominanz, Willensstärke und Optimismus vs. Introversion als Fehlen bzw. geringe Ausprägung dieser Merkmale
- *Offenheit für Erfahrungen*: Interesse an (und Ausmaß der Beschäftigung mit) neuen Erfahrungen, Erlebnissen und Eindrücken vs. Fehlen bzw. geringe Ausprägung dieser Merkmale
- *Verträglichkeit*: Einstellungen und gewohnheitsmäßige Verhaltensweisen in sozialen Beziehungen sind durch Hilfsbereitschaft, Entgegenkommen, Vertrauensbereitschaft und Bemühen, anderen zu helfen (Altruismus) gekennzeichnet vs. Fehlen bzw. geringe Ausprägung dieser Merkmale
- *Gewissenhaftigkeit*: Aktiver Prozess des Planens, der Organisation und Ausführung von Aufgaben, Zielstrebigkeit, Willensstärke, Leistungs- und Pflichtbewusstsein, Genauigkeit, Pünktlichkeit und Zuverlässigkeit vs. Fehlen bzw. geringe Ausprägung dieser Merkmale

Die Items des NEO-PI-R werden auf einem fünfstufigen Antwortformat bearbeitet. Neben der Selbstbeurteilungsform steht auch eine Fremdbeurteilungsform zur Verfügung. Zur Auswertung werden die Rohwerte für jede Facette addiert. Anschließend erfolgt die Addition der Rohwerte der sechs Facetten jeder Dimension zu den Dimensionsrohwerten. Die Bearbeitungsdauer beträgt ca. 30 bis 40 Minuten. Zusätzlich zu den fünf Faktoren wird ein I-

tem erhoben, das das Bemühen nach Ehrlichkeit des Ausfüllers abfragt, so dass insgesamt 241 Items zu beantworten sind.

Anwendung: Das Inventar findet in sehr vielen unterschiedlichen Bereichen der „Forschung und Praxis, z.B. in der Arbeits- und Organisationspsychologie, der Berufsberatung, der Pädagogischen Psychologie, der Gesundheitspsychologie und Verhaltensmedizin und der Klinischen Psychologie“ Einsatz (Ostendorf, 2001, S. 409). Das Instrument wird unter anderem zur Vorhersage beruflicher Erfolgskriterien in der Arbeits- und Organisationspsychologie eingesetzt (siehe hierzu auch unter 3.1).

Gütekriterien: Die internen Konsistenzen der Hauptskalen und Facetten werden in **Tabelle 9** auf Seite 86 berichtet. Die Retest-Reliabilität der Hauptskalen nach fünf Jahren beträgt .76 bis .79, die der Facetten .54 bis .77. Es fanden zahlreiche Untersuchungen zur Überprüfung der Validität statt. Unter anderem wurden Korrelationen mit anderen Fünf-Faktoren-Inventaren, anderen Persönlichkeitstest und Temperamentinventaren berechnet. Des Weiteren wurden Zusammenhänge zwischen Selbst- und Fremdrating und Zusammenhänge zu Intelligenzmaßen und zu Verfahren zur Messung von Persönlichkeitsstörungen berechnet (Ostendorf, 2001). Faktorenanalysen aus zahlreichen Studien mit anderen Verfahren zeigen, dass sich Skalen anderer Verfahren anhand ihrer Ladungen den Dimensionen des NEO-PI-Rs zuordnen lassen (Berth & Goldschmidt, 2006).

Normierung: Es liegen T-Werte, Stanine-Werte und Prozentrangnormen getrennt nach Geschlecht und Altersgruppen vor.

Relevanz für die Fragestellung: Die Verwendung des NEO-PI-R soll ermöglichen, Zusammenhänge der allgemeinen Persönlichkeit mit der Leistung in hoch automatisierten Mensch-Maschine-Teams der Luftfahrt zu erfassen. So soll geprüft werden, ob die unter 3.1 besprochenen Zusammenhänge zwischen den Big Five und beruflicher Leistung auch für zukünftige Luftfahrtoperateure replizierbar sind.

Allgemeine Selbstwirksamkeitserwartung (SWE)

Der von Schwarzer und Jerusalem (1999) entwickelte und revidierte Fragebogen Allgemeine Selbstwirksamkeitserwartung (SWE) ist ein Selbstbeurteilungsverfahren zur Erfassung von allgemeinen optimistischen Selbstüberzeugungen. Unter allgemeiner Selbstwirksamkeitserwartung wird von den Autoren die Erwartung verstanden, mit täglichen Schwierigkeiten und Barrieren zurechtzukommen und kritische Anforderungssituationen aus eigener Kraft erfolgreich bewältigen zu können.

Hintergrund: Die erste Version der Skala wurde mit 20 Items entwickelt. Nach Angaben der Testautoren konnte die Skala jedoch „ohne Qualitätsverluste auf zehn Items gekürzt werden“ (Schwarzer, 1999). Den theoretischen Hintergrund bildet das spezifische Selbstwirksamkeitskonzept (perceived self-efficacy) von Bandura (1977) als ein wesentlicher Aspekt seiner sozial-kognitiven Theorie.

Aufbau: Das Instrument besteht aus einer eindimensionalen Skala mit zehn Items und einem ursprünglich vierstufigen Antwortformat. Auch hier wird für den HTQ-S aus Gründen der Konformität ein fünfstufiges Antwortformat verwendet. Alle Items sind positiv formuliert. Der Testwert einer Person ergibt sich aus der Summe aller zehn Antworten. Ein hoher Wert lässt auf eine hohe optimistische Kompetenzerwartung schließen. Die Bearbeitungsdauer beträgt etwa 4 Minuten.

Anwendung: Der Fragebogen liegt in 27 Sprachen vor. Außerdem gibt es bereichsspezifische Versionen, wie z.B. eine schulbezogene SWE-Skala. Anwendung fand der Fragebogen bisher u.a. im schulischen, kulturellen, medizinischen und psychologischen Kontext (Schwarzer und Jerusalem 1999).

Gütekriterien: Scholz, Gutiérrez-Doña, Sud und Schwarzer (2002) berichten unter Einbezug von 25 Nationen interne Konsistenzen zwischen .75 und .91, in der deutschen Stichprobe einen Wert von .81. Hinsichtlich der Kriteriumsvalidität zeigten sich positive Zusammenhänge zu dispositionalem Optimismus und zur Arbeitszufriedenheit sowie negative Zusammenhänge zu Ängstlichkeit, Depressivität, Burnout und Stresseinschätzungen. Auch eine Vorhersage körperlicher Beschwerden und Depressivität bei DDR-Übersiedlern sowie Genesung und die postoperative Lebensqualität bei Herzpatienten konnte mit dem SWE adäquat getroffen werden (Schwarzer, 2009).

Normierung: Es liegen alters- und geschlechtsspezifische Normwerte (Prozentränge und T-Werte) vor.

Relevanz für die Fragestellung: Die Selbstwirksamkeitserwartung als persönliche Einschätzung der eigenen Kompetenzen könnte im Mensch-Maschine-Team von großer Relevanz sein. Es ist anzunehmen, dass Personen, die davon ausgehen, kritische Anforderungssituationen aus eigener Kraft erfolgreich bewältigen zu können, auch den Anforderungen im Mensch-Maschine-Team besser gerecht werden.

Fragebogen zur Messung von Technikaffinität (TA-EG)

Der Fragebogen zur Messung von Technikaffinität (TA-EG) von Karrer, Glaser, Clemens, und Bruder (2009) erfasst die Einstellung zu und den Umgang mit elektronischen Geräten,

wie z.B. Mobiltelefonen, Computern oder Mp3-Playern. Unter Technikaffinität wird die Anziehung eines Menschen zu technischen Geräten verstanden oder anders formuliert: Technikaffinität ist eine Persönlichkeitseigenschaft „die sich in einer positiven Einstellung, in Begeisterung und Vertrauen einer Person gegenüber Technik ausdrückt. Sie geht mit Interesse an und Akzeptanz von Technik einher und wirkt sich positiv auf das Wissen über und die Erfahrung mit Technik aus“ (Karrer et al., 2009, S. 1).

Hintergrund: Viele Studien und Instrumente beziehen sich auf Affinität zu Computern, aber nur wenige auf Technik allgemein (Karrer et al., 2009). Mit der Entwicklung des TA-EG soll diese Lücke geschlossen werden. Zur Testkonstruktion wurde zunächst ein Itempool durch Studenten und Experten der Mensch-Technik-Interaktion generiert und durch eine Facettenanalyse erweitert. Diese Items wurden einer Hauptkomponentenanalyse mit Varimax-Rotation unterzogen. Es entstanden vier Subskalen. In einer zweiten Untersuchung wurde der Fragebogen nochmals überarbeitet und auf psychometrische Eigenschaften, Struktur und Gütekriterien hin überprüft. Die Endversion enthält 19 Items.
Aufbau: Der TA-EG besteht aus den auch in **Tabelle 7** aufgelisteten vier Subskalen Begeisterung für Technik, Subjektive Kompetenz im Umgang mit Technik, Wahrgenommene positive Technikfolgen und Wahrgenommene negative Technikfolgen. Die Items werden im Fragebogen gemischt. Das Antwortformat ist fünfstufig.

Anwendung: Laut Autorinnen kann der TA-EG z.B. im Rahmen der Evaluation in der Entwicklungsphase technischer Geräte eigesetzt werden. Die Variable Technikaffinität kann als Kontrollvariable, zur Probandenselektion, zur differenzierten Analyse von speziellen Kundensegmenten sowie zur leichtere Interpretation von Ergebnissen aus Studien dienen, in denen elektronische Geräte (z.B. computergestützte Erhebungen) zur Datenerhebung eingesetzt werden (Karrer et al, 2009).

Gütekriterien: Die interne Reliabilität der Skalen liegt zwischen .73 und .86. Die Konstruktvalidität wurde durch Korrelationen zwischen den Skalen des TA-EG und z.B. dem Fragebogen Kontrollüberzeugungen im Umgang mit Technik (KUT, Beier, 2007) belegt. Des Weiteren wurden Zusammenhänge zu Alter, Geschlecht und selbst eingeschätzter Erfahrung berechnet. Die Zusammenhänge sind erwartungskonform und können Befunde der Literatur replizieren: Ältere bzw. weibliche Personen weisen eine geringere selbsteingeschätzte Kompetenz und Begeisterung im Umgang mit EG auf als jüngere bzw. männliche Personen. Die Struktur des TA-EG wurde mittels konfirmatorischer Faktorenanalyse überprüft und bestätigt (Karrer et al, 2009).

Relevanz für die Fragestellung: Aufgrund der stark computer- und technikbezogenen Arbeit ist ein Zusammenhang zwischen Technikaffinität und Leistung im Mensch-Maschine-Team anzunehmen.

8.1.2 Technische Umsetzbarkeit der HTQ-Skalen

Voraussetzung für den Einsatz der soeben beschriebenen Skalen im HTQ-S ist – abgesehen von den psychometrischen Kennwerten, die im Abschnitt 8.1.4 berichtet werden – die Möglichkeit, die Skalen computergestützt umsetzen zu können. Dies gelang für alle bis auf eine der Skalen. Der Fragebogen FEcA (Fragebogen zur Erfassung computerspezifischer Attributionen) war im HTQ-1 enthalten, erwies sich jedoch für die Durchführung am Computer als nicht geeignet und wurde deshalb aus der Testbatterie entfernt. Die Versuchsteilnehmer hatten große Schwierigkeiten, das Aufgabenprinzip sowie die Formulierung der Antwortmöglichkeiten zu verstehen. Zudem erwies sich die Dauer des Tests von durchschnittlich 40 Minuten als unökonomisch.

Alle übrigen Skalen konnten von den Probanden problemlos computerbasiert bearbeitet werden und wurden, wie bereits unter 7.1 erwähnt, über das im DLR eingesetzte Testsystem CAT (Computer Assisted Testing) bzw. auf der Testplattform Inquisit ^{TM} vorgegeben.

8.1.3 Deskriptive Statistiken der HTQ-Skalen

Im Folgenden werden die deskriptiven Statistiken des HTQ-S vorgestellt. **Tabelle 8** auf den nachfolgenden Seiten zeigt die Mittelwerte (M) und Standardabweichungen (SD) des HTQ-S für die gesamte Stichprobe sowie getrennt nach Geschlecht als auch Bewerbertyp. Sofern signifikante Mittelwertsunterschiede zwischen den Gruppen vorliegen, werden zudem die Signifikanzen (p) und Effektstärken (d) der durchgeführten t-Tests dargestellt. Die Histogramme der Variablen hinsichtlich der Gesamtstichprobe sind im Anhang aufgeführt.

Tabelle 8: Deskriptiva des HTQ-S

	Gesamt		*Männer*		*Frauen*		*t-Test*		*Lotsen-Bewerber*		*Piloten-Bewerber*		*t-Test*	
HTQ-S	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*	*d*	*M*	*SD*	*M*	*SD*	*p*	*d*
Balloon Analogue Risk Task	3.49	1.02	3.45	1.00	3.60	1.07			3.40	.99	3.66	1.06		
BFI-10: Neurotizismus	3.08	.44	3.09	.41	3.07	.53			3.12	.42	3.01	.48		
BFI-10: Extraversion	3.18	.46	3.19	.47	3.15	.43			3.15	.48	3.23	.42		
BFI-10: Offenheit	3.33	.65	3.37	.64	3.21	.67			3.29	.65	3.39	.67		
BFI-10: Verträglichkeit	3.50	.45	3.53	.43	3.42	.49			3.44	.42	3.60	.47	*	.36
BFI-10: Gewissenhaftigkt.	3.43	.44	3.41	.47	3.48	.34			3.41	.39	3.46	.51		
BIP: Teamorientierung	2.92	.32	2.92	.31	2.90	.33			2.88	.33	2.98	.30		
BIP: Flexibilität	3.29	.27	3.29	.26	3.28	.28			3.31	.26	3.25	.27		
Bedürfnis nach Teamarbeit	3.88	.59	3.88	.57	3.87	.64			3.84	.59	3.94	.59		
CaP: Risikotoleranz	3.63	.49	3.62	.46	3.68	.56			3.66	.47	3.60	.52		
Cognitive Failures Questionnaire	2.27	.42	2.25	.42	2.32	.40			2.28	.45	2.26	.34		
Computernutzungs-fragebogen	3.31	.47	3.37	.43	3.15	.52	*	.46	3.24	.48	3.42	.42	*	.40
Einstellung ggü. Automation	2.87	.34	2.91	.35	2.76	.31	**	.48	2.82	.35	2.97	.31	**	.45
Einstellung ggü. Arbeit im Mensch-Maschine-Team	2.58	.72	2.50	.72	2.78	.67	*	.41	2.53	.68	2.67	.78		
FIT: Bereitschaft zur Teamarbeit	4.15	.47	4.13	.50	4.20	.37			4.15	.45	4.15	.51		
FIT : Vorbehalte ggü. Teamarbeit	3.00	.62	2.99	.61	3.02	.65			2.96	.65	3.07	.56		

Fortsetzung auf der nächsten Seite

	Gesamt		*Männer*		*Frauen*		*t-Test*		*Lotsen-Bewerber*		*Piloten-Bewerber*		*t-Test*	
HTQ-S	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*	*d*	*M*	*SD*	*M*	*SD*	*p*	*d*
FIDEC B: Pos. Erfahrungen Unterhaltung	3.80	.53	3.84	.51	3.69	.57			3.80	.49	3.80	.61		
FIDEC F: Pos. Erfahrungen Lernen & Arbeit	4.04	.45	4.02	.44	4.08	.47			3.96	.42	4.18	.47	**	.48
FIDEC G: Neg. Erfahrungen Unterhaltung	2.25	.45	2.20	.40	2.38	.53	*	.39	2.27	.44	2.21	.45		
FIDEC H: Neg. Erfahrungen Lernen & Arbeit	2.15	.47	2.08	.40	2.33	.59	**	.51	2.20	.49	2.05	.43		
Sicherheit im Umgang mit Computern	2.72	.24	2.71	.23	2.75	.27			2.72	.23	2.72	.25		
Kontrollüberzeugungen im Umgang m. Technik	3.01	.22	3.01	.21	3.00	.25			3.00	.22	3.03	.23		
NEO-PI-R: Neurotizismus	2.56	.39	2.49	.39	2.76	.35	**	.72	2.58	.40	2.54	.38		
NEO-PI-R: Extraversion	3.59	.33	3.57	.32	3.63	.34			3.59	.32	3.58	.34		
NEO-PI-R: Offenheit	3.41	.35	3.35	.35	3.56	.28	**	.64	3.41	.38	3.42	.29		
NEO-PI-R: Verträglichkeit	3.41	.30	3.38	.30	3.50	.28	*	.40	3.46	.28	3.33	.31	**	.43
NEO-PI-R: Gewissenhaftigk.	3.59	.35	3.58	.37	3.60	.30			3.58	.33	3.59	.39		
Selbstwirksamkeits-erwartung	3.62	.41	3.64	.41	3.57	.40			3.60	.33	3.66	.52		
TA-EG: Begeisterung	3.72	.67	3.79	.66	3.56	.67			3.65	.65	3.85	.69		
TA-EG: Subj. Kompetenz	3.71	.38	3.73	.38	3.66	.38			3.71	.36	3.71	.42		
TA-EG: Pos. Technikfolg.	3.83	.47	3.83	.47	3.81	.47			3.75	.46	3.96	.47	**	.45
TA-EG: Neg. Technikfolg.	2.61	.55	2.62	.55	2.58	.55			2.58	.52	2.66	.60		

Fortsetzung auf der nächsten Seite

	Gesamt		*Männer*		*Frauen*		*t-Test*		*Lotsen-Bewerber*		*Piloten-Bewerber*		*t-Test*	
HTQ-S	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*	*d*	*M*	*SD*	*M*	*SD*	*p*	*d*
Wochenstunden Computer insg.	3.06	.65	3.12	.64	2.93	.67			3.01	.69	3.16	.57		
Jahre Computernutzung	3.46	.75	3.48	.75	3.40	.76			3.42	.72	3.53	.79		
PC-/Laptop-Nutzung	.99	.11	.99	.09	.98	.15			.99	.10	.98	.13		
Smartphone-Nutzung	.75	.43	.77	.42	.70	.46			.67	.47	.89	.31		
Spielkonsolen-Nutzung	.40	.49	.41	.49	.40	.49			.46	.50	.31	.47		
MP3-Player-Nutzung	.83	.38	.79	.41	.93	.26			.83	.38	.82	.39		
Nutzung tragbarer Videospiele	.33	.47	.29	.46	.42	.50			.34	.47	.31	.47		
Computernutzung Schule/Arbeit/ Ausbildung/Uni	.69	.46	.70	.46	.67	.47			.62	.49	.82	.39	**	.44
Computernutzung Zuhause	1.00	.00	1.00	.00	1.00	.00			1.00	.00	1.00	.00		
Computernutzung Öffentlichkeit	.37	.48	.41	.49	.28	.45			.31	.46	.49	.50	*	.38
Computernutzung bei Freunden	.65	.48	.64	.48	.70	.46			.64	.48	.67	.47		
Wochenstunden Computerspiele insg.	1.08	1.10	1.34	1.14	.42	.63			1.16	1.16	.95	.99		
Wochenstunden Computer-Strategiespiele	.52	.76	.65	.81	.19	.45	**	.70	.53	.81	.49	.66		
Wochenstunden Computer-Jump'n-Run	.21	.44	.22	.46	.19	.39			.21	.41	.22	.50		
Wochenstunden Computer-Egoshooter	.38	.78	.52	.88	.00	.00	**	.84	.41	.87	.33	.58		
Wochenstunden Computerschach	.08	.28	.09	.29	.07	.26			.11	.31	.04	.19		
Wochenstunden Computer-Rollenspiele	.28	.61	.35	.67	.09	.37	*	.47	.32	.65	.20	.52		

Fortsetzung auf der nächsten Seite

	Gesamt		*Männer*		*Frauen*		*t-Test*		*Lotsen-Bewerber*		*Piloten-Bewerber*		*t-Test*	
HTQ-S	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*	*d*	*M*	*SD*	*M*	*SD*	*p*	*d*
Wochenstunden Computer-Lernspiele	.19	.41	.16	.39	.28	.45			.19	.39	.20	.45		
Wochenstunden Computer-Sportspiele	.36	.63	.46	.69	.09	.29	**	.69	.42	.68	.25	.52		
Wochenstunden weitere Computer-spiele	.40	.66	.44	.72	.30	.46			.44	.70	.35	.58		
	N=156		N=113		N=43				N=101		N=55			

*Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig),*
M = Mittelwert, SD = Standardabweichung, d = Effektstärke Cohens d

Wie ersichtlich, weisen einzelne Skalen Unterschiede zwischen den Geschlechtern bzw. Bewerbergruppen auf. Aufgrund der Vielzahl der Skalen ohne Mittelwertsunterschiede wird es als legitim betrachtet, die in Kapitel 7 beschriebenen Analysen mit der Gesamtstichprobe durchzuführen. Wie unter 7.3.2 besprochen, werden jedoch im Rahmen der Korrelationsanalysen ergänzend Partialkorrelationen hinsichtlich der Variable Geschlecht durchgeführt.

8.1.4 Psychometrische Kennwerte der HTQ-Skalen

Die psychometrischen Kennwerte der HTQ-Skalen werden in der Arbeit von Knappe (2010) ausführlich besprochen und daher hier nicht in Gänze wiederholt, sondern in einer Zusammenfassung dargestellt. Die internen Konsistenzen zur Schätzung der Reliabilität der Skalen werden im Einzelnen berichtet. **Tabelle 9** fasst die Werte von Cronbachs α zum NEO-PI-R im Vergleich der HTQ-1-Stichprobe mit den von den Testautoren berichteten Werten zusammen. Dabei zeigt sich, dass die von den Testautoren berichteten Werte in der HTQ-1-Stichprobe repliziert werden konnten. Bis auf zwei Facetten („Aktivität“ und „Werte“), deren Cronbachs α im HTQ-1 sichtbar niedriger liegt als in der Normstichprobe, fallen alle anderen Reliabilitätskoeffizienten gleich hoch oder nur etwas niedriger aus.

Tabelle 9: Vergleich der internen Konsistenzen (Cronbachs α) des NEO-PI-R zwischen dem HTQ-1 und den Ergebnissen von Ostendorf und Angleitner (2004)

Merkmalsbereiche und Facetten	*Cronbachs α HTQ-1 (2010)*	*Cronbachs α Ostendorf und Angleitner (2004)*
Neurotizismus	**.91**	**.92**
Ängstlichkeit	.74	.82
Reizbarkeit	.60	.73
Depression	.78	.85
Soziale Befangenheit	.66	.72
Impulsivität	.61	.64
Verletzlichkeit	.75	.79
Extraversion	**.88**	**.89**
Herzlichkeit	.68	.71
Geselligkeit	.77	.80
Durchsetzungsfähigkeit	.78	.80
Aktivität	.49	.70
Erlebnissuche	.57	.60
Frohsinn	.73	.80
Offenheit für Erfahrungen	**.85**	**.89**
Phantasie	.73	.81
Ästhetik	.78	.78
Gefühle	.68	.76
Handlungen	.60	.67
Ideen	.74	.77
Werte	.33	.53
Verträglichkeit	**.87**	**.87**
Vertrauen	.72	.76
Freimütigkeit	.62	.64
Altruismus	.71	.70
Entgegenkommen	.61	.64
Bescheidenheit	.73	.74
Gutherzigkeit	.62	.60
Gewissenhaftigkeit	**.91**	**.90**
Kompetenz	.59	.65
Ordnungsliebe	.68	.70
Pflichtbewusstsein	.69	.67
Leistungsstreben	.71	.70
Selbstdisziplin	.84	.81
Besonnenheit	.74	.74

Die internen Konsistenzen der weiteren Skalen des HTQ-S werden in **Tabelle 10** berichtet. Die Tabelle enthält die anhand Cronbachs α geschätzten Reliabilitäten aus der Erprobung des HTQ-1 im Jahr 2010 sowie zur Überprüfung auch diejenigen aus der Erprobung des HTQ-2 im Jahr 2011. Da sich eine computergestützte Umsetzung des FEcA wie im Abschnitt 8.1.2 beschrieben, als nicht praktikabel erwies, konnte hierzu keine Berechnung von Testkennwerten durchgeführt werden. Der Fragebogen wurde aus der Testbatterie entfernt.

Tabelle 10: Interne Konsistenzen (Cronbachs α) der Skalen des HTQ-1 und HTQ-2

Skala	*Cronbach's α HTQ-1 (2010)*	*Cronbach's α HTQ-2 (2011)*
BIP: Flexibilität	.74	.77
BIP: Teamorientierung	.84	.84
BnT: Bedürfnis nach Teamarbeit		.77
CaP: Risikotoleranz		.45
CFQ: Häufigkeit von Alltagsfehlern	.82	.84
CNFB: Computernutzung	.86	.86
EgA: Einstellung gegenüber Automation		.75
FIT: Bereitschaft zur Teamarbeit	.66	.72
FIT: Vorbehalte gegenüber Teamarbeit	.59	.53
INCOBI-R: Sicherheit im Umgang mit Computeranwendungen (COMA)	.85	.80
INCOBI-R: Positive Erfahrungen mit Computern hinsichtlich Unterhaltung (FIDEC B)	.83	.82
INCOBI-R: Positive Erfahrungen mit Computern hinsichtlich Lernen & Arbeit (FIDEC F)	.84	.83
INCOBI-R: Negative Erfahrungen mit Computern hinsichtlich Unterhaltung (FIDEC G)	.78	.72
INCOBI-R: Negative Erfahrungen mit Computern hinsichtlich Lernen & Arbeit (FIDEC H)	.84	.79
KUT: Kontrollüberzeugungen im Umgang mit Technik	.89	.87
SWE: Allgemeine Selbstwirksamkeitserwartung	.73	.79
TA-EG: Begeisterung für Technik	.78	.78
TA-EG: Subjektive Kompetenz im Umgang mit Technik	.74	.75
TA-EG: Wahrgenommene negative Technikfolgen	.71	.71
TA-EG: Wahrgenommene positive Technikfolgen	.64	.69

Für die Big Five Kurzskala BFI-10 wurde auf eine Berechnung von Cronbachs α verzichtet, da jeder der fünf Faktoren mit nur je zwei Items gemessen wird. Abschnitt 8.1.1 berichtet Analysen der Testautoren.

Bei Betrachtung der Kennwerte fällt die Skala Risikotoleranz aus dem CaP auf, die nur ein α von .45 erreicht. α-Werte ≤ .50 gelten gemeinhin als inakzeptabel, da das Konstrukt der Risikotoleranz jedoch eher breit angelegt ist (siehe unter 8.1.1), was die interne Konsistenz von Skalen mindert, wird die Skala weiter verwendet. Die übrigen in **Tabelle 10** dargestellten α-Werte reichen von α =.53 bis .89. In Anbetracht der teilweise geringen Itemanzahl der Skalen sind diese Werte als zufriedenstellend bis sehr gut einzuschätzen. Zu beachten ist außerdem, dass Persönlichkeitsfragebögen im Allgemeinen ohnehin eine geringere Reliabilität aufweisen als Leistungsverfahren. Laut Maschke (2005) liegt die Reliabilität von Methoden zur Messung von Persönlichkeit meist unter .80, im Falle von Rating-Skalen oft nur bei .60. Auch anhand des Reliabilitätsvergleichs zum NEO-PI-R (siehe oben) wird deutlich, dass trotz der durch verschiedene Merkmale vorselektierten Stichprobe von Luftfahrtbewerbern ähnlich hohe Reliabilitäten wie in der Allgemeinbevölkerung erzielt werden können. Die Befürchtung, es mit einer Stichprobe zu tun zu haben, bei der Persönlichkeitsfragebögen nicht oder anders funktionieren als bei anderen Populationen, erweist sich damit als unbegründet.

Der objektive Persönlichkeitstest BART sowie die Skala EgMMT (Einstellung gegenüber Arbeit im Mensch-Maschine-Team) wurden erst in der Kombinationsstudie zu HTQ-S und HINT eingesetzt. Eine Berechnung von Cronbachs α ist für BART nicht möglich, da in dieser Arbeit lediglich mit einem Gesamtwert der BART und nicht auf Item-Basis gearbeitet wird. Das Cronbach's α für EgMMT liegt bei .62 und damit im zufriedenstellenden Bereich. Auch die Ergebnisse der Itemanalysen bestätigen die Qualität der Skalen. Knappe (2010) berichtet zur Erprobung des HTQ-1, dass die Item-Trennschärfen insgesamt gesehen zufriedenstellend bis sehr gut ausfallen. Damit repräsentieren die Items ihre Skalen auf eine angemessene Weise. Trotz einiger extremer Items sind die Schwierigkeiten der Skalen im akzeptablen bis sehr guten Bereich angesiedelt. Weiterhin wird berichtet, dass der Mittelwert vieler Items bei den meisten technikbezogenen Skalen über dem theoretischen Skalenmittel von drei liegt. Dies entspricht der Erwartung, dass die untersuchte Population Technik gegenüber grundsätzlich positiv eingestellt ist. Dennoch sind die Skalen (mit Ausnahme der biografischen Angaben zur wöchentlichen Computernutzung) mindestens annähernd normalverteilt (siehe Anhang A)

Zusammenfassend sind die Erprobungen der HTQ-Skalen als sehr erfolgreich zu bewerten. Die Ergebnisse lassen auf die Geeignetheit fast aller Fragebögen für tiefergehende Analysen schließen.

