

---

# Personality and Computing: Zum Zusammenspiel von Mensch und Maschine

Johannes Albert-von der Gönna

---



München 2016

---

# **Personality and Computing: Zum Zusammenspiel von Mensch und Maschine**

**Johannes Albert-von der Gönna**

---

Inaugural-Dissertation  
zur Erlangung des Doktorgrades  
der Philosophie an der Ludwig-Maximilians-Universität  
München

vorgelegt von  
Johannes Albert-von der Gönna  
aus Penzberg

München, den 06.10.2016

Erstgutachter: Prof. Dr. Markus Bühner  
Zweitgutachter: Prof. Dr. Moritz Heene  
Tag der mündlichen Prüfung: 07.02.2017

# Inhaltsverzeichnis

Abbildungsverzeichnis	vi
Tabellenverzeichnis	vii
Abkürzungsverzeichnis	viii
Zusammenfassung	x
Einführung	1
<b>1 Persönlichkeit, Technologieakzeptanz und -nutzung</b>	<b>2</b>
1.1 Einleitung . . . . .	2
1.1.1 Technologieakzeptanz und -nutzung . . . . .	3
1.1.2 Persönlichkeitseigenschaften . . . . .	7
1.1.3 Persönlichkeit und Technologieakzeptanz . . . . .	8
1.1.4 Technikbereitschaft . . . . .	17
1.1.5 Forschungsfragen und Annahmen . . . . .	18
1.2 Methode . . . . .	21
1.2.1 Erhebungsablauf . . . . .	21
1.2.2 Stichprobenszusammensetzung . . . . .	22
1.2.3 Verwendete Instrumente . . . . .	22
1.2.4 Auswertungsprozess und technische Umsetzung . . . . .	24
1.3 Ergebnisse . . . . .	26
1.3.1 Initiale Analysen . . . . .	26
1.3.2 Modellprüfung TAM, UTAUT und UTAUT2 . . . . .	31
1.3.3 Zusammenhänge der Persönlichkeit mit TAM-Konstrukten . . . . .	33
1.3.4 Modellerweiterungen . . . . .	36
1.3.5 Technikbereitschaft: Modellprüfung und Persönlichkeit . . . . .	37
1.4 Diskussion . . . . .	40
1.4.1 Modelle der Technologieakzeptanz . . . . .	40
1.4.2 TAM: Zusammenhänge mit Persönlichkeit und Modellerweiterung . . . . .	41

---

1.4.3	Technikbereitschaft und Persönlichkeit . . . . .	43
1.4.4	Limitationen . . . . .	44
1.4.5	Fazit und Ausblick . . . . .	45
<b>2</b>	<b>Computerintensive Methoden, Persönlichkeit und Absentismus</b>	<b>47</b>
2.1	Einleitung . . . . .	47
2.1.1	Kognitive Fähigkeiten und berufliche Leistung . . . . .	49
2.1.2	Persönlichkeit und berufliche Leistung . . . . .	50
2.1.3	Kontraproduktives Verhalten . . . . .	51
2.1.4	Persönlichkeit und Absentismus . . . . .	52
2.1.5	Das Bandwidth-Fidelity Dilemma . . . . .	54
2.1.6	Methodische Abwägungen und maschinelles Lernen . . . . .	56
2.1.7	Annahmen der Untersuchung . . . . .	64
2.2	Methode . . . . .	65
2.2.1	Datenerhebung und Stichprobenszusammensetzung . . . . .	65
2.2.2	Verwendete Instrumente . . . . .	66
2.2.3	Auswertungsprozess . . . . .	67
2.3	Ergebnisse . . . . .	73
2.3.1	Vorbetrachtungen . . . . .	73
2.3.2	Hauptergebnisse . . . . .	75
2.3.3	Zusätzliche Befunde . . . . .	80
2.4	Diskussion . . . . .	83
2.4.1	Überprüfung der zugrunde gelegten Annahmen . . . . .	83
2.4.2	Schätzung der Determinationskoeffizienten . . . . .	84
2.4.3	Limitationen . . . . .	85
2.4.4	Fazit und Ausblick . . . . .	86
	<b>Epilog</b>	<b>88</b>
	<b>Anhang A Technologieakzeptanz und -nutzung</b>	<b>89</b>
A.1	Korrelationstabelle . . . . .	89
	<b>Anhang B Computerintensive Methoden</b>	<b>91</b>
B.1	Paarweise Abbildungen . . . . .	91
B.2	Korrelationstabelle . . . . .	95
	<b>Literaturverzeichnis</b>	<b>97</b>

# Abbildungsverzeichnis

1.1	Darstellung des Technology Acceptance Models (TAM) . . . . .	4
1.2	Darstellung der Unified Theory of Acceptance and Use of Technology (UTAUT) . . . . .	5
1.3	Darstellung der Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) . . . . .	7
1.4	Forest-Plots der meta-analytisch aggregierten Zusammenhänge zwischen Persönlichkeit und TAM . . . . .	16
1.5	Modellprüfung und -darstellung TAM . . . . .	32
1.6	Modellprüfung und -darstellung UTAUT . . . . .	33
1.7	Modellprüfung und -darstellung UTAUT2 . . . . .	34
1.8	Modellprüfung und -darstellung Technikbereitschaft . . . . .	38
2.1	Histogramm krankheitsbedingter und unentschuldigter Fehltage . . . . .	73
2.2	Boxplots MSE primäres Kriterium Poisson Regression, Ridge, Lasso und Elastic Net . . . . .	77
2.3	Boxplots MSE primäres Kriterium Lineare Regression, Zero-inflated, Hurdle, Negativ Binomiale und Random Forest . . . . .	77
2.4	Histogramm Yeo-Johnson transformierter krankheitsbedingter und unentschuldigter Fehltage. . . . .	78
2.5	Boxplots MSE für die Vorhersage Yeo-Johson transformierter Fehltage. . .	79
2.6	Histogramme verschiedener Leistungsmaße. . . . .	80
2.7	Boxplots MSE für die Vorhersage verschiedener Leistungsmaße. . . . .	81
B.1	Paarweise Abbildung von Geschlecht, Alter, kognitiven Fähigkeiten und Fehltagen . . . . .	92
B.2	Paarweise Abbildung: Facetten der Dimension Neurotizismus und Fehltage	92
B.3	Paarweise Abbildung: Facetten der Dimension Extraversion und Fehltage .	93
B.4	Paarweise Abbildung: Facetten der Dimension Offenheit und Fehltage . . .	93
B.5	Paarweise Abbildung: Facetten der Dimension Verträglichkeit und Fehltage	94
B.6	Paarweise Abbildung: Facetten der Dimension Gewissenhaftigkeit und Fehltage . . . . .	94

# Tabellenverzeichnis

1.1	Dimensionen und Facetten NEO-PI-R/BFSI . . . . .	9
1.2	Deskriptive Statistik Nutzungshäufigkeit und Persönlichkeit . . . . .	27
1.3	Deskriptive Statistik und Itemanalyse TAM . . . . .	28
1.4	Deskriptive Statistik und Itemanalyse UTAUT/UTAUT2 . . . . .	29
1.5	Deskriptive Statistik und Itemanalyse TB . . . . .	30
1.6	Korrelationen TAM und Persönlichkeit . . . . .	35
1.7	Korrelationen Technikbereitschaft und Persönlichkeit . . . . .	39
2.1	Deskriptive Statistiken der Variablen im primären Modell . . . . .	74
2.2	Anpassungsgüte unterschiedlicher Modelle für das primäre Kriterium . . . . .	76
2.3	Deskriptive Statistiken weiterer abhängiger Variablen . . . . .	78
2.4	Anpassungsgüte unterschiedlicher Modelle für das Yeo-Johnson transformierte primäre Kriterium . . . . .	79
2.5	Anpassungsgüte unterschiedlicher Modelle für weitere abhängige Variablen . . . . .	82
2.6	Vergleich unterschiedlicher Determinationskoeffizienten bei Leistungsmaßen . . . . .	85
A.1	Korrelationen UTAUT/UTAUT2 und Persönlichkeit . . . . .	90
B.1	Korrelationen demographischer Variablen, kognitiver Fähigkeiten, Persönlichkeitseigenschaften und Kriterien . . . . .	96

# Abkürzungsverzeichnis

BFSI . . . . .	Big-Five Struktur Inventar
BI . . . . .	Behavioral Intention (to Use)
Big Five . . . . .	Big Five Modell der Persönlichkeit
CWB . . . . .	Counterproductive Work Behavior
EE . . . . .	Effort Expectancy
FC . . . . .	Facilitating Conditions
FFM . . . . .	Fünf-Faktoren-Modell der Persönlichkeit
GLM . . . . .	Generalisiertes Lineares Modell
GMA . . . . .	General Mental Ability; Allgemeiner Faktor der Intelligenz
HIT . . . . .	Höchster-Intelligenztest
HM . . . . .	Hedonic Motivation
HT . . . . .	Habit
IPIP . . . . .	International Personality Item Pool
Lasso . . . . .	Least Absolute Shrinkage and Selection Operator
MIS . . . . .	Management Information Systems
MSE . . . . .	Mean Square Error; Mittlere quadratische Abweichung
NEO-PI-R . . . . .	NEO-Persönlichkeitsinventar nach Costa und McCrae, rev. Fassung
PE . . . . .	Performance Expectancy



PEOU . . . . .	Perceived Ease of Use
PU . . . . .	Perceived Usefulness
PV . . . . .	Price Value
RF . . . . .	Random Forest
RMSE . . . . .	Root Mean Square Error
RSS . . . . .	Residual Sum of Squares; Residuenquadratsumme
SI . . . . .	Social Influence
TAM . . . . .	Technology Acceptance Model
UTAUT . . . . .	Unified Theory of Acceptance and Use of Technology
UTAUT2 . . . . .	Unified Theory of Acceptance and Use of Technology 2

# Zusammenfassung

Die vorliegende Arbeit bietet aus zwei unterschiedlichen Perspektiven einen differentialpsychologisch geprägten Zugang zu Phänomenen, die sich aus dem Zusammenspiel von Mensch und Maschine ergeben.

Zunächst erfolgt eine Betrachtung möglicher Zusammenhänge spezifischer Persönlichkeitseigenschaften mit Modellen der Technikbereitschaft und Technologieakzeptanz — u. a. *Technology Acceptance Model (TAM)*. In einer studentischen Stichprobe wurde untersucht, in welcher Weise sich die individuelle Akzeptanz einer statistischen Anwendungssoftware unterschied und inwieweit sich die genannten Ansätze zur Vorhersage einer entsprechenden Nutzungsintention bzw. des Nutzungsverhaltens eigneten. Die Beziehungen der jeweiligen Modellannahmen zu Persönlichkeitseigenschaften wurden erstmals auch auf Facettenebene der Big Five spezifisch ermittelt und hinsichtlich einer etwaigen Modellerweiterung überprüft. Unter den berücksichtigten Modellen wies ausschließlich das TAM eine adäquate Modellpassung und einen relevanten Erklärungswert für die berichtete Handlungsabsicht, nicht jedoch für das erfasste Nutzungsverhalten auf. Ein auf Basis der meta-analytisch aggregierten Vorbefunde vermuteter Zusammenhang zwischen der Persönlichkeitsdimension Offenheit und der Nutzungsintention konnte aufgezeigt werden, führte allerdings nicht zu einer Modelloptimierung. Die Bedeutung von Persönlichkeitseigenschaften und die Verwendung etablierter Ansätze zur Erklärung und Vorhersage des Nutzungsverhaltens werden für die betrachtete Technologie dementsprechend kritisch diskutiert.

Des Weiteren wird mithilfe computerintensiver Methoden des maschinellen Lernens im eignungsdiagnostischen Kontext der Auswahl von Bewerberinnen und Bewerbern eine möglichst präzise Vorhersage des Absentismus während der Berufsausbildung angestrebt. Die entsprechende Analyse erfolgte anhand der in einem Ausbildungs- und Personaldienstleistungsunternehmen erhobenen spezifischen kognitiven Fähigkeiten und differenzierten Persönlichkeitseigenschaften von Auszubildenden. Eine verschachtelte Resampling-Strategie erlaubte den Vergleich unterschiedlicher Modellierungsansätze in Hinblick auf die Genauigkeit der erzielten Vorhersagen. Es gelang nicht, Absentismus in befriedigender Weise vorherzusagen, allerdings deuteten sich bei der beispielhaften Modellierung weiterer Leistungsindikatoren vorteilhafte Eigenschaften der im Fokus stehenden Regularisierungsmethoden an (*Ridge Regression*, *Lasso* und *Elastic Net*). Es wird eine Einordnung dieser Befunde vor dem Hintergrund der möglichen Überanpassung und mangelnden Generalisierbarkeit gängiger Modellierungsprinzipien in der psychologischen Diagnostik und Forschung vorgenommen.

# Einführung

Die Entwicklung von Informations- und Kommunikationstechnologien schreitet in den letzten Jahrzehnten in einer beachtlichen und scheinbar unaufhaltsamen Geschwindigkeit voran. Die in diesem Kontext konstruierten elektronischen Geräte (*Hardware*), die zu ihrer Steuerung entwickelten, datenverarbeitenden Programme (*Software*) und die aus diesen Komponenten gebildeten, weltumspannenden Netzwerke nehmen vermehrt Raum in der Lebenswirklichkeit aller Menschen ein. Sie beeinflussen nicht nur in zunehmenden Maße die menschliche Wahrnehmung der Umwelt und die Interaktion mit anderen Personen, sondern sind als von Menschen geschaffene Maschinen und Technologien selbst Objekte der Aufmerksamkeit und Ziele gerichteten Handelns. Ein derartiger, für das menschliche Erleben und Verhalten zunehmend zentraler Themenkomplex legt eine Auseinandersetzung im Rahmen psychologischer Forschung nahe.

Die im Folgenden berichteten Untersuchungen befassen sich mit differentialpsychologischen Aspekten an der Schnittstelle des Zusammenspiels von Mensch und Maschine. Der in Teilen englischsprachige Titel der Arbeit wurde hierbei bewusst gewählt, um die unterschiedlichen Perspektiven zu verdeutlichen, die dazu eingenommen werden. Diese werden insbesondere im Begriff *Computing* deutlich, dem keine einzelne deutsche Übersetzung umfassend gerecht wird. Er beschreibt einerseits ganz allgemein die elektronische Datenverarbeitung und das Benutzen eines Computers, gleichzeitig steht er aber auch für die konkrete Durchführung von Berechnungen, heutzutage i. d. R. ebenfalls unter Zuhilfenahme eines Computers.

Im Folgenden werden zwei Studien berichtet, die sich mit diesen beiden Komponenten des *Computing* aus persönlichkeitspsychologischer Perspektive auseinandersetzen. Im Kapitel »*Zusammenhänge zwischen Persönlichkeit, Technologieakzeptanz und -nutzung*« wird eine Untersuchung beschrieben, in der *Computing* die Rolle der ersten dieser Bedeutungen annimmt. Es wird eine Annäherung an die Beantwortung der Frage vorgenommen, welche interindividuellen Faktoren in Form spezifischer, zeitlich weitgehend stabiler Persönlichkeitseigenschaften mit der unterschiedlichen Akzeptanz und Nutzungsbereitschaft neuer technologischer Entwicklungen in Zusammenhang stehen und diese ggf. beeinflussen.

Im zweiten Teil der vorliegenden Arbeit wird unter der Überschrift »*Computerintensive Methoden, Persönlichkeit und Absentismus während der Berufsausbildung*« im Sinne des *Computing* als Durchführung von Berechnungen eine neuartige, rechenintensive Perspektive auf die Tradition der Prädiktion von ausbildungs- bzw. berufsrelevanten Kriterien im Rahmen der berufsbezogenen Eignungsdiagnostik eingenommen und vorgestellt.

# Kapitel 1

## Zusammenhänge zwischen Persönlichkeit, Technologieakzeptanz und -nutzung

Die nachfolgend dargestellte Untersuchung widmet sich den interindividuellen Faktoren unterschiedlicher Technologieakzeptanz und -nutzung. Sie versucht durch einen detaillierten, differentialpsychologischen Zugang zum Aufbau eines besseren Verständnisses dieses Themenkomplexes beizutragen.

### 1.1 Einleitung

Hinsichtlich der Art und des Umfangs der Nutzung zeigen unterschiedliche Personen eine teils deutliche Variabilität im Umgang mit der sie umgebenden Hard- und Software. Im Weiteren erfolgt deshalb eine Auseinandersetzung mit grundlegenden Aspekten derartiger Technologienutzung. Zentral wird hierbei die Frage sein, welche wesentlichen Faktoren die Akzeptanz und Nutzungsbereitschaft neuer technologischer Entwicklungen beeinflussen. Erstaunlicherweise finden diese Fragestellungen in der Persönlichkeitspsychologie bisher vergleichsweise wenig Beachtung. Im unternehmerischen Kontext und dem wissenschaftlichen Diskurs von Management-Informationssystemen, bzw. der Wirtschaftsinformatik — engl. *Management Information Systems (MIS)* — werden entsprechende Inhalte aber bereits seit mehreren Jahrzehnten diskutiert (z. B. Swanson, 1974). In der vorliegenden Arbeit werden diese Erkenntnisse aufgegriffen und um eine differentialpsychologische Perspektive erweitert. Als Ausgangspunkte der folgenden Untersuchung können dementsprechend die folgenden Fragen angesehen werden:

In welcher Weise unterscheidet sich die persönliche Bereitschaft, sich mit neuen Technologien auseinanderzusetzen und welche interindividuellen Faktoren führen zur Akzeptanz (oder Ablehnung) neuer technologischer Entwicklungen?

Ein besonderes Augenmerk soll hierbei auf die Bedeutung grundlegender, zeitlich weitgehend stabiler Persönlichkeitseigenschaften in Ergänzung zu möglicherweise situations-

bzw. technikspezifisch veränderlichen Einstellungen und Einschätzungen gerichtet werden. Es erfolgt eine konkrete Betrachtung diesbezüglicher Phänomene im Rahmen der psychologischen Methodenlehre des universitären Psychologiestudiums. Der eingangs kurz skizzierte technologische Fortschritt hat dazu geführt, dass in diesem Kontext das schriftliche Lösen mathematischer Gleichungen oder das Herleiten statistischer Formeln weitgehend in den Hintergrund gerückt ist und zunehmend durch die Schulung in statistischer Analysesoftware ersetzt wird (Tishkovskaya & Lancaster, 2012). Inwieweit derartige Programmpakete aber von Studierenden akzeptiert werden, bzw. welchen Anteil differentialpsychologische — überwiegend als invariant erachtete — Persönlichkeitsmerkmale an der Beschreibung und Erklärung der Bereitschaft zur Nutzung dieser Software einnehmen, wurde bislang nicht systematisch untersucht. Dementsprechend soll der mögliche Einfluss solcher interindividueller Faktoren auf das Nutzungsverhalten den Kern der nachfolgend vorgestellten Untersuchung ausmachen.