8.1.5 Endgültige Skalen des HTQ-S

Die Zusammenstellung der Skalen des HTQ-S entspricht den in **Tabelle 10** aufgelisteten Fragebögen, ergänzt um den beschriebenen objektiven Persönlichkeitstest BART sowie die Skala EgMMT.

8.2 Deskriptive Statistiken der HINT-Variablen

Tabelle 11 stellt die deskriptiven Statistiken der unter 7.2.2 genauer beschriebenen HINT-Variablen vor. Wie dort bereits erwähnt, sind die Variablen „Reaktionszeit", „Target-Abweichungen Alpha" sowie „Target-Abweichungen Bravo" Durchschnittswerte aus jeweils mehreren zuvor z-transformierten Ursprungsvariablen. Sie haben daher zwar noch immer einen Mittelwert von 0, jedoch durch die Kombination aus mehreren Variablen andere Standardabweichungen als 1. Die restlichen Variablen wurden aus jeweils nur einer Ursprungsvariablen gebildet und haben entsprechend der z-Transformation jeweils einen Mittelwert von 0 und eine Standardabweichung von 1. Nichts desto trotz werden die Werte in der Tabelle präsentiert, da diese auch noch einmal gut verdeutlicht, für welche Variablen jeweils welche Stichprobengröße zur Verfügung steht (siehe hierzu auch unter 7.3.1).

Tabelle 11: Deskriptive Statistiken der HINT-Variablen

		Korrekte CAC-Bearbeitung	*Reaktionszeit*	*RQ Alpha an Bravo*	*RQ-Ablehnung durch Alpha*	*RQ-Annahme durch Alpha*	*Target-Abweichungen Alpha*	*Target-Abweichungen Bravo*
N	Gültig	156	154	69	154	154	69	69
	Fehlend	0	2	87	2	2	87	87
Mittelwert		.00	.00	.00	.00	.00	.00	.00
Standardabweichung		1.00	.82	1.00	1.00	1.00	.44	.57

Die Histogramme der HINT-Variablen wurden auf Normalverteilung gesichtet und sind im Anhang aufgeführt. Für die Variable „Korrekte CAC-Bearbeitung" stellte sich heraus, dass

diese aus Mangel an Varianz für weitere Analysen ungeeignet ist. Bis auf einzelne Probanden erzielen alle denselben Testwert. Die Güte der CAC-Bearbeitung lässt sich also scheinbar nicht anhand dessen erfassen, ob CAC korrekt bearbeitet werden – das tun fast alle Probanden gleichermaßen – sondern lediglich anhand der Geschwindigkeit der Bearbeitung. Diese wird mit der Variablen „Reaktionszeit" erfasst. Die Variable „Korrekte CAC-Bearbeitung" wird von weiteren Analysen ausgeschlossen. Für alle anderen Variablen kann anhand der Histogramme eine annähernde Normalverteilung angenommen werden.

Hinsichtlich der Interkorrelationen der Variablen ergibt sich folgendes Bild (**Tabelle 12**, dargestellt sind der Übersicht halber ausschließlich die signifikanten Korrelationen):

Tabelle 12: Interkorrelationen der HINT-Variablen

	Reaktionszeit	*RQ Alpha an Bravo*	*RQ-Ablehnung durch Alpha*	*RQ-Annahme durch Alpha*	*Target-Abweichungen Alpha*	*Target-Abweichungen Bravo*
Reaktionszeit		-.34**		-.32**		.29*
N		68		154		68
RQ Alpha an Bravo			-.25*	.36**	-.46**	-.47**
N			68	68	69	69
RQ-Ablehnung durch Alpha				-.82**		
N				154		
RQ-Annahme durch Alpha						
N						
Target-Abweichungen Alpha						.40**
N						69
Target-Abweichungen Bravo						
N						

*Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig)*

Wie aus der Tabelle ersichtlich, gehen höhere Werte in der Reaktionszeit, d.h. langsameres Arbeiten, mit weniger gestellten Requests, weniger angenommenen Requests sowie höheren Target-Abweichungen im Sektor Bravo einher. Je mehr Requests durch die Probanden selbst gestellt werden, desto weniger lehnen diese auch von Bravo erhaltene Requests ab

und nehmen mehr an. Zusätzlich liegen desto geringere Abweichungen in beiden Sektoren vor.

Letzteres verdeutlicht, dass es sich positiv auf die HINT-Leistung auswirkt (geringe Abweichungen), viele Requests zu stellen, also den Austausch mit dem automatisierten Partner zu suchen. Der hohe Zusammenhang der Variablen RQ-Ablehnung und RQ-Annahme durch Alpha erklärt sich von selbst, da hier lediglich verfallene Requests (siehe 7.2.2) einen vollständigen positiven linearen Zusammenhang verhindern. Zuletzt gehen höhere Target-Abweichungen im Sektor Alpha mit ebenfalls höheren Abweichungen im Sektor Bravo einher.

8.3 Korrelationsergebnisse

In den folgenden Abschnitten werden korrelative Zusammenhänge zwischen HTQ-S und Eignungsdiagnostik sowie HINT berichtet. Ergänzt werden Geschlechtseffekte in den Korrelationen. Aufgrund der Vielzahl der verwendeten Variablen werden in den Ergebnistabellen jeweils nur die signifikanten Korrelationen dargestellt.

8.3.1 Korrelationen mit aktueller Eignungsdiagnostik

Zur Berechnung korrelativer Zusammenhänge mit aktueller Eignungsdiagnostik werden die unter 7.1.4 aufgelisteten Variablen aus den Auswahluntersuchungen von Nachwuchsfluglotsen und –piloten verwendet. Zunächst werden Zusammenhänge mit den Basisleistungen (VA/ BU) in der Auswahl dargestellt.

Die Ergebnisse zur VA (**Tabelle 13**) auf der folgenden Seite zeigen, dass insgesamt drei der Persönlichkeitsskalen aus dem HTQ-S signifikant mit den Basisleistungen der Fluglotsenauswahl zusammenhängen. Höhere Kontrollüberzeugungen im Umgang mit Technik und ein stärkeres Kompetenzgefühl bezüglich Technik gehen mit besserer Leistung in der VA einher. Ebenso geringere Werte hinsichtlich negativer Erfahrungen mit Computern als Lern- und Arbeitsmittel (FIDEC H). Die überwiegende Anzahl der signifikanten Zusammenhänge besteht mit Variablen aus dem Bereich der Computernutzung. Hierbei handelt es sich nicht um Skalen, sondern um einzelne Items, die in biografischer Form abfragen, ob, in welchem Umfang und in welcher Art Computer und computerähnliche Geräte von den Bewerbern verwendet werden. Sie stammen aus dem biografischen Fragebogen (BIO) sowie dem Computernutzungsfragebogen (CNFB). Intensivere Computernutzung korrespondiert mit

jeweils besseren Leistungen in der VA. Das Item „Wochenstunden Computerspiele insgesamt“ zeigt dabei die meisten Zusammenhänge mit den VA-Werten. Personen, die sich mehrere Stunden wöchentlich mit Computerspielen beschäftigen, schneiden also insgesamt besser in der VA ab, als Personen, die dies weniger oder gar nicht tun. Hier lässt sich jedoch die Variable Geschlecht als Einflussfaktor vermuten, die eventuell sowohl auf die Beschäftigung mit Computerspielen als auch auf das Abschneiden in der VA einwirkt. Diese Vermutung wird am Ende dieses Abschnittes noch einmal aufgegriffen.

Tabelle 13: Korrelationen des HTQ-S mit Basisleistungen in der Fluglotsenauswahl

	Mittlere Testleistung	*Teilbereich Konzentration*	*Teilbereich Aufmerksamkeit*	*Teilbereich Merkfähigkeit*	*Teilbereich Raum-orientierung*
Kontrollüberzeugungen im Umgang mit Technik				.16*	
TA-EG: Subjektive Kompetenz	.14*		.15*	.14*	
FIDEC H: Neg. Erfahrungen Lernen & Arbeit			-.15*		
Jahre Computernutzung					.17*
Gameboy / Playstation portable / Nintendo DS				.15*	
Wochenstunden Computer insgesamt	.17*	.14*	.15*		
Wochenstunden Computerspiele insgesamt	.22**	.23**		.18**	.22**
Wochenstunden Computer-Strategiespiele	.15*	.16*			
Wochenstunden Computer-Rollenspiele		.14*			
Wochenstunden Computer-Sportspiele			.16*		

Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig), N = 214

Von der Anzahl der Prädiktoren her wesentlich mehr, jedoch ebenfalls geringe Zusammenhänge ergeben sich mit den Basisleistungen in der Pilotenauswahl (siehe **Tabelle 14** auf der nächsten Seite). Insbesondere der Teilbereich Technikwissen hängt mit verschiedenen Prädiktoren zusammen. Überwiegend handelt es sich dabei um diejenigen, die eine positive Einstellung gegenüber Technik / Computern und hohe Kontrollüberzeugungen sowie Sicherheit in diesem Bereich beinhalten. Hinsichtlich der allgemeinen Persönlichkeit zeigt Extraversion die eindeutigsten Befunde. Hier sind es die introvertierteren Bewerber, die in den Basisleistungen besser abschneiden. Ein auffälliger Befund zeigt sich bezogen auf die spezifische Kategorie der Computer-Rollenspiele. Hierbei handelt es sich um Spiele, in denen Spieler die Rollen fiktiver Charaktere bzw. Figuren übernehmen und soziale Situationen bzw. Abenteuer in einer erdachten Computerwelt erleben. Es ergeben sich negative Zusammenhänge zu drei der fünf Teilbereiche der BU. Bewerber, die sich mehrere Stunden pro Woche mit derartigen Spielen beschäftigen, schneiden schlechter bei den Tests zur Merkfähigkeit, Konzentration und Raumorientierung ab, als Bewerber, die sich nicht damit beschäftigen.

Tabelle 14: Korrelationen des HTQ-S mit Basisleistungen in der Pilotenauswahl

	Mittlere Testleistung	*Teilbereich Technikwissen*	*Teilbereich Mathematik*	*Teilbereich Merkfähigkeit*	*Teilbereich Konzentration*	*Teilbereich Raumorientierung*
BFI-10: Verträglichkeit			-.18*			
BFI-10: Gewissenhaftigkeit					.14*	
BFI-10: Extraversion	-.19**	-.19**				-.20**
Bedürfnis nach Teamarbeit		-.20**				
CaP: Risikotoleranz	-.16*	-.18*				
Sicherheit im Umgang mit Computern		.16*				
TA-EG: Begeisterung		.22**				
TA-EG: Subjektive Kompetenz		.18*				
TA-EG: Positive Technikfolgen		.18**	.16*			
Kontrollüberzeugungen im Umgang mit Technik		.19**				
Einstellung gegenüber Automation		.20**	.21**			
FIDEC B: Pos. Erfahrungen Unterhaltung		.16*	.15*			
FIDEC F: Pos. Erfahrungen Lernen & Arbeit		.18*				
FIDEC G: Neg. Erfahrungen Unterhaltung		-.21**	-.23**			
Wochenstunden Computer insgesamt		.14*				
Jahre Computernutzung	.18*	.22**	.18*			
Wochenstunden Computer-Rollenspiele				-.17*	-.17*	-.18*
Wochenstunden weitere Computerspiele			-.18*			

*Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig), N = 196*

Die signifikanten Korrelationen zwischen der Gerätestufe (Fluglotsen- als auch Pilotenauswahl) und dem HTQ-S (**Tabelle 15** auf der Folgeseite) bewegen sich überwiegend in mittlerer Höhe. Bis auf die Stichprobe zum MIC, der im Rahmen der BU durchgeführt wird, sind die Stichproben bzgl. der Gerätetests deutlich kleiner als hinsichtlich der Basisleistungen. Dies resultiert aus der Ablehnung von Bewerbern auf der vorigen Auswahlstufe.

Wie auch auf Ebene der Basisleistungen zeigt sich, dass Extraversion für das Abschneiden in den Testverfahren negativ wirken kann. Im Strip Display Management Test (SDM) schneiden extravertiertere Bewerber schlechter ab als introvertiertere. Die stärksten Zusammenhänge der HTQ-S-Skalen ergeben sich mit dem Simulator-Test (SIM). Zu angenommenen negativen Technikfolgen liegt mit .51** die deutlichste Korrelation mit der SIM-Leistung vor.

Ähnlich wie der negative Effekt von Computer-Rollenspielen auf die Basisleistungen in der Pilotenauswahl, wirken die sogenannten Jump'n-Run-Spiele kontraproduktiv auf zwei der Gerätetests. Sowohl im DAC als auch im SIM schneiden Bewerber, die sich mit solchen Spielen beschäftigen, schlechter ab als diejenigen, die dies weniger oder gar nicht tun. Positiv wirken dagegen Computer-Schach (positive Korrelation mit SIM-Leistung) und Computer-Sportspiele (positive Korrelation mit SDM-Leistung). Vorteilhaft für das Abschneiden im SDM ist zudem auch eine höhere Anzahl von Wochenstunden am Computer insgesamt, also unabhängig von der Art der Beschäftigung.

Tabelle 15: Korrelationen des HTQ-S mit der Gerätestufe beider Auswahlverfahren

	DAC	*SDM*	*MIC*	*SIM*
BFI-10: Extraversion		-.27*		
BFI-10: Neurotizismus				.35*
BFI-10: Offenheit für Erfahrungen			-.18*	
CaP: Risikotoleranz			-.19**	
BIP: Flexibilität	-.29*			
TA-EG: Positive Technikfolgen	-.28*			
TA-EG: Negative Technikfolgen				.51**
Smartphone-Nutzung				.32*
Wochenstunden Computer insgesamt		.38**		
Wochenstunden Computerspiele insgesamt		.27*	.20**	
Wochenstunden Computer-Jump'n-Run	-.33*			-.36*
Wochenstunden Computer-Sportspiele		.28*		
Wochenstunden Computer-Schach				.37*
Wochenstunden weitere Computerspiele				.35*
	N = 58	N = 57	N = 196	N = 39

*Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig); reduzierte Stichprobengrößen resultieren aus der Ablehnung von Bewerbern auf der vorigen Auswahlstufe*

Wie unter 7.1.5 erläutert, ist Geschlecht als Einflussfaktor auf die soeben berichteten Korrelationen nicht auszuschließen. Um reine Geschlechtseffekte von tatsächlichen Korrelationen des HTQ-2 mit den abhängigen Variablen zu trennen, werden für die gefundenen Zusammenhänge Partialkorrelationen mit der Variablen Geschlecht berechnet.

Tabelle 16 (siehe nächste Seite) stellt zunächst diejenigen Variablen des HTQ-2 vor, die sowohl mit Variablen aus der Eignungsdiagnostik für Fluglotsen als auch mit Geschlecht korrelieren. Variablen, die ausschließlich mit Geschlecht oder ausschließlich mit der Eignungsdiagnostik korrelieren, sind nicht aufgeführt. Ergänzt sind die Korrelationen von HTQ-2 und Eignungsdiagnostik für Fluglotsen unter Auspartialisierung von Geschlecht. Sofern ein Prädiktor erst nach der Auspartialisierung von Geschlecht signifikant wird, ist er in der Spalte zu den ursprünglichen Korrelationen der Prädiktoren mit der Eignungsdiagnostik mit n.s. gekennzeichnet.

Tabelle 16: Skalen des HTQ-2 mit signifikanten Zusammenhängen sowohl zum Geschlecht als auch zu Variablen aus der Eignungsdiagnostik für Fluglotsen

	Korrelation mit Geschlecht	*Korrelation(en) mit Variable(n) aus der Eignungsdiagnostik*	*Partialkorrelation(en) ohne Geschlecht*
BFI-10: Extraversion	.17*	-.27* (SDM) n.s. (Aufmerksamkeit)	-.28* .14*
BFI-10: Gewissenhaftigkeit	.24**	n.s. (Raumvorstellung)	.19**
Kontrollüberzeugungen im Umgang mit Technik	-.20**	.16* (Merkfähigkeit) n.s. (DAC)	.13* .29*
TA-EG: Subjektive Kompetenz	-.27*	.14* (Mittlere Testleistung) .15* (Aufmerksamkeit) .14* (Merkfähigkeit)	n.s. n.s. n.s.
TA-EG: Begeisterung	-.25**	n.s. (SDM)	-.31*
Wochenstunden Computer insgesamt	-.25**	.17* (Mittlere Testleistung) .14* (Konzentration) .15* (Aufmerksamkeit) .38** (SDM)	n.s. n.s. n.s. .39**
Wochenstunden Computerspiele insgesamt	-.31**	.22** (Mittlere Testleistung) .23** (Konzentration) .18** (Merkfähigkeit) .22** (Raumorientierung)	n.s. .18** n.s. n.s.
Wochenstunden Computer-Jump'n-Run	-.13*	-.33* (DAC)	n.s.
Wochenstunden Computer-Sportspiele	-.21**	.16* (Aufmerksamkeit) .28* (SDM)	n.s. n.s.

Anmerkungen: * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), n.s. = nicht signifikant,

N = 214 (für die VA-Variablen), N = 54 (für die Variablen DAC und SDM)

Wie aus der Tabelle ersichtlich, hängt die Variable Extraversion (BFI-10) signifikant mit dem Geschlecht zusammen. Die negative Korrelation zum SDM ist jedoch nicht davon beeinflusst. Bei Auspartialisierung von Geschlecht wird sogar zusätzlich ein Zusammenhang zum Teilbereich Aufmerksamkeit der VA sichtbar. Dieser war offenbar bisher durch einen Geschlechtseffekt verdeckt. Gleiches gilt für den Zusammenhang zwischen dem Teilbereich Raumvorstellung (VA) und Gewissenhaftigkeit (BFI-10). Höhere Extraversion bzw. höhere Gewissenhaftigkeit führen zu besserer Leistung in den Aufmerksamkeits- bzw. Raumvorstellungstests.

Im Bereich der technikbezogenen Persönlichkeitsaspekte korrelieren Kontrollüberzeugungen im Umgang mit Technik (KUT), die subjektive Kompetenz im Umgang mit Technik (TA-EG) sowie die Begeisterung im Umgang mit Technik (TA-EG) mit Geschlecht. Durch dessen Auspartialisierung kommt es zu einem leichten Rückgang des Zusammenhangs des KUT mit dem Teilbereich Merkfähigkeit (VA). Bei den Korrelationen der Variablen aus der Fluglotsen-Eignungsdiagnostik mit der subjektiven Kompetenz im Umgang mit Technik (TA-EG) handelt es sich sogar um reine Scheinkorrelationen, die allein auf das Geschlecht zurückzuführen sind: männliche Probanden schneiden in den Bereichen mittlere Testleistung, Aufmerksamkeit sowie Merkfähigkeit besser ab und sind gleichzeitig überzeugter, kompetent im Umgang mit Technik zu sein. Durch das Geschlecht verdeckt waren dagegen zwei Korrelationen mit den Gerätetests. So gehen höhere Kontrollüberzeugungen im Umgang mit Technik mit besserer DAC-Leistung sowie eine ausgeprägtere Begeisterung im Umgang mit Technik mit schlechterer SDM-Leistung einher.

Die Zusammenhänge der Variablen aus der Eignungsdiagnostik für Fluglotsen mit den Computernutzungsvariablen des HTQ-2 sind bis auf zwei Ausnahmen durchweg reine Geschlechtseffekte. Nach der Auspartialisierung von Geschlecht haben nur der Zusammenhang der insgesamt wöchentlich am Computer verbrachten Stunden mit dem SDM sowie der Stunden, die wöchentlich mit Computerspielen verbracht werden, mit der Konzentrationsleistung in der VA Bestand.

Analog zur vorigen Tabelle stellt **Tabelle 17** auf der Folgeseite diejenigen Variablen des HTQ-2 vor, die sowohl mit Variablen aus der Eignungsdiagnostik für Piloten als auch mit Geschlecht korrelieren. Variablen, die ausschließlich mit Geschlecht oder ausschließlich mit der Eignungsdiagnostik korrelieren, sind erneut nicht aufgeführt. Ergänzt sind die Korrelationen von HTQ-2 und Eignungsdiagnostik für Piloten unter Auspartialisierung von Geschlecht.

Tabelle 17: Skalen des HTQ-2 mit signifikanten Zusammenhängen sowohl zum Geschlecht als auch zu Variablen aus der Eignungsdiagnostik für Piloten

	Korrelation mit Geschlecht	*Korrelation(en) mit Variable(n) aus der Eignungsdiagnostik*	*Partialkorrelation(en) ohne Geschlecht*
BFI-10: Gewissenhaftigkeit	.15*	.14* (Konzentration)	.14*
BFI-10: Extraversion	.17*	-.19** (Mittlere Testleistung) -.19** (Technikwissen) -.20** (Raumorientierung) ..n.s. (Konzentration)	-.16* -.15* -.18** -.14*
Kontrollüberzeugungen im Umgang mit Technik	-.20**	.19** (Technikwissen)	.15*
TA-EG: Begeisterung	-.21**	.22** (Technikwissen)	.17*
TA-EG: Subjektive Kompetenz	-.27**	.18* (Technikwissen)	n.s.
FIDEC B: Pos. Erfahrungen Unterhaltung	-.29*	.16* (Technikwissen) .15* (Mathematik)	n.s. n.s.
FIDEC G: Neg. Erfahrungen Unterhaltung	.23**	-.21** (Technikwissen) -.23** (Mathematik)	-.17* -.19**
Wochenstunden Computer insgesamt	-.25**	.14* (Technikwissen)	n.s.
Wochenstunden Computerspiele insgesamt	-.31**	n.s. (Merkfähigkeit)	-.16*
Wochenstunden Computer-Jump'n-Run	-.14*	-.36* (SIM)	-.35*
Wochenstunden Computer-Sportspiele	-.21**	n.s. (Merkfähigkeit)	-.14*
Bedürfnis nach Teamarbeit	.18**	-.20** (Technikwissen)	-.17*

Anmerkungen: * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), n.s. = nicht signifikant,

N = 196 (für die BU-Variablen und MIC), N = 36 (für die Variable SIM)

Wie auch in der Stichprobe der Fluglotsenbewerber zeigen im Bereich der allgemeinen Persönlichkeit die Variablen Extraversion und Gewissenhaftigkeit signifikante Zusammenhänge mit dem Geschlecht. Die Korrelationen zu den Variablen der Eignungsdiagnostik sind davon jedoch kaum beeinflusst. Verzeichnen lässt sich ein nur sehr geringer Rückgang der Korrelationen. Mit der Partialkorrelation aufgedeckt wird ein zusätzlicher Zusammenhang

zwischen Extraversion und dem BU-Teilbereich Konzentration: extravertiertere Probanden schneiden in den Konzentrationstests tendenziell schlechter ab.

Hinsichtlich der technikbezogenen Persönlichkeitsaspekte lässt sich in der Stichprobe der Pilotenbewerber ein leichter Rückgang in der Höhe der Zusammenhänge zu den Variablen der Eignungsdiagnostik feststellen. Zwei der Prädiktoren, subjektive Kompetenz im Umgang mit Technik (TA-EG) und positive Erfahrungen mit Computern als Unterhaltungsmedien (FIDEC B), korrelieren nur durch einen Geschlechtseffekt mit Kriterien aus der Auswahl. Auch in dieser Stichprobe sind die männlichen Teilnehmer überzeugter von ihrer Kompetenz im Umgang mit Technik und haben zusätzlich besserer Erfahrungen mit Computern als Unterhaltungsmedien. Gleichzeitig schneiden sie besser im Technikwissen und den Mathematiktests ab.

Im Gegensatz zur Stichprobe der Fluglotsenbewerber haben bei den Pilotenbewerbern die Zusammenhänge zwischen Computernutzungsvariablen und der Leistung in den Auswahlverfahren auch nach Auspartialisierung der Variable Geschlecht fast durchweg Bestand. Lediglich der Zusammenhang zwischen den Wochenstunden am Computer insgesamt und dem BU-Teilbereich Technikwissen scheint auf einen reinen Geschlechtseffekt zurückzuführen zu sein. Sogar zwei bisher verdeckte Korrelationen im Bereich der Computernutzung werden nun sichtbar: die mit Computerspielen wöchentlich insgesamt verbrachte Zeit geht mit tendenziell schlechterer Merkfähigkeit einher, ebenso die mit Computer-Sportspielen verbrachte Zeit.

Zusammenfassend lässt sich feststellen, dass sich zwar Geschlechtseffekte finden, ein Großteil der ermittelten Zusammenhänge zwischen dem HTQ-2 und der aktuellen Eignungsdiagnostik im DLR jedoch von der Variable Geschlecht unbeeinflusst bestehen.

8.3.2 Korrelationen mit der der ATM-Simulation HINT

Zur Darstellung der Korrelationen des HTQ-S mit der ATM-Simulation HINT werden die Skalen des HTQ-S in Kategorien eingeordnet (allgemeine Persönlichkeit, technikbezogene Persönlichkeitsaspekte / technikbezogene Einstellung, Neigung zu risikoreichem Entscheidungsverhalten, Computernutzung). Zu jeder Kategorie von Skalen werden die Ergebnisse separat berichtet.

Tabelle 18 verdeutlicht, dass Offenheit aus der Kurzskala BFI sowie einzelne Facetten der Big Five geringe signifikante Zusammenhänge mit HINT zeigen. Auf der Ebene der Primärfaktoren lassen sich keine weiteren Prädiktoren ausmachen. Inhaltlich ist festzustellen, dass reizbarere Personen bei der Bearbeitung von HINT tendenziell langsamer reagieren

(positive Korrelation mit Reaktionszeit) und weniger von der Automation gestellte Requests annehmen (negative Korrelation mit RQ-Annahme durch Alpha) als weniger reizbare Probanden. Die Facette Erlebnissuche hängt mit drei der Leistungsvariablen von HINT zusammen. Personen mit größerem Erlebnisdrang agieren langsamer (positive Korrelation mit Reaktionszeit) und erzielen größere Abweichungen sowohl in ihrem eigenen Sektor Alpha als auch in dem automatisierten Partnersektor Bravo (positive Korrelationen mit den Abweichungsvariablen). Offenheit wie auch Gutherzigkeit korrelieren ebenfalls mit langsamerer Bearbeitung von CAC (positive Korrelation mit Reaktionszeit) die Facette Offenheit für Handlungen zudem mit höheren Target-Abweichungen im eigenen Sektor. Signifikanten Einfluss auf die Request-Stellung an den automatisierten Partner Bravo zeigt das Pflichtbewusstsein von Probanden (positive Korrelation mit RQ Alpha an Bravo). Pflichtbewusstere Personen kommunizieren demnach häufiger mit Bravo als weniger pflichtbewusste. Weitere Effekte hinsichtlich der allgemeinen Persönlichkeit ergeben sich nicht.

Tabelle 18: Korrelationen HTQ-S & HINT: allgemeine Persönlichkeit

	Reaktionszeit	*RQ Alpha an Bravo*	*RQ-Ablehnung durch Alpha*	*RQ-Annahme durch Alpha*	*Target-Abweichungen Alpha*	*Target-Abweichungen Bravo*
NEO-PI-R: Reizbarkeit (Neurotizismus)	.19*			-.16*		
NEO-PI-R: Erlebnissuche (Extraversion)	.18*				.25*	.28*
BFI: Offenheit	.19*					
NEO-PI-R: Offenheit für Handlungen (Offenheit)					.24*	
NEO-PI-R: Gutherzigkeit (Verträglichkeit)	.18*					
NEO-PI-R: Pflichtbewusstsein (Gewissenhaftigkeit)		.25*				
	N=154	N=69	N=154	N=154	N=69	N=69

*Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig)*

Die nächste Skalenkategorie beinhaltet technikbezogene Persönlichkeits- und Einstellungsaspekte (siehe **Tabelle 19**). Drei der entsprechenden Skalen ergeben sich als statistisch bedeutsam für die Leistung in der Simulation. Eine ausgeprägtere Sicherheit im Umgang mit Computern geht mit höherer Reaktionszeit einher. Ein weiterer Effekt zeigt sich hinsichtlich Technikaffinität. Ein stärkerer Eindruck, Technik habe negative Folgen, korrespondiert mit seltenerer Ablehnung der Requests des automatisierten Partners. Hinsichtlich der Einstellung gegenüber Automation zeigt sich, dass Probanden mit positiverer Einstellung kooperativer mit der Automation agieren. Requests des automatisierten Sektors Bravo werden seltener abgelehnt und häufiger angenommen.

Tabelle 19: Korrelationen HTQ-S & HINT: Technikbezogene Persönlichkeitsaspekte / Technikbezogene Einstellung

	Reaktionszeit	*RQ Alpha an Bravo*	*RQ-Ablehnung durch Alpha*	*RQ-Annahme durch Alpha*	*Target-Abweichungen Alpha*	*Target-Abweichungen Bravo*
Sicherheit im Umgang mit Computern	.20*					
TA-EG: Negative Folgen von Technik			-.21*			
Einstellung gegenüber Automation			-.19*	.16*		
	N=154	N=69	N=154	N=154	N=69	N=69

*Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig)*

Tabelle 20 stellt die Korrelationen hinsichtlich der Variable Neigung zu risikoreichem Entscheidungsverhalten dar, die mit der Balloon Analogue Risk Task (BART) gemessen werden. Hier zeigt sich, dass eine höhere Neigung zu risikoreichem Entscheidungsverhalten mit höheren Target-Abweichungen korreliert. Probanden mit niedrigerer Risikoneigung arbeiten demnach regelkonformer (Minimierung der Diskrepanz zwischen Target- und Actual-Werten).