Einzelne Vorbefunde zur gemeinsamen Betrachtung der angesprochenen, nachfolgend in Abschnitt 1.1.1 vertieften Modelle der Technologieakzeptanz und grundlegender Persönlichkeitseigenschaften (vgl. Kapitel 1.1.2) werden in Abschnitt 1.1.3 behandelt, bevor in Kapitel 1.1.4 ein alternatives, in der weiteren Untersuchung ebenfalls berücksichtigtes, integratives Modell der Technikbereitschaft vorgestellt wird. Diese Untersuchungsergebnisse bilden schließlich die gemeinsame Grundlage für die in Abschnitt 1.1.5 konkretisierten Forschungsfragen und untersuchungsleitenden Annahmen der vorliegenden Studie.

### 1.1.1 Technologieakzeptanz und -nutzung

Das prominenteste Modell, das in der Wirtschaftsinformatik bzw. den MIS zur Beantwortung der initialen, oben wiedergegebenen Fragestellung entwickelt wurde, ist das sog. *Technology Acceptance Model (TAM)* (Davis, 1989, 1993; Davis & Venkatesh, 1996). Abgeleitet aus der ursprünglich in der sozialpsychologischen Forschung entwickelten und in den Wirtschaftswissenschaften adaptierten Theorie des überlegten Handelns (engl. *Theory of Reasoned Action*) (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) führt es die individuelle Ausprägung der Nutzung neuer Technologien auf eine vorausgehende Handlungsabsicht, bzw. Nutzungsintention zurück — engl. *Behavioral Intention (to Use) (BI)*. Für diese Handlungsabsicht werden wiederum die Einschätzungen der empfundenen Nützlichkeit — engl. *Perceived Usefulness (PU)* — und der empfundenen Leichtigkeit der Nutzung im Sinne des Bedienkomforts bzw. der Benutzerfreundlichkeit — engl. *Perceived Ease of Use (PEOU)* — als ausschlaggebend angesehen. Diese Einstellungen und Erwartungen werden in spezifischen Situationen jeweils in Hinblick auf eine konkret benannte Technologie erfasst. In ihrem Zusammenwirken sind diese drei Faktoren schließlich als Maße der Technologieakzeptanz interpretierbar. Eine Darstellung dieses Modells ist Abbildung 1.1 zu entnehmen.

In mehreren Weiterentwicklungen wurde dieses ursprüngliche TAM um verschiedene Antezedenzen der empfundenen Nützlichkeit (PU) und Leichtigkeit der Nutzung (PEOU) sowie moderierende Einflussfaktoren ergänzt. Basierend auf theoretischen Überlegungen und initial an vier unterschiedlichen Stichproben (jeweils  $N \leq 51$ ) in längsschnittlichen



2008). Es wird deutlich, dass es sich hierbei zwar schon um einzelne, teils womöglich auch wenig veränderbare, aber domänenspezifische Persönlichkeitsmerkmale handelt, um die das ursprüngliche TAM ergänzt wurde.

Parallel zu diesen Weiterentwicklungen des ursprünglichen TAM wurde — teilweise von den gleichen Autoren — der Bedarf einer Vereinheitlichung der unterschiedlichen Ansätze erkannt, welche bis dato zur Erklärung von Akzeptanz und Nutzung neuartiger Informationstechnologien herangezogen wurden (Venkatesh, Morris, Davis & Davis, 2003). Nach einer umfangreichen Durchsicht der bestehenden Literatur identifizierten und analysierten sie deshalb acht entsprechende Modelle, unterzogen diese einer empirischen Prüfung und schlugen schließlich ein weiteres Modell vor, das auf Basis dieser Betrachtungen und theoretischen Überlegungen einzelne Bestandteile der vorausgegangenen Annahmen zusammenführte. Anhand weiterer empirischer Überprüfungen wurde die Überlegenheit in der Vorhersage der Nutzungsintention dieses als *Unified Theory of Acceptance and Use of Technology (UTAUT)* bezeichneten Modells aufgezeigt, welches Abbildung 1.2 entnommen werden kann.

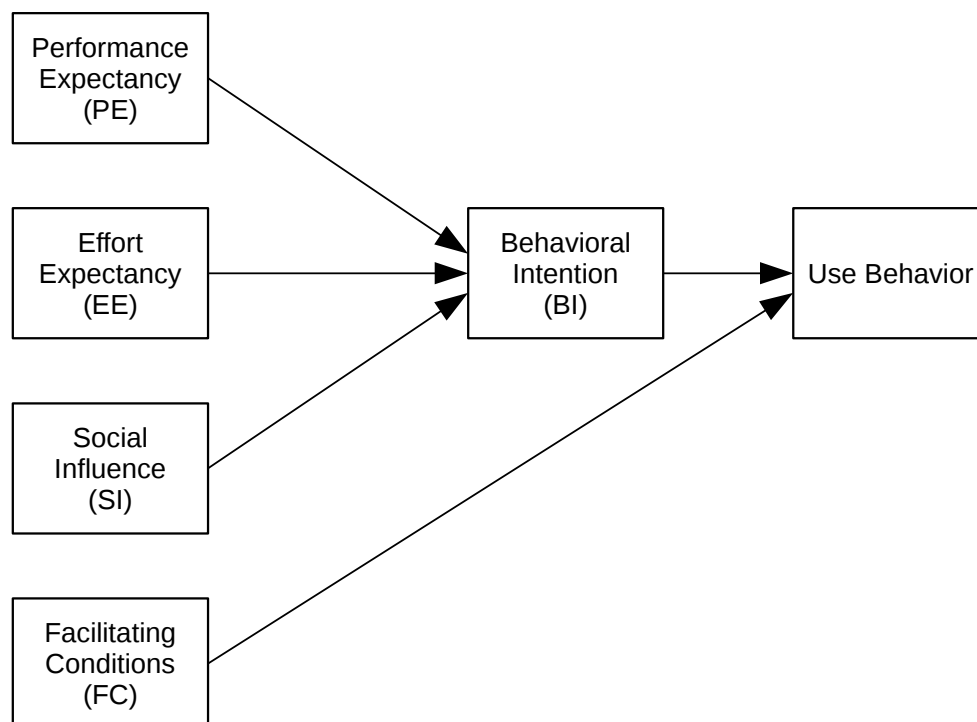


Abbildung 1.2: Das Modell der *Unified Theory of Acceptance and Use of Technology (UTAUT)* nach Venkatesh, Morris, Davis und Davis (2003).

Als Einflussfaktoren auf die Nutzungsintention bzw. das Nutzungsverhalten stellen die Erwartungen an die Leistung der Technologie (engl. *Performance Expectancy (PE)*; konzeptuell vergleichbar mit PU aus dem TAM), bzw. den Aufwand der Nutzung (engl. *Effort Expectancy (EE)*; konzeptuell vergleichbar mit PEOU aus dem TAM), soziale Einflüsse

(engl. *Social Influence (SI)*; alle mit angenommenen Einfluss auf die Nutzungsintention) und förderliche Umgebungsbedingungen wie die organisatorische und technische Infrastruktur (engl. *Facilitating Conditions (FC)* mit direktem Einfluss auf das Nutzungsverhalten) die zentralen Komponenten dieses Modells dar. Geschlecht, Alter, Erfahrung und Freiwilligkeit der Nutzung wurden als mögliche moderierende Einflussgrößen identifiziert (Venkatesh et al., 2003).

Auch dieses Modell wurde einige Jahre später mit einem besonderen Blick auf die Akzeptanz und Nutzung neuartiger Technologien im Konsumentenbereich erweitert. Waren die vorausgegangenen Modelle vornehmlich im Unternehmenskontext entwickelt und an Angestellten validiert worden, entstand ein zunehmender Bedarf, die erarbeiteten Erklärungsansätze auch auf ein breiteres Anwendungsfeld zu übertragen (Venkatesh, Thong & Xu, 2012). Zu diesem Zwecke wurde das ursprüngliche UTAUT um weitere Determinanten der postulierten Handlungsabsicht (und des Nutzungsverhaltens) ergänzt und als *Unified Theory of Acceptance and Use of Technology 2 (UTAUT2)* präsentiert. Für die förderlichen Umgebungsbedingungen des UTAUT wurde nun auch ein direkter Einfluss auf die Nutzungsintention postuliert. Weiterhin wurde das Preis-Leistungs-Verhältnis der Technologie — engl. *Price Value (PV)* — als zusätzlicher Prädiktor der Handlungsabsicht identifiziert. Schließlich wurde das Modell um zwei weitere Einflussfaktoren ergänzt, die von besonderem psychologischen Interesse sein könnten: der hedonistischen Motivation (engl. *Hedonic Motivation (HM)*, im Sinne eines bei der Nutzung erwarteten Vergnügens) wurde ein Einfluss auf die Nutzungsintention zugesprochen, ebenso der selbst eingeschätzten Gewohnheit im Umgang mit der neuen Technologie — engl. *Habit (HT)*. Für letztere wurde zusätzlich ein direkter Einfluss auf das Nutzungsverhalten angenommen (Venkatesh et al., 2012). Das entsprechend ergänzte Modell ist in Abbildung 1.3 dargestellt.

Da die Auseinandersetzung mit diesen Inhalten vornehmlich in den Fachbereichen Wirtschaftsinformatik bzw. MIS erfolgte, ist es vermutlich wenig verwunderlich, dass interindividuelle Unterschiede grundlegender, zeitlich weitgehend als stabil angenommener Persönlichkeitsmerkmale in den zuvor besprochenen, ursprünglichen Betrachtungen keine direkte Beachtung erfahren haben. Obgleich diese Modelle — wie dargestellt — unmittelbar an einer prominenten, sozialpsychologischen Annahme über das Zusammenwirken von Einstellungen und Verhaltensweisen anknüpfen. Erst in jüngster Vergangenheit wurde in einzelnen, weitgehend als explorativ zu beurteilenden Arbeiten die Ergänzung dieser Modelle um gewisse Persönlichkeitseigenschaften in Erwägung gezogen. Die hier vorliegende Arbeit soll an diese Veröffentlichungen anknüpfen und sie durch eine detaillierte Betrachtungsweise erweitern. Hierzu wird im nächsten Abschnitt zunächst das verwendete Persönlichkeitsmodell vorgestellt, das sich u. a. durch eine entsprechend differenzierte Beschreibung der menschlichen Persönlichkeit auszeichnet. Die im Hinblick auf eine gemeinsame Betrachtung von derartigen Persönlichkeitseigenschaften und den vorgestellten Technologieakzeptanz-Modellen verfügbaren Vorbefunde werden dann im darauf folgenden Abschnitt thematisiert.



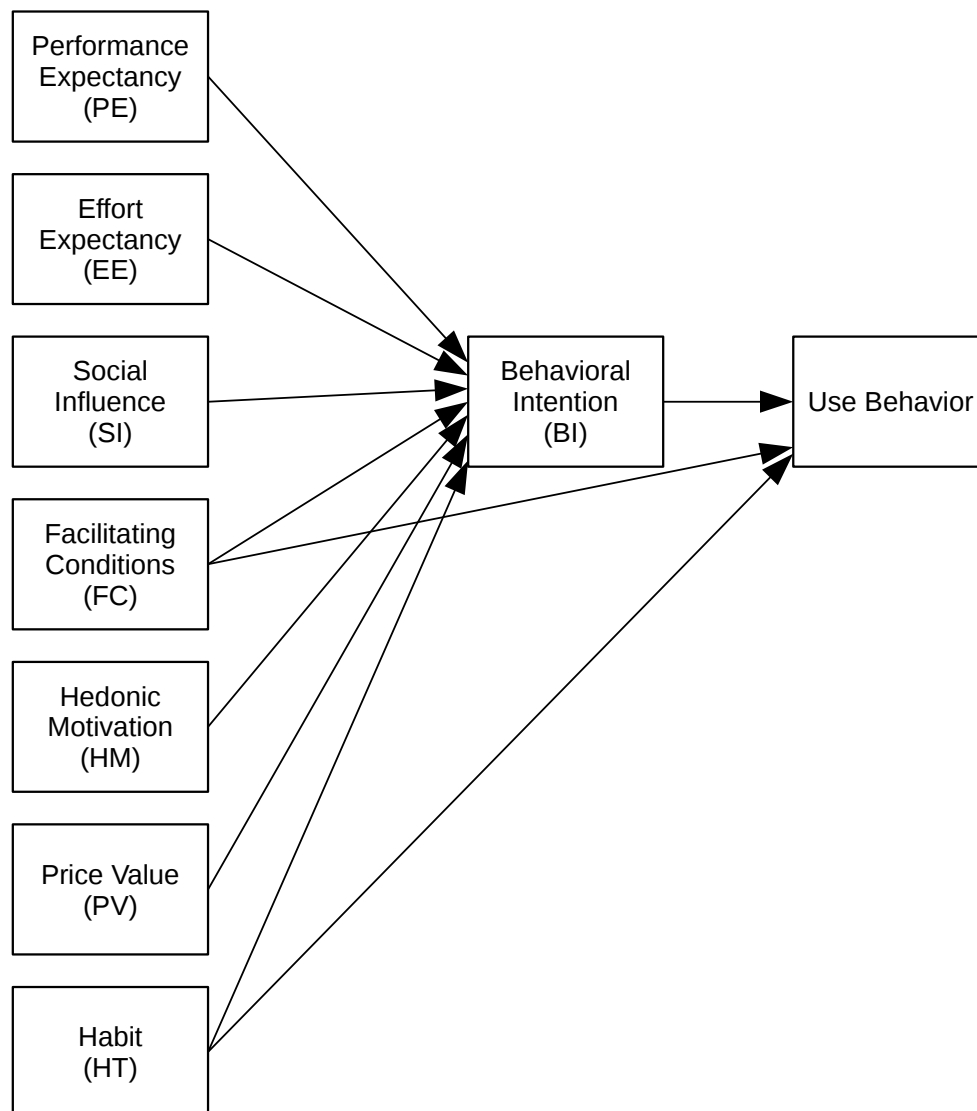


Abbildung 1.3: Das Modell der erweiterten *Unified Theory of Acceptance and Use of Technology* (UTAUT2) im Konsumentenbereich nach Venkatesh, Thong und Xu (2012).

### 1.1.2 Persönlichkeitseigenschaften

Anknüpfend an die lexikographische Forschungstradition (u. a. Allport & Odbert, 1936; Cattell, 1946) und auf Basis ausführlicher faktorenanalytischer Untersuchungen zur Struktur von Persönlichkeitsfragebögen (u. a. Tupes & Christal, 1961, 1992) wurde in der zweiten Hälfte des vergangenen Jahrhunderts ein umfassender und einheitlicher Beschreibungsrahmen der menschlichen Persönlichkeit entwickelt, das sog. *Fünf-Faktoren-Modell der Persönlichkeit* (FFM). Gemäß diesem Modell lässt sich die Gesamtheit der menschlichen Persönlichkeit anhand von fünf Persönlichkeitsdimensionen erfassen und beschreiben, die als relativ unabhängig voneinander angenommen werden: *Neurotizismus* (bzw. bei der Definiti-

on über die positive Polung erfolgt eine Benennung als *Emotionale Stabilität*), *Extraversion*, *Offenheit* (für Erfahrungen), *Gewissenhaftigkeit* und *Verträglichkeit* (Digman, 1990; McCrae & Costa, 1985, 1987). Dieses Modell wird daher auch als das *Big Five Modell der Persönlichkeit* (*Big Five*) bezeichnet (Goldberg, 1981). In der Konzeption umfangreicher psychologischer Persönlichkeitsfragebögen wird es oftmals als hierarchisches Persönlichkeitssystem spezifischer Eigenschaften weiter ausgebaut. Das *NEO-Persönlichkeitsinventar nach Costa und McCrae, rev. Fassung (NEO-PI-R)* ist ein solches, im deutschen Sprachraum etabliertes Verfahren (Ostendorf & Angleitner, 2004). Es differenziert die fünf Dimensionen in einer Unterebene in spezifische Persönlichkeitseigenschaften, die sog. Facetenebene der Persönlichkeit aus. Eine Übersicht und Zuordnung dieser Facetten zu den fünf genannten Persönlichkeitsdimensionen des FFM kann Tabelle 1.1 entnommen werden. Dieser sind auch die teilweise abweichenden Bezeichnungen für diese spezifischen Persönlichkeitskonstrukte aus dem *Big-Five Struktur Inventar (BFSI)* zu entnehmen, einem alternativen, strukturell vergleichbaren Verfahren zur Erfassung des FFM (Arendasy, Sommer & Feldhammer, 2009).

### 1.1.3 Persönlichkeit und Technologieakzeptanz

In einem wenig beachteten Konferenzbeitrag stellten Sharma und Citurs (2004) erstmals einen Zusammenhang zwischen den zuvor beschriebenen Technologieakzeptanzmodellen und Persönlichkeitseigenschaften in Form der Big Five her. Sie entwarfen ein Modell, das die fünf Dimensionen der Persönlichkeit als Vermittler unterschiedlicher Zusammenhänge im zu diesem Zeitpunkt neu vorgestellten UTAUT vorsah und skizzierten ein Forschungsvorhaben, um dieses zu überprüfen.

Eine gemeinsame Diskussion der TAM- und Persönlichkeitskonstrukte, die mehr Beachtung erhielt, erfolgte durch McElroy, Hendrickson, Townsend und DeMarie (2007). In ihrer Arbeit konnte ein direkter Zusammenhang zwischen einzelnen Faktoren des NEO-PI-R und der selbst eingeschätzten, allgemeinen Internetnutzung, bzw. der Durchführung von Online-Verkäufen aufgezeigt werden. Die Autoren leiteten daraus eine weiterreichende Bedeutsamkeit der Persönlichkeitseigenschaften ab und rieten eine Kombination mit den etablierten Modellen der Technologieakzeptanz an. Sie mutmaßten weiterhin, dass die Beachtung spezifischerer Persönlichkeitseigenschaften, etwa wie sie in Form der Facetten des NEO-PI-R erfasst werden, zusätzliche Erkenntnisse zur Rolle der Persönlichkeit bei der Akzeptanz und Nutzung technologischer Entwicklungen bieten könnten (McElroy et al., 2007).