Tabelle 20: Korrelationen HTQ-S & HINT: Neigung zu risikoreichem Entscheidungsverhalten

	Reaktionszeit	*RQ Alpha an Bravo*	*RQ-Ablehnung durch Alpha*	*RQ-Annahme durch Alpha*	*Target-Abweichungen Alpha*	*Target-Abweichungen Bravo*
Balloon Analogue Risk Task					.27*	
	N=154	N=69	N=154	N=154	N=69	N=69

*Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig)*

Die letzte Kategorie in der sich signifikante Effekte ergeben, ist die Kategorie der Computernutzung. Wie **Tabelle 21** zeigt, bewirkt die Nutzung von Smartphones und MP3-Playern signifikant negative Effekte hinsichtlich der Kooperation mit dem maschinellen Partner (negative Korrelation mit RQ Alpha an Bravo bzw. RQ-Annahme durch Alpha). Zudem erweist sich Smartphone-Nutzung als Prädiktor für höhere Target-Abweichungen sowohl im eigenen als auch im automatisierten Partnersektor. Betrachtet man die wöchentliche Beschäftigung mit Computeranwendungen, ergeben sich für verschiedene Anwendungen unterschiedliche Ergebnisse. Eine höhere Anzahl an Wochenstunden am Computer über alle Beschäftigungsmöglichkeiten hinweg hat einen negativen Einfluss auf die HINT-Leistung (höhere Target-Abweichungen bei Bravo). Die Gesamtanzahl der Spielstunden hat einen positiven Effekt. Mehr Zeit pro Woche mit Computerspielen zu verbringen, geht mit geringeren Target-Abweichungen bei Bravo einher. Sofern Schach- oder Lernspiele am Computer gespielt werden, korrespondiert ein höherer wöchentlicher Zeitaufwand in diesem Bereich mit geringerer Reaktionszeit.

Tabelle 21: Korrelationen HTQ-S & HINT: Computernutzung

	Reaktionszeit	*RQ Alpha an Bravo*	*RQ-Ablehnung durch Alpha*	*RQ-Annahme durch Alpha*	*Target-Abweichungen Alpha*	*Target-Abweichungen Bravo*
Smartphone-Nutzung ♦		-.24*			.29*	.29*
MP3-Player-Nutzung ♦		-.28*		-.19*		
Wochenstunden Computer insgesamt						.30*
Wochenstunden Computerspiele insgesamt						-.26*
Wochenstunden Computer-Schach	-.16*					
Wochenstunden Computer-Lernspiele	-.21**					
Wochenstunden Computer-Sportspiele			.17*			
	N=154	N=69	N=154	N=154	N=69	N=69

Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig);
♦ Spearman-Korrelation, da dichotome Variable

Auch für die Zusammenhänge zwischen HTQ-S und HINT werden Geschlechtseffekte auspartialisiert. **Tabelle 22** stellt diejenigen Variablen des HTQ-S vor, die sowohl mit HINT-Variablen als auch mit dem Geschlecht korrelieren. Variablen, die ausschließlich mit dem Geschlecht oder ausschließlich mit HINT korrelieren, sind nicht aufgeführt. Ergänzt sind die Partialkorrelationen von HTQ-S und HINT ohne Geschlecht.

Tabelle 22: Skalen des HTQ-S mit signifikanten Zusammenhängen sowohl zum Geschlecht als auch zu Variablen aus der HINT-Simulation

	Korrelation mit Geschlecht	*Korrelation(en) mit HINT-Variable(n)*	*Partialkorrelation ohne Geschlecht*
NEO-PI-R: Reizbarkeit (Neurotizismus)	.18*	.19* (Reaktionszeit) n.s. (RQ-Annahme durch Alpha)	.16* -.17*
NEO-PI-R: Impulsivität (Neurotizismus)	.27**	n.s. (Reaktionszeit)	.16*
NEO-PI-R: Offenheit	.27**	.19* (Reaktionszeit)	n.s.
NEO-PI-R: Gutherzigkeit (Verträglichkeit)	.21**	.18* (Reaktionszeit)	n.s.
Einstellung gegenüber Automation	-.21**	-.19* (RQ-Ablehnung durch Alpha) .16* (RQ-Annahme durch Alpha)	-.18* n.s.
MP3-Player-Nutzung	.17*	-.28* (RQ Alpha an Bravo) -.19* (RQ-Annahme durch Alpha)	n.s. -.21**
Wochenstunden Computerspiele insgesamt	-.37**	-.26* (Target-Abweichungen Bravo)	n.s.
Wochenstunden Computer-Sportspiele	-.26**	.17* (RQ-Ablehung durch Alpha)	.20*

Anmerkungen: * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), N =156 (für die Variablen Reaktionszeit, RQ-Ablehnung durch Alpha, RQ-Annahme durch Alpha), N =64 (für die Variablen RQ Alpha an Bravo, Target-Abweichungen Alpha, Target-Abweichungen Bravo)

Wie in der Tabelle ersichtlich, sind die Zusammenhänge der HINT-Variablen mit Offenheit, Gutherzigkeit, und Wochenstunden Computerspiele nach Auspartialisierung von Geschlecht nicht mehr signifikant (n.s.). Sie sind in Wahrheit nur auf das Geschlecht und nicht auf den HTQ-S zurück zu führen. Durch die Variable Geschlecht bisher verdeckt wurde eine Korrelation zwischen Impulsivität und Reaktionszeit. Durch die Auspartialisierung wird ein (wenn auch nur geringer) Zusammenhang aufgezeigt: impulsivere Probanden agieren bei der Bearbeitung von HINT langsamer.

Keinerlei Korrelationen ergeben sich für die Skalen zu Cognitive Failures (CFQ), Flexibilität und Teamorientierung aus dem BIP, Bedürfnis nach Teamarbeit (BnT), Risikotoleranz aus dem CaP, Einstellung gegenüber Arbeit im Mensch-Maschine-Team (EgMMT), Individuelle Einstellung gegenüber Teamarbeit (FIT) sowie allgemeine Selbstwirksamkeitserwartung. Damit scheiden unter anderem alle Skalen, die Aspekte von Teamarbeit betreffen, als Prädiktoren für Leistung in HINT aus. Mögliche Hintergründe für das Fehlen der erwarteten Effekte werden in Kapitel 9 diskutiert.

8.4 Ergebnisse der Gruppenvergleiche

Wie unter 7.3.3 beschrieben, wird ergänzend zu korrelativen Zusammenhängen überprüft, ob bessere und schlechtere Leistung in der HINT-Simulation mit Unterschieden in der Beantwortung der HTQ-S-Skalen einhergehen. Der Vergleich ergibt mehrere Skalen, für die dies der Fall ist. **Tabelle 23** bis **Tabelle 27** listen die Variablen mit signifikanten Unterschieden in den HINT-Parametern auf. Enthalten sind jeweils die Mittelwerte (z-Werte) der gebildeten Gruppe 1 (niedrigere Merkmalsausprägung) und Gruppe 2 (höhere Merkmalsausprägung), das Signifikanzniveau (p) des gefundenen Unterschiedes sowie dessen Effektstärke (d). Alle ermittelten Effekte liegen bei .31 bis .65 und damit nach Bortz (2005) im schwachen bis mittleren Bereich. Die Variablen mit eindeutigem Leistungsbezug (Reaktionszeit und Target-Abweichungen) werden zuerst betrachtet. Ergänzend werden die Variablen zur Stellung und Bearbeitung von Requests analysiert. Wie bereits unter 7.2.2 erläutert, ist für diese Variablen nicht eindeutig definierbar, ob eine große Anzahl gestellter bzw. abgelehnter oder angenommener Requests für gute Leistung steht, da dieses jeweils sehr vom individuellen Kontext jedes Requests abhängt. Jedoch sind die Variablen Indikatoren für die Bereitschaft von Probanden, mit ihrem automatisierten Partner zusammen zu arbeiten (d.h. für Kooperationsleistung) und daher auch hinsichtlich möglicher Gruppenunterschiede in der Persönlichkeit interessant zu analysieren. Für die Variable „RQ Alpha an Bravo" ergeben sich keine signifikanten Gruppenunterschiede, so dass hierzu keine Tabelle aufgeführt wird.

Zunächst wird die Variable Reaktionszeit betrachtet (**Tabelle 23**). Hier zeigt sich, dass Probanden mit niedrigeren Werten, sprich schnelle Operateure (Gruppe 1), signifikant niedrigere Ausprägungen hinsichtlich Erlebnissuche, Offenheit für Handlungen, Begeisterung im Umgang mit Technik, angenommenen positiven Folgen von Technik sowie Smartphone-Nutzung haben. Signifikant höher ist der Wert der wöchentlichen Nutzung von Computer-Lernspielen.

Tabelle 23: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren Werten und höheren Werten hinsichtlich *Reaktionszeit*

	Mittelwerte		*t-Test*	
	Gruppe 1	*Gruppe 2*	*p*	*d*
NEO-PI-R: Erlebnissuche (Extraversion)	3.42	3.62	*	.39
NEO-PI-R: Offenheit für Handlungen (Offenheit)	3.20	3.37	*	.36
TA-EG: Begeisterung im Umgang mit Technik	3.62	3.84	*	.32
TA-EG: Positive Folgen von Technik	3.75	3.92	*	.36
Smartphone-Nutzung	.69	.82	*	.32
Wochenstunden Computer-Lernspiele	.29	.08	**	.53
	N=83	N=73		

Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig), d = Effektstärke Cohens d

Hinsichtlich der Target-Abweichungen im Sektor Alpha (**Tabelle 24**) zeigt sich, dass Probanden mit geringeren Target-Abweichungen (Gruppe 1) signifikant höhere Werte bezüglich Entgegenkommen, Pflichtbewusstsein, Leistungsstreben, Gewissenhaftigkeit und Kompetenz im Umgang mit Technik aufweisen. Smartphone-Nutzer befinden sich signifikant öfter in der Gruppe mit höheren Target-Abweichungen.

Tabelle 24: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren und höheren *Target-Abweichungen im Sektor Alpha*

	Mittelwerte		*t-Test*	
	Gruppe 1	*Gruppe 2*	*p*	*d*
NEO-PI-R: Pflichtbewusstsein (Gewissenhaftigkeit)	4.00	3.81	**	.47
NEO-PI-R: Leistungsstreben (Gewissenhaftigkeit)	3.76	3.55	**	.47
NEO-PI-R: Gewissenhaftigkeit	3.68	3.55	*	.36
NEO-PI-R: Entgegenkommen (Verträglichkeit)	3.43	3.24	*	.41
TA-EG: Kompetenz im Umgang mit Technik	3.81	3.67	*	.38
Smartphone-Nutzung	.59	.81	**	.50
	N=41	N=115		

Anmerkungen: * p < .05 (2-seitig). ** p < .01 (2-seitig). d = Effektstärke Cohens d

Ein ähnliches Bild wie für die Target-Abweichungen im Sektor Alpha zeigt sich für die Abweichungen im Sektor Bravo (**Tabelle 25**). Erneut präsentiert sich Gewissenhaftigkeit mit mehreren Einzelfacetten sowie auch als Primärfaktor als Persönlichkeitsaspekt, der förderlich für gute Leistung ist. Gruppe 1 mit niedrigeren Target-Abweichungen im Sektor Bravo zeigt signifikant höhere Werte in diesem Bereich. Signifikant niedrigere Werte liegen dagegen erneut bei der Smartphone-Nutzung sowie zusätzlich bei der Computernutzung in Unterricht, Studium usw. sowie der Computernutzung in der Öffentlichkeit vor.

Tabelle 25: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren und höheren *Target-Abweichungen im Sektor Bravo*

	Mittelwerte		*t-Test*	
	Gruppe 1	*Gruppe 2*	*p*	*d*
NEO-PI-R: Pflichtbewusstsein (Gewissenhaftigkeit)	4.05	3.80	**	.65
NEO-PI-R: Leistungsstreben (Gewissenhaftigkeit)	3.79	3.55	**	.56
NEO-PI-R: Besonnenheit (Gewissenhaftigkeit)	3.36	3.15	*	.38
NEO-PI-R: Gewissenhaftigkeit	3.71	3.55	**	.48
Smartphone-Nutzung	.58	.81	**	.50
Computernutzung Unterricht / Studium/ Ausbildung / Arbeitsplatz	.55	.74	*	.39
Computernutzung Öffentlichkeit	.24	.42	*	.38
	N=38	N=118		

Anmerkungen: * p < .05 (2-seitig). ** p < .01 (2-seitig). d = Effektstärke Cohens d

Für die Variable zur Ablehnung von an Alpha gestellten Requests (**Tabelle 26**) ergeben sich drei Persönlichkeitsbereiche mit signifikanten Gruppenunterschieden. Probanden, die weniger Requests ablehnen (Gruppe 1), beschreiben sich im HTQ-S als signifikant weniger herzlich, erwarten signifikant mehr negative Folgen von Technik, aber haben eine signifikant positivere Einstellung gegenüber Automation.

Tabelle 26: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren und höheren Werten hinsichtlich der *Ablehnung von an sie (Alpha) gestellten Requests*

	Mittelwerte		*t-Test*	
	Gruppe 1	*Gruppe 2*	*p*	*d*
NEO-PI-R: Herzlichkeit (Extraversion)	3.83	3.97	*	.32
TA-EG: Negative Folgen von Technik	2.71	2.49	**	.40
Einstellung gegenüber Automation	2.92	2.81	*	.33
	N=82	N=74		

Anmerkungen: * p < .05 (2-seitig). ** p < .01 (2-seitig). d = Effektstärke Cohens d

In **Tabelle 27** werden die Gruppenunterschiede hinsichtlich der Annahme von erhaltenen Requests dargestellt. Für drei Variablen lassen sich signifikante Unterschiede feststellen. Probanden, die weniger Requests annehmen (Gruppe 1), sind signifikant extravertierter, risikotoleranter und antizipieren signifikant weniger negative Folgen von Technik.

Tabelle 27: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren und höheren Werten hinsichtlich der *Annahme von an sie (Alpha) gestellten Requests*

	Mittelwerte		*t-Test*	
	Gruppe 1	*Gruppe 2*	*p*	*d*
NEO-PI-R: Extraversion	3.64	2.18	*	.31
CAP: Risikotoleranz	3.71	3.56	*	.31
TA-EG: Negative Folgen von Technik	2.50	2.71	*	.40
	N=79	N=77		

Anmerkungen: * p < .05 (2-seitig), ** p < .01 (2-seitig), d = Effektstärke Cohens d

Nicht bei zweiseitiger, jedoch bei einseitiger Testung signifikant ist ein Gruppenunterschied hinsichtlich der Variable Einstellung gegenüber Arbeit im Mensch-Maschine-Team. Aufgrund der in Abschnitt 8.1.1 postulierten Annahme, dass Personen mit positiver Einstellung gegenüber Arbeit im Mensch-Maschine-Team auch bessere Leistungen bei der tatsächlichen Arbeit mit einem maschinellen Teampartner erzielen bzw. gewillter sind, mit diesem zu kooperieren, ist in diesem Fall eine gerichtete Testung angemessen. Wie **Tabelle 28** zeigt, sind Probanden der Gruppe 1 mit weniger angenommenen Requests weniger positiv gegenüber Arbeit im Mensch-Maschine-Team eingestellt. Probanden, die mehr Requests des maschinellen Teampartners annehmen (Gruppe 2) sind solcher Zusammenarbeit gegenüber signifikant positiver eingestellt.

Tabelle 28: Signifikanter Unterschied im EgMMT zwischen Probanden mit niedrigeren und höheren Werten hinsichtlich der *Annahme von an sie (Alpha) gestellten Requests*

	Mittelwerte		*t-Test*	
	Gruppe 1	*Gruppe 2*	*p*	*d*
Einstellung gegenüber Arbeit im Mensch-Maschine-Team	2.47	2.69	♦	.31
	N=76	N=74		

Anmerkungen: ♦ $p < .05$ (1-seitig), d = Effektstärke Cohens d

Der nun folgende Abschnitt 8.5 befasst sich über einfache Korrelationen und Gruppenunterschiede hinaus mit der inkrementellen Validität des HTQ-S.

8.5 Inkrementelle Validität des HTQ-S

Wie unter 7.3.4 erläutert, wird die inkrementelle Validität des HTQ-S mittels multipler Korrelationsstatistik geschätzt. Hierfür werden zunächst die für diese Methode notwendigen Voraussetzungen (siehe 7.3.4) geprüft sowie für die HINT-Variablen analysiert, welche Leistungsvariablen aus der Vorauswahl bzw. Berufsgrunduntersuchung der Bewerber signifikant zur Kriteriumsvarianz R^2 beitragen. Anschließend werden die Variablen des HTQ-S als zusätzliche Prädiktoren in die Analysen aufgenommen. Damit wird festgestellt, ob die HTQ-S-Variablen den erklärbaren Anteil der Kriteriumsvarianz R^2 vergrößern können.

Zur Voraussetzungsprüfung wurde als erstes die Verteilung der relevanten Variablen mit Histogrammen begutachtet. Diese sind in Anhang A und B aufgelistet und zeigen, dass für einen Großteil der Variablen annähernde Normalverteilung vorliegt. Jedoch werden die Variablen zu den Wochenstunden verschiedener Computeraktivitäten und ebenso die HINT-Variable „Korrekte CAC-Bearbeitung“ (siehe hierzu auch unter 8.2) aus den Analysen ausgeschlossen, da hier keine Normalverteilung vorliegt. Die Annahme der Homoskedastizität wurde mit Streudiagrammen kontrolliert und kann bestätigt werden. Zudem wurden die Histogramme der standardisierten Residuen dahingehend gesichtet, ob deren Fehler zufällig auftreten und normalverteilt sind. Auch dies ist der Fall. Die Voraussetzung von Non-Kollinearität der einbezogenen Variablen kann mit Einschränkungen bestätigt werden. Die Interkorrelationsanalysen zu den HTQ-S-Variablen ergeben geringe bis mittlere Zusammenhänge einzelner Prädiktoren, insbesondere sofern verwandte Konstrukte korreliert werden. Jedoch betonen Backhaus, Erichson, Plinke und Weiber (2003), dass sich keine exakten Grenzen für Multikollinearität angeben lassen.

Da die Voraussetzungen zur Durchführung der multiplen linearen Regression als ausreichend erfüllt betrachtet werden, werden die Analysen vorgenommen. Die Ergebnisse werden im Folgenden jeweils pro HINT-Variable und getrennt nach Fluglotsen- und Pilotenauswahl (VA und BU) dargestellt. Für die Bewerbergruppen liegen als kognitive Prädiktoren jeweils Variablen der VA *oder* der BU vor, so dass keine gemeinsame Analyse stattfinden kann. Die jeweilige Stichprobengröße der Analysen ergibt sich in den Tabellen aus der Summe der Freiheitsgrade in einer Zelle. Da die Anzahl der Probanden mit gültigen Werten in der Stichprobe der BU-Teilnehmer teilweise sehr gering ist, muss in manchen Fällen auf eine Regressionsanalyse verzichtet werden. Aufgrund der Vielzahl der Prädiktoren ist mit zu geringen Stichprobengrößen keine sinnvolle Berechnung möglich. Sofern dies zutrifft, wird an der entsprechenden Stelle darauf hingewiesen.

Als erste Variable wird die Reaktionszeit betrachtet. Bei der Einbeziehung der kognitiven Leistungsvariablen aus der VA, zeigt sich der Teilbereich Konzentration mit einem R^2 von .19 als Prädiktor relevant, die anderen Variablen werden ausgeschlossen. Von den in Schritt 2 hinzugefügten Prädiktoren des HTQ-S wird ausschließlich die Variable zur Sicherheit im Umgang mit Computeranwendungen ins Modell aufgenommen. Bei erster Betrachtung können demnach 4% zusätzlicher Varianz durch diese HTQ-S-Skala aufgeklärt werden. Wird jedoch, wie unter 7.3.4 angekündigt, für R^2 die Schrumpfungskorrektur nach Olkin und Pratt (1958) durchgeführt, schrumpft $R^2 = .19$ im ersten Schritt auf $\widehat{R}^2 = .09$. Im zweiten Schritt ist kein Zusammenhang mehr nachweisbar. Durch die Hinzunahme der HTQ-S-Variablen bei gleichbleibender Stichprobengröße im zweiten Schritt schrumpft auch die Varianzaufklärung durch den kognitiven Prädiktor in sich zusammen. **Tabelle 29** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 29: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *Reaktionszeit*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\widehat{R}^2$	$\Delta \widehat{R}^2$
Schritt 1						
Teilbereich Konzentration (VA)	.19		1, 96	**	.09	
Schritt 2						
Sicherheit im Umgang mit Computeranwendungen	.23	.04	2, 95	**	--	--

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2, df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), $\widehat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \widehat{R}^2$ = Differenz zu $\widehat{R}^2$ von Schritt 1, -- = kein Zusammenhang nachweisbar

Hinsichtlich der kognitiven Leistungsvariablen aus der BU, zeigt sich der Teilbereich Raumvorstellung mit einem R^2 von .22 als relevant für das Kriterium Reaktionszeit, die anderen Variablen werden ausgeschlossen. Von den in Schritt 2 hinzugefügten Prädiktoren des HTQ-S wird die Skala Flexibilität aus dem BIP ins Modell aufgenommen. 9% zusätzlicher Varianz können bei erster Betrachtung durch diese HTQ-S-Skala aufgeklärt werden. Wird die Schrumpfungskorrektur durchgeführt, schrumpft jedoch $R^2 = .22$ im ersten Schritt auf $\widehat{R}^2 = .12$ und analog zur VA-Stichprobe ist auch für die BU-Stichprobe im zweiten Schritt kein Zusammenhang mehr nachweisbar. **Tabelle 30** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 30: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der BU für das Kriterium *Reaktionszeit*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\hat{R}^2$	$\Delta \hat{R}^2$
Schritt 1						
Teilbereich Raumvorstellung (BU)	.22		1, 49	**	.12	
Schritt 2						
BIP: Flexibilität	.31	.09	2, 48	**	--	--

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2, df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), $\widehat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \widehat{R}^2$ = Differenz zu $\widehat{R}^2$ von Schritt 1, -- = kein Zusammenhang nachweisbar

Als nächstes wird das Kriterium RQ Alpha an Bravo untersucht, folglich die Anzahl der Requests, die von den Probanden an den maschinellen Partner Bravo gestellt werden. Hier zeigt sich der Teilbereich Konzentration mit einem R^2 von .18 für die VA-Stichprobe als relevant. Die anderen Variablen der VA werden ausgeschlossen. Bei zusätzlicher Eingabe der Variablen des HTQ-S im zweiten Schritt, präsentieren sich die Variablen subjektive Kompetenz im Umgang mit Technik (TA-EG), positive persönliche Erfahrungen mit Computern hinsichtlich Lernen und Arbeit (FIDEC F) sowie Begeisterung im Umgang mit Technik (TA-EG) als zusätzlich varianzaufklärend. Insgesamt steigt der Anteil von 18% aufgeklärter Varianz auf 41%, ΔR^2 beträgt entsprechend .23. Wird die Schrumpfungskorrektur durchgeführt, sinkt $R^2 = .18$ für den ersten Schritt auf $\widehat{R}^2 = .09$. Für den zweiten

Schritt inklusive HTQ-S schrumpft $R^2 = .41$ auf $\widehat{R}^2 = .19$. Somit ist jedoch auch nach der Schrumpfungskorrektur immer noch ein Zugewinn an Varianzaufklärung von 10% nachweisbar. Für die Stellung von Requests kann somit festgestellt werden, dass neben dem kognitiven Prädiktor Konzentrationsfähigkeit auch das Kompetenzgefühl der Probanden gegenüber Technik, ihre positiven Erfahrungen mit Computeranwendungen im Bereich Lernen und Arbeit sowie ihre Begeisterung im Umgang mit Technik valide Prädiktoren für die Bereitschaft sind, mit einem automatisierten Partner, sprich im Mensch-Maschine-Team, zu arbeiten. **Tabelle 31** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 31: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *RQ Alpha an Bravo*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\widehat{R}^2$	$\Delta \widehat{R}^2$
Schritt 1						
Teilbereich Konzentration (VA)	.18		1, 50	**	.09	
Schritt 2						
TA-EG: Subjektive Kompetenz	.27		2, 49	**	--	
FIDEC F: Positive Erfahrungen Lernen & Arbeit	.35		3, 48	**	.10	
TA-EG: Begeisterung	.41	.23	4, 47	**	.19	.10

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2, df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), $\widehat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \widehat{R}^2$ = Differenz zu $\widehat{R}^2$ von Schritt 1, -- = kein Zusammenhang nachweisbar

Für die Stichprobe der BU-Teilnehmer wird auf eine Regressionsanalyse verzichtet, da die Anzahl der Probanden mit gültigen Werten für diese Variable bei lediglich N=16 liegt. Aufgrund der Vielzahl der Prädiktoren ist mit dieser geringen Stichprobengröße keine sinnvolle Berechnung möglich (siehe oben).

Für das Kriterium RQ Ablehnung durch Alpha, also die relative Anzahl der von Alpha abgelehnten Requests des maschinellen Teampartners, zeigt sich keiner der möglichen kognitiven Prädiktoren aus der VA signifikant. Wird der HTQ-S hinzugefügt, ergibt sich eine geringe Varianzaufklärung von 5% durch den Prädiktor wahrgenommene negative Technikfolgen (TA-EG). Diese hat jedoch nach der Schrumpfungskorrektur keinen Bestand und verschwindet bei der Anwendung. **Tabelle 32** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 32: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *RQ Ablehnung durch Alpha*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\hat{R}^2$	$\Delta \hat{R}^2$
Schritt 1						
VA	--				--	
Schritt 2						
TA-EG: Negative Technikfolgen	.05	.05	1, 97	*	--	--

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2, df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), $\widehat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \widehat{R}^2$ = Differenz zu $\widehat{R}^2$ von Schritt 1, -- = kein Zusammenhang nachweisbar

Auch die kognitiven Prädiktoren der BU können keine Varianz des Kriteriums RQ Ablehnung durch Alpha aufklären. Wird der HTQ-S hinzugefügt, ergibt sich eine Varianzaufklärung von 11% durch den Prädiktor Extraversion (NEO-PI-R). Diese hat jedoch nach der Schrumpfungskorrektur ebenfalls keinen Bestand und verschwindet bei der Anwendung. **Tabelle 33** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 33: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der BU für das Kriterium *RQ Ablehnung durch Alpha*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\hat{R}^2$	$\Delta \hat{R}^2$
Schritt 1						
BU	--				--	
Schritt 2						
TA-EG: Negative Technikfolgen	.05	.05	1, 97	*	--	--

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2,
df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig),
** $p < .01$ (2-seitig), $\widehat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \widehat{R}^2$ = Differenz zu $\widehat{R}^2$ von Schritt 1,
-- = kein Zusammenhang nachweisbar

Für das Kriterium RQ Annahme durch Alpha, also die relative Anzahl der von Alpha angenommenen Requests des maschinellen Teampartners, zeigt sich der Teilbereich Konzentration mit einem R^2 von .04 als geringfügig relevant, die anderen Variablen werden aus dem Modell ausgeschlossen. Wird der HTQ-S hinzugefügt, ergibt sich eine zusätzliche Varianzaufklärung von $\Delta R^2 = .04$ durch den Prädiktor wahrgenommene negative Technikfolgen (TA-EG). Auch in diesem Fall hat jedoch das daraus resultierende R^2 von .08 nach der Schrumpfungskorrektur keinen Bestand und verschwindet bei deren Anwendung. **Tabelle 34** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 34: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *RQ Annahme durch Alpha*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\hat{R}^2$	$\Delta \hat{R}^2$
Schritt 1						
Teilbereich Konzentration (VA)	.04		1, 96	*	--	
Schritt 2						
TA-EG: Negative Technikfolgen	.08	.04	2, 95	*	--	--

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2, df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), $\widehat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \widehat{R}^2$ = Differenz zu $\widehat{R}^2$ von Schritt 1, -- = kein Zusammenhang nachweisbar

Von den kognitiven Prädiktoren der BU zeigt sich keiner für das Kriterium RQ Annahme durch Alpha relevant. Nimmt man im zweiten Schritt die HTQ-S-Skalen als Prädiktoren hinzu, ergibt sich ein R^2 von .08 für die Skala Bedürfnis nach Teamarbeit. Wie erwartet, hat auch diese nach der Schrumpfungskorrektur keinen Bestand und verschwindet. **Tabelle 35** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 35: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der BU für das Kriterium *RQ Annahme durch Alpha*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\hat{R}^2$	$\Delta \hat{R}^2$
Schritt 1						
BU	--				--	
Schritt 2						
Bedürfnis nach Teamarbeit	.08	.08	1, 53	*	--	--

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2, df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), $\widehat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \widehat{R}^2$ = Differenz zu $\widehat{R}^2$ von Schritt 1, -- = kein Zusammenhang nachweisbar

Für das nächste Kriterium der Target-Abweichungen im Sektor Alpha, zeigt sich erneut keiner der möglichen kognitiven Prädiktoren aus der VA signifikant. Werden die Variablen des HTQ-S hinzugenommen, ergibt sich eine Varianzaufklärung von 12% durch den Prädiktor Neigung zu risikoreichem Entscheidungsverhalten (Balloon Analogue Risk Task), die jedoch nach Anwendung der Schrumpfungskorrektur verschwindet. **Tabelle 36** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 36: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *Target-Abweichungen Alpha*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\hat{R}^2$	$\Delta \hat{R}^2$
Schritt 1						
VA	--				--	
Schritt 2						
Balloon Analogue Risk Task	.12	.12	1, 51	**	--	--

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2, df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), $\widehat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \widehat{R}^2$ = Differenz zu $\widehat{R}^2$ von Schritt 1, -- = kein Zusammenhang nachweisbar

Für die Stichprobe der BU-Teilnehmer muss analog zum Kriterium RQ Alpha an Bravo auch für die Variable Target-Abweichungen Alpha auf Analysen verzichtet werden. Die Anzahl der Probanden mit gültigen Werten liegt bei lediglich N=16. Wie bereits oben erwähnt, ist aufgrund der Vielzahl der Prädiktoren mit dieser geringen Stichprobengröße keine sinnvolle Berechnung möglich.