Die erste Erhebung, in der diese Überlegungen tatsächlich umgesetzt wurden, wurde nahezu parallel von Devaraj, Easley und Crant (2008) durchgeführt. Die Autoren vermerkten, dass das FFM in der bestehenden Literatur zur Technologieakzeptanz und -nutzung bisher so gut wie nicht behandelt worden war. Aufgrund des konzeptuellen Ursprungs dieser Modelle in der Theorie des überlegten Handelns (Ajzen & Fishbein, 1980), in deren Rahmen Persönlichkeitseigenschaften aber explizit als sog. externale Faktoren diskutiert wurden, die Einfluss auf die Herausbildung von individuellen Überzeugungen haben (und damit den Einstellungen und einer möglichen Handlungsintention vorausgehen), nah-

Tabelle 1.1: Persönlichkeitsdimensionen und -facetten des revidierten NEO-Persönlichkeitsinventars (NEO-PI-R) und Big-Five Struktur Inventars (BFSI).

Dimension	Facette
N: Neurotizismus [Emotionale Stabilität] (neuroticism)	N1: Ängstlichkeit [Unbekümmertheit] (anxiety) N2: Reizbarkeit [Gelassenheit] (angry hostility) N3: Depression [Positive Grundstimmung] (depression) N4: Soziale Befangenheit[/Selbstsicherheit] (self-consciousness) N5: Impulsivität [Selbstbeherrschung] (impulsiveness) N6: Verletzlichkeit [Emotionale Robustheit] (vulnerability)
E: Extraversion (extraversion)	E1: Herzlichkeit [Freundlichkeit] (warmth) E2: Geselligkeit (gregariousness) E3: Durchsetzungsfähigkeit (assertiveness) E4: Aktivität [Tatendrang] (activity) E5: Erlebnishunger [Abenteuerlust] (excitement-seeking) E6: Frohsinn [Fröhlichkeit] (positive emotions)
O: Offenheit für Erfahrungen [Offenheit] (openness to experience)	O1: Offenheit für Fantasie (fantasy) O2: Offenheit für Ästhetik (aesthetics) O3: Offenheit für Gefühle (feelings) O4: Offenheit für Handlungen (actions) O5: Offenheit für Ideen (ideas) O6: Offenheit des Werte- und Normensystems (values)
A: Verträglichkeit (agreeableness)	A1: Vertrauen[sbereitschaft] (trust) A2: Freimütigkeit [Aufrichtigkeit] (straightforwardness) A3: Altruismus [Hilfsbereitschaft] (altruism) A4: Entgegenkommen (compliance) A5: Bescheidenheit (modesty) A6: Gutherzigkeit[-mütigkeit] (tender-mindedness)
C: Gewissenhaftigkeit (conscientiousness)	C1: Kompetenz (competence) C2: Ordnungsliebe (order) C3: Pflichtbewusstsein (dutifulness) C4: Leistungsstreben [Ehrgeiz] (achievement striving) C5: Selbstdisziplin [Disziplin] (self-discipline) C6: Besonnenheit [Bedachtsamkeit] (deliberation)

*Anmerkung.* Vom NEO-PI-R abweichende Bezeichnungen des BFSI sind in eckigen, die des amerikanischen Originals in runden Klammern angegeben.

men sie an, dass ein vergleichbarer Einfluss auch im Rahmen von TAM-Betrachtungen zu erwarten sei. Sie postulierten dementsprechend primär direkte Einflüsse der Big Five-Persönlichkeitsmerkmale auf die empfundene Nützlichkeit der betrachteten Technologie (PU) sowie einen weiteren direkten Einfluss von Neurotizismus auf das ergänzend miteinbezogene Konstrukt subjektiver Normen. Weiterhin wurden moderierende Einflüsse ein-

zelter Persönlichkeitseigenschaften auf die Zusammenhänge zwischen der empfundenen Nützlichkeit, bzw. den subjektiven Normen und der Handlungsabsicht vermutet. Das Nutzungsverhalten sollte dementsprechend auf unterschiedliche Weise indirekt durch stabile Persönlichkeitsmerkmale beeinflusst werden. Diese Annahmen wurden an einer Stichprobe von  $N = 180$ , vornehmlich männlicher (83%) MBA-Studenten (Alter:  $M = 30$  Jahre) überprüft, die während des Verlaufs eines Semesters die Möglichkeit hatten, veranstaltungsbegleitende, teils gruppenorientierte Arbeitsaufträge mithilfe einer kollaborativen, projektbasierten Online-Plattform zu bearbeiten. Zur Erfassung der TAM-Konstrukte wurde auf eine Selbstauskunft analog zu Davis (1989) zurückgegriffen, das Big Five-Persönlichkeitsmodell wurde mit dem NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae (1992) erhoben. Das Nutzungsverhalten wurde über die Logfiles der Online-Plattform objektiv erfasst (alle Nutzeraktionen außerhalb der Kernkurse, alle Nutzeraktionen in zugewiesenen Gruppenprojekten und Größe des Kontaktnetzwerkes) und anschließend in aggregierter Form analysiert. Mit Ausnahme des angenommenen Zusammenhangs zwischen Offenheit und der empfundenen Nützlichkeit der Online-Plattform wurden bedeutsame Zusammenhänge in der postulierten (und zuvor knapp wiedergegebenen) Form beobachtet. Die Autoren berichteten auch von einem bedeutsamen, positiven Zusammenhang zwischen der erhobenen Handlungsabsicht und dem tatsächlichen Nutzungsverhalten. Da entgegen der entsprechenden Annahme kein Zusammenhang zwischen Offenheit und der empfundenen Nützlichkeit aufgezeigt werden konnte, prüften die Autoren die Rolle dieser Persönlichkeitseigenschaft mithilfe eines weiteren, explorativen Modells. In diesem wurde ein direkter Einfluss von Offenheit auf die Nutzungsintention zugelassen, der sich auch als bedeutsam erwies. Die Autoren schlossen daraus, dass einzelne Persönlichkeitsmerkmale womöglich auch einen direkten, nicht notwendigerweise durch die anderen TAM-Konstrukte vermittelten Zusammenhang mit der Handlungsabsicht aufweisen könnten, welche wiederum der Technologienutzung vorausgeht (Devaraj et al., 2008).

In einer Studie zur Akzeptanz eines sozialen Online-Netzwerkes schlugen Rosen und Kluemper (2008) eine ähnliche Erweiterung des TAM um die Big Five-Persönlichkeitseigenschaften vor. Sie adaptierten die Erfassung der empfundenen Nützlichkeit, um dem Einsatzzweck der genannten Technologie abseits einer unternehmerischen Umgebung gerecht zu werden. Weiterhin wurden mit Flow und Netzwerkgröße — neben Verträglichkeit, Offenheit und Extraversion — zwei weitere Determinanten der solcherart eingeschätzten Nützlichkeit postuliert. Für Extraversion wurde außerdem ein Einfluss auf den wahrgenommenen Bedienkomfort angenommen, ebenso von Neurotizismus und Gewissenhaftigkeit. Zur Überprüfung dieser Annahmen wurden einer studentischen Stichprobe ( $N = 522$ , 32% männlich, Alter:  $M = 21$  Jahre) eine nicht explizit bezeichnete, aus fünfzig Items bestehende Fassung des *International Personality Item Pool (IPIP)* (Goldberg et al., 2006) vorgelegt.<sup>1</sup> Die Erfassung der TAM-Konstrukte erfolgte (mit Ausnahme der empfundenen Nützlichkeit, s.o.) mithilfe der von Davis (1989) vorgeschlagenen Skalen ebenfalls per Selbstauskunft. Eine Einschätzung des tatsächlichen Nutzungsverhaltens wurde weder sub-

---

<sup>1</sup>Aufgrund des erwähnten Umfangs und einer gewissen Popularität könnte es sich um die 50-Item Variante der Goldberg (1992) Marker für die Big Five-Faktorstruktur gehandelt haben.

jektiv noch objektiv vorgenommen. Die zahlreich formulierten Annahmen wurden allesamt auf Basis der bivariaten Korrelationen zwischen den erwähnten Konstrukten geprüft.

Punnoose (2012) analysierte in ähnlicher Absicht die Zusammenhänge der Big Five-Persönlichkeitseigenschaften mit den TAM-Konstrukten im Kontext der studentischen Nutzung eines universitären E-Learning-Angebots. Sein, an vorausgehende Untersuchungen angelehntes Modell enthielt neben den aufgeführten Konstrukten zusätzlich Maße der computerbezogenen Selbstwirksamkeit, subjektiver Normen und des empfundenen Vergnügens bei der Nutzung dieser Technologie. Er vermutete diverse Einflüsse verschiedener Big Five-Persönlichkeitseigenschaften auf das empfundene Vergnügen, die Nützlichkeit dieser spezifischen Technologie und die subjektive Normwahrnehmung, welche ihrerseits wiederum die nachgeordnete Handlungsabsicht beeinflussen sollten. Auf der Basis von Selbstausskünften einer multinationalen, studentischen Stichprobe an einer thailändischen Universität ( $N = 249$ , 45% männlich, keine Altersangabe) wurden alle möglichen Modelle mit diesen Variablen angepasst und anhand der Minimierung des genormten  $\chi^2$ -Wertes ein (ausschließlich an den vorhandenen Daten optimiertes) Modell ausgewählt. Die genannten Konstrukte wurden mithilfe verschiedener Kurzfragebögen erfasst, die in vorausgegangenen Untersuchungen bereits Verwendung gefunden hatten und in Anlehnung an etablierte Skalen entwickelt oder von solchen abgeleitet wurden. Im Falle der empfundenen Nützlichkeit der Technologie erfolgte eine Entfernung einzelner Items, die mit zu hohen Nebenladungen begründet wurde. Die berichteten Befunde wurden von Punnoose (2012) in weitgehender Unterstützung seiner Vorannahmen interpretiert.

Die in der Folgezeit von Svendsen, Johnsen, Almås-Sørensen und Vittersø (2013) veröffentlichte Arbeit erlaubte es, eine zusätzliche Abschätzung der Zusammenhänge von Persönlichkeitseigenschaften und den etablierten Einflussgrößen der Technologieakzeptanz und -nutzung vorzunehmen. Aus ihrer Aufarbeitung der Vorbefunde zu interindividuellen Unterschieden (ohne Berücksichtigung des FFM) im erweiterten Kontext der TAM-Forschung schlussfolgerten die Autoren, dass die Annahme einer vollständigen Vermittlung oder Mediation entsprechender Persönlichkeitsmerkmale durch die etablierten TAM-Konstrukte (im Sinne externaler Faktoren der Theorie des überlegten Handelns) vermutlich nicht aufrechtzuerhalten sei. Stattdessen legten sie nahe, zwischen unterschiedlichen Szenarien der Technologienutzung zu unterscheiden, in denen von unterschiedlichen Ausgangserwartungen an die betrachteten Persönlichkeitseigenschaften auszugehen sei. Sie schlugen vor, dass im sog. *Standard-Fall* (engl. *standard case*), von dem in TAM-Studien üblicherweise ausgegangen wird, die Nutzer sowohl mit der zu bewertenden Technologie, als auch mit ihren eigenen Interessen und Fähigkeiten weitgehend vertraut sind. Die Bewertung der Technologie erfolge dann vor dem Hintergrund interindividueller Unterschiede in Persönlichkeitsmerkmalen und führt zu unterschiedlichen Neigungen im Umgang mit der untersuchten Technologie — welche sich gemäß TAM in variierenden Handlungsabsichten und einem entsprechenden Nutzungsverhalten niederschlägt. Nur in diesem Fall sollte ein möglicher Einfluss stabiler, interindividuell unterschiedlicher Persönlichkeitseigenschaften auf die Technologienutzung vollständig durch die TAM-Konstrukte der empfundenen Nützlichkeit und Benutzerfreundlichkeit vermittelt werden. Die Autoren hielten fest, dass diese Annahme nahelegt, dass sich die resultierenden Zusammenhänge zwischen verschiedenen

Studien in Abhängigkeit der betrachteten Technologie teils deutlich unterscheiden sollten. Eine andere Erwartung bestünde im Falle einer *schlechten Justierung* (engl. *misalignment case*): dient die Technologienutzung einem sekundären Zweck (z. B. einer Emotionsregulierung), der unabhängig von der jeweils betrachteten Technologie wirksam und durch die TAM-Konstrukte nicht erfasst werden kann, sollten diese die etwaigen Einflüsse zeitstabiler Persönlichkeitsmerkmale auch nicht vermitteln. In diesem Fall würde nach Svendsen et al. (2013) ein direkter Zusammenhang der Persönlichkeitsfaktoren mit der erfassten Handlungsabsicht sichtbar werden. Schließlich wurde argumentiert, dass im sog. *Beurteilungsfall* (engl. *rating case*) und insbesondere dann, wenn keine oder nur eine geringe Vertrautheit mit der zu bewertenden Technologie besteht, unterschiedliche Ausprägungen in verschiedenen Persönlichkeitseigenschaften generell zu unterschiedlichen Bewertungen der primären TAM-Konstrukte führen könnten, ohne dass ein weiterführender Zusammenhang mit der Handlungsabsicht anzunehmen sei. Basierend auf diesen Überlegungen stellten die Autoren mehrere Annahmen auf, die verschiedene Zusammenhänge zwischen vier der fünf Dimensionen des FFM (Verträglichkeit wurde aufgrund einer niedrigen Reliabilität der Messung nicht in die Analysen aufgenommen) und den erweiterten Konstrukten des TAM (wahrgenommene Nützlichkeit, Bedienkomfort und die subjektiv empfundene Norm) betrafen. Eine Überprüfung dieser Annahmen erfolgte anhand einer umfangreichen Befragung von  $N = 1004$  Panel-Mitgliedern eines norwegischen Statistik-Instituts (50% männlich, Alter:  $M = 45$  Jahre). Die Erhebung wurde webbasiert durchgeführt und schloss eine Beschreibung der fraglichen Technologie mit ein. Eine Erhebung tatsächlichen Nutzungsverhaltens erfolgte nicht. Die TAM-Konstrukte wurden mithilfe etablierter Maße erfasst (vgl. Ajzen & Fishbein, 1980; Davis, 1989; Venkatesh et al., 2003), wobei die Beurteilung der Nützlichkeit auf den Kontext der privaten Technologienutzung angepasst wurde. Zur Messung der Big Five wurde auf eine 20-Item Skala des IPIP (Goldberg, 1999) zurückgegriffen, die Svendsen et al. (2013, Anhang 2) entnommen werden kann.

Ein kürzlich veröffentlichter Konferenzbeitrag von Özbek, Alınacı, Koc, Akkılıç und Kaş (2014) bot ergänzende Hinweise auf die Zusammenhänge der benannten Konstrukte. Die Autoren erweiterten ein Modell der beiden zentralen TAM-Konstrukte, d. h. der Beurteilung des Bedienkomforts und der empfundenen Nützlichkeit (als Determinanten der Handlungsabsicht in Bezug auf die Nutzung von Smartphones) um die Dimensionen des FFM. Hierbei wurden zunächst uneingeschränkt Einflüsse aller fünf Persönlichkeitseigenschaften auf diese beiden TAM-Variablen zugelassen, die sich jedoch nicht allesamt als bedeutsam erwiesen. Das Modell wurde ebenfalls an einer studentischen Stichprobe ( $N = 401$ , keine Angaben zum Geschlechterverhältnis und dem Durchschnittsalter) überprüft. Hierzu kamen, teils auf Basis faktoranalytischer Betrachtungen erneut modifizierte, zuvor bereits sprachlich adaptierte und inhaltlich vereinfachte IPIP-Items sowie auf Davis (1989) zurückführbare Fragen zum TAM zum Einsatz. Weder für Extraversion noch für Gewissenhaftigkeit zeigten sich die erwarteten Zusammenhänge. Verträglichkeit und Offenheit nahmen ausschließlich Einfluss auf die Wahrnehmung der Benutzerfreundlichkeit, Neurotizismus auf die Einschätzung der empfundenen Nützlichkeit (Özbek et al., 2014).

Eine weitere Betrachtung des Zusammenwirkens der Big Five-Persönlichkeitseigenschaften und Konstrukte des UTAUT wurde schließlich von Barnett, Pearson, Pearson und Kel-

lermanns (2015) vorgenommen. Eine interaktionale Perspektive einnehmend, vermuteten diese Autoren ebenfalls eine Bedeutsamkeit der Persönlichkeitseigenschaften (als dispositionale Faktoren), die sich nicht nur in einer durch die übrigen TAM- und UTAUT-Konstrukte (als vornehmlich situative Einflussgrößen) vermittelten Form äußern, sondern auch in direkten Zusammenhängen sowohl mit der Handlungsabsicht, als auch dem Nutzungsverhalten sichtbar werden sollte. Sie prüften dementsprechend ein erweitertes UTAUT-Modell, das um diese Einflüsse der Big Five ergänzt wurde. Anhand einer studentischen Stichprobe ( $N = 382$ , 58% männlich, Alter:  $M = 22$  Jahre) wurden diese Annahmen wiederum am Nutzungsverhalten einer kursbegleitend eingesetzten Online-Plattform überprüft. Die Erfassung der TAM-Größen erfolgte mit etablierten Maßen (vgl. Venkatesh et al., 2003). Das FFM wurde ursprünglich mit 50 IPIP-Items erhoben, dann aber mithilfe faktorenanalytischer Verfahren auf insgesamt 30 Items reduziert. Für die Ermittlung des Nutzungsverhalten wurde sowohl auf eine Selbstauskunft, als auch auf objektive Maße aus den Logfiles der Online-Plattform zurückgegriffen (Häufigkeit der Zugriffe, E-Mail-Postfächer und Notenverzeichnisse). Zusammengefasst zeigte sich zunächst, dass die Grundannahmen des UTAUT weitgehend aufrechterhalten werden konnten. Lediglich die förderlichen Umgebungsbedingungen wiesen keinen direkten Zusammenhang mit dem Nutzungsverhalten auf. Für Neurotizismus und Gewissenhaftigkeit konnte erwartungsgemäß jeweils ein bedeutsamer, direkter Zusammenhang mit dem Nutzungsverhalten aufgezeigt werden, der jedoch nicht über die Handlungsabsicht vermittelt wurde. Offenheit wies keine Zusammenhänge mit dem Nutzungsverhalten auf. Im Falle der Extraversion entsprach die Richtung des beobachteten, bedeutsamen Zusammenhangs mit dem Nutzungsverhalten nicht den Erwartungen. Schließlich zeigte sich, dass Verträglichkeit gemäß den Erwartungen keinen Einfluss auf das Nutzungsverhalten aufwies. Für keine der Persönlichkeitseigenschaften wurde ein bedeutsamer Einfluss mit der Nutzungsintention sichtbar. Die diesbezüglichen Annahmen einer möglichen, ggf. auch nur partiellen Mediation des Zusammenhangs mit dem Nutzungsverhalten musste dementsprechend verworfen werden (Barnett et al., 2015).