Für das Kriterium Target-Abweichungen im Sektor Bravo wird der VA-Teilbereich Raumvorstellung mit einem R^2 von .09 signifikant, die anderen Variablen werden aus dem Modell ausgeschlossen. Bei Hinzunahme des HTQ-S im zweiten Schritt, ergibt sich die Variable positive persönliche Erfahrungen mit Computern hinsichtlich Lernen und Arbeit (FIDEC F) als signifikanter Prädiktor. Der kognitive Prädiktor Raumvorstellung wird in diesem Modell jedoch nicht mehr signifikant. Die verbleibenden 9% Varianzaufklärung durch den Prädiktor FIDEC F haben auch in diesem Fall nach der Schrumpfungskorrektur keinen Bestand und verschwinden bei der Anwendung. **Tabelle 37** enthält eine Zusammenfassung der wichtigsten Regressionsstatistiken.

Tabelle 37: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *Target-Abweichungen Bravo*

	Regressionsstatistiken					
Prädiktoren	R^2	ΔR^2	*df*	*p*	$\hat{R}^2$	$\Delta \hat{R}^2$
Schritt 1						
Teilbereich Raumvorstellung (VA)	.09		1, 50	*	--	
Schritt 2						
FIDEC F: Positive Erfahrungen Lernen & Arbeit	.09	.00	1, 50	*	--	--

Anmerkungen: R^2 = Bestimmtheitsmaß, ΔR^2 = Differenz von R^2 Schritt 1 und R^2 Schritt 2, df = Freiheitsgrade (Hinweis: N ergibt sich aus der Addition der df in einer Zelle), * $p < .05$ (2-seitig), ** $p < .01$ (2-seitig), $\hat{R}^2 = R^2$ nach Schrumpfungskorrektur, $\Delta \hat{R}^2$ = Differenz zu $\hat{R}^2$ von Schritt 1, -- = kein Zusammenhang nachweisbar

Erneut muss aufgrund der geringen Stichprobengröße von N=16 für die Variable Target-Abweichungen Bravo für die BU-Teilnehmer auf Analysen verzichtet werden.

8.6 Zusammenfassung der Ergebnisse

In den vorangegangenen Abschnitten 8.3 (Korrelationsergebnisse), 8.4 (Gruppenvergleiche) und 8.5 (Inkrementelle Validität) wurde ausführlich auf die verschiedenen Ergebnisse zu dem Fragebogen HTQ-S Bezug genommen. An dieser Stelle wird noch einmal grafisch zusammengefasst, welche der untersuchten Variablen für die Ergebnisse der heutigen Eignungsdiagnostik als auch die Leistungen in der ATM-Simulation HINT relevant sind. Zunächst werden die Korrelationen zur aktuellen Eignungsdiagnostik erneut beleuchtet. In einer Abbildung zusammengefasst sind jeweils alle Skalen mit positiven, negativen oder ambivalenten Zusammenhängen. **Abbildung 11** skizziert die Skalen mit positiven Zusammenhängen.

Abbildung 11: Positive Zusammenhänge zu aktueller Eignungsdiagnostik
(*Anmerkung*: grün=Persönlichkeitsvariablen, hellgrün=Computernutzungsvariablen)

In dieser Zusammenschau auffällig sind die große Anzahl der technikbezogenen Persönlichkeits- und Einstellungsskalen sowie die große Relevanz der Computernutzungsvariablen. Andere Persönlichkeitsaspekte sind hier zunächst nur wenig vertreten. Lediglich die beiden Big-Five-Primärfaktoren Gewissenhaftigkeit und Neurotizismus zählen zu den Variablen mit positivem Einfluss.

Abbildung 12 veranschaulicht die Skalen mit negativen Zusammenhängen. Hier sind auch die übrigen drei der Big Five Persönlichkeitsfaktoren zu finden. Alle Primärfaktoren des Modells beeinflussen also entweder in positiver oder in negativer Form die Leistung in der aktuellen Eignungsdiagnostik für Fluglotsen und / oder Piloten. Zusätzlich wirken auch ein stärkeres Bedürfnis nach Teamarbeit, stärkere Flexibilität und höhere Risikotoleranz als weitere Variablen ohne Computer- und/oder Technikbezug negativ auf die Leistung.

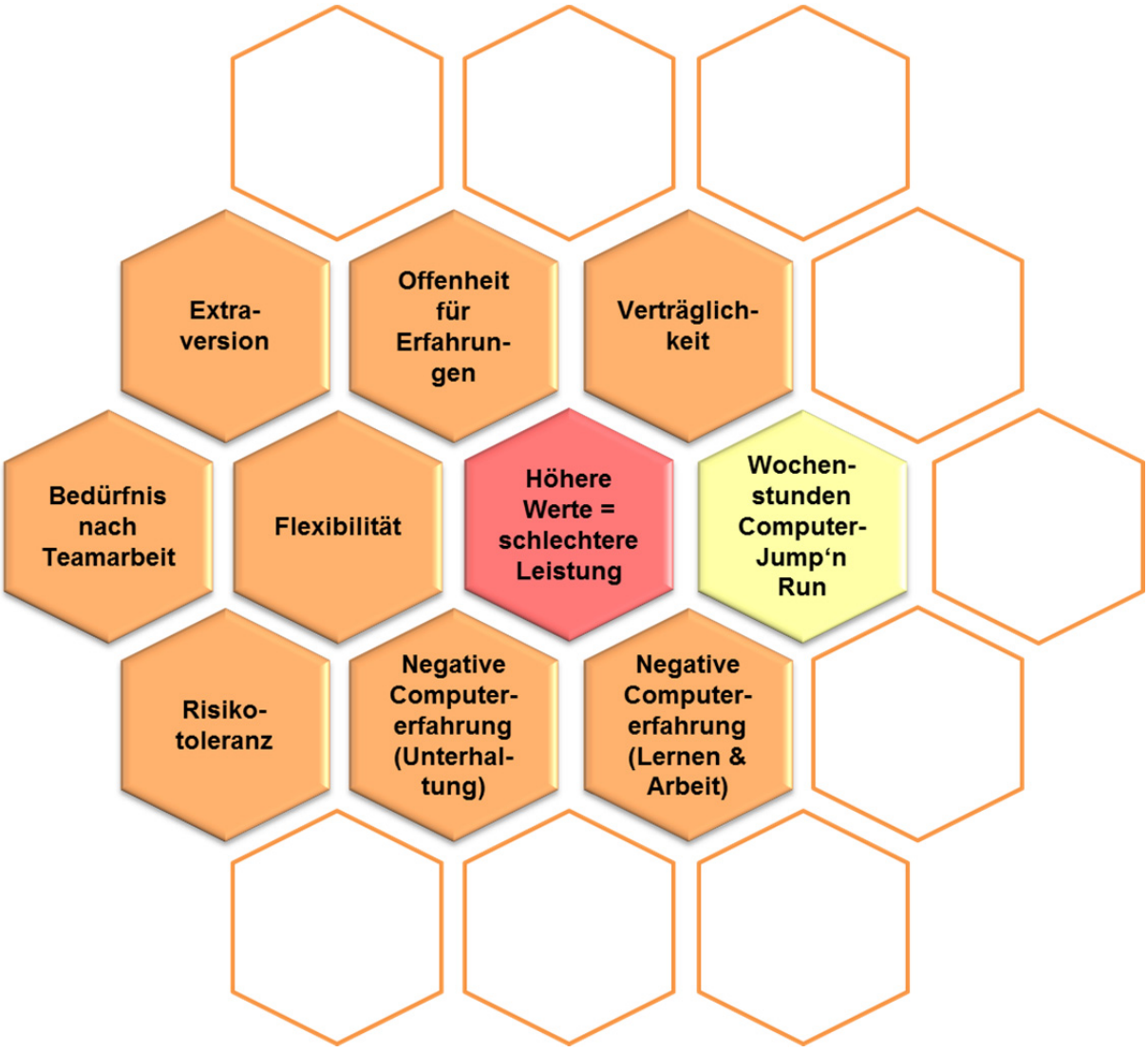


Abbildung 12: Negative Zusammenhänge zu aktueller Eignungsdiagnostik
(*Anmerkung*: orange=Persönlichkeitsvariablen, gelb=Computernutzungsvariablen)

Ambivalente Zusammenhänge, d.h. sowohl positive als auch negative Korrelationen mit verschiedenen Leistungsvariablen aus der Eignungsdiagnostik zeigen sich für zwei der HTQ-S-Skalen (siehe **Abbildung 13**). So wirken sich höhere Werte in der Skala zur Annahme positiver Technikfolgen (TA-EG) auf zwei der Leistungsvariablen aus der BU positiv aus, auf die DAC-Leistung jedoch negativ. Die wöchentliche Beschäftigung mit Rollenspielen am Computer wirkt positiv auf einen VA-Teilbereich, jedoch negativ auf mehrere BU-Teilbereiche.

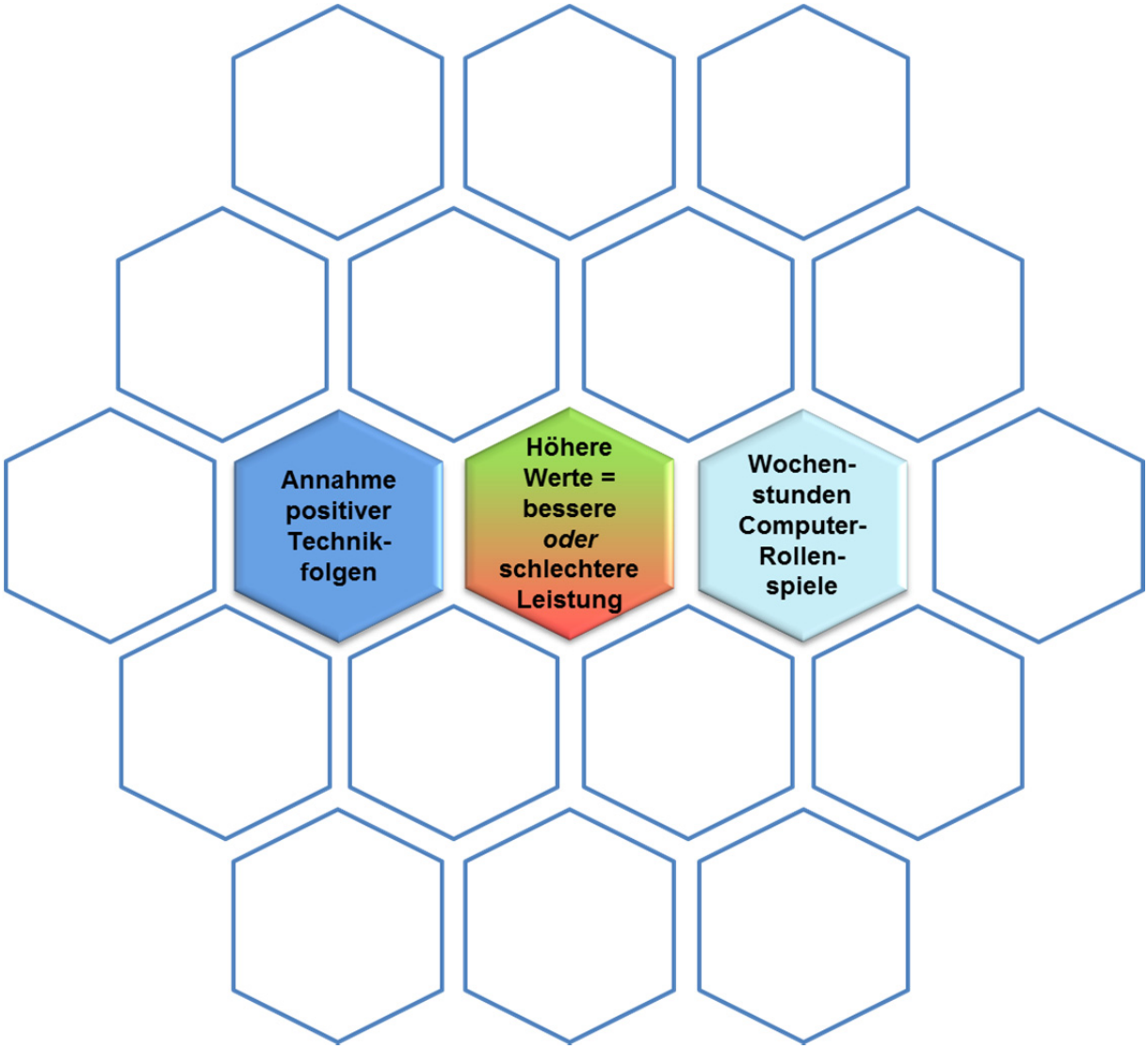


Abbildung 13: Ambivalente Zusammenhänge zu aktueller Eignungsdiagnostik
(*Anmerkung*: blau=Persönlichkeitsvariablen, hellblau=Computernutzungsvariablen)

Welche der untersuchten Variablen für die Leistung in der ATM-Simulation HINT relevant sind, wird in den folgenden Abbildungen verdeutlicht. Zunächst werden analog zu den obigen Grafiken die Skalen mit korrelativen Zusammenhängen visualisiert. **Abbildung 14** skizziert die Skalen mit positiven Zusammenhängen.

Abbildung 14: Positive Zusammenhänge zu HINT
(*Anmerkung*: grün=Persönlichkeitsvariablen, hellgrün=Computernutzungsvariablen)

Im Gegensatz zur Eignungsdiagnostik sind bezüglich HINT unter den positiv auf die Leistung wirkenden Variablen überwiegend Facetten der Big Five. Technikbezogene Persönlichkeitsvariablen als auch Computernutzung spielen eine kleinere Rolle. Letztere wirkt sich sogar in vier Fällen negativ auf die HINT-Leistung aus, wie **Abbildung 15** auf der nächsten Seite zeigt.

Von den Merkmalen der allgemeinen Persönlichkeit wirkt sich Reizbarkeit als Neurotizismus-Facette negativ auf die HINT-Leistung aus. Zusätzlich wirkt auch eine stärkere Neigung zu risikoreichem Entscheidungsverhalten negativ.

Abbildung 15: Negative Zusammenhänge zu HINT
(*Anmerkung*: orange=Persönlichkeitsvariablen, gelb=Computernutzungsvariablen)

Neben den Zusammenhangsanalysen wurde ebenfalls geprüft, welche HTQ-S-Skalen signifikant unterschiedliche Merkmalsausprägungen zeigen, sofern Probanden anhand ihrer HINT-Leistung in Gruppen eingeteilt werden. **Abbildung 16** visualisiert die betreffenden Skalen. Wie hier ersichtlich, handelt es sich bei den Variablen, die für die jeweiligen Leistungsgruppen unterschiedlich ausfallen, überwiegend um Merkmale der allgemeinen Persönlichkeit sowie technikbezogene Persönlichkeitsaspekte. Besonders gut zwischen den Gruppen zu differenzieren, scheinen die Faktoren Extraversion (mit zwei Facetten und dem Primärfaktor) sowie Gewissenhaftigkeit (mit drei Facetten und dem Primärfaktor). Die Computernutzungsvariablen spielen nur vereinzelt eine Rolle.

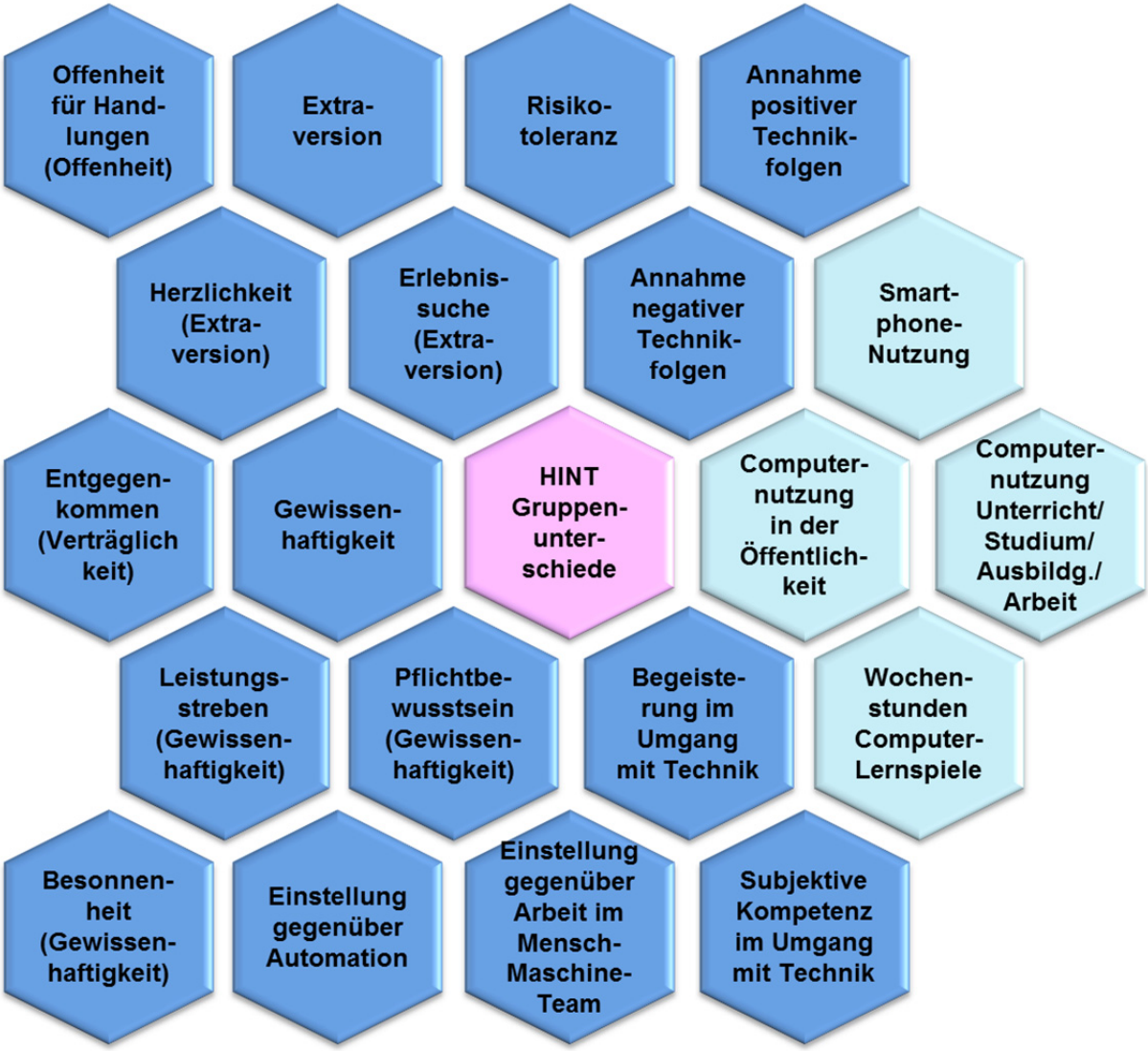


Abbildung 16: Gruppenunterschiede in HINT-Gruppen
(*Anmerkung*: blau=Persönlichkeitsvariablen, hellblau=Computernutzungsvariablen)

Hinsichtlich der inkrementellen Validität der HTQ-S-Skalen konnten drei Merkmale identifiziert werden, die zur Varianzaufklärung beitragen. Wie in Abschnitt 8.5 beschrieben, hatten weitere Ergebnisse der Regressionsanalysen nach Durchführung einer Schrumpfungskorrektur keinen Bestand. Wie in **Abbildung 17** ersichtlich, handelt es sich bei allen drei Merkmalen um technikbezogene Persönlichkeitsvariablen. Deren Einfluss war in den einfachen Korrelationsanalysen zu den HINT-Variablen nicht signifikant geworden (siehe **Abbildung 14** und **Abbildung 15**), zeigt sich aber in der multiplen Korrelationsanalyse, deren Ergebnisse unter 8.5 beschrieben werden.

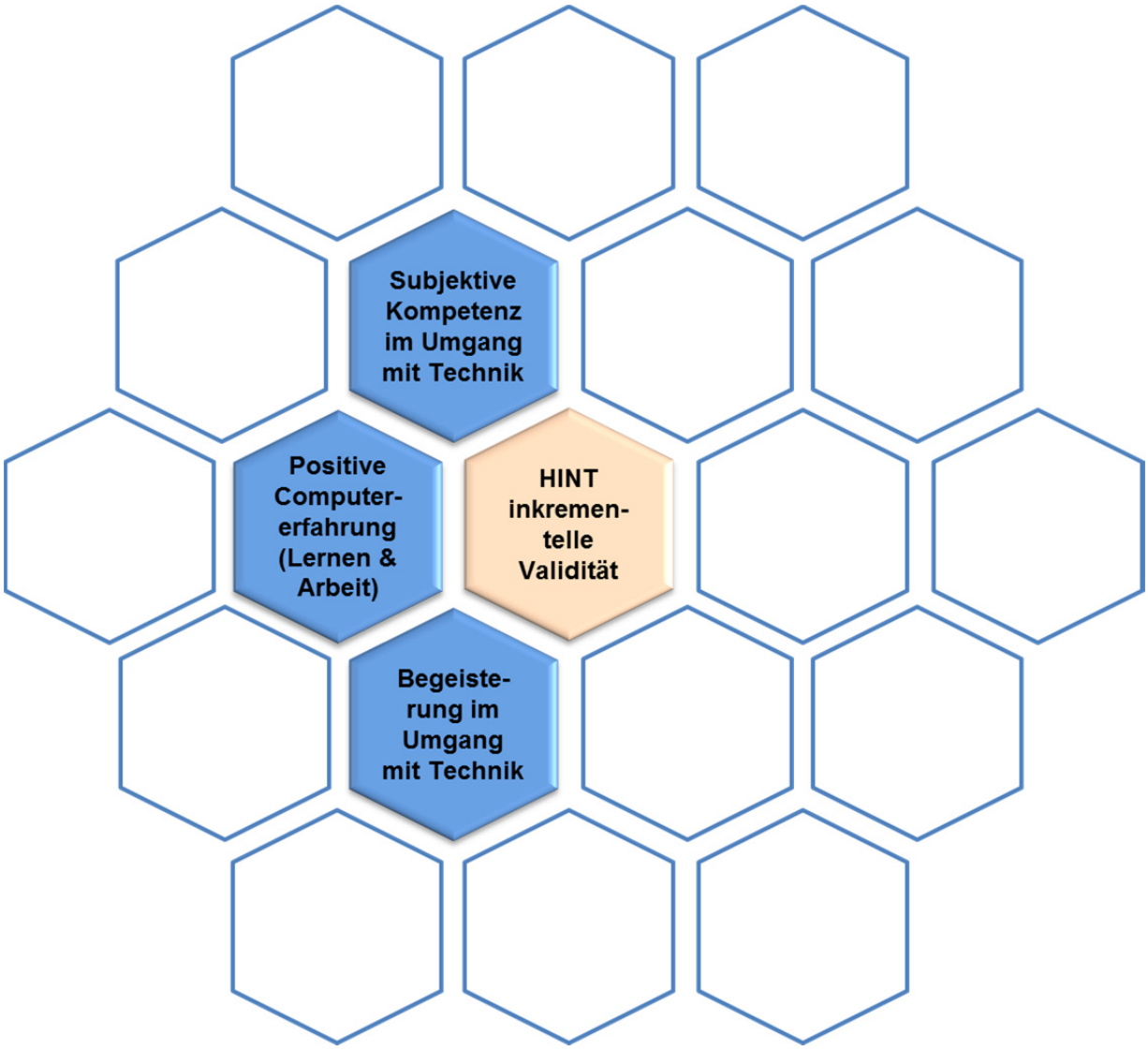


Abbildung 17: Inkrementelle Validität
(*Anmerkung*: blau=Persönlichkeitsvariablen)

9 Diskussion

In den folgenden Abschnitten werden die im vorigen Kapitel berichteten Ergebnisse hinsichtlich ihrer Bedeutung diskutiert. Vorweg sei erwähnt, dass die erzielten Effekte klein bis maximal mittel ausfallen. Wie bereits unter 3.1 beschrieben, lässt sich bei Studien zum Thema Persönlichkeit und Leistung allerdings generell feststellen, dass eher kleine bis maximal mittlere Effekte erzielt werden. Dennoch zeigen sich diese konsistent und sollten daher auch in zukünftiger Eignungsdiagnostik nicht unbeachtet bleiben. Bei gleicher kognitiver Leistungsfähigkeit von Bewerbern kann die Persönlichkeit ausschlaggebend dafür sein, ob eine Eignung besteht oder nicht.

9.1 Zusammenhänge zu aktueller Eignungsdiagnostik – Diskussion zu Hypothese 1

In Kapitel 6 wurde die Hypothese 1 „*Es bestehen Zusammenhänge zwischen den Skalen zur Persönlichkeit und Einstellung des im Rahmen der Dissertation konzipierten Forschungsfragebogens HTQ-S und den Leistungen von Bewerbern in der heutigen Eignungsdiagnostik für Fluglotsen und Piloten.“* formuliert. Bereits in der Einleitung wurde auf die Relevanz der Zusammenhänge hingewiesen, da große Teile der aktuellen Testbatterie Fähigkeiten überprüfen, die auch in Zukunft bedeutsam sein können. Zusammenhänge des HTQ-S sowohl mit heutigen Variablen als auch mit Variablen der Zukunft sind demnach wünschenswert, wenn das Verfahren in zukünftige Eignungsdiagnostik einbezogen werden soll.

Anhand der in Abschnitt 8.3.1 geschilderten Ergebnisse kann Hypothese 1 für mehrere Aspekte der Persönlichkeit und Einstellung bestätigt werden. Wie **Abbildung 11** sowie **Abbildung 12** verdeutlichen, zeigen alle Primärfaktoren des Big Five Persönlichkeitsmodells entweder positiven oder negativen Einfluss auf kognitive Leistungsvariablen der heutigen Eignungsdiagnostik. Auch Variablen der technikbezogenen Persönlichkeits- und Einstellungsskalen sowie Computernutzungsvariablen zeigen prädiktive Validität hinsichtlich der Leistung im Auswahlverfahren. Für andere Aspekte wie z.B. Selbstwirksamkeitserwartung oder Neigung zu risikoreichem Entscheidungsverhalten ergibt sich keine Relevanz für die Leistungsvariablen. Für diese muss die Hypothese verworfen werden. Im Folgenden werden die wichtigsten Ergebnisse zur Eignungsdiagnostik diskutiert.

Aussagekräftig ist der HTQ vor allem für Variablen der Piloten-Grunduntersuchung (BU) sowie für Variablen der zweiten Auswahlstufe (Gerätetests) beider Verfahren (Fluglotsen- und Pilotenauswahl). Hinsichtlich der Basisleistungen der Fluglotsenauswahl (VA) lassen sich nur wenige Prädiktoren ausmachen. So hängen nur drei der Persönlichkeitsskalen signi-

fikant mit den Basisleistungen der Fluglotsenauswahl zusammen und erzielen zudem nur geringe Effekte, diese jedoch in vermuteter Richtung. Erwartungskonform gehen höhere Kontrollüberzeugungen im Umgang mit Technik und ein stärkeres subjektives Kompetenzgefühl bezüglich Technik mit besserer Leistung in der VA einher. Wie bereits unter 8.3.1 geschildert, handelt es sich bei letzterem Effekt jedoch nur um eine Scheinkorrelationen, die allein auf das Geschlecht zurückzuführen ist: männliche Probanden schneiden in den Bereichen mittlere Testleistung, Aufmerksamkeit sowie Merkfähigkeit besser ab und sind gleichzeitig überzeugter, kompetent im Umgang mit Technik zu sein. Geschlechtsunabhängig geht höhere VA-Leistung mit geringeren Werten hinsichtlich negativer Erfahrungen mit Computern als Lern- und Arbeitsmittel (FIDEC H) einher. Dies entspricht der Annahme (siehe 8.1.1), dass computerbezogene Einstellungen die Leistung bei der Arbeit mit Computern beeinflussen können.

Die überwiegende Anzahl der signifikanten Zusammenhänge mit der VA besteht mit Variablen aus dem Bereich der Computernutzung. Intensivere Computernutzung korrespondiert mit jeweils besseren Leistungen in der VA. Das Item „Wochenstunden Computerspiele insgesamt“ zeigt dabei die meisten Zusammenhänge mit den VA-Werten. Personen, die sich mehrere Stunden wöchentlich mit Computerspielen beschäftigen, schneiden also insgesamt besser in der VA ab, als Personen, die dies weniger oder gar nicht tun. Prüft man die deskriptiven Statistiken der Computernutzungsvariablen, fällt auf, dass im Vergleich zur weiblichen Stichprobe in diesem Bereich durchweg höhere Nutzungswerte in der männlichen Stichprobe zu finden sind (auch wenn nur einige Unterschiede im T-Test signifikant sind). Bei genauerem Hinsehen handelt es sich bei den genannten Zusammenhängen dann auch tatsächlich fast ausschließlich um Geschlechtseffekte. Nach der Auspartialisierung von Geschlecht hat lediglich der Zusammenhang der Stunden, die wöchentlich mit Computerspielen verbracht werden, mit der Konzentrationsleistung in der VA Bestand.

Der im Ergebnisteil beschriebene Befund, dass insbesondere der Teilbereich Technikwissen aus den Basisleistungen der Pilotenauswahl (BU) mit verschiedenen Prädiktoren zusammenhängt, verwundert nicht. Betrachtet man, welche Persönlichkeitsbereiche hier aussagekräftig sind, wird deutlich, dass es sich überwiegend um diejenigen handelt, die eine positive Einstellung gegenüber Technik / Computern und hohe Kontrollüberzeugungen sowie Sicherheit in diesem Bereich beinhalten. Dass sich entsprechende Eigenschaften und Einstellungen auch in größerem Wissen in diesem Bereich niederschlagen, erscheint logisch. Hinsichtlich der allgemeinen Persönlichkeit zeigt Extraversion die eindeutigsten Befunde bei den Pilotenbewerbern. Hier sind es die introvertierteren Kandidaten, die in den Basisleistungen besser abschneiden. Setzt man dieses Ergebnis mit Studien in Bezug, die einen negativen Zusammenhang zwischen Extraversion und beispielsweise Internetnutzung erga-

ben (siehe 3.2), erscheint dieser Zusammenhang sinnhaft. Hintergrund könnte sein, dass introvertiertere Personen sich häufiger mit Computeranwendungen befassen als extravertiertere und daher bessere Leistung in Computertests erzielen.

Ein auffälliger Befund zeigt sich bezogen auf die spezifische Kategorie der Computer-Rollenspiele. Wie bereits erläutert, handelt es sich hierbei um Spiele, in denen Spieler die Rollen fiktiver Charaktere bzw. Figuren übernehmen und soziale Situationen bzw. Abenteuer in einer erdachten Computerwelt erleben. Es ergeben sich negative Zusammenhänge zu drei der fünf Teilbereiche der Basisleistungen aus der Pilotenauswahl (BU). Bewerber, die sich mehrere Stunden pro Woche mit derartigen Spielen beschäftigen, schneiden schlechter bei den Tests zur Merkfähigkeit, Konzentration und Raumorientierung ab, als Bewerber, die sich nicht damit beschäftigen. Vorstellbar ist, dass die Verwendung von Computer-Rollenspielen eine bestimmte Herangehensweise an computerbasierte Aufgaben trainiert, die sich kontraproduktiv auf die Leistung in den Auswahltests auswirkt.

Die Korrelationen zu den Arbeitsproben (Gerätestufe) beider Auswahlverfahren fallen insgesamt höher aus, als diejenigen zu den Basisleistungen und bewegen sich durchschnittlich im mittleren Bereich. Einhergehend mit dem Effekt von Extraversion auf Leistungsvariablen der BU schneiden introvertiertere Bewerber auch im Strip Display Management Test (SDM) besser ab als extravertiertere. Eventuell spielt auch hier die häufigere Beschäftigung mit Computern von introvertierteren Personen eine Rolle. Hinsichtlich der in der heutigen Eignungsdiagnostik beinhalteten Untersuchungsabschnitte VerDi und biografisches Interview (siehe 4.1) dürfte sich der Effekt des positiveren Abschneidens introvertierterer Personen jedoch umkehren. Hier sind es die Extravertierten, die mit ihrer interpersonalen Kompetenz punkten können und nach dem heutigen Anforderungsprofil (das gute Teamarbeit mit anderen Menschen verlangt) als geeignet bewertet werden. Da in der Teamarbeit mit maschinellen Teampartnern derartige Fähigkeiten höchstwahrscheinlich in den Hintergrund treten, ist es positiv zu sehen, dass introvertiertere Personen bessere Ergebnisse in den Computertests erzielen.

Die stärksten Zusammenhänge der HTQ-S-Skalen ergeben sich mit dem Simulator-Test (SIM). Auffällig ist die hohe Korrelation von angenommenen negativen Technikfolgen (TA-EG) mit guter SIM-Leistung. Zunächst eher denkbar wäre hier eher ein gegenteiliger Effekt gewesen. Gutes Abschneiden im SIM verlangt jedoch eine gute Psychomotorik und erfordert im Vergleich zu anderen Tests der Auswahlverfahren „Handarbeit“. Vorstellbar ist, dass Personen, die eher antizipieren, dass Technik negative Folgen hat, in der Interaktion mit selbiger aktiver sind, als Personen, die davon ausgehen, dass das Verlassen auf Technik allein auch zu positiven Resultaten führen wird. Unterstützend für diese Erklärung

ist der Befund, dass das gegenteilige Konstrukt zur Annahme positiver Technikfolgen (TA-EG) negativ auf die Arbeitsprobe DAC wirkt. Auch hierfür könnte die gleiche Erklärung gelten, dass weniger angenommene positive Technikfolgen dazu führen, dass Personen in Interaktion mit einer Maschine (hier während der DAC-Simulation) aktiver sind und bessere Leistungen erzielen. Dem entgegen steht die positive Korrelation des Merkmales mit zwei der kognitiven Leistungsvariablen aus der BU. Wie bereits beschrieben, gehen mehr angenommene positive Technikfolgen mit besseren Ergebnissen in den Teilbereichen Technikwissen und Mathematik einher. Möglicherweise macht eine durch den Arbeitsprobencharakter des DAC hervorgerufene stärkere Involvierung des Probanden in die Mensch-Maschine-Interaktion den Unterschied.

Auch die positive Korrelation zwischen stärkerem Neurotizismus und besserem Abschneiden im SIM erscheint auf den ersten Blick kontraintuitiv. Der mit dem BFI erfasste Neurotizismus beinhaltet jedoch die Selbstbeschreibungen, sich leichter aus der Ruhe bringen zu lassen sowie nervöser und unsicherer zu sein. Ebenso wie antizipierte negative Technikfolgen möglicherweise dazu führen, dass Personen in der Mensch-Maschine-Interaktion aktiver sind, könnten Personen, die sich leichter aus der Ruhe bringen lassen und nervöser sind, im SIM sorgfältiger arbeiten und schneller agieren – angetrieben durch die Angst, Fehler zu machen. Diese Vermutung erfordert jedoch nähere Untersuchungen.

Zunächst durch die Variable Geschlecht verdeckt waren zwei weitere Korrelationen mit den Gerätetests. So gehen nach Auspartialisierung der Variable Geschlecht höhere Kontrollüberzeugungen im Umgang mit Technik mit besserer DAC-Leistung sowie eine ausgeprägtere Begeisterung im Umgang mit Technik mit schlechterer SDM-Leistung einher. Der erste Befund ist hinsichtlich der Annahme, dass höhere Kontrollüberzeugungen dazu führen, dass Personen lieber mit Technik zusammenarbeiten (und damit darin wohl auch besser sind), erwartungskonform. Weshalb jedoch eine ausgeprägtere Begeisterung im Umgang mit Technik zu einer schlechteren SDM-Leistung führt, ist nicht nachvollziehbar und bleibt unklar.

Die im Bereich der Arbeitsproben auftretenden Effekte der Computernutzungs- und Computerspielvariablen sind fast durchweg positiv (siehe 8.3.1). Auffällig ist jedoch der immerhin mittelgroße negative Effekt der Verwendung von Jump'n-Run-Spielen. Je mehr dieser Spieltyp verwendet wird, desto schlechter ist das Abschneiden in der Flugsimulation SIM. Vorstellbar ist, dass das Spielen von Jump'n Run, das schnelles und zwangsweise wenig überlegtes Reagieren erfordert, zu vorschnellen Entscheidungen im Simulator führt. Selbstverständlich muss auch hier zügig entschieden werden, doch unterscheidet sich die Vorgehensweise, die eine sorgfältige Arbeitsweise verlangt, deutlich von der schnellen, intuitiven

und eher wilden Spielweise bei Jump'n Run. Goeters und Lorenz (1985) untersuchten den Einfluss des Videospiels „Pac-Man" auf die damaligen Fähigkeitstests in der Luftfahrtpsychologie des DLR. Hier zeigten sich positive Auswirkungen des „Pac-Man"-Spiels z.B. auf die Wahrnehmungsgeschwindigkeit, Raumvorstellung und Psychomotorik. Weiterführende Untersuchungen zu den Auswirkungen verschiedener Computerspieltypen auf die Leistungsfähigkeit in computerbasierten Tests und Berufen wären sicherlich sinnvoll.

9.2 Zusammenhänge mit zukünftigen Anforderungen – Diskussion zu Hypothese 2

Zur Prüfung der Hypothese 2 *„Es bestehen Zusammenhänge zwischen den Skalen zur Persönlichkeit und Einstellung des im Rahmen der Dissertation konzipierten Forschungsfragebogens HTQ-S und den Leistungen menschlicher Operateure im Mensch-Maschine-Team"* wurde der HTQ-S in Kombination mit der ATM-Simulation HINT eingesetzt. Im Folgenden werden nun die Zusammenhänge zu den einzelnen Teilbereichen des Fragebogens diskutiert.

Wie unter 8.1.1 beschrieben, soll die Einbeziehung der Big Five (gemessen mit NEO-PI-R und BFI) in den HTQ-S ermöglichen, Zusammenhänge der allgemeinen Persönlichkeit mit der Leistung in hoch automatisierten Mensch-Maschine-Teams der Luftfahrt zu erfassen. So sollte geprüft werden, ob die unter 3.1 besprochenen Zusammenhänge zwischen den Big Five und beruflicher Leistung auch für zukünftige Luftfahrtoperateure replizierbar sind.

Wie bereits unter 8.6 zusammengefasst, sind bei HINT im Gegensatz zur Eignungsdiagnostik anstelle der Primärfaktoren ausschließlich Einzelfacetten der Big Five unter den auf die Leistung wirkenden Variablen. Auf der Ebene der Primärfaktoren scheint auf den ersten Blick Offenheit als Prädiktor zu wirken, das Ergebnis stellt sich jedoch nach Auspartialisierung der Variable Geschlecht als Scheinkorrelation heraus. Auf Ebene der Facetten wurde festgestellt, dass reizbarere, d.h. stärker neurotische Personen bei der Bearbeitung von HINT tendenziell langsamer reagieren (positive Korrelation mit Reaktionszeit) und weniger von der Automation gestellte Requests annehmen (negative Korrelation mit RQ-Annahme durch Alpha) als weniger reizbare Probanden. Die Befunde der aktuellen Studie können damit in der Literatur beschriebene Zusammenhänge (siehe 3.1) für HINT teilweise replizieren, die emotionale Stabilität (= geringer Neurotizismus) als guten Prädiktor für allgemeine Arbeitsleistung beschreiben. Die in diesem Bereich korrelierende Variable Reaktionszeit profitiert ebenso wie die Request-Annahme erwartungskonform von geringerem Neurotizismus. Zusätzlich erwartbar wären hier jedoch auch positive Zusammenhänge mit den Abweichungsvariablen (hoher Neurotizismus = hohe Abweichungen) sowie ein negativer Zu-

sammenhang mit der Variable zur Request-Stellung an Bravo (hoher Neurotizismus = wenig Requests) gewesen. Diese Variablen stehen für ein aktives Bearbeiten der Simulation und sind dadurch gut mit der erwähnten allgemeinen Arbeitsleistung vergleichbar.

Hinsichtlich der unter 3.1 besprochenen Zusammenhänge zwischen den Big Five und beruflicher Leistung aus der Literatur zeigt sich, dass auch zum Merkmal Gewissenhaftigkeit Befunde replizierbar sind. Der Faktor Gewissenhaftigkeit, welcher als zuverlässiger Prädiktor für Leistungsvariablen gilt, zeigt mittels seiner Facette Pflichtbewusstsein einen Zusammenhang zur Kooperationsleistung von Probanden (RQ Alpha an Bravo). Jedoch muss erwähnt werden, dass Zusammenhänge zu den anderen Leistungsvariablen nicht bestehen. Mindestens für die beiden Variablen der Target-Abweichungen, die unter anderem für eine gewissenhafte Bearbeitung der Simulation stehen, wären Zusammenhänge zu Gewissenhaftigkeit bzw. den zugehörigen Facetten erwartungskonform gewesen.

Im Bereich der Extraversion wirkt die Facette Erlebnissuche konsistent auf mehrere der HINT-Parameter. Höhere Werte wirken negativ auf die Reaktionszeit (d.h. es liegt eine langsame Arbeitsweise vor) und zudem auf das Aufrechterhalten eines ausgeglichenen Luftverkehrs entsprechend der Anweisungen in der Simulation (gekennzeichnet durch höhere Target-Abweichungen). Wie bereits im Bereich heutiger Eignungsdiagnostik vorgefunden, wirkt sich höhere Extraversion also auch negativ auf die Leistung in der Simulation aus. Dies widerspricht keineswegs bisherigen Befunden aus der Literatur, da diese von einer positiven Wirkung von Extraversion nur berichten, sofern Berufe eine starke interpersonale Komponente haben (siehe 3.1). So mag geringere Extraversion in der Zusammenarbeit mit einem maschinellen statt einem menschlichen Teampartner durchaus positiv sein. Hinzu kommt, dass es in sicherheitsrelevanten Berufen wie Fluglotse oder Pilot schon rein intuitiv wünschenswert erscheint, dass die Berufsinhaber nicht zu abenteuerlustig, draufgängerisch, risikofreudig und waghalsig sind, womit die Facette Erlebnissuche gekennzeichnet ist.

Weitere Effekte hinsichtlich der allgemeinen Persönlichkeit ergeben sich nicht. Hypothese 2 kann in diesem Bereich damit als eingeschränkt gültig bezeichnet werden. Für die meisten Facetten sowie auf der Ebene der Primärfaktoren zeigen sich keine Zusammenhänge zu den Leistungsvariablen aus HINT, die vorliegenden Befunde entsprechen jedoch Ergebnissen aus der Literatur.

Technikbezogene Persönlichkeitsvariablen spielen im Vergleich zu Zusammenhängen mit heutiger Eignungsdiagnostik eine kleinere Rolle für die HINT-Leistung. Unter 8.1.1 wurde zu den entsprechenden Skalen postuliert, warum diese jeweils relevant für die Arbeit in einem Mensch-Maschine-Team sein könnten. Zusammenfassend lässt sich wiederholen, dass aufgrund der stark computer- und technikbezogenen Arbeit im Mensch-Maschine-Team

angenommen wird, dass sowohl computerbezogene Einstellungen, die Sicherheit im Umgang mit Computeranwendungen, Technikaffinität als auch Kontrollüberzeugungen im Umgang mit Technik die Leistung beeinflussen können. Ausgeprägtere Sicherheit, positivere Einstellungen und Kontrollüberzeugungen sowie stärkere Technikaffinität sollten sich demnach in besserer HINT-Leistung niederschlagen. Drei der entsprechenden Skalen ergeben sich als tatsächlich statistisch bedeutsam für die Leistung in der Simulation.

Eine ausgeprägtere Sicherheit im Umgang mit Computern geht mit höherer Reaktionszeit einher. Dies wirkt entsprechend obiger Erläuterungen zunächst nicht erwartungskonform. Allerdings ist vorstellbar, dass sich im Umgang mit dem Computer sicherere Probanden mehr Zeit bei der Bearbeitung von Requests lassen und mehr nachdenken, bevor sie einen Request bearbeiten, während unsicherere Probanden lieber schnell reagieren, um nichts falsch zu machen. Ein weiterer Effekt zeigt sich hinsichtlich Technikaffinität. Ein stärkerer subjektiver Eindruck, Technik habe negative Folgen, geht mit seltenerer Ablehnung der Requests des automatisierten Partners einher. Anzunehmen wäre jedoch eigentlich ein Effekt in umgekehrter Richtung: die Überzeugung, Technik habe negative Folgen, sollte sich in stärkerer Ablehnung von Anfragen der Automation (deren negative Auswirkungen man befürchtet) ausdrücken. Dies ist hier jedoch nicht der Fall.

Der eigens für die Fragestellung dieser Arbeit zusammengestellte Fragebogen zur Einstellung gegenüber Automation dagegen liefert Ergebnisse in angenommener Richtung. Entsprechend der Vermutung, dass Personen mit positiver Einstellung gegenüber einem maschinellen Teampartner auch die Zusammenarbeit mit selbigem besser gelingt, zeigen sich Probanden mit entsprechender Einstellung kooperativer gegenüber der Automation. Requests des automatisierten Sektors Bravo werden seltener abgelehnt und häufiger angenommen.

Für drei der technikbezogenen Skalen kann Hypothese 2 also angenommen werden. Für Aspekte wie Kontrollüberzeugungen im Umgang mit Technik oder computerbezogene Einstellungen ergibt sich allerdings entgegen vorheriger Annahmen keine Relevanz für die HINT-Variablen. Entsprechend muss die Hypothese für diese verworfen werden.

Hinsichtlich des Merkmals der Neigung zu risikoreichem Entscheidungsverhalten (BART) wurde unter 8.1.1 postuliert, risikoreicheres Entscheidungsverhalten könne mit schlechterer HINT-Leistung einhergehen. Dies kann, wie bereits unter 8.3.2 beschrieben, anhand der Korrelation zwischen BART und den Target-Abweichungen von Alpha bestätigt werden: eine höhere Neigung zu risikoreichem Entscheidungsverhalten korreliert mit höheren Target-Abweichungen. Probanden mit niedrigerer Risikoneigung scheinen also regelkonformer (Minimierung der Diskrepanz zwischen Target- und Actual-Werten) zu arbeiten. Denkbar

ist auch ein weniger sorgfältiges Überwachungsverhalten, das zu den höheren Abweichungen führt.

Die letzte Kategorie, in der sich signifikante Effekte ergeben, ist die Kategorie der Computernutzung. Wie unter 8.1.1 beschrieben, ist aufgrund der stark computerisierten Tätigkeit im Mensch-Maschine-Team ein Zusammenhang zwischen bisheriger Computernutzung und Leistung in solch einem Team anzunehmen. Ausgeprägtere Computernutzung sollte also mit besserer Leistung in HINT einhergehen. Dies ergibt sich jedoch nicht für die Nutzung von Smartphones und MP3-Playern. Beide Variablen zeigen signifikant negative Effekte hinsichtlich der Kooperation mit dem maschinellen Partner (negative Korrelation mit RQ Alpha an Bravo bzw. RQ-Annahme durch Alpha). Zudem präsentiert sich Smartphone-Nutzung als Prädiktor für höhere Target-Abweichungen sowohl im eigenen als auch im automatisierten Partnersektor. Dies steht dem zuvor geschilderten Befund gegenüber, dass Smartphone-Nutzung sich positiv auf die SIM-Leistung in der Pilotenauswahl auswirkt. An dieser Stelle sei jedoch erwähnt, dass der Effekt der Smartphone-Nutzung sicherlich stark von gesellschaftlichen Wandlungen bzw. dem technischen Fortschritt in der Mobilfunkindustrie beeinflusst ist. Gegebenenfalls wird dieser Befund in einigen Jahren hinfällig sein.

Betrachtet man den Bereich der wöchentlichen Beschäftigung mit Computeranwendungen, fällt auf, dass diese nicht uneingeschränkt positiv auf die HINT-Leistung wirkt. Stattdessen scheint es darauf anzukommen, welcher Art die Beschäftigung ist. Eine höhere Anzahl an wöchentlichen Spielstunden am Computer zeigt einen positiven Einfluss auf die HINT-Leistung. Insbesondere wirken sich Schach- und Lernspiele am Computer positiv auf die Reaktionsgeschwindigkeit aus, so dass hier geringere Reaktionszeiten entstehen. Möglicherweise wird durch Schach sowie Lernspiele eingeübt, zügige und dennoch durchdachte Entscheidungen zu treffen, wovon die Bewerber bei der Request-bearbeitung profitieren könnten. Sportspiele sowie die insgesamt wöchentlich am Computer verbrachte Zeit wirken sich dagegen negativ aus.

Anhand der geschilderten Ergebnisse kann Hypothese 2 hinsichtlich der Computernutzung für einige der Items bestätigt werden. Für andere Aspekte ergibt sich keine Relevanz für die HINT-Variablen, so dass die Hypothese für diese verworfen werden muss. Abschnitt 9.5 verdeutlicht, welche Variablen demnach beibehalten werden sollten und welche nicht.

In der Gesamtschau auf die korrelativen Zusammenhänge von HTQ-S und HINT zeigt sich, dass Korrelationen für die Skalen zu Cognitive Failures (CFQ), Flexibilität und Teamorientierung aus dem BIP, Bedürfnis nach Teamarbeit (BnT), Risikotoleranz aus dem CaP, Einstellung gegenüber Arbeit im Mensch-Maschine-Team (EgMMT), Individuelle Einstellung gegenüber Teamarbeit (FIT) sowie allgemeine Selbstwirksamkeitserwartung gänzlich feh-

len. Damit scheiden unter anderem alle Skalen, die Aspekte von Teamarbeit betreffen, als Prädiktoren für Leistung in HINT aus.

Dennoch könnte das Bedürfnis nach Teamarbeit in Zukunft ausschlaggebend für die Eignung für Mensch-Maschine-Teams sein. Vorstellbar ist, dass die Persönlichkeitsaspekte zu Teamarbeit erst relevant werden, sofern über einen langen Zeitraum wie im alltäglichen Berufsleben ohne menschliche Teampartner gearbeitet werden muss. Die Bearbeitung von HINT nahm für die Studienteilnehmer einen vergleichsweise kleinen Zeitraum eines einzelnen Tages in Anspruch. Dieser fällt auch für sehr teamorientierte Personen mit einem starken Bedürfnis nach Teamarbeit möglicherweise nicht so stark ins Gewicht, dass darunter ihre Leistungsfähigkeit leiden würde. Nur in Langzeitstudien könnte solch ein Effekt sichtbar gemacht werden.

9.3 Gruppenunterschiede – Diskussion zu Hypothese 3

Zum Test von Hypothese 3 *„Probanden mit besseren Leistungen im Mensch-Maschine-Team unterscheiden sich in den Skalen des HTQ-S signifikant von Probanden mit schlechteren Leistungen im Mensch-Maschine-Team.“* wurden, wie in Abschnitt 7.3.3 beschrieben, mittels Median-Split Gruppen mit besseren bzw. schlechteren Leistungswerten in den verschiedenen HINT-Variablen gebildet.

Die Gruppenvergleiche dienen dazu, erste Unterschiede zwischen Probanden im HTQ-S je nach Performance in der HINT-Simulation aufzuzeigen. Die Einteilung in zwei Gruppen anhand eines Median Splits kann methodenkritisch betrachtet werden, da bei solch einer künstlichen Dichotomisierung Probanden mit mittleren Leistungen einer der beiden Gruppen (niedrige oder hohe Ausprägung) zugeordnet und dadurch eventuelle Unterschiede zwischen Personen mittlerer Ausprägung und solchen mit niedriger oder hoher Ausprägung verdeckt werden (MacCallum, Zhang, Preacher, & Rucker, 2002). Für einen ersten Eindruck davon, ob Probanden sich in Abhängigkeit von ihrer Leistung signifikant im HTQ-S unterscheiden, wird die Verwendung eines Median Splits jedoch als legitim angesehen. Sollen Skalen des HTQ-S für zukünftige Eignungsdiagnostik eingesetzt werden, ist es jedoch selbstverständlich notwendig, begründete Grenzwerte zu bestimmen und diese zu validieren.

Alle ermittelten Effekte liegen bei .31 bis .65 und damit nach Bortz (2005) im schwachen bis mittleren Bereich. Dies wird im Kontext bisheriger Forschung zu Persönlichkeit und Leistung als erfolgreich bewertet, da Effekte in diesem Gebiet üblicherweise in dieser Größenordnung und oft sogar ausschließlich im schwächeren Bereich liegen (siehe auch 3.1).

Als erstes wird nun die Variable Reaktionszeit hinsichtlich Gruppenunterschieden betrachtet. Wie unter 8.4 berichtet, zeigt sich, dass Probanden mit niedrigeren Werten (schnelle Operateure, Gruppe 1), signifikant niedrigere Werte hinsichtlich Erlebnissuche, Offenheit für Handlungen, Begeisterung im Umgang mit Technik, angenommenen positiven Folgen von Technik sowie Smartphone-Nutzung haben. Signifikant höher ist der Wert der wöchentlichen Nutzung von Computer-Lernspielen. Einschränkend muss an dieser Stelle erwähnt werden, dass die dargestellten Unterschiede zwar signifikant sind, sich jedoch im Bereich von maximal 0.22 Testpunkten der verwendeten Likert-Skala mit fünfstufigem Antwortformat bewegen. Die im Folgenden angestellten inhaltlichen Überlegungen sind dementsprechend als Tendenzen zu verstehen.

Erlebnissuche deckt als Facette der Extraversion unter anderem Abenteuerlust und Wagemut ab. Hinsichtlich Reaktionszeit kann vermutet werden, dass wagemutigere Probanden sich trauen, die auftauchenden critical Aircraft (CAC) mit Verzögerung zu bearbeiten und ggf. erst eine aktuelle Aktion beenden, bevor sie sich einem CAC zuwenden. Weniger wagemutige Probanden versuchen dagegen die Gefahr, nötige Korrekturen an CAC zu spät vorzunehmen, zu vermeiden und reagieren schneller. Ähnliches könnte für den Gruppenunterschied bezüglich der Variable Offenheit für Handlungen gelten, die als Facette der Offenheit für neue Erfahrungen Probierfreude und Flexibilität beinhaltet. Flexiblere Charaktere lassen sich möglicherweise von drohenden CAC nicht so leicht aus der Ruhe bringen wie weniger flexible. Die genannten Effekte korrespondieren mit dem oben beschriebenen Ergebnis, dass eine ausgeprägtere Sicherheit im Umgang mit Computern korrelativ mit höherer Reaktionszeit einhergeht. Vorstellbar ist hierzu, wie bereits beschrieben, dass im Umgang mit dem Computer sicherere Probanden sich mehr Zeit bei der Bearbeitung von Requests lassen und mehr nachdenken, bevor sie einen Request bearbeiten, während unsicherere Probanden lieber schnell reagieren, um nichts falsch zu machen.

Bezogen auf die Variablen Begeisterung im Umgang mit Technik und positive Folgen von Technik beschreiben sich langsamere Probanden als signifikant begeisterter und nehmen signifikant mehr positive Folgen an als schnellere. Denkbar ist, dass derartige Antizipationen zu einem entspannteren und dadurch langsameren Vorgehen im Umgang mit Technik führen. Die dichotome Variable Smartphone-Nutzung (0=nein / 1=ja) zeigt, dass sich in der Gruppe der langsameren Probanden mehr Smartphone-Nutzer befinden als in der Gruppe der schnelleren Probanden. Hier könnte ein durch den regelmäßigen Umgang mit einem technischen Gerät entspannteres und dadurch langsameres Vorgehen auschlaggebend sein.

Die wöchentliche Stundenanzahl der Verwendung von Computer-Lernspielen liegt insgesamt auf einem sehr niedrigen Niveau. Ein Wert von 1 bedeutet in diesem Fall eine Nutzung

von bis zu zwei Stunden pro Woche, ein Wert von 0 gar keine Nutzung. Während in der Gruppe der langsameren Probanden fast gar keine Lernspiele gespielt werden, profitieren die Probanden aus der Gruppe der schnelleren Probanden eventuell von solchem (wenn auch wenig genutzten) Training und können daher schneller reagieren.

Hinsichtlich der Target-Abweichungen im Sektor Alpha wurde berichtet, dass Probanden mit geringeren Target-Abweichungen (Gruppe 1) signifikant höhere Werte bezüglich Entgegenkommen, Pflichtbewusstsein, Leistungsstreben, Gewissenhaftigkeit und Kompetenz im Umgang mit Technik aufweisen. Smartphone-Nutzer befinden sich signifikant öfter in der Gruppe mit höheren Target-Abweichungen. Auch hier ist jedoch zu beachten, dass sich die Mittelwertsunterschiede auf vergleichsweise niedrigem Niveau bewegen. Erneut sind maximal 0.22 Testpunkte Unterschied festzustellen.

Aus dem Bereich der Gewissenhaftigkeit finden sich Gruppenunterschiede sowohl für die Facetten Pflichtbewusstsein und Leistungsstreben als auch den Generalfaktor. Entsprechend der Befunde der in Kapitel 3 dargestellten Literatur zeigt sich dieser Persönlichkeitsbereich also auch hier als aussagekräftig für Leistung: pflichtbewusstere Probanden haben signifikant niedrigere Target-Abweichungen in ihrem Sektor als weniger pflichtbewusste. Die vermissten Bezüge zwischen Gewissenhaftigkeit und den Abweichungsvariablen lassen sich also durch Gruppenunterschiede doch noch feststellen.

Entgegenkommen erfasst charakteristische Reaktionen auf zwischenmenschliche Konflikte als Facette der Verträglichkeit. Entgegenkommendere, also sanftere und nachgiebigere Probanden erzielen bessere Leistungen in ihrem Sektor als weniger entgegenkommende Probanden. Die Ursache dieses Gruppenunterschiedes bleibt offen. Nachvollziehbar ist das bessere Abschneiden von Probanden mit stärker gefühlter Kompetenz im Umgang mit Technik. Diese scheint sich auch in ausgeprägterer tatsächlicher Kompetenz im Umgang mit Technik niederzuschlagen.