Um die offensichtliche Variabilität der Befunde aus diesen unterschiedlichen Betrachtungen mit teils deutlich variierenden und modifizierten Modellannahmen erstmals einer systematischen Überprüfung zu unterziehen und dadurch ggf. verallgemeinernde Aussagen über die zeitliche und situative Stabilität der in diesen Studien berichteten Zusammenhänge zwischen den Big Five-Persönlichkeitseigenschaften und zentralen TAM-Konstrukten zu ermöglichen, wurden die bivariaten Korrelationskoeffizienten aus den vorliegenden Arbeiten extrahiert — sofern enthalten, siehe Barnett et al. (2015), Devaraj et al. (2008), Punnoose (2012), Rosen und Kluemper (2008), Svendsen et al. (2013) — und meta-analytisch aggregiert. Gemäß der zuvor geschilderten konzeptuellen Vergleichbarkeit zwischen den Konstrukten PEOU und PU aus dem TAM sowie EE und PE aus dem UTAUT wurden diese gemeinsam analysiert. Vor dem Hintergrund der Annahme, dass auch methodische und stichprobenbedingte Einflüsse zur beobachteten Variation zwischen den vorliegenden Koeffizienten beigetragen haben und eine verallgemeinernde Abschätzung der wahren Zusammenhänge über die gegebenen Studien hinaus möglich sein soll, wurden *random-effects* Modelle zur Schätzung der wahren Effektstärken verwendet (vgl. Viechtbauer, 2010). Die *Forest-Plots* in Abbildung 1.4 enthalten die kombinierten, Fisher-z-transformierten Korre-

lationskoeffizienten der berichteten Studien und die meta-analytischen Schätzer der wahren Effekte für die spezifizierten Zusammenhänge, sowie die jeweiligen 95%-Konfidenzintervalle.

Für den Großteil der geschätzten random-effects Modelle schließen die ermittelten 95%-Konfidenzintervalle um den berichteten Punktschätzwert des Effekts den Wert Null nicht ein. Dies suggeriert, dass es sich bei den meisten dieser Effekte zwar um relativ kleine, aber substantielle Zusammenhänge handeln könnte. Die Zusammenhänge zwischen den Persönlichkeitseigenschaften des Big Five und der empfundenen Leichtigkeit der Nutzung (PEOU) erweisen sich in dieser Interpretation allesamt als bedeutsam. In inhaltlich plausibler Weise scheint der wahre Korrelationskoeffizient für Neurotizismus negativ, für die anderen Persönlichkeitseigenschaften positiv auszufallen. Die jeweiligen Zusammenhänge von Neurotizismus und Offenheit mit der wahrgenommenen Nützlichkeit der Technologie (PU) erweisen sich nicht als bedeutsam, die drei verbleibenden Persönlichkeitsdimensionen scheinen jeweils positive Zusammenhänge mit PU aufzuweisen. Für Neurotizismus wurde ein bedeutsamer, negativer Zusammenhang mit der Nutzungsintention (BI) ermittelt, für Extraversion, Offenheit und Verträglichkeit jeweils positive. Einzig zwischen Gewissenhaftigkeit und BI scheint zeit- und situationsübergreifend kein ausgeprägter Zusammenhang zu bestehen.

Einschränkend muss allerdings beachtet werden, dass beim Großteil der geschätzten Modelle von einer beträchtlichen Heterogenität der zugrundeliegenden, wahren Effekte ausgegangen werden muss, d. h. dass offenbar eine hohe Variabilität zwischen den wahren Zusammenhängen vorliegt, die in den individuellen Studien ermittelt wurden: die inferenzstatistische Überprüfung der Annahme, dass in allen Studien der selbe Effekt Betrachtung fand, wurde mithilfe des  $Q$ -Tests auf Heterogenität vorgenommen (Higgins, Thompson, Deeks & Altman, 2003). Die Prüfgröße dieses Verfahrens (Cochran's  $Q$ ) ergibt sich aus der Summe der gewichteten Abweichungen der studienspezifischen Effektschätzer vom meta-analytischen Gesamtschätzwert und folgt einer  $\chi^2$ -Verteilung. Ein signifikantes Testergebnis spricht gegen die Annahme der erwähnten Nullhypothese. Einzig bei den Zusammenhängen von Neurotizismus und Benutzerfreundlichkeit, Neurotizismus und Nützlichkeit, Extraversion und Handlungsabsicht sowie Offenheit und Handlungsabsicht spricht die Überschreitungswahrscheinlichkeit der Prüfgröße für die Annahme von vergleichsweise homogenen Effekten. Allerdings ist auch in diesen Fällen das 95%-Konfidenzintervall des  $I^2$ -Kennwertes (prozentualer Anteil der Gesamtvariabilität zwischen den berücksichtigten Zusammenhängen, der auf Unterschiede zwischen den geschätzten wahren Effekten zurückgeht, vgl. Higgins & Thompson, 2002) sehr breit und die obere Grenze liegt meist bei einem Wert von etwa 95%. Die einzige Ausnahme stellt hierbei der Zusammenhang zwischen Offenheit und der Handlungsabsicht BI dar: der Punktschätzwert für  $I^2$  beträgt für diesen Effekt 0% und das obere Limit des 95%-Konfidenzintervalls liegt knapp unter 85%. Folglich kann wohl nur für diesen Zusammenhang angenommen werden, dass es sich um einen zwar kleinen, aber relativ homogenen, direkt über die Studien vergleichbaren und bedeutsamen Effekt handelt.

Es ist jedoch festzuhalten, dass weder das Vorliegen einer derartigen Heterogenität zwischen den Effekten, noch Intervallschätzer der Effekte in Form von Konfidenzintervallen, die den Wert Null einschließen, einer weiterführenden, spezifischen Betrachtung auf Ebene



der Persönlichkeitsfacetten entgegen sprechen. Ganz im Gegenteil könnte die Einnahme einer entsprechend differenzierten Perspektive zusätzliche Aufschlüsse über die grundlegende Natur der hier beschriebenen Zusammenhänge liefern.

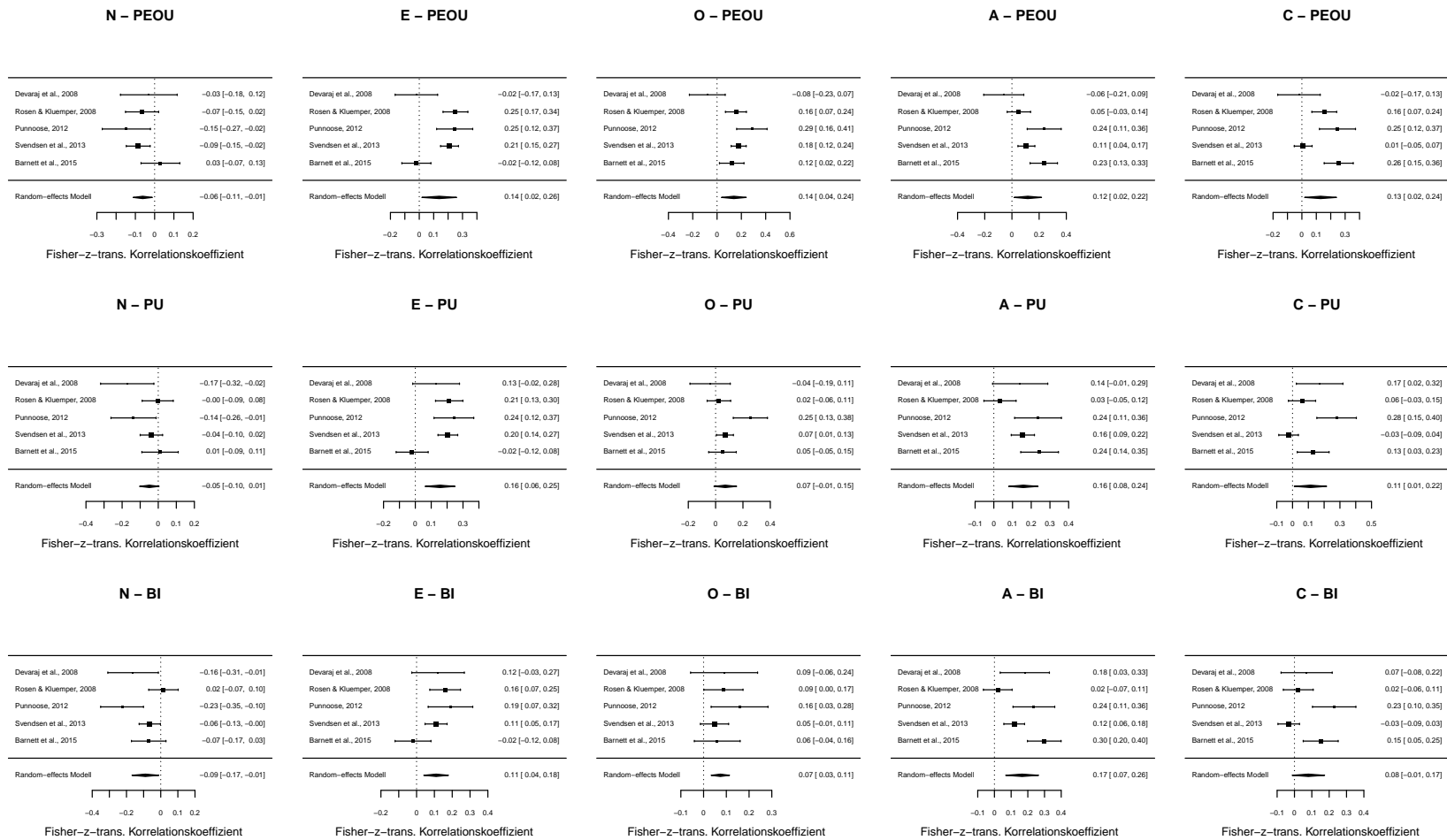


Abbildung 1.4: *Forest-Plots* der meta-analytisch aggregierten Zusammenhänge zwischen den Persönlichkeitsbereichen des FFM und den zentralen Konstrukten des TAM. Abgebildet sind die Fisher-z-transformierten Korrelationskoeffizienten für die Effekte der berichteten Studien und meta-analytische Schätzer der *random-effects* Modelle sowie die jeweiligen 95%-Konfidenzintervalle.

### 1.1.4 Technikbereitschaft

Einen eigenständigen Ansatz, das primär einstellungsorientierte TAM um eine persönlichkeitspsychologische Perspektive zu erweitern, stellten Neyer, Felber und Gebhardt (2012) vor. Sie postulierten, auch unter Verweis auf neuere Entwicklungen der Theorie des geplanten Handelns (Ajzen, 1991), dass — neben den im TAM berücksichtigten Einstellungen, d. h. den Faktoren der Technikakzeptanz — für einen erfolgreichen, persönlich gewinnbringenden Umgang mit technischen Entwicklungen zusätzliche Überzeugungen des kompetenten Umgangs und der Kontrolle ausschlaggebend sein sollten. Vor dem Hintergrund dieser Überlegungen entwickelten sie ein integratives Modell der Technikbereitschaft. Gemäß diesen Modellannahmen wird Technikbereitschaft sowohl durch Technikakzeptanz, als auch durch Technikkompetenz- und Technikkontrollüberzeugungen definiert. In Anlehnung an Davis (1989) wurde Technikakzeptanz hierbei als Einstellungsmerkmal betrachtet, das die subjektive Bewertung des technologischen Fortschritts explizit repräsentiert. Als solches sollte es sich primär im persönlichen Bezug zu modernen Technologien und dem individuellen Interesse an technischen Innovationen manifestieren (in Abgrenzung zu einer Bewertung der Technik in Hinblick auf ihre gesamtgesellschaftliche Bedeutung). Daneben wurden Technikkompetenzüberzeugungen mit Bezug auf Krampen (1991) als die subjektive Einschätzung der eigenen Handlungsmöglichkeiten in technikrelevanten Situationen verstanden. Sie sollten das auf Basis kumulierter Erfahrungen im Umgang mit bekannten Technologien sowie der angenommenen Anpassungsfähigkeit an technologische Innovationen gebildete Selbstkonzept diesbezüglicher Fähigkeiten abbilden. Als individuelle Kontingenzerwartungen hinsichtlich der Resultate technikrelevanter Handlungen wurden schließlich Technikkontrollüberzeugungen definiert. Diese sollten das Ausmaß der wahrgenommenen Einflussenerwartungen über technische Prozesse und ihre direkten Folgen in der individuellen Umgebung zum Ausdruck bringen (Neyer et al., 2012).

Es wurde weiterhin das Ziel verfolgt, eine in Forschung und Praxis nutzbare Kurzsкала zur Erfassung dieser Konstrukte zu entwickeln. Hierzu wurde eine umfassende Literaturrecherche vorgenommen und auf etablierte Verfahren zur allgemeinen Erhebung von Einstellungen, Selbstwirksamkeits- und Kontrollüberzeugungen sowie spezifischere Instrumente zurückgegriffen, die den Zugang zu und Umgang mit modernen Technologien erfassen sollen. Auf dieser Basis wurde ein Fragenpool mit 80 Items gebildet, welche als Indikatoren der beschriebenen Konstrukte angesehen wurden. Eine Itemanalyse an den Daten einer umfangreichen Online-Erhebung ( $N = 639$ , Alter:  $M = 36$  Jahre) führte zu einer Vorauswahl von 29 Items, die nachfolgend einer Stichprobe im höheren Erwachsenenalter vorgelegt wurden ( $N = 134$ , Alter:  $M = 68$  Jahre). Reliabilitätsschätzungen und insbesondere eine explorative Faktorenanalyse (an der Gesamtstichprobe) führten zu einer endgültigen Auswahl von jeweils vier Items mit den höchsten Faktorladungen auf den beschriebenen Facetten der Technikbereitschaft. Inhaltlich wurden diese Konstrukte als spezifische Persönlichkeitsmerkmale angesehen und sollten u. a. substantielle Zusammenhänge mit den grundlegenden Persönlichkeitseigenschaften des FFM aufweisen.

In drei weiterführenden Untersuchungen wurde deshalb eine erste Validierung dieses neuen Verfahrens vorgenommen. Hierzu wurden eine Gruppe junger ( $N_1 = 513$ , 52% männ-

lich, Alter:  $M = 25$  Jahre) und zwei weitere Stichproben älterer Erwachsener ( $N_2 = 150$ , 35% männlich, Alter:  $M = 67$  Jahre;  $N_3 = 162$ , 65% männlich, Alter:  $M = 65$  Jahre) rekrutiert. In diesen Erhebungen wurden außerdem auch Annahmen über mögliche Zusammenhänge zwischen den Konstrukten der Technikbereitschaft und dem tatsächlichen Nutzungsverhalten technologischer Entwicklungen überprüft. Letzteres wurde jeweils über eine einmalige Selbstauskunft zum Nutzungsverhalten unterschiedlicher, vorgegebener technischer Geräte über einen mehrwöchigen Zeitraum erfasst. Für die Technikbereitschaft zeigten sich mittlere Zusammenhänge mit dem berichteten Nutzungsverhalten ( $r_1 = .31$ ,  $r_2 = .40$ ). Hieraus wurde abgeleitet, mithilfe von Technikbereitschaft einen erfolgreichen Umgang mit neuen Technologien vorhersagen zu können, insbesondere im höheren Lebensalter. Neurotizismus und Offenheit wurden in Anlehnung an Devaraj et al. (2008) außerdem als relevante Persönlichkeitseigenschaften der Big Five identifiziert, die Zusammenhänge mit den Konstrukten der Technikbereitschaft aufweisen sollten. Diese Dimensionen der Persönlichkeit wurden mit einer (in Studie 1 gekürzten) Fassung der deutschen Version des Big Five Inventory (BFI) von F. R. Lang, Lüdtke und Asendorpf (2001) erfasst. Neurotizismus wies erwartungsgemäß in beiden Studien einen negativen Zusammenhang mit Technikbereitschaft auf ( $r_1 = -.21$ ,  $r_2 = -.23$ ), bei Offenheit zeigte sich ein positiver Zusammenhang, der allerdings nur in der ersten Untersuchung signifikant von Null abwich ( $r_1 = .29$ ,  $r_2 = .11$ ). Diese, wie auch die zuvor berichteten Befunde führten zur Ableitung der nachfolgend berichteten Fragestellungen.

### 1.1.5 Forschungsfragen und Annahmen

Vor dem Hintergrund der hier wiedergegebenen Vorbefunde lässt sich festhalten, dass die systematische Betrachtung spezifischer, stabiler Persönlichkeitseigenschaften und ihrer Rolle im Zusammenspiel mit etablierten Konstrukten der Technologieakzeptanz und -nutzung in mindestens zweierlei Hinsicht aussichtsreich erscheint. Einerseits sollte die differenzierte Betrachtung der Persönlichkeit auf Facettenebene erstmals eine präzisierende Auskunft über die Relevanz spezifischer Persönlichkeitseigenschaften geben. Gleichzeitig gibt die in den Vorbefunden zu beobachtende Variabilität der Zusammenhänge Hinweise auf das vermutliche Wirken zeit-, situations- und technikspezifischer Einflussfaktoren. Die hier berichtete Untersuchung nähert sich der Thematik im Rahmen des eingangs beschriebenen Kontextes deshalb teils über einen explorativen Zugang, um auch diesen zusätzlichen Einflussgrößen angemessen gerecht zu werden. Dadurch soll ein Beitrag zur verbesserten Einordnung dieser Faktoren im Zusammenspiel mit weitgehend invarianten Einflüssen geleistet werden. Es sollen dazu folgende Forschungsfragen und Annahmen geprüft werden:

#### Forschungsfrage 1

Eignen sich die etablierten Modelle der Technologieakzeptanz zur Vorhersage der Nutzungsintention bzw. des Nutzungsverhaltens von statistischer Anwendungssoftware bei Studierenden der Psychologie bzw. Schulpsychologie? Da der Einsatz technologischer Entwicklungen im universitären Umfeld weder strengen unternehmerischen Vorgaben folgt, noch vollständig mit der im Verbrauchermarkt gegebenen Entschei-

dungsfreiheit vergleichbar ist, wird eine Gegenüberstellung zwischen dem etablierten TAM sowie dem UTAUT in seiner ursprünglichen wie auch überarbeiteten und speziell für den Konsumentenbereich ergänzten Form (UTAUT2) vorgenommen.

Weiterhin soll ein möglicher Einfluss grundlegender, zeitlich weitgehend unveränderlicher Persönlichkeitsmerkmale untersucht werden.

#### **Annahme 1a**

Es können Zusammenhänge zwischen grundlegenden, spezifischen Persönlichkeitsmerkmalen und den zentralen TAM- bzw. UTAUT-Konstrukten identifiziert werden. Diese werden insbesondere zwischen der Persönlichkeitsdimension Offenheit bzw. ihren Facetten und der Nutzungsintention (BI) erwartet.

#### **Annahme 1b**

Unterschiede in der Nutzungsintention und dem Nutzungsverhalten lassen sich in Teilen durch die ermittelten Zusammenhänge mit Persönlichkeitseigenschaften erklären, so dass eine entsprechende Erweiterung der betrachteten Modelle zu einer bedeutsamen Verbesserung in der Vorhersage der Nutzungsintention, bzw. des Nutzungsverhaltens führt.

In einer zusätzlichen Betrachtung soll das integrative Modell der Technikbereitschaft (Neyer et al., 2012) Beachtung finden:

#### **Forschungsfrage 2**

Stellt dieses kürzlich vorgestellte Erklärungsmodell im vorliegenden Kontext eine valide Alternative zur Vorhersage des Nutzungsverhaltens dar?

#### **Annahme 2**

Es können Zusammenhänge zwischen grundlegenden, spezifischen Persönlichkeitsmerkmalen und der Technikbereitschaft bzw. ihren Facetten Technikakzeptanz, Technikkompetenz- und Technikkontrollüberzeugungen identifiziert werden. Diese werden vornehmlich für die Dimensionen Offenheit (positiv) und Neurotizismus (negativ, bzw. für die Emotionale Stabilität des BFSI ebenfalls positiv) sowie deren jeweilige Persönlichkeitsfacetten erwartet.

Im Bestreben einer initialen Beantwortung dieser Fragestellungen und der Überprüfung dieser Annahmen wurde eine umfangreiche Erhebung an den Besuchern einer Statistik-Vorlesung im ersten Studiensemester durchgeführt. Inhaltlicher Bestandteil dieser Veranstaltung bzw. ihrer Begleitseminare war u. a. die Vermittlung einer statistischen Anwendungssoftware, die den Studierenden in den ersten Vorlesungswochen einführend vorgestellt wurde. Einige Wochen nach Vorlesungsbeginn wurden mit den Teilnehmenden der Studie in kontrollierter Laborumgebung sowohl Verfahren zur Persönlichkeitsdiagnostik, als auch Befragungen (Fragebögen zur Erfassung der eingangs vorgestellten Aspekte der

Technologieakzeptanz und Technikbereitschaft) durchgeführt.<sup>2</sup> Im Anschluss an diesen Erhebungstermin wurde das individuelle Nutzungsverhalten der Statistiksoftware über den Verlauf von 20 Tagen mithilfe einer täglich online abzugebenden Selbstauskunft erfasst. Eine detaillierte Beschreibung dieses Vorgehens folgt auf den nächsten Seiten.

---

<sup>2</sup>Neben den hier angeführten und nachfolgend detaillierter beschriebenen Verfahren kamen weiterhin ein Instrument zur Intelligenzdiagnostik, Fragebögen zur Computernutzung, der Einstellung zu und Vertrautheit mit Computern sowie der allgemeinen Techniknutzung zum Einsatz (s. Abschnitt 1.2.3). Die derart erhobenen Maße finden in der vorliegenden Arbeit jedoch keine weitere Berücksichtigung.

## 1.2 Methode

Die nachfolgenden Kapitel enthalten zunächst eine Beschreibung des Ablaufs der vorliegenden Untersuchung (1.2.1), der erhobenen Stichprobe (1.2.2) und eine Vorstellung der verwendeten Messinstrumente (1.2.3). Schließlich erfolgt in Abschnitt 1.2.4 die Schilderung des Vorgehens bei der vorgenommenen Datenauswertung.

### 1.2.1 Erhebungsablauf

Die hier beschriebene Erhebung erfolgte an den Teilnehmerinnen und Teilnehmern der Vorlesung *Statistik I* im ersten Fachsemester des Bachelorstudiengangs Psychologie und des Studiengangs Schulpsychologie (im Rahmen eines Lehramtsstudiums) an der Ludwig-Maximilians-Universität München im Wintersemester 2013/14. Zu Semesterbeginn war ein wesentlicher Inhalt dieser Veranstaltung bzw. der parallel stattfindenden Begleitseminare die Vorstellung und Einführung der statistischen Anwendungssoftware *IBM SPSS Statistics* für deskriptiv- und inferenzstatistische Analysen. Die Studierenden waren mit diesem Programmpaket zuvor nicht vertraut, es bot für sie dementsprechend eine neuartige Erfahrung mit einer bis dato unbekanntem technologischen Entwicklung. Nach dem Besuch der einführenden Sitzungen wurde die hier beschriebene Untersuchung ausschließlich im Rahmen der genannten Vorlesung als Studie zu interindividuellen Unterschieden im Kontext der Mensch-Computer-Interaktion beworben. In einem Kurzvortrag wurden den prospektiven Studienteilnehmenden dazu die beiden Phasen der Erhebung vorgestellt. Alle Probandinnen und Probanden nahmen demnach zunächst an einem einmaligen Erhebungstermin in einem Laborraum der Universität teil. In diesen etwa zweistündigen Sitzungen wurden die nachfolgend beschriebenen Verfahren in Kleingruppen von maximal 7 Personen an individuellen Arbeitsplätzen computergestützt durchgeführt. Diese Untersuchungstermine fanden an Werktagen im Zeitraum von 10.00 Uhr bis ca. 18.00 Uhr statt. Im Anschluss an eine mündliche Begrüßung und Einführung unterzeichneten die Teilnehmenden eine schriftliche Einverständniserklärung. Die Erhebung erfolgte softwaregestützt mithilfe des *Wiener Testsystems* (SCHUHFRIED GmbH, 2011) sowie der Online-Plattform *SoSci Survey* (Leiner, 2013; <https://www.sosicurvey.de/>). Letztere wurde ebenfalls für die zweite, für alle Beteiligten direkt an diesen Erhebungstermin anschließende Erhebungsphase genutzt. Über einen Zeitraum von 20 Tagen sollten die Untersuchungsteilnehmenden demnach täglich ihr individuelles Nutzungsverhalten von Statistiksoftware übermitteln.

Die ersten Probandinnen und Probanden begannen ihren Erhebungszeitraum (mit der Teilnahme an der anfänglichen Erhebung unter kontrollierten Bedingungen) Mitte Oktober — etwa einen Monat nach Semesterbeginn und dem entsprechenden Erstkontakt mit der relevanten Anwendungssoftware. Die Möglichkeit zur Untersuchungsteilnahme wurde über mehrere Wochen aufrechterhalten, der letzte Erhebungstermin in den Laborräumen der Universität fand Anfang Dezember statt. Die Datenerfassung des Nutzungsverhaltens aller Teilnehmenden war dementsprechend nach 20-tägiger Erhebungsperiode auch für die letzten Teilnehmerinnen und Teilnehmer bis zum Ende des Jahres 2013 abgeschlossen.

### 1.2.2 Stichprobenzusammensetzung

Insgesamt nahmen  $N = 102$  Studentinnen und Studenten an der ersten Untersuchungsphase, d. h. dem einmaligen Erhebungstermin in einem Laborraum der Universität teil.  $N = 96$  Teilnehmende der Untersuchung schlossen auch den daran anschließenden, 20-tägigen Erhebungszeitraum vollständig ab (Drop-out-Rate  $< 6\%$ ).

Es handelte sich um  $n = 81$  Frauen (entspricht ca. 84%) und  $n = 15$  Männer (ca. 16%) mit einem durchschnittlichen Alter von  $M = 22.03$  Jahren ( $SD = 6.70$ ). Etwas mehr als die Hälfte dieser Teilnehmerinnen und Teilnehmer waren im Bachelorstudiengang Psychologie eingeschrieben ( $n = 57$ ), alle verbleibenden in einem Studiengang der Psychologie mit schulpyschologischem Schwerpunkt (vereinzelt als Erweiterungsfach). Knapp 94% waren Studierende im ersten Fachsemester. Die vorausgegangene, (teils angeleitete, teils eigenständige) aktive Nutzung der statistischen Anwendungssoftware SPSS betrug zum Zeitpunkt der initialen Erhebung laut Selbstauskunft im Mittel  $M = 4.95$  Wochen ( $SD = 2.27$ ). Drei Personen gaben zu diesem Erhebungszeitpunkt an, noch kein *IBM SPSS Statistics* zu nutzen. Der längste, selbst berichtete Nutzungszeitraum betrug 12 Wochen.

Als Anreiz für die Teilnahme an der Erhebung wurden den Studierenden nach vollständiger Bearbeitung Versuchspersonenstunden (als anteilige Studienleistung) in Aussicht gestellt und gutgeschrieben. Optional wurde außerdem die Möglichkeit angeboten, Rückmeldung zu den bearbeiteten Persönlichkeits- und Leistungstests zu erhalten sowie an einer Verlosung von insgesamt drei Einkaufsgutscheinen teilzunehmen.

### 1.2.3 Verwendete Instrumente

Die nachfolgend beschriebenen Verfahren wurden beim initialen Erhebungstermin in einem Laborraum der Universität in der hier angegebenen Reihenfolge computergestützt vorgegeben.

#### Persönlichkeit

Nach der Bearbeitung eines Testverfahrens zur Erfassung der allgemeinen, fluiden Intelligenz (über die Subtests numerisch-induktives, figural-induktives und verbal-deduktives Denken der Intelligenz-Struktur-Batterie (INSBAT); Arendasy et al., 2010), das in der vorliegenden Arbeit keine weitere Berücksichtigung findet und nach einer anschließenden, ca. zweiminütigen Pause wurden die Ausprägungen der Untersuchungsteilnehmenden in den Persönlichkeitseigenschaften des Big Five-Modells erhoben.

Hierzu kam das BFSI (Arendasy et al., 2009) zum Einsatz. Es handelt sich um ein etabliertes, computergestütztes Testverfahren, das bei vergleichsweise geringer Durchführungsdauer die Erfassung der fünf grundlegenden Persönlichkeitsdimensionen sowie jeweils sechs untergeordnete -facetten bei hoher psychometrischer Qualität über Personenparameter nach dem *Partial Credit Modell* (Masters, 1982) erlaubt (vgl. Tabelle 1.1). Dadurch soll auch eine eng umgrenzte Beschreibung grundlegender, spezifischer Persönlichkeitsaspekte ermöglicht werden. Bei der Durchführung dieses Verfahrens hatten die Probandinnen und



Probanden insgesamt 300 Adjektive bzw. Kurzaussagen auf einer vierstufigen Likert-Skala wie folgt zu bewerten: „untypisch für mich“, „eher untypisch für mich“, „eher typisch für mich“ oder „typisch für mich“. Die gesamte Bearbeitungszeit dieser im Wiener Testsystem durchgeführten Testbatterie betrug etwa  $M = 78$  Minuten ( $Min = 40$ ;  $Max = 176$ ), wobei die Erhebung der Persönlichkeitseigenschaften durchschnittlich etwa  $M = 15$  Minuten beanspruchte ( $Min = 8$ ;  $Max = 30$ ).

### Technikbereitschaft

Nach einer erneuten kurzen Pause erfolgte die Vorgabe aller weiteren Erhebungsinstrumente über die zuvor bereits erwähnte Online-Plattform *SoSci Survey*. Einführend wurden zunächst erneut die Modalitäten der Untersuchungsteilnahme und das Prozedere der noch bevorstehenden, wiederholten Erhebungen in der zweiten Erhebungsphase präsentiert. Danach erfolgte eine Abfrage demographischer Daten. Für die im Weiteren erfassten Inhalte wurde gemäß gängiger Praxis nach Möglichkeit auf bereits veröffentlichte, etablierte Verfahren zurückgegriffen. Dies beinhaltete eine Erfassung des allgemeinen Techniknutzungsverhaltens, der Computernutzung im Speziellen (*Computer Usage Questionnaire (CUQ)*; Schroeders & Wilhelm, 2011) sowie den Einstellungen zu und der Vertrautheit mit Computern (*Computer Aversion, Attitudes, and Familiarity Index (CAAFI)*; Schulenberg & Melton, 2008), welche allesamt in der vorliegenden Arbeit jedoch ebenfalls keine Berücksichtigung finden. Daran anschließend erfolgte schließlich die Abfrage der Technikbereitschaft mithilfe der von Neyer et al. (2012) veröffentlichten Kurzsкала. Die einzelnen Aussagen des Verfahrens wurden von den Teilnehmenden auf jeweils 5-stufigen Likert-Skalen („stimmt gar nicht“, „stimmt wenig“, „stimmt teilweise“, „stimmt ziemlich“ und „stimmt völlig“) hinsichtlich ihres Zutreffens bewertet. Die Vorgabe der einzelnen Fragen erfolgte für dieses, wie alle weiteren berichteten Verfahren, in randomisierter Abfolge.

### Technologieakzeptanz

Die Technologieakzeptanz wurde in zweierlei Ausprägungen erfasst: zunächst wurde auf die von Venkatesh et al. (2012) im Rahmen der Postulierung des UTAUT2 verwendete Skala zurückgegriffen. Hierdurch wurde eine umfassende Betrachtung sowohl des ursprünglichen UTAUT-Modells sowie der Erweiterungen des UTAUT2 ermöglicht. Die entsprechenden Items wurden im vorliegenden Kontext für das Nutzungsverhalten von *IBM SPSS Statistics* adaptiert und waren von den Teilnehmerinnen und Teilnehmern auf 7-stufigen Likert-Skalen (von „trifft gar nicht zu“ über „trifft teils-teils zu“ bis „trifft völlig zu“) zu bewerten. Bei diesem und allen anderen im Original auf Englisch verfügbaren Instrumenten übersetzte der Autor der vorliegenden Arbeit das Ausgangsmaterial zunächst ins Deutsche, bevor es von einer staatlich geprüften Übersetzerin ohne Kenntnis der ursprünglichen Vorlagen wieder ins Englische zurück übersetzt wurde. Im Großteil der Fälle stimmten diese Rückübersetzungen mit den Originalversionen der entsprechenden Aussagen überein. Bei einzelnen Abweichungen erfolgte in Absprache der beiden Beteiligten eine klärende Anpassung der deutschen Übersetzung.

Zusätzlich wurde eine Erhebung der TAM-Kernkonstrukte PEOU, PU und BI mithilfe der von Venkatesh und Bala (2008) publizierten, auf Davis (1989) zurückgehenden Skalen vorgenommen. Die entsprechenden Aussagen waren auf den gleichen 7-stufigen Likert-Skalen wie die UTAUT-Aussagen zu bewerten.

### Nutzungsverhalten

Die Probandinnen und Probanden wurden abschließend dazu aufgefordert, ihr Nutzungsverhalten der Statistiksoftware über die anschließenden 20 Tage regelmäßig und gegen Ende des jeweiligen Tages zu berichten (eine Bearbeitung war jedoch auch nachträglich möglich, was kein Ausschlusskriterium darstellte). Zu diesem Zweck nutzten die Teilnehmenden ihre individuell eindeutigen, aber anonymen Zugangsdaten, mit denen sie sich weiterhin auf der Online-Plattform anmelden konnten, um tagesaktuell jeweils maximal zwei Fragen zu beantworten. Primär wurde erfasst, ob im Laufe des vergangenen Tages eine statistische Anwendungssoftware genutzt wurde. Die entsprechende Fragestellung lautete: „Welche Statistik-Software haben Sie heute verwendet?“ mit den Mehrfachwahlmöglichkeiten „SPSS“, „R“, „andere Statistik-Software“ und „keine“ (wurde die Nutzung mindestens einer Statistik-Software vermerkt, erfolgte in einem zweiten Schritt die Abfrage des jeweiligen Nutzungsumfangs: „Wie viel Zeit haben Sie heute mit ... verbracht? (Angabe in Stunden)“, wobei die Frage je nach vorausgegangener Angabe wiederholt und der entsprechende Programmname eingefügt wurde; das Eingabeformat erlaubte zwei Dezimalstellen). Die nachfolgenden Analysen beruhen auf der berichteten Häufigkeit der Nutzung von *IBM SPSS Statistics* während des 20-tägigen Erhebungszeitraums.

### 1.2.4 Auswertungsprozess und technische Umsetzung

Die im Folgenden beschriebenen Schritte der Datenaufbereitung und -analyse wurden mit der statistischen Programmiersprache *GNU R* (R Core Team, 2016) realisiert. Zunächst erfolgte eine deskriptiv-statistische Betrachtung zentraler Variablen der berichteten Erhebung sowie die psychometrische Überprüfung der verwendeten, adaptierten Befragungsinstrumente mithilfe von Item- und Reliabilitätsanalysen. Hierzu fanden — neben der R-Basisinstallation — die Pakete *psych* (Revelle, 2016) und *car* (Fox & Weisberg, 2016) Verwendung. Anschließend wurden die zuvor diskutierten Modelle der Technologieakzeptanz und Technikbereitschaft mithilfe von Strukturgleichungsmodellen auf ihre Passung im bestehenden Kontext überprüft (vgl. Forschungsfragen 1 und 2). Die Schätzung der Modellparameter erfolgte nach Ausschluss vereinzelter Fälle mit fehlenden Werten mithilfe der Maximum-Likelihood-Methode (mit robuster Schätzung der Standardfehler; MLMR). Hierzu, bzw. für die grafische Darstellung dieser Modelle wurden die Pakete *lavaan* (Rosseel, 2015) und *semPlot* (Epskamp, 2014) verwendet. Weiterhin wurden in einer erstmaligen Betrachtung die Analyse möglicher Zusammenhänge grundlegender und zeitlich stabiler, spezifischer Persönlichkeitseigenschaften mit den zentralen Konstrukten des TAM, UTAUT (vgl. Annahme 1a) und der Technikbereitschaft (vgl. Annahme 2) vorgenommen. Die diesbezüglich erfolgte und bereits berichtete meta-analytische Aggregation erfolgte mithilfe des

---

Paketes `metafor` (Viechtbauer, 2015). Außerdem wurden gemäß Annahme 1b die Möglichkeiten angemessener Erweiterungen der zuvor berichteten und empirisch geprüften Modelle um differenzierte Persönlichkeitseigenschaften evaluiert. Alle Tabellen und zusätzlichen Grafiken wurden mit den Paketen `Hmisc` (Harrell, 2016), `xtable` (Dahl, 2016) und `ggplot2` (Wickham & Chang, 2016) erstellt.