Erneut spielt auch die Smartphone-Nutzung der Probanden eine Rolle. In der Gruppe mit geringeren Target-Abweichungen im Sektor Alpha befinden sich signifikant weniger Personen, die ein Smartphone nutzen, als in der anderen Gruppe. Smartphone-Nutzer erzielen demnach tendenziell schlechtere Leistungen als Nichtnutzer, was sich bereits in den Korrelationsanalysen herauskristallisierte.

Ein ähnliches Bild wie für die Target-Abweichungen im Sektor Alpha zeigt sich für die Abweichungen im Sektor Bravo (siehe unter 8.4). Erneut präsentiert sich Gewissenhaftigkeit mit mehreren Einzelfacetten sowie auch als Primärfaktor als Persönlichkeitsaspekt, der förderlich für gute Leistung ist. Gruppe 1 mit niedrigeren Target-Abweichungen im Sektor Bravo zeigt signifikant höhere Werte in diesem Bereich. Signifikant niedrigere Wer-

te liegen dagegen erneut bei der Smartphone-Nutzung sowie zusätzlich bei der Computernutzung in Unterricht, Studium usw. sowie der Computernutzung in der Öffentlichkeit vor. Wie zuvor ist zu beachten, dass sich die Mittelwertsunterschiede auf vergleichsweise niedrigem Niveau bewegen. Maximal sind 0.25 Testpunkte Unterschied festzustellen.

Die signifikant höheren Werte hinsichtlich der Gewissenhaftigkeitsvariablen entsprechen den Befunden zu den Target-Abweichungen im Sektor Alpha. Die Betrachtungen hierzu werden daher nicht erneut ausgeführt. Zusätzlich zur Smartphone-Nutzung gibt es für den Sektor Bravo Gruppenunterschiede in zwei weiteren Computernutzungsvariablen. In der Gruppe mit besserer Leistung befinden sich weniger Probanden, die Computer in Unterricht, Studium usw. sowie in der Öffentlichkeit nutzen. Die Ursache dieses Effekts erschließt sich aus der aktuellen Untersuchung nicht und sollte in weiterführender Forschung analysiert werden.

Für die Variable zur Ablehnung von an Alpha gestellten Requests (siehe auch hierzu 8.4) ergeben sich drei Persönlichkeitsbereiche mit signifikanten Gruppenunterschieden. Probanden, die weniger Requests ablehnen (Gruppe 1), beschreiben sich im HTQ-S als signifikant weniger herzlich, erwarten signifikant mehr negative Folgen von Technik und haben eine signifikant positivere Einstellung gegenüber Automation. Erneut ist zu beachten, dass sich die Mittelwertsunterschiede auf vergleichsweise niedrigem Niveau bewegen. In diesem Fall sind maximal 0.22 Testpunkte Unterschied festzustellen.

Herzlichkeit beinhaltet als Facette der Extraversion warmherziges und freundliches Verhalten anderen gegenüber, herzliche Personen gehen leicht Bindungen zu anderen ein. Dass herzlichere Probanden Requests öfter ablehnen als weniger herzliche, widerspricht der Annahme, das automatisierte System werde als wirklicher Teampartner wahrgenommen. In diesem Fall sollte ausgeprägtere Herzlichkeit – Bindung zum Teampartner – mit weniger Ablehnung von dessen Vorschlägen einhergehen. Hingewiesen sei jedoch auf den sehr geringen Testwertunterschied von lediglich 0.14 Punkten, so dass das Ergebnis nur eine kleine Tendenz darstellt. Auch das Ergebnis hinsichtlich der antizipierten negativen Folgen von Technik fällt eher entgegen den Erwartungen aus. Probanden, die mehr Requests ablehnen (Gruppe 2), gehen von geringeren negativen Folgen von Technik aus als die Vergleichsgruppe. Nachvollziehbar wäre stattdessen gewesen, dass die Antizipation negativer Folgen von Technik dazu führt, die Vorschläge eines technischen Systems vermehrt abzulehnen. Erwartungskonform ist dagegen der Gruppenunterschied im Bereich der Einstellung gegenüber Automation. In der Gruppe der weniger ablehnenden Probanden ist eine positivere Einstellung vorhanden. Auffällig ist hier der Vergleich der beiden Konstrukte: die Gruppe, welche mehr Requests ablehnt, erwartet weniger negative Folgen von Technik, hat aber eine

negativere Einstellung gegenüber Automation als die Vergleichsgruppe. Ein Grund mag darin liegen, dass die Items der TA-EG-Skala zu negativen Folgen von Technik eher auf elektronische Geräte des Alltags bezogen sind, während die Skala zur Einstellung gegenüber Automation spezifisch auf Automation, z.B. in der Luftfahrt oder Medizin eingeht. Letztere Bereiche haben mit der Bearbeitung von HINT offensichtlich und nachvollziehbarerweise mehr gemeinsam als das Bedienen elektronischer Alltagsgeräte. Dieses wird von den Probanden möglicherweise gar nicht mit der Bearbeitung von HINT in Beziehung gesetzt.

Gruppenunterschiede zur Annahme von erhaltenen Requests ergeben sich für drei Variablen. Probanden, die weniger Requests annehmen (Gruppe 1), sind signifikant extravertierter, risikotoleranter und antizipieren signifikant weniger negative Folgen von Technik. Im Vergleich zu den bisherigen Gruppenunterschieden ergibt sich für den Bereich Extraversion der bisher höchste Testwertunterschied von 1.46 Punkten. Die beiden anderen Variablen bewegen sich im üblichen Bereich von bis zu 0.21 Punkten Unterschied. Der Primärfaktor Extraversion zeigt also die deutlichste Differenz zwischen den beiden Gruppen. Extravertierte Persönlichkeiten zeichnen sich z.B. durch Geselligkeit, Freundlichkeit, Unternehmungsfreude, Dominanz, Willensstärke und Optimismus aus (Abschnitt 8.1.1 geht näher darauf ein). Probanden der Gruppe 1, die weniger der an sie gestellten Requests annehmen als die Probanden der Vergleichsgruppe, sind signifikant extravertierter. Hintergrund könnte sein, dass der Dominanzaspekt dieser Persönlichkeitseigenschaft dazu führt, dass die Probanden lieber eigene Handlungen/ Konzepte ausführen möchten und aufgrund ihres Optimismus auch davon ausgehen, dass diese funktionieren werden. Die Vorschläge des automatisierten Partners könnten dadurch sorgloser abgelehnt werden. Der Gruppenunterschied im Bereich Risikotoleranz lässt vermuten, dass risikotolerantere Probanden sich eher trauen, Requests abzulehnen. Ihre Angst, etwas falsch zu machen, indem eigene Entscheidungen verfolgt werden, dürfte geringer sein. Das Ergebnis hinsichtlich der antizipierten negativen Folgen von Technik fällt analog zu der Variable für Request-Ablehnung durch Alpha eher entgegen den Erwartungen aus. Probanden, die mehr Requests annehmen (Gruppe 2), gehen von stärkeren negativen Folgen von Technik aus als die Vergleichsgruppe. Nachvollziehbar wäre stattdessen, dass die Antizipation negativer Folgen von Technik dazu führt, die Vorschläge eines technischen Systems seltener anzunehmen.

Nicht bei zweiseitiger, jedoch bei einseitiger Testung signifikant ist ein Gruppenunterschied hinsichtlich der Variable Einstellung gegenüber Arbeit im Mensch-Maschine-Team. Aufgrund der in Abschnitt 8.1.1 postulierten Annahme, dass Personen mit positiver Einstellung gegenüber Arbeit im Mensch-Maschine-Team auch bessere Leistungen bei der tatsächlichen Arbeit mit einem maschinellen Teampartner erzielen bzw. gewillter sind, mit diesem zu ko-

operieren, wurde in diesem Fall eine gerichtete Testung als angemessen betrachtet. Probanden der Gruppe 1 mit weniger angenommenen Requests sind erwartungskonform weniger positiv gegenüber Arbeit im Mensch-Maschine-Team eingestellt. Probanden, die mehr Requests des maschinellen Teampartners annehmen (Gruppe 2), sind solcher Zusammenarbeit gegenüber signifikant positiver eingestellt.

Die Ergebnisse zu den Gruppenunterschieden sind für die genannten Merkmale als Erfolg zu werten. Auch wenn die Ergebnisse nicht immer in erwarteter Richtung ausfallen, sind die entsprechenden Variablen dennoch aussagekräftig und können zur Differenzierung von Probanden herangezogen werden. Hypothese 3, in der die Gruppenunterschiede postuliert wurden (s.o.), kann also für die entsprechenden Variablen bestätigt werden. Besonders erfreulich sind die Effekte für die Skalen Einstellung gegenüber Automation und Einstellung gegenüber Arbeit im Mensch-Maschine-Team, da diese eigens entwickelt wurden, um für die Teamarbeit mit automatisierten Systemen zwischen Probanden zu differenzieren.

9.4 Inkrementelle Validität – Diskussion zu Hypothese 4

Hypothese 4 *„Die Skalen des HTQ-S klären gegenüber kognitiven Variablen inkrementelle Validität zur Vorhersage der Leistung im Mensch-Maschine-Team auf.“* wurde mittels multipler Regressionsanalyse getestet. Die Ergebnisse fallen in der Gesamtschau unerwartet gering aus. Die Varianzaufklärung durch die Skalen ist mager und nur für die HINT-Variable RQ Alpha an Bravo überhaupt signifikant. Das subjektive Kompetenzgefühl der Probanden gegenüber Technik, ihre positiven Erfahrungen mit Computeranwendungen im Bereich Lernen und Arbeit sowie ihre Begeisterung im Umgang mit Technik präsentieren sich als valide Prädiktoren und klären insgesamt 10% zusätzlicher Varianz auf. Weitere Ergebnisse haben bei Anwendung der Schrumpfungskorrektur nach Olkin und Pratt (1958) keinen Bestand (siehe 8.5). Jedoch fallen auch die prädiktiven Eigenschaften der zur Referenz verwendeten kognitiven Variablen gering bis bedeutungslos aus. Kritisch betrachtet werden sollte daher, ob die angewandte Methode der multiplen Regression mit den Gegebenheiten hinsichtlich der Anzahl der Prädiktoren und der vorliegenden Stichprobengröße überhaupt andere Werte als die vorliegenden ergeben konnte. Auffällig ist, dass in fast jeder der vorgenommenen Analysen zunächst Prädiktoren signifikant wurden, nach der Anwendung der Schrumpfungskorrektur jedoch nicht mehr bedeutsam waren.

Ziel der Schrumpfungskorrektur ist (wie unter 7.3.4 beschrieben), die Überschätzung multipler Korrelationen durch eine große Anzahl von Prädiktoren zu relativieren. Betrachtet man jedoch die im Methodenkapitel dargestellte Formel auf Seite 59, wird deutlich, dass im vorliegenden Fall mit einer großen Anzahl von Prädiktoren und einer im Vergleich dazu

sehr geringen Stichprobengröße die ermittelten multiplen Korrelationen schon rein mathematisch kaum, bzw. nur bei sehr hohen Werten, die Schrumpfungskorrektur überstehen können. Empfohlen wird im Allgemeinen ein N von mindestens 20 pro Prädiktor, das im vorliegenden Fall nicht annähernd erreicht wird. Da die Stichprobengröße für die Studie mit HTQ-S und HINT durch äußere Faktoren wie Bereitschaft der Probanden zur Teilnahme und zeitliche Begrenzung der Datenerhebung durch Vorgaben des Projektplanes nicht erhöht werden konnte, musste mit den damit einhergehenden Begebenheiten konstruktiv umgegangen werden. So werden die Ergebnisse zur inkrementellen Validität als erster Schritt in die richtige Richtung gewertet. In weiterführender Forschung sollte geprüft werden, ob die Varianzaufklärung der HTQ-S-Prädiktoren, die nach Schrumpfungskorrekturen keinen Bestand hat, zu Recht ihre Signifikanz verliert oder ob sie sich bei einer größeren Stichprobe und dem Einsatz nur ausgewählter Prädiktoren doch als valide erweist.

9.5 Fazit und Ausblick

In den beschriebenen Studien konnten kleine bis mittlere Effekte von Persönlichkeitsmerkmalen auf die Leistung sowohl in heutiger Eignungsdiagnostik für Fluglotsen und Piloten als auch hinsichtlich zukünftiger Anforderungen ermittelt werden. Die inkrementelle Validität des eingesetzten Fragebogens HTQ-S konnte für drei der Prädiktoren nachgewiesen werden. Auch wenn dies gemessen an der großen Anzahl der Prädiktoren ein kleiner Prozentsatz ist, wird die Arbeit als Erfolg gewertet. Erstmalig wurde ein spezifischer Persönlichkeits- und Einstellungsfragebogen für zukünftige Anforderungen an Luftfahrtoperateure zusammengestellt, der sich mit relevanten Aspekten für die Arbeit mit hoch automatisierten Systemen befasst. Dieser wurde ausführlich auf seine psychometrischen Eigenschaften überprüft, zudem erfolgte eine erste Prüfung der prädiktiven Validität. Dass nicht alle der verwendeten Skalen aussagekräftig sind, ist erwartet und erwünscht worden. Wie bereits in der Einleitung erwähnt, wurde ein sehr offenes und exploratives Design mit vielen Prädiktoren gewählt. Ziel dieser ersten Arbeit zu dem Thema war entsprechend, relevante Merkmale zu identifizieren und irrelevante zu eliminieren, so dass ein ökonomischer Einsatz des Fragebogens HTQ-S möglich wird. Voraussichtliche Einsatzmöglichkeiten bestehen, wie bereits erwähnt, in der Human-Factors- und Automationsforschung sowie auch ganz konkret in zukünftiger Eignungsdiagnostik für Luftfahrtoperateure.

Höhere als die berichtete prädiktive Validität des Fragebogens könnte wahrscheinlich in einer Stichprobe aus der Allgemeinbevölkerung mit größerer Varianz in vielen der verwendeten Persönlichkeitsskalen erzielt werden. Bewerber, die sich für eine Ausbildung in dem schon heute sehr technischen Bereich der Luftfahrt interessieren, sind mit ihrem Interesse an diesem Berufsbereich bereits durch Selbstselektion vorausgewählt. Dies dürfte sich beispielsweise an ihrer Einstellung gegenüber Technik bemerkbar machen. Ein Hinweis darauf ist der unter 8.1.4 geschilderten Sachverhalt, dass der Mittelwert der meisten technikbezogenen Skalen über dem theoretischen Skalenmittel von drei liegt. Da jedoch auch zukünftige Bewerber aus der gleichen Population stammen dürften und daher eine Differenzierung innerhalb dieser Gruppe vorrangig ist, wird dies als unkritisch betrachtet.
Welche der verwendeten HTQ-S-Skalen für weitere Forschung bzw. langfristig für die Eignungsdiagnostik für zukünftige Luftfahrtoperateure eingesetzt werden sollten und welche mangels Aussagekraft aus dem HTQ-S entfernt werden sollten, fasst **Tabelle 38** auf den folgenden Seiten in einem Fazit zusammen. Die Merkmale, zu denen neben „Ja“ oder „Nein“ weitere Kommentare in der Tabelle enthalten sind, werden untenstehend mit ergänzenden Erklärungen noch einmal aufgegriffen.

Tabelle 38: Fazit für die einzelnen Skalen des HTQ-S

HTQ-S	***Beibehalten: ja oder nein***
Balloon Analogue Risk Task	Ja, jedoch schwierig in Eignungsdiagnostik (s.u.)
BFI-10: Neurotizismus	Ja
BFI-10: Extraversion	Ja
BFI-10: Offenheit	Ja
BFI-10: Verträglichkeit	Ja
BFI-10: Gewissenhaftigkeit	Ja
BIP: Teamorientierung	Ja, da eventuell in Langzeitstudien wichtig
BIP: Flexibilität	Ja
Bedürfnis nach Teamarbeit	Ja
CaP: Risikotoleranz	Ja
Cognitive Failures Questionnaire	Nein
Einstellung gegenüber Automation	Ja
Einstellung gegenüber Arbeit im Mensch-Maschine-Team	Ja
FIT: Bereitschaft zur Teamarbeit	Ja, da eventuell in Langzeitstudien wichtig
FIT : Vorbehalte gegenüber Teamarbeit	Ja, da eventuell in Langzeitstudien wichtig
FIDEC B: Pos. Erfahrungen Unterhaltung	Ja
FIDEC F: Pos. Erfahrungen Lernen & Arbeit	Ja
FIDEC G: Neg. Erfahrungen Unterhaltung	Ja

HTQ-S	***Beibehalten: ja oder nein***
FIDEC H: Neg. Erfahrungen Lernen & Arbeit	Ja
Sicherheit im Umgang mit Computern	Ja
Kontrollüberzeugungen im Umgang mit Technik	Ja
NEO-PI-R: Neurotizismus	Nein, s.u.
NEO-PI-R: Extraversion	Nein, s.u.
NEO-PI-R: Offenheit	Nein, s.u.
NEO-PI-R: Verträglichkeit	Nein, s.u.
NEO-PI-R: Gewissenhaftigkeit	Nein, s.u.
Selbstwirksamkeitserwartung	Nein
TA-EG: Begeisterung	Ja
TA-EG: Subjektive Kompetenz	Ja
TA-EG: Positive Technikfolgen	Ja
TA-EG: Negative Technikfolgen	Ja
Wochenstunden Computer insgesamt	Ja
Jahre Computernutzung	Ja
PC-/Laptop-Nutzung	Nein
Smartphone-Nutzung	Nein, nur evtl. in Forschungssettings mit entsprechender Fragestellung
Spielkonsolen-Nutzung	Nein
MP3-Player-Nutzung	Nein, nur evtl. in Forschungssettings mit entsprechender Fragestellung

HTQ-S	***Beibehalten: ja oder nein***
Nutzung tragbarer Videospiele (Handheld-Nutzung)	Ja
Computernutzung Schule/Arbeit/Ausbildung/Uni	Ja
Computernutzung Zuhause	Nein
Computernutzung Öffentlichkeit	Ja
Computernutzung bei Freunden	Nein
Wochenstunden Computerspiele insgesamt	Ja
Wochenstunden Computer-Strategiespiele	Ja
Wochenstunden Computer-Jump'n-Run	Ja
Wochenstunden Computer-Egoshooter	Nein
Wochenstunden Computerschach	Ja
Wochenstunden Computer-Rollenspiele	Ja
Wochenstunden Computer-Lernspiele	Ja
Wochenstunden Computer-Sportspiele	Ja
Wochenstunden weitere Computerspiele	Nein

Wie bereits angekündigt, soll zum Abschluss Bezug auf die mit Kommentaren versehenen Merkmale genommen werden. So ist die Balloon Analogue Risk Task BART als Prädiktor aussagekräftig, aber in der vorliegenden Form nur zu Forschungszwecken und nicht zur Eignungsdiagnostik nutzbar. Die Wirkung der BART basiert auf monetären Anreizen, die zwar in Forschungssettings mit Probandengeldern, im Rahmen von Auswahluntersuchungen jedoch selbstverständlich nicht möglich sind.

Die eingesetzten Skalen zu Teamarbeit zeigten zwar in den aktuellen Studien keine signifikanten Effekte, sollten jedoch trotzdem weiter im Fragebogen beibehalten werden, da diese

eventuell in langfristigen Arbeitssituationen relevant werden. Sofern Personen mit einem starken Bedürfnis nach Teamarbeit über einen langen Zeitraum ohne menschliche Teampartner auskommen müssen, könnte sich deren Fehlen negativ auf die Leistung auswirken. Da die Studie zu zukünftigen Anforderungen an Luftfahrtoperateure in Gruppen von jeweils bis zu vier Probanden gleichzeitig durchgeführt wurde (siehe 7.3) und damit z.B. auch die Fragen der Teilnehmer in der Gruppe besprochen wurden, war die Zeit ausschließlicher Mensch-Maschine-Teamarbeit in der Studie begrenzt. Möglicherweise führte dies dazu, dass das Bedürfnis der Teilnehmer nach Teamarbeit stets ausreichend erfüllt war. Auswirkungen eines nicht erfüllten Bedürfnisses nach Teamarbeit kämen damit gar nicht zum Tragen.

Hinsichtlich des NEO-PI-R ist festzuhalten, dass dieser zwar eindeutig aussagekräftige Prädiktoren enthält, jedoch mit insgesamt 241 Items sehr zeitintensiv in der Bearbeitung ist. Als ökonomischere Zwischenlösung zwischen dem sehr kurzen Fragebogen BFI-10 und dem NEO-PI-R könnte zur zukünftigen Messung der Big Five stattdessen z.B. der BFI-44 (John, Donahue, & Kentle, 1991) verwendet werden.

Die Computernutzungsvariablen Smartphone- und MP3-Player-Nutzung waren in den aktuellen Studien aussagekräftig, sind jedoch stark von gesellschaftlich-technologischen Entwicklungen abhängig und daher für eignungsdiagnostische Zwecke nicht brauchbar. Sie können je nach Fragestellung jedoch in Forschungssettings sinnvoll sein.

Die in Kapitel 5 aufgeworfene Frage, ob im Vergleich zu heute in zukünftigen Arbeitsumfeldern von Luftfahrtoperateuren veränderte Persönlichkeitsprofile gefordert sind, kann mit dieser Arbeit nicht abschließend beantwortet werden. Das Ziel der Entwicklung eines Forschungsfragebogens zu Persönlichkeits- und Einstellungsaspekten, die für die Arbeit im Mensch-Maschine-Team eine Rolle spielen, wurde jedoch erreicht. Ein wichtiger Schritt in diesem Forschungsbereich ist damit getan.

Wie bereits erwähnt, ist der Einsatz des HTQ-S sowohl in weiteren Forschungsprojekten als auch langfristig in der Eignungsdiagnostik für zukünftige Luftfahrtoperateure vorstellbar. Bereits in zwei DLR-Forschungsprojekten, die zum einen Flugverkehrskontrolle in einem hochautomatisierten Setting und zum anderen die Zusammenarbeit von Operateuren verschiedener Leitstände betreffen, ist ein Einsatz des HTQ-S geplant. Wünschenswert ist zudem die Entwicklung einer Testkurzform, die allerdings weitere Studien, vor allem hinsichtlich der Validität des Fragebogens, erfordert.

Betont werden muss zum Abschluss, dass die ATM-Simulation HINT selbstverständlich nur ein kurzer Ausschnitt aus dem Alltag im Mensch-Maschine-Team sein kann. Langzeitstu-

dien sind notwendig, um festzustellen, wer dauerhaft gut damit zurechtkommt, seinen Arbeitsalltag ohne menschliche Kollegen zu erleben.

Literatur

Aklin, W. M., Lejuez, C. W., Zvolensky, M. J., Kahler, C. W., & Gwadz, M. (2005). Evaluation of behavioral measures of risk taking propensity with inner city adolescents. *Behaviour Research and Therapy, 43,* 215-228.

Asendorpf, J. (2007). *Psychologie der Persönlichkeit*. Berlin: Springer.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2003). *Multivariate Analyseverfahren - Eine anwendungsorientierte Einführung (10. Auflage)*. Heidelberg: Springer.

Bahner, J. E. (2008). *Übersteigertes Vertrauen in Automation: Der Einfluss von Fehlerfahrungen auf Complacency und Automation Bias.* Berlin: TU Berlin.

Bainbridge, L. (1983). Ironies of Automation. *Automatica, 19,* 775-779.

Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, Vol 84(2),* 191-215.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and Performance at the beginning of the New Millennium: What do We Know and Where Do We Go Next? *International Journal of Selection and Assessment, 9,* 9-30.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50,* 7-15.

Beier, G. (2004). *Kontrollüberzeugungen im Umgang mit Technik: Ein Persönlichkeitsmerkmal mit Relevanz für die Gestaltung technischer Systeme*. Berlin: dissertation.de.

Beier, G., Spiekermann, S., & Rothensee, M. (2006). Die Akzeptanz zukünftiger Ubiquitious Computing Anwendungen. In A.M.Heinecke & H. Paul (Eds.), *Mensch & Computer 2006: Mensch und Computer im Strukturwandel* (pp. 145-154). Gelsenkirchen: Oldenbourg.

Berth, H. & Goldschmidt, S. (2006). NEO-PI-R. NEO-Persönlichkeitsinventar nach Costa und McCrae. Revidierte Fassung von Fritz Ostendorf und Alois Angleitner (2004). *Diagnostica, 52,* 95-103.

Bibby, K. S., Margulies, F., Rijnsdorp, J. E., & Withers, R. M. J. (1975). Man's role in control systems. In *Proceedings of the 6th IFAC Congress*, Boston.

Bortz, J. (2005). *Statistik für Human-und Sozialwissenschaftler. 6. Auflage*. Heidelberg: Springer.

Bradshaw, J. M., Feltovich, P., & Johnson, M. (2011). Human-agent interaction. In G.A.Boy (Ed.), *Handbook of Human-Machine Interaction* (pp. 283-302). Ashgate.

Broach, D. (2002). Air Traffic Controller Ability Requirements in the U.S. National Airspace System. In H. Eißfeldt, M. C. Heil, & D. Broach (Eds.), *Staffing the ATM System* (pp. 3-12). Aldershot: Ashgate.

Broach, D. (2005). A Singular Success: Air Traffic Control Specialist Selection 1981 - 1992. In B.Kirwan, M. Rodgers, & D. Schäfer (Eds.), *Human Factors Impacts in Air Traffic Management* (pp. 177-206). Aldershot: Ashgate.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21,* 1-16.

Bruder, C., Eißfeldt, H., Maschke, P., & Hasse, C. (2013). Differences in Monitoring between Experts and Novices. *Prodeedings of the HFES 57th Annual Meeting, October 22 - 26, 2013, Boston,* (in press).

Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality, 11,* 303-315.

Byrdorf, P. (1998). Military Pilot Selection. In K.-M.Goeters (Ed.), *Aviation Psychology: A Science and a Profession* (pp. 63-72). Aldershot: Ashgate.

CAST (1999). *CAST: Final Report* (Rep. No. AI-97-SC. 2029). Brussels: European Commission.

Cattell, H. E. P. (1996). The original big five: A historical perspective. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquee., 46,* 5-14.

Corsini, R. J. (1977). *Current personality theories*. Itasca, Illinois: FE Peacock Publishers.

Costa, P. T. & McCrae, R. R. (1992). *Professional manual: revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.

Deshmukh, A. V., McComb, S. A. & Wernz, Ch. (2008). Agents as Collaborating Team Members. In M. P. Letsky, N. W. Warner, S. M. Fiore & C. A. P. Smith (Eds.), *Macrocognition in Teams* (pp. 106-125). Aldershot: Ashgate

Dette, D. E., Abele, A. E., & Renner, O. (2004). Zur Definition und Messung von Berufserfolg. *Zeitschrift für Personalpsychologie, 3,* 170-183.

Dickhaeuser, O. & Stiensmeier-Pelster, J. (2000). Entwicklung eines Fragebogens zur Erfassung computerspezifischer Attributionen. *Diagnostica, 46,* 103-111.

Domke, B. (2008). Keine Karriere ohne Assessment Center. *Harvard Business Manager,*6-11.

Dyer, J. L. (1984). Team research and team training: A state-of-the-art review. In F. A. Muckler (Ed.), *Human factors review* (pp. 285-323). Santa Monica, CA: Human Factors Society.

Eißfeldt, H. (1991). Automation in ATC: How Does it Affect the Selection of Controllers? In J.A.Wise, V. D. Hopkin, & M. L. Smith (Eds.), *Automation and System Issues in Air Traffic Control* (pp. 461-465). Berlin: Springer Verlag.

Eißfeldt, H. (2002). *Der Einsatz einer computergestützten Arbeitsprobe nach dem Lerntestkonzept zur Vorhersage des Erfolges in berufspraktischen Abschnitten der Fluglotsenausbildung* (Rep. No. DLR FB 2002-13). Hamburg: DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.

Eißfeldt, H. (2008). *Projektantrag Hybrid: Entscheidungsverhalten in hybriden Teams (unveröffentlichtes Dokument)*. Hamburg: DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.

Eißfeldt, H., Goeters, K.-M., Lorenz, B., & Pecena, Y. (1997). *Eignungsauswahl für den Flugverkehrskontrolldienst* (Rep. No. DLR-IB-316-97-01). Hamburg: DLR Deutsches Zentrum für Luft- und Raumfahrt.

Eißfeldt, H., Goeters, K.-M., Lorenz, B., Scholz, B., & Stelling, D. (1996). *Eignungsauswahl für den Flugverkehrskontrolldienst* (Rep. No. DLR-IB-316-96-01). Hamburg: DLR - Deutsches Zentrum für Luft- und Raumfahrt e.V.

Eißfeldt, H., Grasshoff, D., Hasse, C., Hörmann, H.-J., Schulze Kissing, D., Stern, C. et al. (2009). *Aviator 2030 - Ability Requirements in Future ATM Systems II: Simulations and Experiments* (Rep. No. ISRN DLR-FB-2009-28). Hamburg: DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.

Eißfeldt, H. & Heintz, A. (2002). Ability requirements for DFS controllers: Current and future. In H. Eißfeldt, M. C. Heil, & D. Broach (Eds.), *Staffing the ATM System: The Selection of Air Traffic Controllers* (pp. 13-24). Aldershot: Ashgate.

Eißfeldt, H., Keye, D., Conzelmann, K., Grasshoff, D., Pecena, Y., & Heintz, A. (2013). *Validity Report: Selection of ab initio air traffic controllers for DFS Deutsche Flugsicherung GmbH* (Rep. No. ISRN DLRFB-2013-02). Hamburg: DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.

Eschen, S., Knappe, K., & Eißfeldt, H. (2012). Performance in hybrid teams: development of a research questionnaire and simulation tool. In A. Droog (Ed.), *Working towards zero Accidents* (pp. 32-38). Villasimius: EAAP.

Eschen-Léguedé, S. (2007). *Beobachtertraining im Kontext von Assessment Centern: Entwicklung, Evaluation und Akzeptanz*. Saarbrücken: VDM-Verlag.

Faber, G. (1992). Bereichsspezifische Beziehungen zwischen leistungsthematischen Schülerselbstkonzepten und Schulleistungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 24,* 66-82.

Feuerberg, B. V., Bahner, J. E., & Manzey, D. (2005). Interindividuelle Unterschiede im Umgang mit Automation: Entwicklung eines Fragebogens zur Erfassung des Complacency-Potentials. In L. Urbas & C. Steffens (Hrsg.), *6. Berliner Werkstatt*

Mensch-Maschine-Systeme: Zustandserkennung und Systemgestaltung (S. 199-202). Düsseldorf: VDI Verlag GmbH.

Fitts, P. M. (1951). *Human engineering for an effective air-navigation and traffic-control system*. Washington, D.C.: National Research Council.

Fleishman, E. A. (1995). *Administrator´s Guide to Fleishman Job Analysis Survey (F-JAS)*. Potomac, Maryland: Management Research Institue, Inc.

Flin, R. (2005). The NOTECHS System. In K.-M.Goeters (Ed.), *Aviation Psychology: Practice and Research* (pp. 245-250). Aldershot: Ashgate.

Gerrig, R. J. & Zimbardo, P. G. (2008). *Psychologie*. München: Pearson.

Goeters, K.-M. & Lorenz, B. (1985). *Erfahrung und Leistung in Video-Spielen und der Einfluß auf Fähigkeitstests in der Luftfahrtpsychologie* (Rep. No. ISRN DFVLR-FB-85-43). Hamburg: DFVLR Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt.

Goeters, K.-M., Maschke, P., & Eißfeldt, H. (2004). Job Analyses of Airline Pilots and Air Traffic Controllers. In K.-M.Goeters (Ed.), *Aviation Psychology: Practice and Research* (pp. 99-119). Aldershot: Ashgate.

Grasshoff, D. (2001). *Der Strip Display Management Test (SDM)* (Rep. No. DLR FB 2001-03). Hamburg: DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.

Häcker, H. & Stapf, K. H. (1998). *Dorsch Psychologisches Wörterbuch (13. Auflage)*. Bern: Huber.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload, 1,* 139-183.

Hansman, R. J. (2001). The Effect of Shared Information on Pilot/Controller and Controller/ Controller Interactions. In L.Bianco, P. Dell'Olmo, & A. R. Odoni (Eds.), *New Concepts and Methods in Air Traffic Management* (pp. 143-159). Springer Verlag.

Hasse, C., Grasshoff, D., & Bruder, C. (2012). Eye-tracking parameters as a predictor of human performance in the detection of automation failures. In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Roettger, D. Manzey, S. Biede, F. Reuzeau, & P. Terrier (Eds.), *Human Factors: A view from an integrative perspective*, *Proceedings HFES Europe Chapter Conference* (pp. 133-144). Toulouse: HFES

Hoc, J. M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics, 43,* 833-843.

Hoffman, G., & Breazeal, C. (2004). Collaboration in human-robot teams. In *Proceedings of the AIAA 1st Intelligent Systems Technical Conference*. Chicago, IL, USA.

Höft, S. & Pecena, Y. (2005). Behaviour-Oriented Evaluation of Aviation Personnel: An Assessment Center Approach. In K.-M.Goeters (Ed.), *Aviation Psychology: Practice and Research* (pp. 153-170). Aldershot: Ashgate.

Hollnagel, E. & Woods, D. D. (1983). Cognitive Systems Engineering: New wine in new bottles. *International Journal of Man-Machine Studies, 18,* 583-600.

Hopkin, V. D. (1991). The Impact of Automation on Air Traffic Control Systems. In J.A.Wise, V. D. Hopkin, & M. L. Smith (Eds.), *Automation and System Issues in Air Traffic Control* (pp. 3-19). Berlin: Springer.

Hossiep, R. & Paschen, M. (1998). *Das Bochuner Invertar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP): Handanweisung*. Göttingen: Hogrefe Verlag.

Huber, S., Kain, S., & Nachtwei, J. (2008). Effekte sicherer nachweisen: Persönlichkeitsmerkmale als Kontrollvariablen in der Human Factors Forschung. In M. Grandt & A. Bauch (Eds.), *50. Fachausschusssitzung Anthropotechnik: Beiträge der Ergonomie zur Mensch-System-Integration* (pp. 143-159). Bonn: Deutsche Gesellschaft für Luft- und Raumfahrt e.V.

Janneck, M. & Guczka, S. (2013). The Resigned, the Confident, and the Humble: A Typology of Computer-Related Attribution Styles. In A. Holzinger, M. Ziefle, M. Hitz, & M. Debevc (Eds.), *Human Factors in Computing and Informatics* (pp. 373-390). Springer Berlin Heidelberg.

Janneck, M., Vincent-Höper, S., & Ehrhardt, J. (2012). Das computerbezogene Selbstkonzept: Eine gender-sensitive Studie. In H.Reiter & O. Deussen (Eds.), *Mensch & Computer 2012: 12. fachübergreifende Konferenz für interaktive und kooperative Medien. Interaktiv informiert - allgegenwärtig und allumfassend!?* (pp. 243-252). München: Oldenbourg Wissenschaftsverlag.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory - Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Kain, S. & Nachtwei, J. (2009). Die Rolle von Kontrollvariablen in der Human Factors Forschung: Ein bewährtes Konzept in einem modernen Anwendungsfeld. In B.Krause & P. Metzler (Eds.), *Empirische Evaluationsmethoden (13. ed.)*. Berlin: ZeE Verlag.

Kalinina, V. (2012). *Prognostische Validität der DLR-Eignungstests für Nachwuchspiloten: Rolle der Wissenstests bei der Vorhersage des Ausbildungserfolgs*. Unveröffentlichte Diplomarbeit, Universität Hamburg.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen: Der Fragebogen TA-EG. In A. Lichtenstein, C. Stößel, & C. Clemens (Eds.), *8. Berliner Werkstatt Mensch-Maschine-Systeme: Der Mensch im Mittelpunkt technischer Systeme* (S. 196-201). Düsseldorf: VDI Verlag GmbH.

Kerns, K. (1991). Data-Link Communication Between Controllers and Pilots: A Review and Synthesis of the Simulation Literature. *The International Journal of Aviation Psychology, 1,* 181-204.

Keye, D. & Eschen, S. (2013). A low-fidelity simulation of human-automation interaction (Publikation in Vorbereitung).

Kleinmann, M. (2003). *Assessment-Center*. Göttingen: Hogrefe.

Klumb, P. L. (1995). Cognitive failures and performance differences: validation studies of a German version of the cognitive failures questionnaire. *Ergonomics, 38,* 1456-1467.

Knappe, K. (2010). *Messung von Persönlichkeitsaspekten mit Einfluss auf die Leistungsentfaltung in hochautomatisierten Systemen: Erprobung eines Forschungsfragebogens*. Unveröffentlichte Diplomarbeit, Universität Hamburg.

Landers, R. N. & Lounsbury, J. W. (2006). An investigation of Big Five and narrow personality traits in relation to Internet usage. *Computers in Human Behavior, 22,* 283-293.

Larson, G. E. & Merritt, C. R. (1991). Can Accidents be Predicted? An Empirical Test of the Cognitive Failures Questionnaire. *Applied Psychology, 40,* 37-45.

Lee, J. D. & Moray, N. (1994). Trust, self-confidence, and operators adaption to automation. *International Journal of Human-Computer Studies, 40,* 153-184.

Lee, J. D. & See, K. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors, 46,* 50-80.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L. et al. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8,* 75-84.

Leonhardt, J. (2005). Implementation of Critical Incident Stress Management at the German Air Navigation Services. In B.Kirwan, M. Rodgers, & D. Schäfer (Eds.), *Human Factors Impacts in Air Traffic Management* (pp. 207-226). Aldershot: Ashgate.

Li, A. & Bagger, J. (2006). Using the BIDR to Distinguish the Effects of Impression Management and Self-Deception on the Criterion Validity of Personality Measures: A Meta-Analysis. *International Journal of Selection and Assessment, 14,* 131-141.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19-40.

Manning, C. A. & Broach, D. (1992). *Identifying Ability Requirements for Operators of Future Automated Air Traffic Control Systems* (Rep. No. DOT/FAA/AM-92/26). Oklahoma: FAA Civil Aeromedical Institute.

Mahlfeld, W. (2011). *Der Einfluss des Complacency-Potentials auf das menschliche Überwachungsverhalten in der Luftfahrt*. Unveröffentlichte Diplomarbeit, TU Braunschweig.

Maner, J. K., Gailliot, M. T., Butz, D. A., & Peruche, B. M. (2007). Power, Risk, and the Status Quo: Does Power Promote Riskier or More Conservative Decision Making? *Personality and Social Psychology Bulletin, 33,* 451-462.

Maner, J. K., Richey, J. A., Cromer, K., Mallott, M., Lejuez, C. W., Joiner, T. E. et al. (2007). Dispositional anxiety and risk-avoidant decision-making. *Personality and Individual Differences, 42,* 665-675.

Marcus, B. & Schuler, H. (2006). Leistungsbeurteilung. In H. Schuler (Ed.), *Lehrbuch der Personalpsychologie* (S. 433-469). Göttingen: Hogrefe.

Maschke, P. (1987). *Temperament Structure Scales* (Rep. No. Tech.Rep.ESA-TT-1069). Oberpfaffenhofen: European Space Agency.

Maschke, P. (2005). Personality Evaluation of Applicants in Aviation. In K.-M.Goeters (Ed.), *Aviation Psychology: Practice and Research* (pp. 141-152). Aldershot: Ashgate.

Maschke, P. & Goeters, K.-M. (1999). *Anforderungen an Flugschüler in der Ab-Initio-Ausbildung im Vergleich zu aktiven Linienflugzeugführern* (Rep. No. ISRN DLR-FB-1999-16). Hamburg: DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.

Maschke, P. & Goeters, K.-M. (2003). Ab-initio Flight Training and Airline Pilot Performance: Differences in Job Requirements. In G.Edkins & P. Pfister (Eds.), *Innovation and Consolidation in Aviation* (pp. 189-196). Aldershot: Ashgate.

Mohiyeddini, C. (2001). FIT Fragebogen zur Erfassung induividueller Einstellungen zur Teamarbeit. In W.Sarges & H. Wottawa (Eds.), *Handbuch wirtschaftspsychologischer Testverfahren* (pp. 251-254). Lengerich: Pabst Science Publishers.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60,* 683-729.

Mosier, K. L. & Skitka, L. J. (1996). Human Decision Makers and Automated Decision Aids: Made for Each Other? In R.Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 201-220). New Jersey: Lawrence Erlbaum Associates.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45,* 669-678.

Neerincx, M. A., Bos, A., Olmedo-Soler, A., Brauer, U., Breebaart, L., Smets, N. et al. (2008). The mission execution crew assistant: improving human-machine team

resilience for long duration missions. In *Proceedings of the 59th International Astronautical Congress (IAC 2008)*.

Nowitzke, S. (2013). *Zusammenhänge der generellen Selbstwirksamkeitserwartung sowie technikbezogener Persönlichkeitskonstrukte mit der Leistungsfähigkeit in computerbasierten Testverfahren.* Diplomarbeit in Vorbereitung, Christian-Albrechts-Universität zu Kiel.

Olkin, I. & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of the Mathematical Statistics, 29,* 201-211.

Ones, D. S., Viswesvaran, C., Judge, T. A., & Dilchert, S. (2007). In Support of Personality Assessment in Organizational Settings. *Personnel Psychology, 60,* 995-1027.

Ones, D. S. & Viswesvaran, C. (1998). The Effects of Social Desirability and Faking on Personality and Integrity Assessment for Personnel Selection. *Human Performance, 11,* 245-269.

Ostendorf, F. (2001). NEO-Persönlichkeits-Inventar - revidierte Form. In W.Sarges & H. Wottawa (Eds.), *Handbuch wirtschaftspsychologischer Testverfahren* (pp. 409-415). Lengerich: Pabst Science Publishers.

Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae. Revidierte Fassung (NEO-PI-R)*. Göttingen: Hogrefe.

Parasuraman, R. (1987). Human-Computer Monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 29,* 695-706.

Parasuraman, R. & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 52,* 381-410.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of Adaptive Task Allocation on Monitoring of Automated Systems. *Human Factors, 38,* 665-679.

Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39,* 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30,* 286-297.

Pecena, Y. (2002). Behaviour-Oriented Asessment of Interpersonal Skills in ATCO Selection. In H.Eißfeldt, M. C. Heil, & D. Broach (Eds.), *Staffing the ATM System* (pp. 97-105). Aldershot: Ashgate.

Rammstedt, B. (2007). The 10-Item Big Five Inventory. *European Journal of Psychological Assessment, 23,* 193-201.

Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41,* 203-212.

Rathje, H., Golany, Z., & Eißfeldt, H. (2004). Pan-European Selection Test Battery for Air Traffic Control Applicants. In K.-M.Goeters (Ed.), *Aviation Psychology: Practice and Research* (pp. 171-202). Aldershot: Ashgate.

Richter, T., Naumann, J., & Groeben, N. (2001). Das Inventar zur Computerbildung (INCOBI): Ein Instrument zur Erfassung von Computer Literacy und computer-bezogenen Einstellungen bei Studierenden der Geistes- und Sozialwissenschaften. *Psychologie in Erziehung und Unterricht, 48,* 1-13.

Richter, T., Naumann, J., & Horz, H. (2010). Eine Revidierte Fassung des Inventars zur Computerbildung (INCOBI-R). *Zeitschrift für Pädagogische Psychologie, 24,* 23-37.

Riley, V. (1996). Operator Reliance on Automation. In R.Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Application* (pp. 19-35). Mahwah: LEA.

Rotter, J. B. (1954). *Social learning and clinical psychology*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied, 80,* 1-28.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50,* 540-547.

Sarter, N. B. & Woods, D. D. (1991). Situation Awareness: A Critical But Ill-Defined Phenomenon. *The International Journal of Aviation Psychology, 1,* 45-57.

Schmidt, L. R. (1975). *Objektive Persönlichkeitsmessung in diagnostischer und klinischer Psychologie*. Weinheim: Beltz.

Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik*. (5 ed.) Heidelberg: Springer.

Schmidt-Atzert, L., Deter, B., & Jaeckel, S. (2004). Prädiktion von Ausbildungserfolg: Allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten? *Zeitschrift für Personalpsychologie, 3,* 147-158.

Scholz, U., Gutiérrez-Doña, B., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18(3),* 242-251.

Schroeders, U. & Wilhelm, O. (2010). Computer usage questionnaire: Structure, correlates, and gender differences. *Computers in Human Behavior, 27,* 899-904.

Schwarzer, R. (2009). Gesundheitspsychologie. In G. Krampen (Hrsg.): *Psychologie - Experten als Zeitzeugen* (S. 240-249). Göttingen: Hogrefe.

Schwarzer, R. & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen.* Berlin: Modellversuch selbstwirksame Schulen.

SESAR Consortium (2007). *The ATM Target Concept - D3* (Rep. No. DLM-0612-001-02-00a). Toulouse.

Sheridan, T. B. & Parasuraman, R. (2005). Human-Automation Interaction. *Reviews of Human Factors and Ergonomics, 1,* 89-129.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-Induced "Complacency": Development of the Complacency-Potential Rating Scale. *The Interantional Journal of Aviation Psychology, 3,* 111-122.

Smith, A., Peck, D., & Clatworthy, T. (1989). After-effects of working at visual display units. *Work & Stress, 3,* 195-201.

Soll, H. (2012). Der Simulatortest - LH Ab-initio 2011. In D.Stelling (Ed.), *Jahresbericht 2011 des Arbeitsbereichs Cockpit* (DLR-IB-316-2012-02, S. 39-42). Hamburg: DLR Deutsches Zentrum für Luft- und Raumfahrt e.V.

Szalma, J. L. & Taylor, G. S. (2011). Individual Differences in Response to Automation: The Five Factor Model of Personality. *Journal of Experimental Psychology: Applied, 17,* 71-96.

Tupes, E. C. & Christal, R. E. (1961). Recurrent Personality Factors based on Trait Ratings. *Technical Report ASD-TR-61-97*. Lackland Air Force Base, Texas, USA, United States Air Force.

Volz-Sidiropoulou, E., Bäcker, M., Niemann, H., Privou, C., Zimmermann, P., & Gauggel, S. (2007). Skala zur Erfassung von Aufmerksamkeitsdefiziten (SEA). *Zeitschrift für Neuropsychologie, 18,* 299-309.

Voller, L. & Fowler, A. (2005). Human Factors Longitudinal Study to Support the Improvement of Air Traffic Controller Training. In B.Kirwan, M. Rodgers, & D. Schäfer (Eds.), *Human Factors Impacts in Air Traffic Management* (pp. 153-176). Aldershot: Ashgate.

White, T. L., Lejuez, C. W., & de Wit, H. (2008). Test-retest characteristics of the Balloon Analogue Risk Task (BART). *Experimental and clinical psychopharmacology, 16,* 565-570.

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon Becker, S. (2004). *An Introduction to Human Factors Engineering*. (2 ed.) New Jersey: Pearson Prentice Hall.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R.Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Ziegler, M. & Bühner, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.

Abbildungsverzeichnis

Abbildung 1: Das Mensch-Maschine-Team in der Flugsicherung 19

Abbildung 2: Automatisierungslevel 20

Abbildung 3: Das computerbezogene Selbstkonzept nach Janneck et al. (2012) 27

Abbildung 4: Stufen der Eignungsuntersuchung für Nachwuchsfluglotsen 33

Abbildung 5: Stufen der Eignungsuntersuchung für Nachwuchspiloten 35

Abbildung 6: Logo des HYBRID-Projekts 39

Abbildung 7: Übersicht über die angewandte Methode 45

Abbildung 8: Durchführung der Balloon Analogue Risk Task (BART) 51

Abbildung 9: HINT (Hybrid Interaction Scenario) 53

Abbildung 10: Durchführung der Kombinationsstudie zu HTQ-S und HINT 57

Abbildung 11: Positive Zusammenhänge zu aktueller Eignungsdiagnostik 124

Abbildung 12: Negative Zusammenhänge zu aktueller Eignungsdiagnostik 125

Abbildung 13: Ambivalente Zusammenhänge zu aktueller Eignungsdiagnostik 126

Abbildung 14: Positive Zusammenhänge zu HINT 127

Abbildung 15: Negative Zusammenhänge zu HINT 128

Abbildung 16: Gruppenunterschiede in HINT-Gruppen 129

Abbildung 17: Inkrementelle Validität 130

Tabellenverzeichnis

Tabelle 1: Persönlichkeit in der Vorhersage von HF-relevanten Außenkriterien 25

Tabelle 2: Überblick über Persönlichkeitsmerkmale und ihre Zuordnung zu wissenschaftlichen Begriffen der Persönlichkeitsbeschreibung (nach Huber et al., 2008) 26

Tabelle 3: Relevante Persönlichkeitsaspekte für Piloten und Fluglotsen.....28

Tabelle 4: Heutige (Baseline) vs. zukünftige (Future) Anforderungen an Fluglotsen und Piloten (nach Eißfeldt et al., 2009).....38

Tabelle 5: Variablen aus der Eignungsdiagnostik und ihre Bedeutung 50

Tabelle 6: Variablen aus der HINT-Simulation und ihre Bedeutung.....55

Tabelle 7: Übersicht über die HTQ-Skalen 61

Tabelle 8: Deskriptiva des HTQ-S 82

Tabelle 9: Vergleich der internen Konsistenzen (Cronbachs α) des NEO-PI-R zwischen dem HTQ-1 und den Ergebnissen von Ostendorf und Angleitner (2004) 86

Tabelle 10: Interne Konsistenzen (Cronbachs α) der Skalen des HTQ-1 und HTQ-2.....87

Tabelle 11: Deskriptive Statistiken der HINT-Variablen.....89

Tabelle 12: Interkorrelationen der HINT-Variablen 90

Tabelle 13: Korrelationen des HTQ-S mit Basisleistungen in der Fluglotsenauswahl.....92

Tabelle 14: Korrelationen des HTQ-S mit Basisleistungen in der Pilotenauswahl 94

Tabelle 15: Korrelationen des HTQ-S mit der Gerätestufe beider Auswahlverfahren 96

Tabelle 16: Skalen des HTQ-2 mit signifikanten Zusammenhängen sowohl zum Geschlecht als auch zu Variablen aus der Eignungsdiagnostik für Fluglotsen 98

Tabelle 17: Skalen des HTQ-2 mit signifikanten Zusammenhängen sowohl zum Geschlecht als auch zu Variablen aus der Eignungsdiagnostik für Piloten 100

Tabelle 18: Korrelationen HTQ-S & HINT: allgemeine Persönlichkeit.....102

Tabelle 19: Korrelationen HTQ-S & HINT: Technikbezogene Persönlichkeitsaspekte / Technikbezogene Einstellung.....103

Tabelle 20: Korrelationen HTQ-S & HINT: Neigung zu risikoreichem Entscheidungsverhalten 104

Tabelle 21: Korrelationen HTQ-S & HINT: Computernutzung 105

Tabelle 22: Skalen des HTQ-S mit signifikanten Zusammenhängen sowohl zum Geschlecht als auch zu Variablen aus der HINT-Simulation 106

Tabelle 23: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren Werten und höheren Werten hinsichtlich *Reaktionszeit* 108

Tabelle 24: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren und höheren *Target-Abweichungen im Sektor Alpha*.....109

Tabelle 25: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren und höheren *Target-Abweichungen im Sektor Bravo* 110

Tabelle 26: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren und höheren Werten hinsichtlich der *Ablehnung von an sie (Alpha) gestellten Requests* 111

Tabelle 27: Signifikante Unterschiede im HTQ-S zwischen Probanden mit niedrigeren und höheren Werten hinsichtlich der *Annahme von an sie (Alpha) gestellten Requests* .112

Tabelle 28: Signifikanter Unterschied im EgMMT zwischen Probanden mit niedrigeren und höheren Werten hinsichtlich der *Annahme von an sie (Alpha) gestellten Requests* .113

Tabelle 29: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *Reaktionszeit*.....115

Tabelle 30: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der BU für das Kriterium *Reaktionszeit*.....116

Tabelle 31: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *RQ Alpha an Bravo*.....117

Tabelle 32: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *RQ Ablehnung durch Alpha*.....118

Tabelle 33: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der BU für das Kriterium *RQ Ablehnung durch Alpha*.....119

Tabelle 34: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *RQ Annahme durch Alpha* 120

Tabelle 35: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der BU für das Kriterium *RQ Annahme durch Alpha* 121

Tabelle 36: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *Target-Abweichungen Alpha* 122

Tabelle 37: Zusammenfassung der schrittweisen multiplen Regression zur Untersuchung der inkrementellen Validität des HTQ-S hinsichtlich der VA für das Kriterium *Target-Abweichungen Bravo* 123

Tabelle 38: Fazit für die einzelnen Skalen des HTQ-S 147

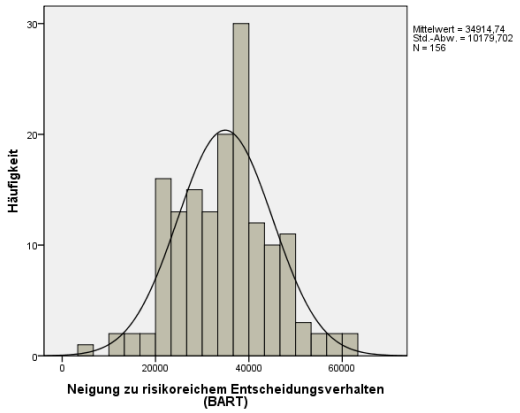
Abkürzungsverzeichnis

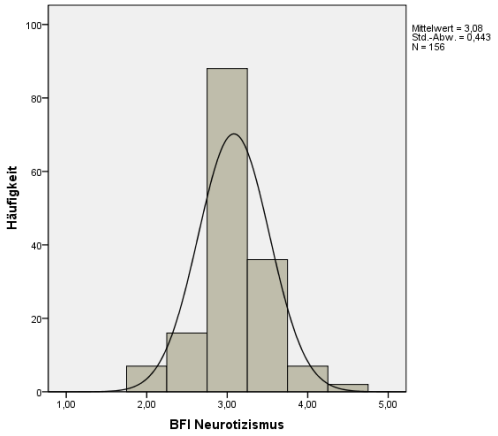
AC	Assessment Center
ATM	Air Traffic Management
BART	Balloon Analogue Risk Task
BFI-10	Big Five Inventory mit 10 Items
BFI-44	Big Five Inventory mit 44 Items
BIO	Biografische Angaben
BIP	Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung
BnT	Bedürfnis nach Teamarbeit
BU	Berufsgrunduntersuchung
CAC	Critical Aircraft
CaP	Complacency als Potential
CAT	Computer Assisted Testing
CAST	Consequences of Future ATM Systems for Air Traffic Controller Selection and Training
CFQ	Cognitive Failures Questionnaire
CNFB	Fragebogen zur Computernutzung
COMA	Fragebogen zur Sicherheit im Umgang mit Computeranwendungen
CPRS	Complacency Potential Rating Scale
CSK	Computerbezogenes Selbstkonzept
DAC	Dynamischer Airtraffic Control Test
DC	Development Center
DFS	Deutsche Flugsicherung
DLH	Deutsche Lufthansa
DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V.
EgA	Einstellung gegenüber Automation
EgMMT	Einstellung gegenüber Arbeit im Mensch-Maschine-Team
FEcA	Fragebogen zur Erfassung computerspezifischer Attributionen

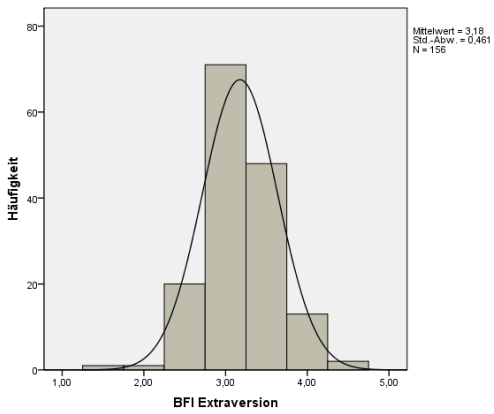
FIDEC	Fragebogen zur inhaltlich differenzierten Erfassung von computerbezogenen Einstellungen
FIT	Fragebogen zur Erfassung individ. Einstellungen gegenüber Teamarbeit
F-JAS	Fleishman Job Analysis Survey
FQ	Firmenqualifikation
HF	Human Factors
HINT	Hybrid Interaction Simulation
HTQ	Hybrid Team Questionnaire
HTQ-S	Hybrid Team Questionnaire – Supplemented
HU	Hauptuntersuchung
INCOBI-R	Inventar zur Computerbildung, revidierte Form
KUT	Kontrollüberzeugungen im Umgang mit Technik
MIC	Multi Instrument Coordination Test
NEO-PI-R	NEO-Persönlichkeits-inventar, revidierte Fassung
RQ	Request
SD	Standard Deviation (Standardabweichung)
SDM	Strip Display Management Test
SES	Single European Sky
SESAR	Single European Sky ATM Research Programme
SIM	Simulatortest (Flugsimulation)
SWE	Allgemeine Selbstwirksamkeitserwartung
TA-EG	Fragebogen zur Messung von Technikaffinität
TSS	Temperament-Struktur-Skalen
VA	Vorauswahl
VerDi	Verhaltensorientierte Diagnostik