## 1.3 Ergebnisse

Nachfolgend erfolgt in Abschnitt 1.3.1 zunächst die Präsentation der Resultate deskriptiv-statistischer Vorbetrachtungen. Daran anschließend werden die besprochenen Modelle in ihrer Passung sowie in Hinblick auf etwaige, in beobachteten Zusammenhängen mit Persönlichkeitseigenschaften begründeten Erweiterungen, überprüft — TAM, UTAUT und UTAUT2 in den Kapiteln 1.3.2, 1.3.3 und 1.3.4, Technikbereitschaft in Abschnitt 1.3.5.

### 1.3.1 Initiale Analysen

In Tabelle 1.2 sind Mittelwert, Standardabweichung, Minimum, Maximum, Schiefe und Kurtosis aller Persönlichkeitsdimensionen und -facetten sowie der erfassten Nutzungshäufigkeit enthalten (letztere wird in allen nachfolgenden Tabellen und Abbildungen mit *U* (von engl. *use*) abgekürzt).<sup>3</sup>

Eine Zusammenfassung deskriptiv-statistischer sowie itemanalytischer Kennwerte für alle von den Teilnehmenden beurteilten Aussagen der betrachteten Technologieakzeptanz-Modelle und der Technikbereitschaft finden sich in den Tabellen 1.3 (TAM), 1.4 (UTAUT, UTAUT2) und 1.5 (Technikbereitschaft).

---

<sup>3</sup>Ein relativ geringer Mittelwert und ein positiver Wert der Schiefe sind Indikatoren einer tendenziell linkssteilen Verteilung der von den Untersuchungsteilnehmerinnen und -teilnehmern berichteten Nutzungshäufigkeit. Um etwaige, durch Abweichungen von der Normalverteilung bedingte Verzerrungen bei der Interpretation der nachfolgend berichteten Analysen ausschließen zu können, wurden die entsprechenden Betrachtungen zusätzlich jeweils mit einer transformierten Variante dieser Variable durchgeführt und überprüft (für eine detaillierte Erläuterung des diesbezüglichen Vorgehens s. »*Yeo-Johnson-Transformation*« in Abschnitt 2.2.3). Diese Transformation führte allerdings zu keinen wesentlichen Veränderungen in den Interpretationen der resultierenden Ergebnisse.

Tabelle 1.2: Deskriptiv-statistische Kennwerte der Nutzungshäufigkeit und Persönlichkeitseigenschaften.

Variable	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	<i>Schiefe</i>	<i>Kurtosis</i>
U: Nutzungshäufigkeit	3.14	3.01	0.00	14.00	1.43	1.89
N: Emotionale Stabilität	-0.22	0.71	-3.27	1.74	-0.76	2.94
E: Extraversion	0.04	0.75	-1.84	2.11	0.21	0.24
O: Offenheit	0.14	0.77	-1.50	2.24	0.60	0.25
A: Verträglichkeit	-0.01	0.83	-1.79	2.28	0.18	-0.24
C: Gewissenhaftigkeit	-0.08	0.77	-1.73	2.29	0.29	0.31
N1: Unbekümmertheit	-0.33	1.20	-3.70	3.24	-0.25	0.30
N2: Gelassenheit	0.35	1.03	-2.61	2.85	-0.11	0.38
N3: Positive Grundstimmung	0.85	1.43	-5.78	4.39	-1.00	3.39
N4: Soziale Selbstsicherheit	0.56	1.25	-3.12	5.06	0.20	1.49
N5: Selbstbeherrschung	0.66	1.18	-3.10	3.36	-0.21	0.74
N6: Emotionale Robustheit	0.06	1.49	-5.62	4.32	-0.53	1.64
E1: Freundlichkeit	1.57	1.27	-2.13	5.41	-0.08	0.46
E2: Geselligkeit	1.77	1.95	-3.02	5.64	0.18	-0.33
E3: Durchsetzungsfähigkeit	0.40	1.52	-2.95	5.61	0.50	0.82
E4: Tatendrang	1.23	1.55	-2.61	5.94	0.41	1.05
E5: Abenteuerlust	0.30	1.43	-3.25	5.27	0.43	1.22
E6: Fröhlichkeit	1.97	1.78	-2.90	6.09	0.09	0.12
O1: Fantasie	1.47	1.47	-2.90	5.33	0.50	0.70
O2: Ästhetik	1.04	1.29	-2.05	4.61	0.48	0.44
O3: Gefühle	3.03	2.09	-1.71	6.04	-0.23	-0.73
O4: Handlungen	1.48	1.41	-2.18	5.42	0.37	0.49
O5: Ideen	1.76	1.45	-0.85	5.51	0.70	0.05
O6: Werte- und Normensystem	1.06	0.93	-0.78	3.68	0.54	0.05
A1: Vertrauensbereitschaft	0.25	1.48	-3.53	5.42	0.14	1.02
A2: Aufrichtigkeit	1.06	0.99	-1.20	3.58	0.34	-0.45
A3: Hilfsbereitschaft	1.82	1.60	-1.39	6.04	0.41	-0.18
A4: Entgegenkommen	1.19	1.49	-2.63	5.55	0.35	0.07
A5: Bescheidenheit	0.72	1.04	-1.47	3.91	0.38	0.31
A6: Gutmütigkeit	2.23	1.86	-2.06	6.40	0.14	-0.22
C1: Kompetenz	0.56	1.14	-2.10	5.67	0.98	2.89
C2: Ordnungsliebe	0.97	1.57	-2.52	4.42	-0.04	-0.40
C3: Pflichtbewusstsein	2.12	1.46	-1.21	5.50	0.43	0.01
C4: Ehrgeiz	2.02	1.79	-2.13	5.86	0.22	-0.01
C5: Disziplin	1.46	1.43	-1.35	5.75	0.19	-0.39
C6: Bedachtsamkeit	1.55	1.41	-1.55	5.75	0.67	0.82

*Anmerkung.* Häufigkeit der Nutzung von *IBM SPSS Statistics* während des 20-tägigen Erhebungszeitraums gemäß Selbstauskunft sowie mit dem BFSI auf Basis des *Partial Credit Modells* ermittelte Personenparameter der Big Five Persönlichkeitsbereiche und -facetten.

Tabelle 1.3: Deskriptiv-statistische und itemanalytische Kennwerte der TAM-Variablen.

Konstrukt	Item	$M$	$SD$	$r_{it}$	$\alpha$	$\omega$
PU	PU1: Die Verwendung von SPSS verbessert meine Leistung bei meiner Tätigkeit.	4.21	1.63	0.77	0.91	0.92
	PU2: Die Verwendung von SPSS bei meiner Tätigkeit erhöht meine Produktivität.	4.06	1.74	0.85		
	PU3: Die Verwendung von SPSS verbessert meine Effektivität bei meiner Tätigkeit.	4.42	1.55	0.87		
	PU4: Ich finde SPSS nützlich für meine Tätigkeit.	4.61	1.53	0.73		
PEOU	PEOU1: Meine Interaktion mit SPSS ist klar und verständlich.	3.40	1.64	0.69	0.83	0.83
	PEOU2: Die Interaktion mit SPSS erfordert keine große mentale Anstrengung.	3.94	1.59	0.61		
	PEOU3: Ich halte SPSS für einfach in der Anwendung.	3.94	1.65	0.71		
	PEOU4: Meiner Ansicht nach lässt sich SPSS leicht dazu bringen, das zu tun, was ich möchte.	3.96	1.61	0.63		
BI	BI1: Vorausgesetzt ich habe weiter Zugang zu SPSS, habe ich vor es zu nutzen.	4.98	1.92	0.75	0.86	0.86
	BI2: Da ich Zugang zu SPSS hatte, nehme ich an, dass ich es auch nutzen werde.	5.16	1.63	0.75		
	BI3: Ich plane, SPSS in den nächsten Monaten zu nutzen.	5.91	1.44	0.72		

*Anmerkungen.* PU = Empfundene Nützlichkeit; PEOU = Empfundener Bedienkomfort; BI = Nutzungsintention;  $r_{it}$  = „part-whole“-korrigierte Trennschärfe;  $\alpha$  = Cronbachs  $\alpha$ ;  $\omega$  = McDonalds  $\omega$ .

Tabelle 1.4: Deskriptiv-statistische und itemanalytische Kennwerte der UTAUT- und UTAUT2-Variablen.

Konstrukt	Item	<i>M</i>	<i>SD</i>	<i>r<sub>it</sub></i>	$\alpha$	$\omega$
PE	PE1: Ich finde SPSS nützlich für meine Tätigkeit.	4.61	1.53	0.65	0.82	0.83
	PE3: Die Verwendung von SPSS hilft mir, Aufgaben schneller zu erledigen.	4.25	1.67	0.73		
	PE4: Die Verwendung von SPSS erhöht meine Produktivität.	3.30	1.77	0.65		
EE	EE1: Das Erlernen von SPSS fällt mir leicht.	3.92	1.62	0.81	0.89	0.89
	EE2: Meine Interaktion mit SPSS ist klar und verständlich.	3.40	1.64	0.75		
	EE3: Ich halte SPSS für einfach in der Anwendung.	3.94	1.65	0.73		
	EE4: Es fällt mir leicht, den Umgang mit SPSS zu meistern.	3.70	1.57	0.77		
SI	SI1: Leute die mir wichtig sind denken, ich soll SPSS verwenden.	2.63	1.77	0.73	0.85	0.85
	SI2: Leute, die mein Verhalten beeinflussen, denken, ich soll SPSS verwenden.	3.01	1.95	0.72		
	SI3: Leute, deren Meinungen ich schätze, ziehen vor, dass ich SPSS verwende.	2.85	1.71	0.72		
FC	FC1: Ich habe die notwendigen Ressourcen, um SPSS zu nutzen.	4.82	1.80	0.48	0.56	0.58
	FC2: Ich habe das notwendige Wissen, um SPSS zu nutzen.	3.33	1.55	0.32		
	FC3: SPSS ist kompatibel mit anderen Technologien, die ich nutze.	3.58	1.77	0.22		
	FC4: Ich kann Hilfe von anderen bekommen, wenn ich Schwierigkeiten habe, SPSS zu nutzen.	5.04	1.79	0.35		
HM	HM1: Die Verwendung von SPSS macht Spaß.	3.06	1.62	0.88	0.92	0.92
	HM2: Die Verwendung von SPSS macht Freude.	2.86	1.53	0.84		
	HM3: Die Verwendung von SPSS ist sehr unterhaltsam.	2.51	1.43	0.76		
PV	PV1: Der Preis von SPSS ist angemessen.	2.86	1.43	0.74	0.89	0.89
	PV2: SPSS bietet ein gutes Preis-Leistungs-Verhältnis.	3.06	1.47	0.79		
	PV3: Zum derzeitigen Preis bietet SPSS eine gute Leistung.	3.18	1.32	0.79		
HT	HT1: Die Verwendung von SPSS ist zur Gewohnheit für mich geworden.	2.24	1.47	0.04	0.30	0.43
	HT2: Ich bin von der Verwendung von SPSS abhängig.	3.43	2.08	0.23		
	HT3: Ich muss SPSS verwenden.	6.08	1.43	0.28		
BI	BI1: Ich habe vor, SPSS in Zukunft weiter zu verwenden.	4.66	1.80	0.70	0.80	0.83
	BI2: Ich werde immer versuchen, SPSS in meiner Tätigkeit zu nutzen.	3.12	1.63	0.49		
	BI3: Ich plane, SPSS auch weiterhin häufig zu nutzen.	4.07	1.77	0.81		

*Anmerkungen.* PE = Erwartete Leistung; EE = Erwarteter Aufwand; SI = Sozialer Einfluss; FC = Förderliche Umgebungsbedingungen; HM = Hedonistische Motivation; PV = Preis-Leistungs-Verhältnis; HT = Gewohnheit; BI = Nutzungsintention;  $r_{it}$  = „part-whole“-korrigierte Trennschärfe;  $\alpha$  = Cronbachs  $\alpha$ ;  $\omega$  = McDonalds  $\omega$ .

Tabelle 1.5: Deskriptiv-statistische und itemanalytische Kennwerte der TB-Variablen.

Konstrukt	Item	$M$	$SD$	$r_{it}$	$\alpha$	$\omega$
TB-Akz	1. Hinsichtlich technischer Neuentwicklungen bin ich sehr neugierig.	2.90	1.22	0.74	0.86	0.86
	2. Ich finde schnell Gefallen an technischen Neuentwicklungen.	3.22	1.09	0.73		
	3. Ich bin stets daran interessiert, die neuesten technischen Geräte zu verwenden.	2.46	1.17	0.69		
	4. Wenn ich Gelegenheit dazu hätte, würde ich noch viel häufiger technische Produkte nutzen, als ich das gegenwärtig tue.	2.26	1.30	0.66		
TB-Komp	5. Im Umgang mit moderner Technik habe ich oft Angst zu versagen.	1.95	1.04	0.59	0.86	0.86
	6. Für mich stellt der Umgang mit technischen Neuerungen zumeist eine Überforderung dar.	2.17	1.02	0.74		
	7. Ich habe Angst, technische Neuentwicklungen eher kaputt zu machen, als dass ich sie richtig benutze.	1.76	0.94	0.72		
	8. Den Umgang mit neuer Technik finde ich schwierig – ich kann das meistens einfach nicht.	2.13	1.02	0.74		
TB-Kont	9. Ob ich erfolgreich in der Anwendung moderner Technik bin, hängt im Wesentlichen von mir ab.	3.83	0.79	0.50	0.69	0.69
	10. Es liegt in meiner Hand, ob mir die Nutzung technischer Neuentwicklungen gelingt – mit Zufall oder Glück hat das wenig zu tun.	3.73	0.94	0.49		
	11. Wenn ich im Umgang mit Technik Schwierigkeiten habe, hängt es schlussendlich allein von mir ab, dass ich sie löse.	2.92	1.08	0.39		
	12. Das, was passiert, wenn ich mich mit technischen Neuentwicklungen beschäftige, obliegt letztlich meiner Kontrolle.	3.61	0.86	0.50		

*Anmerkungen.* TB-Akz = Technikakzeptanz; TB-Komp = Technikkompetenzüberzeugungen; TB-Kont = Technikkontrollüberzeugungen;  $r_{it}$  = „part-whole“-korrigierte Trennschärfe;  $\alpha$  = Cronbachs  $\alpha$ ;  $\omega$  = McDonalds  $\omega$ .



Die aufgeführten Trennschärfe- und Reliabilitätskoeffizienten deuten insgesamt auf eine vertretbare Passung der verwendeten Instrumente hin. Lediglich das Konstrukt FC des UTAUT2 weist grenzwertige, die Skala HT deutlich zu niedrige Trennschärfe und Reliabilitätsschätzungen auf. Dieser Umstand wird im Weiteren berücksichtigt und an späterer Stelle diskutiert (s. Abschnitt 1.4.1).

### 1.3.2 Modellprüfung TAM, UTAUT und UTAUT2

Um eine Beantwortung der ersten Forschungsfrage vorzunehmen, wurde die Passung der TAM-, UTAUT- und UTAUT2-Modelle im gegebenen Kontext mithilfe von Strukturgleichungsmodellen vergleichend überprüft. Besondere Beachtung erhielt hierbei auch die jeweilige Vorhersagegüte der Modelle in Hinblick auf die, von den Untersuchungsteilnehmenden berichtete Nutzungshäufigkeit der statistischen Anwendungssoftware.

Das TAM wies mit  $\chi^2(51) = 55.19$ ,  $p = .32$ ,  $RMSEA = .03$  ( $KI_{90}: .00-.08$ ),  $SRMR = .07$  und  $CFI = .99$  ( $n = 83$ ) eine hervorragende Passung auf (vgl. Hu & Bentler, 1999). Über die TAM-Konstrukte hinweg erwiesen sich erwartungsgemäß nahezu alle geschätzten Pfade (mit Ausnahme des Einflusses von PEOU auf BI) als bedeutsam. Der an der Nutzungsintention BI erklärte Varianzanteil lag bei  $R^2 = .41$ , jener der Nutzungshäufigkeit allerdings nur bei  $R^2 = .01$ . Eine grafische Darstellung des Modells ist Abbildung 1.5 zu entnehmen, bei den dort angegebenen Parametern handelt es sich um die vollständig standardisierten Schätzungen (dies gilt auch für alle weiteren Modelldarstellungen).

Die Passung des UTAUT konnte nicht als gegeben angenommen werden ( $n = 87$ ):  $\chi^2(131) = 293.38$ ,  $p < .01$ ,  $RMSEA = .12$  ( $KI_{90}: .10-.14$ ),  $SRMR = .18$  und  $CFI = .76$  (zusätzlich wurde die offenbar sehr geringe Fehlervarianz des Items BI3 in dieser Betrachtung sogar minimal negativ geschätzt). Die Ladungen aller Items von FC zeigten keine bedeutsame Abweichung von Null, was zusammen mit den zuvor berichteten geringen Trennschärfe und der niedrigen Reliabilität auf eine misslungene Erfassung dieses Konstrukts hinweist. In Übereinstimmung mit der entsprechenden Beobachtung im TAM erwies sich auch der Einfluss von EE auf BI als nicht bedeutsam. Auch der Koeffizient des Einflusses von SI auf BI überschritt die konventionelle Schwelle statistischer Bedeutsamkeit ( $p = .06$ ). Die Varianzaufklärung der Nutzungsintention BI betrug  $R^2 = .44$ , die der Nutzungshäufigkeit sechs Prozent ( $R^2 = .06$ ). Abbildung 1.6 veranschaulicht dieses Modell. Im Rahmen einer explorativen Betrachtung erfolgte zusätzlich die Schätzung eines Modells, das der unzureichenden Güte der Skala FC Rechnung tragen sollte und diese, ohne weitere inhaltliche Legitimation, nicht mitberücksichtigte. Das resultierende Modell wies mit  $\chi^2(74) = 133.21$ ,  $p < .01$ ,  $RMSEA = .10$  ( $KI_{90}: .07-.12$ ),  $SRMR = .13$  und  $CFI = .90$  ( $n = 88$ ) allerdings auch keine adäquate Passung auf.

Die Modellschätzung des UTAUT2 erwies sich ebenfalls als unpassend. Bei  $n = 81$  ergaben sich für die betrachteten Kennwerte folgende Resultate:  $\chi^2(315) = 654.78$ ,  $p < .01$ ,  $RMSEA = .12$  ( $KI_{90}: .10-.13$ ),  $SRMR = .19$  und  $CFI = .70$ . Wie bei der vorausgegangenen Schätzung des UTAUT wiesen keine Ladungen der FC-Items eine bedeutsame Abweichung von Null auf, ebensowenig die Ladungen aller Items des Konstrukts HT — dies ebenfalls in Einklang mit den zuvor berichteten, geringen Trennschärfe und der niedrigen Reliabilität

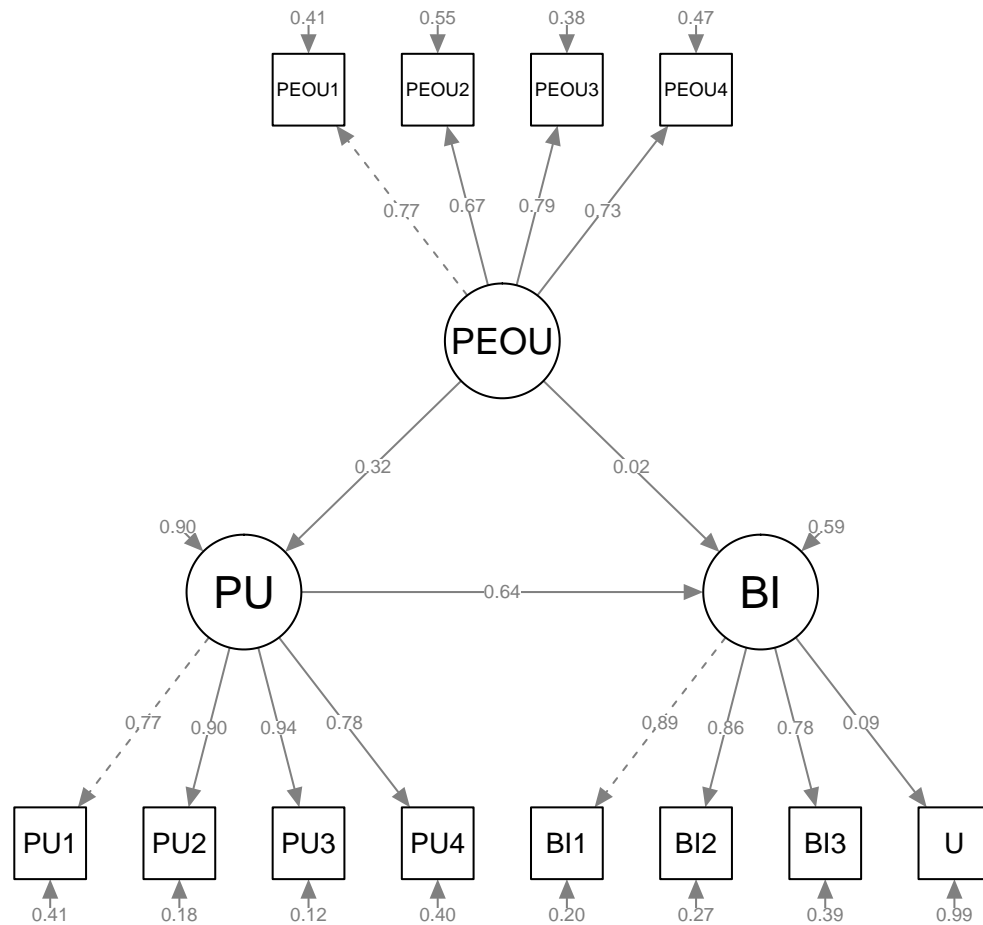


Abbildung 1.5: Strukturgleichungsmodell des *Technology Acceptance Model (TAM)*. Dargestellt sind die vollständig standardisierten, robusten Maximum-Likelihood Parameter-Schätzungen (MLMR). PU = Empfundene Nützlichkeit; PEOU = Empfundener Bedienkomfort; BI = Nutzungsintention; U = Nutzungshäufigkeit.

auch dieser Skala. Als bedeutsame Einflussfaktoren auf BI zeichneten sich außerdem nur PE und EE ab. Der aufgeklärte Varianzanteil von BI betrug  $R^2 = .68$ . Mit  $R^2 = .49$  fiel die vermeintliche Varianzaufklärung der Nutzungshäufigkeit vergleichsweise hoch aus. Wie aus Abbildung 1.7 hervorgeht, war dies maßgeblich auf einen von HT ausgehenden Einfluss zurückzuführen (eine inhaltliche Einordnung dieser Beobachtung erfolgt in Abschnitt 1.4.1). Auch für dieses Modell erfolgte eine weitere explorative Adaption durch die Entfernung der messtheoretisch fragwürdigen Skalen FC und HT. Allerdings führte diese — unabhängig vom inhaltlichen Erklärungswert — ebenfalls nicht zu einer akzeptablen Passung des derart modifizierten Modells:  $\chi^2(165) = 282.02$ ,  $p < .01$ ,  $RMSEA = .09$  ( $KI_{90}$ : .08–.11),  $SRMR = .17$  und  $CFI = .86$  ( $n = 81$ ).

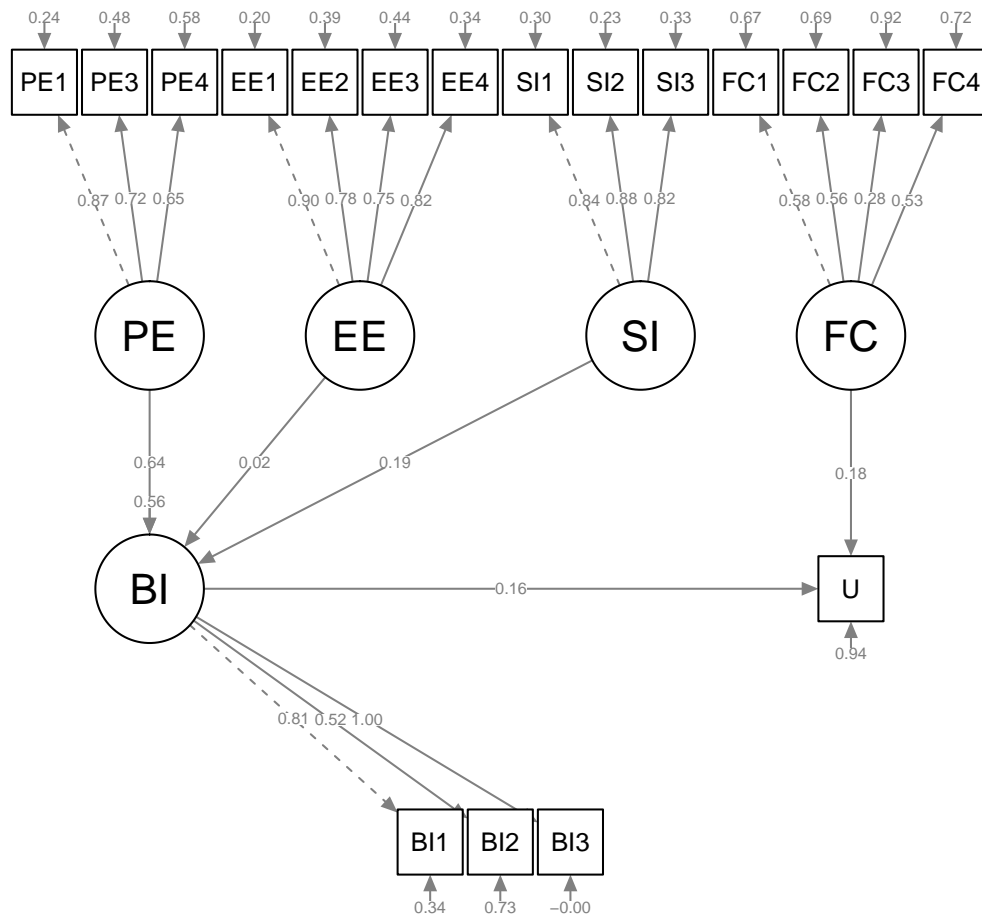


Abbildung 1.6: Strukturgleichungsmodell der *Unified Theory of Acceptance and Use of Technology (UTAUT)*. Dargestellt sind die vollständig standardisierten, robusten Maximum-Likelihood Parameter-Schätzungen (MLMR). PE = Erwartete Leistung; EE = Erwarteter Aufwand; SI = Sozialer Einfluss; FC = Förderliche Umgebungsbedingungen; BI = Nutzungsintention; U = Nutzungshäufigkeit.

### 1.3.3 Zusammenhänge der Persönlichkeit mit TAM-Konstrukten

Anknüpfend an diese Befunde unterschiedlicher Modellpassung — sowohl UTAUT als auch UTAUT2 erwiesen sich als unpassend — wurde in Hinblick auf die untersuchungsleitende Annahme 1a bei der Überprüfung möglicher Zusammenhänge zwischen spezifischen Persönlichkeitsmerkmalen und zentralen Konstrukten der Technologieakzeptanz der weitere Fokus ausschließlich auf das TAM gerichtet. In Tabelle 1.6 sind entsprechend die bivariaten Zusammenhänge der Nutzungshäufigkeit und TAM-Konstrukte mit den differenzierten Persönlichkeitseigenschaften des FFM in Form der Dimensionen und Facetten des BFSI abgetragen. Unter den Persönlichkeitsdimensionen wies Offenheit bedeutsame, mittlere positive Zusammenhänge mit allen drei TAM-Konstrukten auf, Gewissenhaftigkeit ebenfalls mittlere positive Zusammenhänge mit PU und BI (die konventionelle Schwelle statistischer

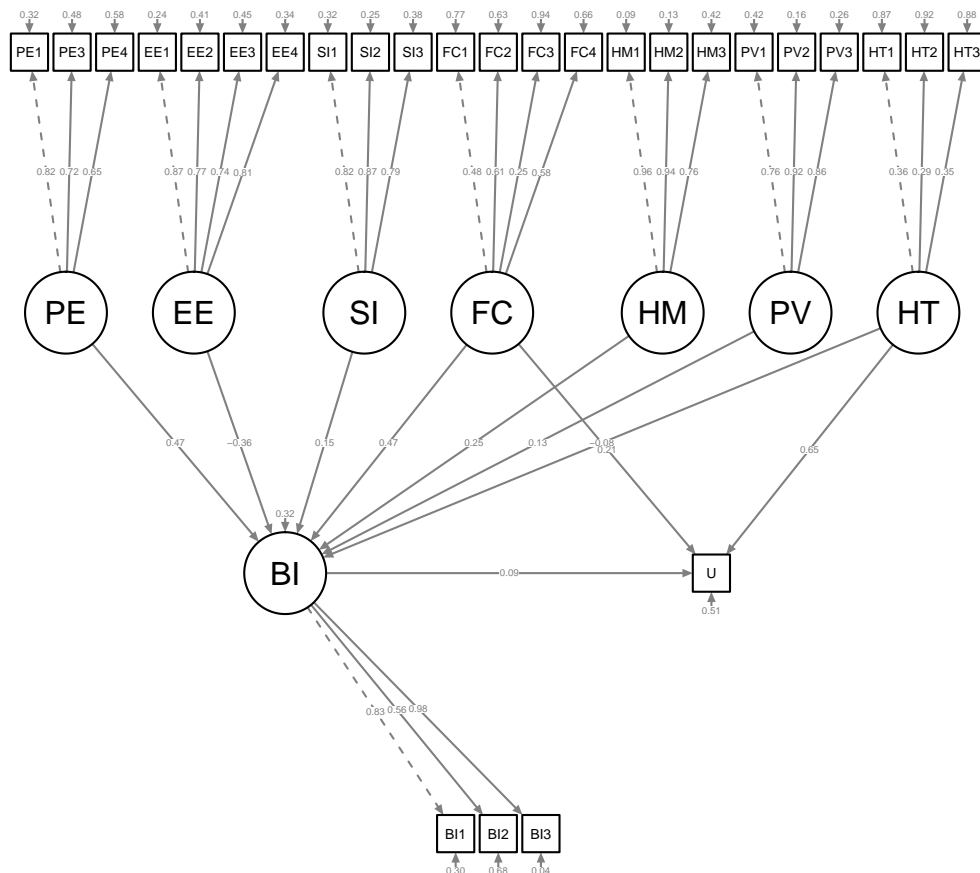


Abbildung 1.7: Strukturgleichungsmodell der erweiterten *Unified Theory of Acceptance and Use of Technology (UTAUT2)* im Konsumentenbereich. Dargestellt sind die vollständig standardisierten, robusten Maximum-Likelihood Parameter-Schätzungen (MLMR). PE = Erwartete Leistung; EE = Erwarteter Aufwand; SI = Sozialer Einfluss; FC = Förderliche Umgebungsbedingungen; HM = Hedonistische Motivation; PV = Preis-Leistungs-Verhältnis; HT = Gewohnheit; BI = Nutzungsintention; U = Nutzungshäufigkeit.

Signifikanz mit einem kritischen  $\alpha$ -Niveau von 5% koinzidierte bei der vorliegenden Stichprobengröße mit einem Korrelationskoeffizienten von mind. .20, was nach J. Cohen (1992) einer tendenziell mittleren Effektstärke entspricht und deshalb hier und nachfolgend als bedeutsamer Zusammenhang erachtet wird).<sup>4</sup>

<sup>4</sup>Im Sinne einer explorativen Annäherung wird hier und im Weiteren die potentielle Problematik der  $\alpha$ -Fehler-Inflation und gehäufte Fehlentscheidungen zugunsten der Alternativhypothese nur nachrangig diskutiert. Ein Zugang unter strengerer statistischer Kontrolle würde angesichts der implizierten Globalannahme, dass (eigentlich beliebige) Zusammenhänge zwischen Persönlichkeitsmerkmalen und den Konstrukten des TAM auftreten können z. B. eine Bonferroni-Korrektur zur inferenzstatistischen Beurteilung der möglichen Zusammenhänge nahelegen. Bei einer Analyse auf Ebene der Persönlichkeitsdimensionen könnte dann etwa für ein korrigiertes kritisches  $\alpha$ -Niveau von  $.05/20 = .0025$  (bei fünf Persönlichkeitsbereichen und vier TAM-Konstrukten) argumentiert werden. In diesem Falle würde einzig der Zusammenhang zwischen Offenheit und BI statistische Signifikanz aufweisen ( $p < .001$ ). Auf Facettenebene wäre wohl eine noch

Tabelle 1.6: Korrelationen der TAM-Konstrukte mit den Persönlichkeitsdimensionen und -facetten des BFSI.

Persönlichkeitsmerkmal	PU	PEOU	BI	U
N: Emotionale Stabilität	0.10	0.12	-0.05	0.05
E: Extraversion	0.13	-0.04	0.00	-0.14
O: Offenheit	0.28**	0.20*	0.36***	-0.07
A: Verträglichkeit	-0.01	0.16	0.08	0.07
C: Gewissenhaftigkeit	0.26*	0.13	0.25*	0.07
N1: Unbekümmertheit	0.04	0.02	-0.10	0.08
N2: Gelassenheit	0.00	0.09	0.04	0.08
N3: Positive Grundstimmung	0.08	0.06	-0.04	-0.12
N4: Soziale Selbstsicherheit	0.08	0.08	-0.13	-0.02
N5: Selbstbeherrschung	0.17	0.27**	0.23*	0.21*
N6: Emotionale Robustheit	0.11	0.08	-0.11	0.05
E1: Freundlichkeit	0.10	0.02	0.06	-0.07
E2: Geselligkeit	-0.04	-0.05	-0.04	-0.12
E3: Durchsetzungsfähigkeit	0.15	-0.06	-0.02	-0.18
E4: Tatendrang	0.19	-0.01	-0.03	-0.15
E5: Abenteuerlust	0.29**	-0.01	0.11	-0.08
E6: Fröhlichkeit	0.03	-0.07	-0.02	-0.04
O1: Fantasie	0.17	0.19	0.34***	-0.16
O2: Ästhetik	0.19	0.07	0.29**	-0.11
O3: Gefühle	0.15	0.04	0.19	0.03
O4: Handlungen	0.28**	0.18	0.23*	0.03
O5: Ideen	0.26*	0.13	0.29**	-0.04
O6: Werte- und Normensystem	0.16	0.26*	0.26*	-0.07
A1: Vertrauensbereitschaft	-0.03	-0.02	-0.11	-0.13
A2: Aufrichtigkeit	-0.05	0.16	-0.02	0.08
A3: Hilfsbereitschaft	-0.01	0.23*	0.09	0.08
A4: Entgegenkommen	0.00	0.05	0.10	0.06
A5: Bescheidenheit	-0.03	0.17	0.11	0.26*
A6: Gutmütigkeit	0.03	0.09	0.10	-0.01
C1: Kompetenz	0.27**	0.00	0.13	0.00
C2: Ordnungsliebe	0.21*	0.04	0.21*	0.04
C3: Pflichtbewusstsein	0.22*	0.15	0.24*	0.06
C4: Ehrgeiz	0.27**	0.08	0.18	0.00
C5: Disziplin	0.20*	0.10	0.21*	0.14
C6: Bedachtsamkeit	0.13	0.22*	0.25*	0.08

Anmerkungen. PU = Empfundene Nützlichkeit; PEOU = Empfundener Bedienkomfort; BI = Nutzungsintention; U = Nutzungshäufigkeit.

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

Für die Persönlichkeitsfacetten der Dimension Offenheit ließen sich (mit Ausnahme der Facette O3: Offenheit für Gefühle) durchweg ebenfalls mittlere positive Zusammenhänge mit der Nutzungsintention BI beobachten. Offenheit für Handlungen (O4) und Offenheit für

strengere Korrektur vorzunehmen, welche dazu führen würde, dass keine der beobachteten Korrelationen als statistisch von Null abweichend zu beurteilen wäre.

Ideen (O5) wiesen außerdem mittlere positive Zusammenhänge mit PU auf, die Offenheit des Werte- und Normensystems (O6) mit PEOU.