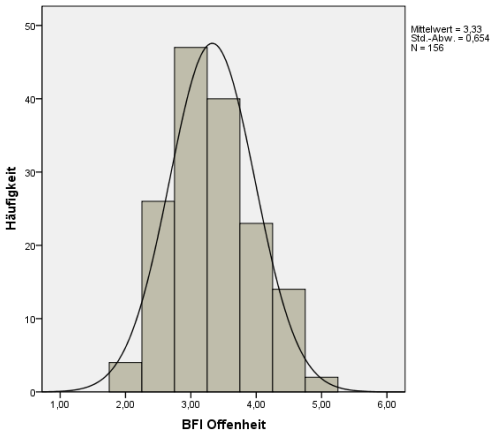
Anhang

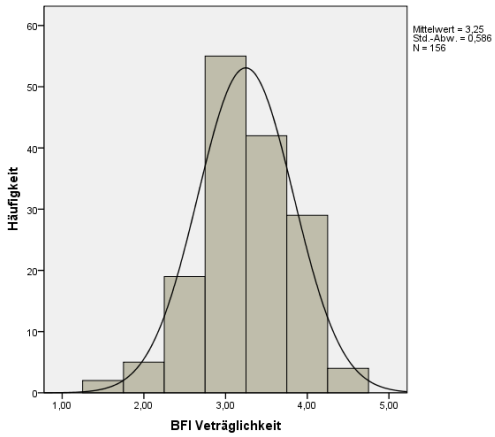
Anhang A: Histogramme zum HTQ-S

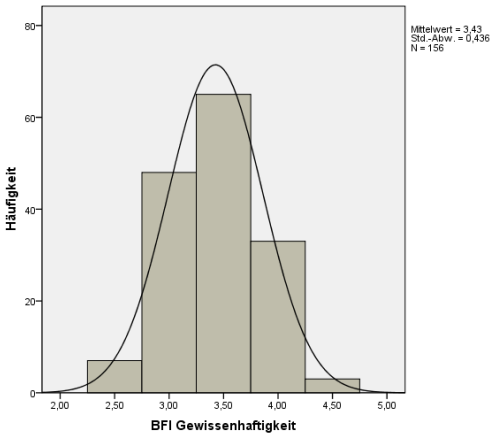


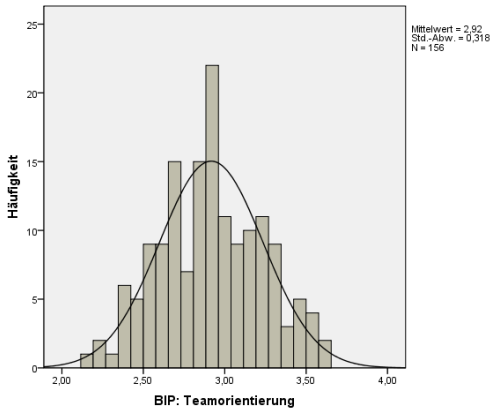


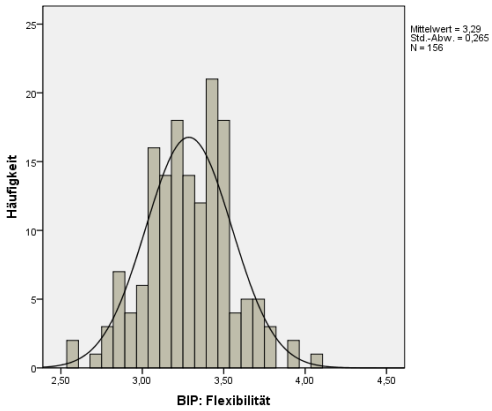


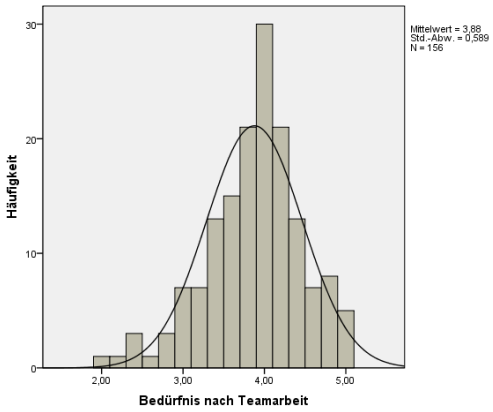


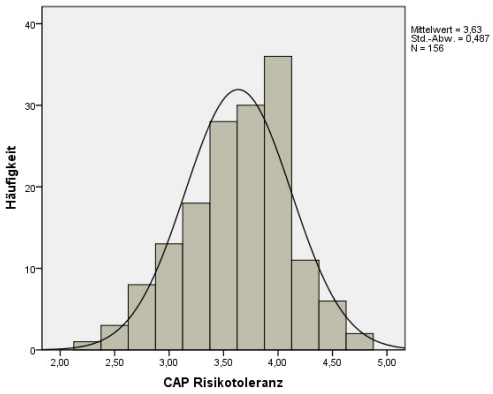


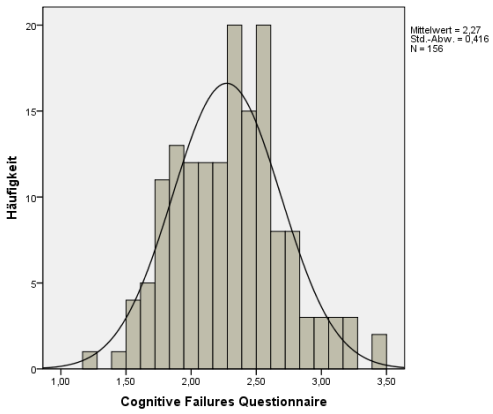


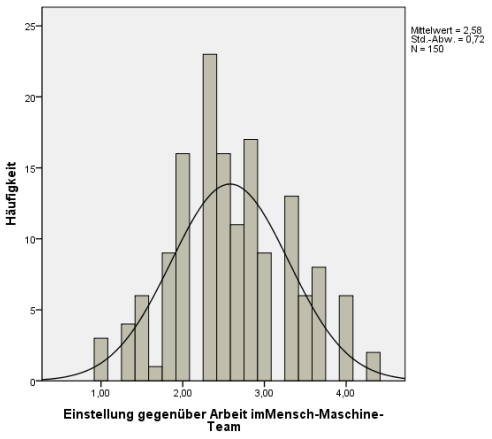

Mittelwert = 2,92
Std.-Abw. = 0,318
N = 156
Häufigkeit
BIP: Teamorientierung

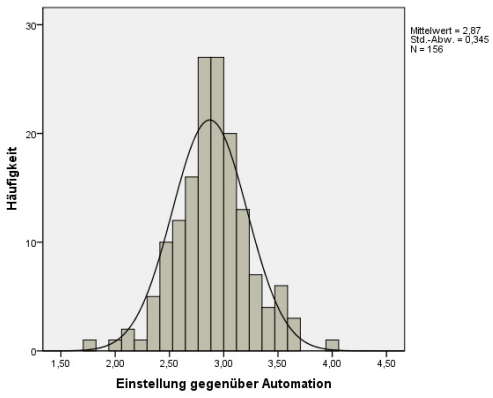

Mittelwert = 3,29
Std.-Abw. = 0,265
N = 156
Häufigkeit
BIP: Flexibilität

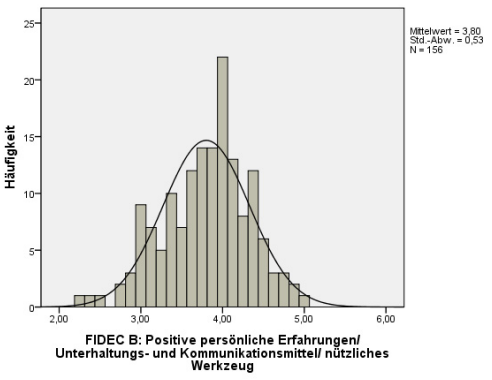

Mittelwert = 3,88
Std.-Abw. = 0,589
N = 156
Häufigkeit
Bedürfnis nach Teamarbeit

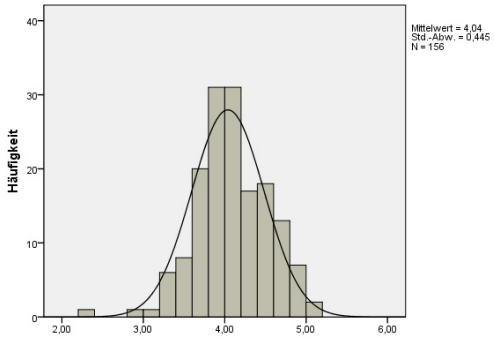

Mittelwert = 3,63
Std.-Abw. = 0,487
N = 156
Häufigkeit
CAP Risikotoleranz


Mittelwert = 2,27
Std.-Abw. = 0,416
N = 156
Häufigkeit
Cognitive Failures Questionnaire

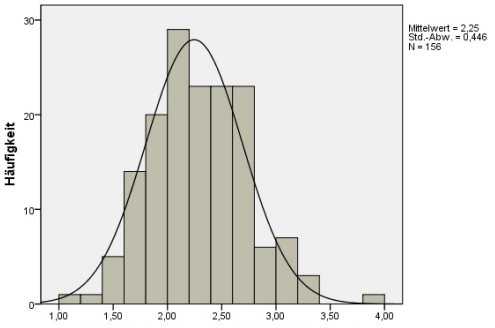

Mittelwert = 2,58
Std.-Abw. = 0,72
N = 150
Häufigkeit
Einstellung gegenüber Arbeit imMensch-Maschine-Team


Mittelwert = 2,87
Std.-Abw. = 0,345
N = 156
Häufigkeit
Einstellung gegenüber Automation

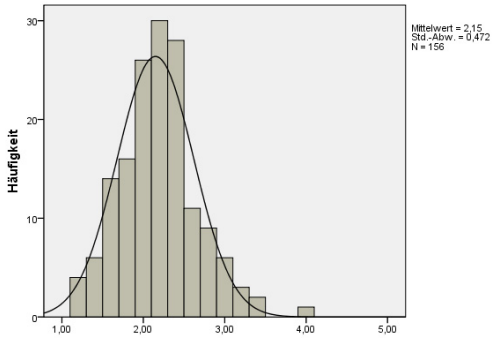

Mittelwert = 3,80
Std.-Abw. = 0,53
N = 156
Häufigkeit
FIDEC B: Positive persönliche Erfahrungen/ Unterhaltungs- und Kommunikationsmittel/ nützliches Werkzeug



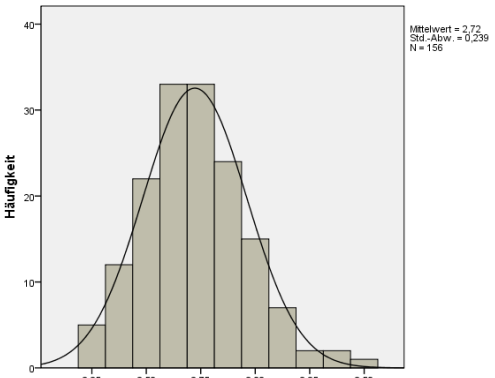
FIDEC F: Positive persönliche Erfahrungen/ Lern- und Arbeitsmittel/ nützliches Werkzeug



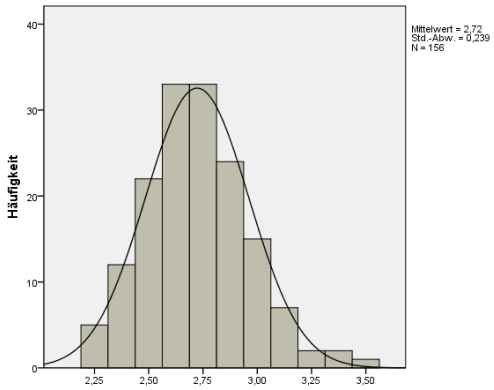
FIDEC G: Negative persönliche Erfahrungen/ Unterhaltungs- und Kommunikationsmittel/ unbeeinflussbare Maschine



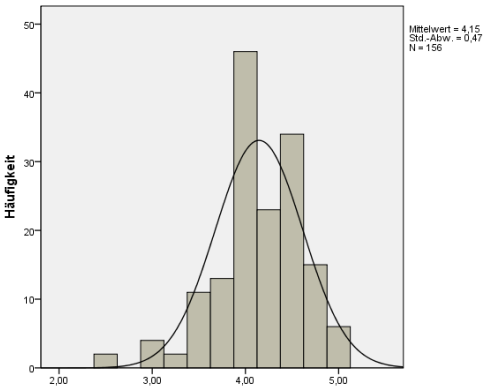
FIDEC H: Negative persönliche Erfahrungen/ Lern- und Arbeitsmittel/ unbeeinflussbare Maschine



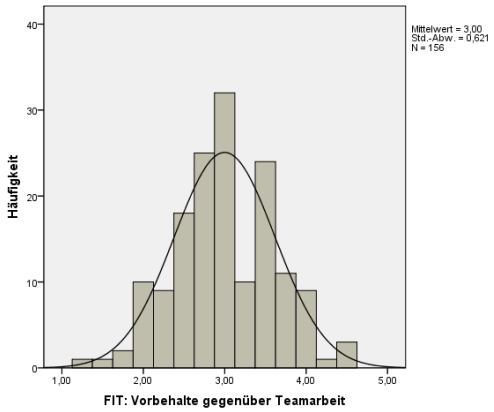
Sicherheit im Umgang mit Computern (COMA)



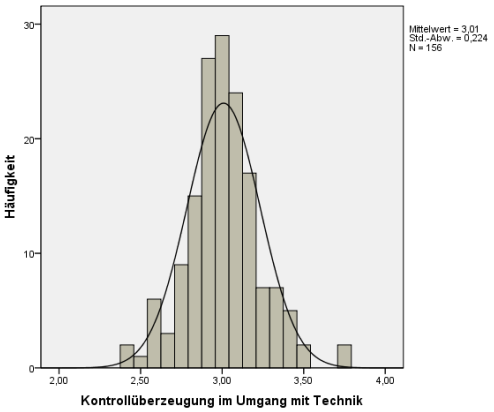
Computerängstlichkeit (COMA)



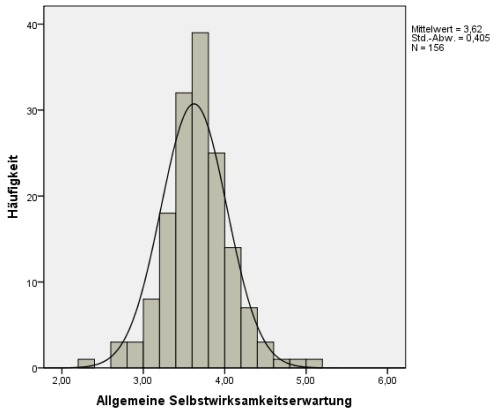
FIT: Bereitschaft zur Teamarbeit

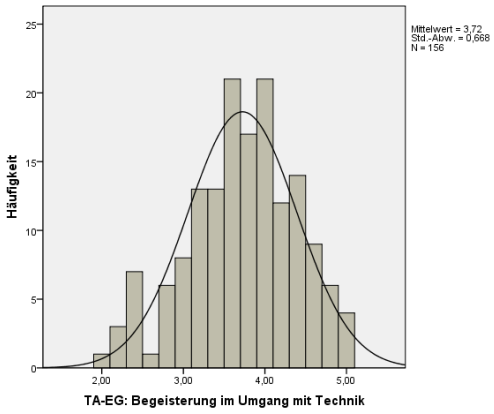


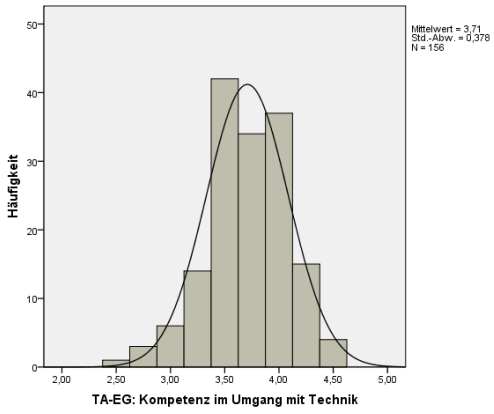
FIT: Vorbehalte gegenüber Teamarbeit

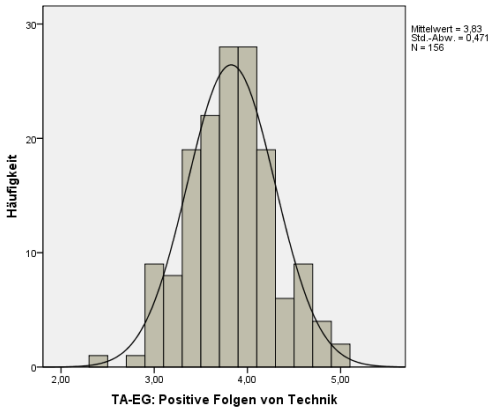


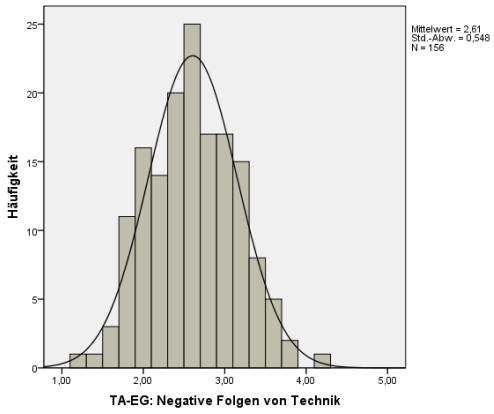
Kontrollüberzeugung im Umgang mit Technik

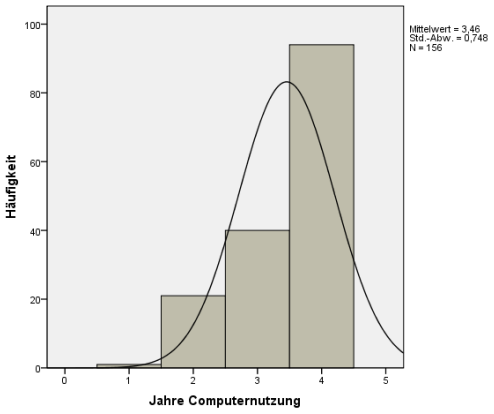

Mittelwert = 3,62
Std.-Abw. = 0,405
N = 156
Häufigkeit
Allgemeine Selbstwirksamkeitserwartung

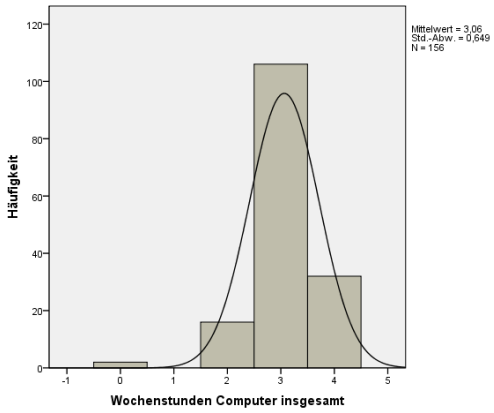

Mittelwert = 3,72
Std.-Abw. = 0,668
N = 156
Häufigkeit
TA-EG: Begeisterung im Umgang mit Technik

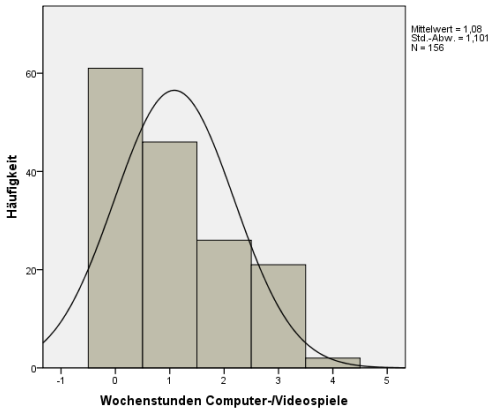

Mittelwert = 3,71
Std.-Abw. = 0,378
N = 156
Häufigkeit
TA-EG: Kompetenz im Umgang mit Technik

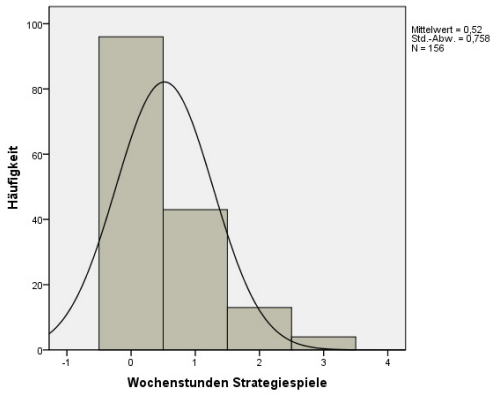

Mittelwert = 3,83
Std.-Abw. = 0,471
N = 156
Häufigkeit
TA-EG: Positive Folgen von Technik

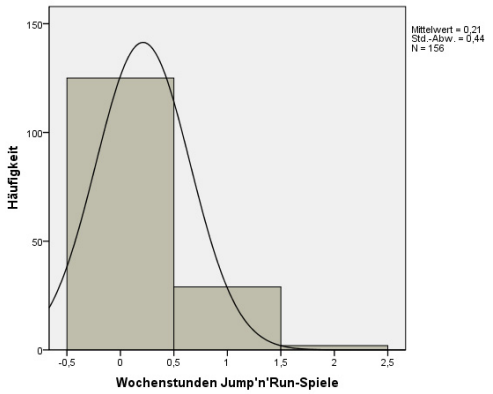

Mittelwert = 2,61
Std.-Abw. = 0,548
N = 156
Häufigkeit
TA-EG: Negative Folgen von Technik

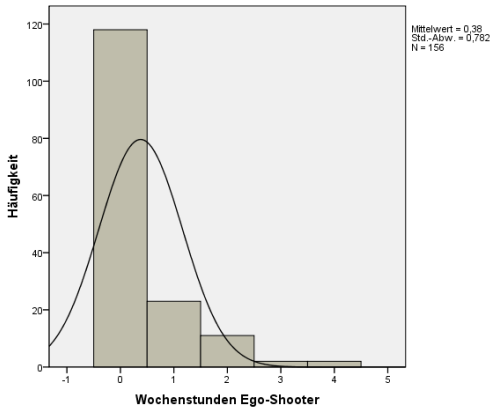

Mittelwert = 3,46
Std.-Abw. = 0,748
N = 156
Häufigkeit
Jahre Computernutzung

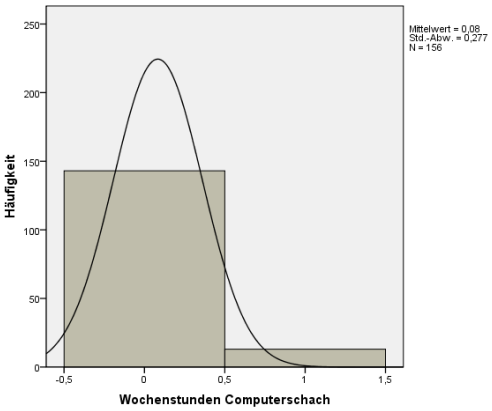

Mittelwert = 3,06
Std.-Abw. = 0,649
N = 156
Häufigkeit
Wochenstunden Computer insgesamt

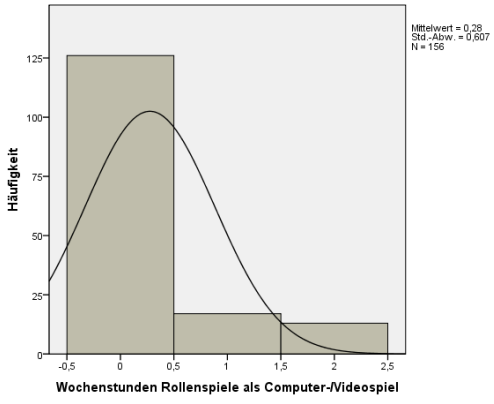

Mittelwert = 1,08
Std.-Abw. = 1,101
N = 156
Häufigkeit
Wochenstunden Computer-/Videospiele

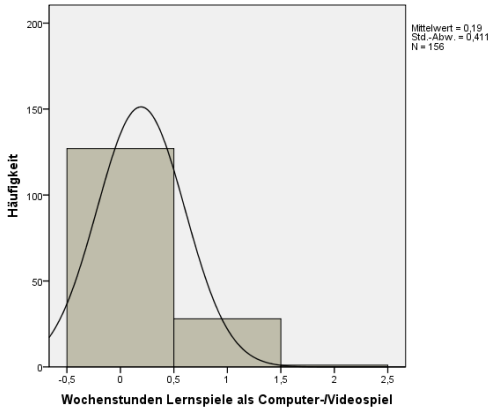

Mittelwert = 0,52
Std.-Abw. = 0,758
N = 156
Häufigkeit
Wochenstunden Strategiespiele

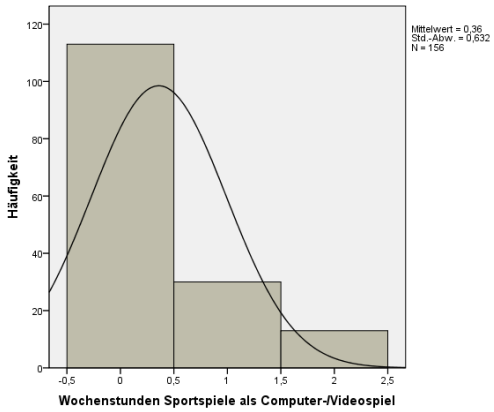

Mittelwert = 0,21
Std.-Abw. = 0,44
N = 156
Häufigkeit
Wochenstunden Jump'n'Run-Spiele

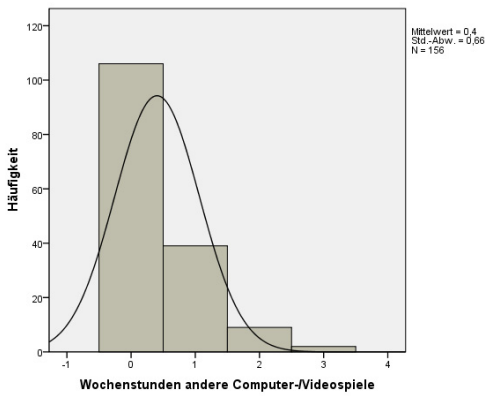
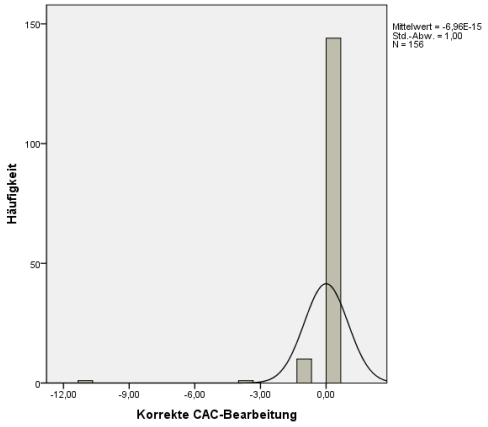
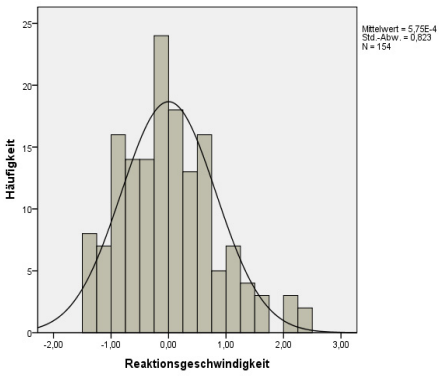
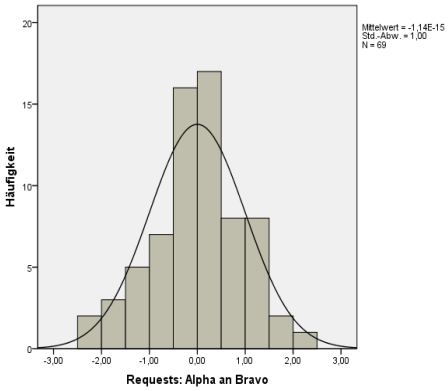
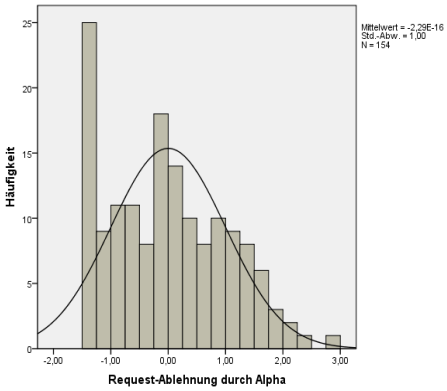
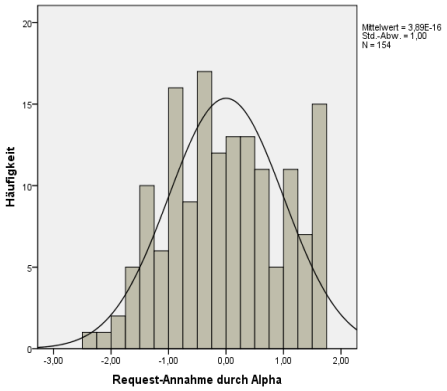
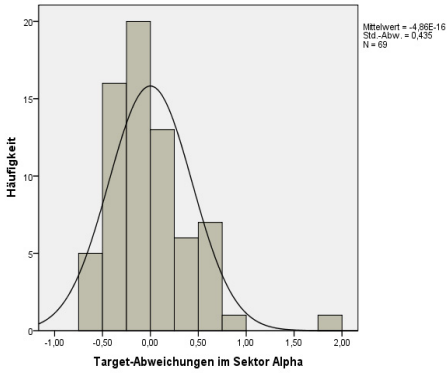
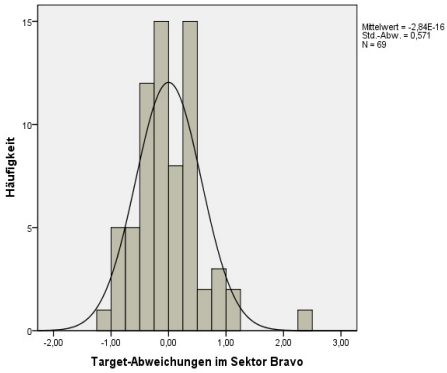

Mittelwert = 0,38
Std.-Abw. = 0,782
N = 156
Häufigkeit
Wochenstunden Ego-Shooter


Mittelwert = 0,08
Std.-Abw. = 0,277
N = 156
Häufigkeit
Wochenstunden Computerschach


Mittelwert = 0,28
Std.-Abw. = 0,607
N = 156
Häufigkeit
Wochenstunden Rollenspiele als Computer-/Videospiel


Mittelwert = 0,19
Std.-Abw. = 0,411
N = 156
Häufigkeit
Wochenstunden Lernspiele als Computer-/Videospiel


Mittelwert = 0,36
Std.-Abw. = 0,632
N = 156
Häufigkeit
Wochenstunden Sportspiele als Computer-/Videospiel


Mittelwert = 0,4
Std.-Abw. = 0,66
N = 156
Häufigkeit
Wochenstunden andere Computer-/Videospiele

Anhang B: Histogramme der HINT-Variablen