Mit Ausnahme der Bedachtsamkeit (C6) ergaben sich für alle Gewissenhaftigkeits-Facetten mittlere positive Zusammenhänge mit PU. Bedachtsamkeit wies hingegen als einzige dieser Facetten einen ebenfalls positiven Zusammenhang mit PEOU auf. Für alle sechs Facetten lagen außerdem positive Zusammenhänge mit BI vor, wobei die Korrelationskoeffizienten von Kompetenz (C1) und Ehrgeiz (C4) die Schwelle statistischer Signifikanz nicht überschritten. Weiterhin ließen sich auf Ebene der Persönlichkeitsfacetten einzelne, jeweils positive Zusammenhänge zwischen Abenteuerlust (E5) und PU sowie zwischen PEOU und Hilfsbereitschaft (A3) bzw. Selbstbeherrschung (N5) beobachten. Letztere korrelierte zusätzlich jeweils positiv mit BI und der Nutzungshäufigkeit. Einen vergleichbaren Zusammenhang mit der Nutzungshäufigkeit wies außerdem Bescheidenheit (A5) auf. Alle anderen Korrelationen der Persönlichkeitsfacetten bzw. -dimensionen mit der Nutzungshäufigkeit fielen deutlich kleiner aus und lagen nahe Null oder teilweise im niedrigen, negativen Wertebereich.

Eine vergleichbare Aufstellung der Zusammenhänge zwischen den einzelnen Konstrukten des UTAUT und UTAUT2 mit den Persönlichkeitsdimensionen und -facetten des BFSI kann — unter Berücksichtigung der eingeschränkten Passung dieser Modelle sowie der Hinweise auf eine mögliche  $\alpha$ -Fehler-Inflation — Tabelle A.1 im entsprechenden Anhang entnommen werden.

### 1.3.4 Modellerweiterungen

Aufbauend auf den geschilderten Befunden erfolgte die Überprüfung von Annahme 1b ebenfalls auf Basis des TAM, da es sich hierbei um das einzige Ausgangsmodell mit hervorragender Passung handelte. Eine Ergänzung dieses Modells erfolgte jeweils auf Ebene der Persönlichkeitsdimensionen sowie der -facetten des FFM. Hierbei wurde auf die zuvor ermittelten und als bedeutsam berichteten, bivariaten Zusammenhänge der TAM-Konstrukte mit den erfassten Persönlichkeitseigenschaften zurückgegriffen (vgl. Tabelle 1.6).

Ein entsprechend um Offenheit und Gewissenhaftigkeit ergänztes und geschätztes Modell wies eine gute Passung auf:  $\chi^2(70) = 76.24$ ,  $p = .29$ ,  $RMSEA = .03$  ( $KI_{90}$ : .00–.07),  $SRMR = .07$  und  $CFI = .99$  ( $n = 83$ ). Für die angenommenen Einflüsse von Offenheit auf PEOU, PU und BI sowie von Gewissenhaftigkeit auf PU und BI ergaben sich jedoch durchweg keine bedeutsamen Abweichungen von Null. Die Varianzaufklärung der Nutzungsintention BI erhöhte sich durch die ergänzend hinzugefügten Persönlichkeitsmerkmale nur minimal auf  $R^2 = .44$  (vgl. Abschnitt 1.3.2). Da kein direkter Einfluss auf die Nutzungshäufigkeit angenommen wurde, verblieb die entsprechende Vorhersage unverändert. Eine vergleichende Betrachtung der Informationskriterien zwischen dem ursprünglichen und dem derart erweiterten TAM-Modell wies für das unmodifizierte Ausgangsmodells niedrigere Werte aus:  $AIC_{orig} = 3408.93$  vs.  $AIC_{erw} = 3767.66$  und  $BIC_{orig} = 3503.26$  vs.  $BIC_{erw} = 3874.08$ .<sup>5</sup>

<sup>5</sup>Ein im Vergleich konservativeres Modell, das — der vorausgegangenen Diskussion zur

Zur Erweiterung des Modells auf Facettenebene der Persönlichkeit wurden ebenfalls alle zuvor als bedeutsam identifizierten Zusammenhänge der entsprechenden Eigenschaften mit den Konstrukten des TAM und der berichteten Nutzungshäufigkeit berücksichtigt. Die Schätzung dieses erweiterten Modells führte zu uneindeutigen Resultaten ( $n = 83$ ). Der globale Modelltest legte die Ablehnung des Modells nahe ( $\chi^2(207) = 271.99$ ,  $p < .01$ ), die Fit-Indizes waren jedoch größtenteils als akzeptabel zu beurteilen:  $RMSEA = .06$  ( $KI_{90}: .04-.08$ ),  $SRMR = .07$  und  $CFI = .89$ ). Allerdings wiesen auch in diesem Modell alle von den Persönlichkeitsfacetten ausgehenden Pfade keine bedeutsam von Null abweichenden Koeffizienten auf. In der Varianzaufklärung von BI war eine gewisse Zunahme zu beobachten ( $R^2 = .52$ ), für die Nutzungshäufigkeit fiel diese mit  $R^2 = .02$  nur minimal aus (vgl. Abschnitt 1.3.2). Die Informationskriterien  $AIC$  und  $BIC$  waren für das unmodifizierte Ausgangsmodells erneut deutlich niedriger ( $AIC_{orig} = 3408.93$  vs.  $AIC_{erw} = 6720.11$ ,  $BIC_{orig} = 3503.26$  vs.  $BIC_{erw} = 6872.50$ ).

### 1.3.5 Technikbereitschaft: Modellprüfung und Persönlichkeit

In einer mit dem vorausgehend beschriebenen Vorgehen übereinstimmenden Verfahrensweise wurde zur Beantwortung der zweiten Forschungsfrage die Passung des integrativen Modells der Technikbereitschaft mithilfe eines weiteren Strukturgleichungsmodells überprüft. Erneut wurde auch hierbei gezielt die Vorhersagegüte des Modells für die berichtete Nutzungshäufigkeit der statistischen Anwendungssoftware betrachtet (siehe Abbildung 1.8).

Die Passung des derart geschätzten Modells erwies sich als uneindeutig ( $n = 90$ ). Während der globale Modelltest für eine Ablehnung des Modells sprach ( $\chi^2(62) = 92.94$ ,  $p = .01$ ), lagen die betrachteten Fit-Indizes größtenteils innerhalb der konventionellen Annahmebereiche ( $RMSEA = .07$ , ( $KI_{90}: .04-.11$ ),  $SRMR = .08$  und  $CFI = .93$ ). Die Ladungen der verschiedenen Facetten von Technikbereitschaft auf den jeweils zugeordneten Items erwiesen sich allesamt als bedeutsam, ebenso diejenigen der globalen Technikbereitschaft auf den drei untergeordneten Facetten. Der Anteil aufgeklärter Varianz in der Nutzungshäufigkeit betrug  $R^2 = .01$ , wobei der für die Technikbereitschaft geschätzte Regressionskoeffizient deskriptiv einen negativen Wert besaß und im Sinne einer inferenzstatistischen Interpretation nicht bedeutsam von Null abwich ( $p = .51$ ).

Eine Überprüfung der gemäß Annahme 2 erwarteten Zusammenhänge zwischen grundlegenden, im BFSI erfassten Persönlichkeitsmerkmalen und der Technikbereitschaft bzw. ihren Facetten erfolgte mithilfe der in Tabelle 1.7 abgetragenen Korrelationskoeffizienten.

Als einzige der Persönlichkeitsdimensionen wies Emotionale Stabilität bedeutsame positive Zusammenhänge mittlerer Höhe mit der Technikbereitschaft und deren Facette Technikkompetenzüberzeugungen auf.

Eine spezifische Betrachtung auf Ebene der entsprechenden Persönlichkeitsfacetten legte nahe, dass sich diese Zusammenhänge maßgeblich in der Facette Emotionale Robustheit

---

$\alpha$ -Fehler-Inflation folgend — ausschließlich einen zusätzlichen Pfad von Offenheit auf BI beinhaltete, wies ebenfalls eine adäquate Passung auf. Es führte aber ebenso wenig zu einer verbesserten Varianzaufklärung der Nutzungsintention, noch war es gemäß den betrachteten Informationskriterien dem Ausgangsmodell überlegen.

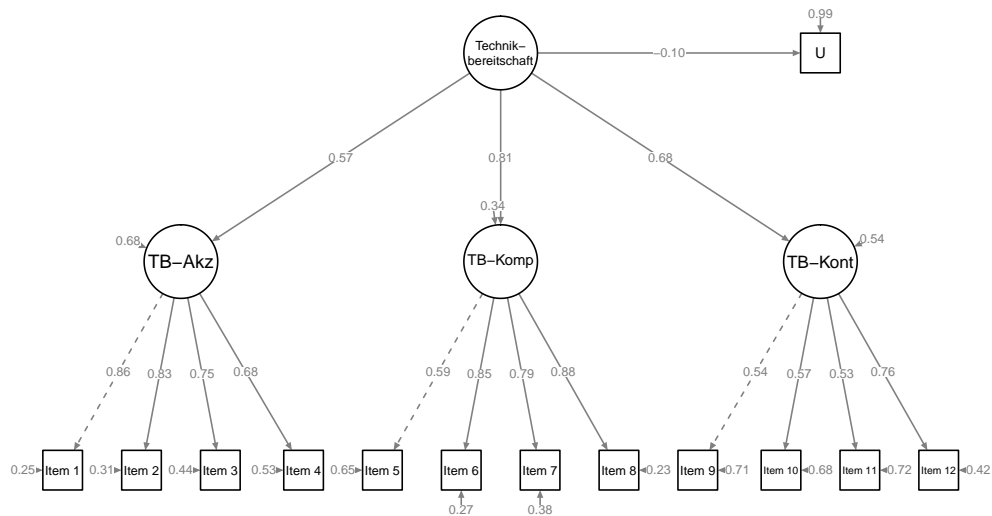


Abbildung 1.8: Strukturgleichungsmodell der Technikbereitschaft. Dargestellt sind die vollständig standardisierten, robusten Maximum-Likelihood Parameter-Schätzungen (MLMR). TB-Akz = Technikakzeptanz; TB-Komp = Technikkompetenzüberzeugungen; TB-Kont = Technikkontrollüberzeugungen; U = Nutzungshäufigkeit.

(N6) wiederfinden ließen, welche sogar in etwas stärker ausgeprägter Form mit den beiden genannten Konstrukten der Technikbereitschaft korrelierte. Als weitere Facette dieser Dimension wies außerdem Selbstbeherrschung (N5) einen mittleren, positiven Zusammenhang mit den Technikkontrollüberzeugungen auf. Die Zusammenhänge zwischen der Persönlichkeitsdimension Offenheit und allen Konstrukten der Technikbereitschaft verfehlten zwar die Schwelle statistischer Signifikanz, lagen aber durchweg im niedrigen bis mittleren positiven Bereich. Unter den Facetten dieses Persönlichkeitsmerkmals waren infolgedessen auch statistisch bedeutsame Zusammenhänge mit den Konstrukten der Technikbereitschaft zu beobachten: Offenheit des Werte- und Normensystems (O6) korrelierte positiv mit den Technikkompetenz- und Technikkontrollüberzeugungen sowie der Technikbereitschaft, Offenheit für Fantasie (O1) wies ebenfalls einen mittleren, positiven Zusammenhang mit letztgenanntem Faktor auf. Durchsetzungsfähigkeit (E3) wies als einzige Facette der Dimension Extraversion einen substantiellen Zusammenhang mit einer Facette der Technikbereitschaft auf, und zwar den Technikkompetenzüberzeugungen. Weiterhin ließ sich ein mittlerer, ebenfalls positiver Zusammenhang der Facette Kompetenz (C1) mit dem Faktor Technikbereitschaft beobachten.



Tabelle 1.7: Korrelationen der Technikbereitschaft mit den Persönlichkeitsdimensionen und -facetten des BFSI.

Persönlichkeitsmerkmal	TB-Akz	TB-Komp	TB-Kont	TB
N: Emotionale Stabilität	0.10	0.21*	0.11	0.20*
E: Extraversion	0.09	0.01	-0.08	0.04
O: Offenheit	0.17	0.14	0.13	0.18
A: Verträglichkeit	0.08	0.03	-0.01	0.04
C: Gewissenhaftigkeit	0.00	0.12	0.14	0.11
N1: Unbekümmertheit	0.01	0.15	0.12	0.11
N2: Gelassenheit	0.10	0.11	0.11	0.16
N3: Positive Grundstimmung	0.04	0.08	-0.06	0.05
N4: Soziale Selbstsicherheit	0.10	0.16	-0.02	0.14
N5: Selbstbeherrschung	0.05	0.13	0.21*	0.16
N6: Emotionale Robustheit	0.14	0.25*	0.16	0.26*
E1: Freundlichkeit	0.16	0.02	-0.07	0.08
E2: Geselligkeit	0.13	-0.04	-0.09	0.02
E3: Durchsetzungsfähigkeit	0.05	0.21*	0.08	0.18
E4: Tatendrang	0.04	0.01	-0.09	0.01
E5: Abenteuerlust	-0.07	-0.16	-0.12	-0.12
E6: Fröhlichkeit	0.06	0.02	-0.06	0.01
O1: Fantasie	0.18	0.18	0.17	0.22*
O2: Ästhetik	0.10	0.08	0.03	0.09
O3: Gefühle	0.02	-0.07	-0.11	-0.07
O4: Handlungen	0.14	0.04	0.00	0.09
O5: Ideen	0.14	0.13	0.19	0.19
O6: Werte- und Normensystem	0.13	0.22*	0.20*	0.23*
A1: Vertrauensbereitschaft	0.11	-0.03	-0.07	0.01
A2: Aufrichtigkeit	-0.05	0.12	0.01	0.02
A3: Hilfsbereitschaft	0.10	0.09	0.04	0.10
A4: Entgegenkommen	0.04	-0.10	0.02	-0.03
A5: Bescheidenheit	-0.13	0.06	-0.02	-0.07
A6: Gutmütigkeit	0.17	0.01	-0.07	0.06
C1: Kompetenz	0.13	0.18	0.18	0.26*
C2: Ordnungsliebe	-0.03	0.03	0.09	0.04
C3: Pflichtbewusstsein	-0.05	0.03	0.16	0.04
C4: Ehrgeiz	-0.01	0.08	0.07	0.05
C5: Disziplin	-0.06	0.14	0.11	0.08
C6: Bedachtsamkeit	0.05	0.13	0.12	0.13

Anmerkungen. TB-Akz = Technikakzeptanz; TB-Komp = Technikkompetenzüberzeugungen; TB-Kont = Technikkontrollüberzeugungen; TB = Technikbereitschaft.

\* $p < .05$ .

## 1.4 Diskussion

In den folgenden Passagen erfolgt eine interpretative Auseinandersetzung mit den berichteten Befunden vor dem Hintergrund der zuvor formulierten Forschungsfragen und Annahmen (Abschnitte 1.4.1–1.4.3). Abschließend werden in Kapitel 1.4.4 einschränkende Faktoren der vorliegenden Untersuchung sowie in Abschnitt 1.4.5 weiterführende Perspektiven theoretischer und praktischer Natur diskutiert.

### 1.4.1 Modelle der Technologieakzeptanz

Im Vergleich der drei Technologieakzeptanz-Modelle TAM, UTAUT und UTAUT2 wies im betrachteten Kontext einzig das ursprüngliche TAM eine adäquate Modellpassung auf. Offenbar tragen die strukturellen Anpassungen und ergänzend postulierten Faktoren sozialer Einflüsse (SI) und förderlicher Umgebungsbedingungen (FC) des UTAUT unter den gegebenen Bedingungen nicht angemessen zur Erklärung der Nutzungsintention bzw. tatsächlichen Nutzungshäufigkeit bei. Diese Beobachtung scheint in ähnlicher Form auf das UTAUT2 zuzutreffen. Dessen, vor dem Hintergrund der Verbreitung neuer Technologien im Verbrauchermarkt zusätzlich eingeführten Modellerweiterungen (HM, PV und HT) besitzen im universitären Umfeld vermutlich ebenfalls nachrangige Bedeutung.

Eine Ausnahme könnte hierbei allerdings der Aspekt Gewohnheit (HT) darstellen: zwar deuteten die niedrigen Trennschärfen und Ladungen der zugrunde liegenden Aussagen sowie die mangelhafte Reliabilität der Skala auf eine testtheoretisch ungenaue Erfassung dieser Einschätzung hin (vgl. Tabelle 1.4), gleichzeitig wies das so erfasste Konstrukt aber nicht unbedeutende Zusammenhänge mit der berichteten Nutzungshäufigkeit auf (Korrelation der ermittelten Summenwerte  $r_{SW} = .29$ , bzw. der latenten Variablen des berichteten Modells  $r_{LV} = .89$ , siehe auch Abbildung 1.7). Eine dedizierte, inhaltliche Analyse der entsprechenden Items legt nahe, dass dies vermutlich durch eine — teils der deutschen Übersetzung geschuldete — Bedeutungsverschiebung der betroffenen Aussagen verursacht wurde. Während im englischen Original die ersten beiden Ausführungen zweifelsfrei den Aspekt der individuellen Gewohnheit bzw. Abhängigkeit von der bewerteten Technologie adressieren („... has become a habit for me ...“ und „I am addicted to ...“) bleibt wenn überhaupt bei der dritten Aussage die Frage nach der Quelle des thematisierten Zwanges offen („I must use ...“). In der verwendeten deutschen Fassung lag dagegen sowohl für die zweite, als auch die dritte Aussage die Interpretation in Hinblick auf eine vermeintlich externe Quelle der Abhängigkeit bzw. des Zwanges nahe. Die besonders geringe Trennschärfe des ersten Items unterstützt die Interpretation, dass dieser externe Einfluss schließlich auch vornehmlich im resultierenden Konstrukt erfasst wurde. Der dann zu beobachtende, bedeutsame Zusammenhang mit der Nutzungshäufigkeit geht folglich vermutlich primär auf die Wahrnehmung und die Empfänglichkeit für entsprechende, systemische Vorgaben und weniger die individuelle Gewohnheit zurück. Offenbar waren ähnliche Überlegungen auch schon in die Ausgestaltung des TAM2 und der ursprünglichen UTAUT eingeflossen. Bei unternehmensübergreifenden Erhebungen war die „Freiwilligkeit der Nutzung“ als Einflussgröße berücksichtigt worden, allerdings nur als möglicher Moderator (dichotom: Nutzung

un-/freiwillig) des Zusammenhangs zwischen sozialen Einflüssen/sozialen Normen SI und der Handlungsabsicht BI. Womöglich sollte dieser Aspekt in weiterführenden Betrachtungen prominenter hervorgehoben und gezielt erfasst werden.

Schließlich muss festgehalten werden, dass aber selbst das TAM als augenscheinlich passendes Modell daran scheitert, eine substantielle Varianzaufklärung der Nutzungshäufigkeit zu erzielen. Die erste Forschungsfrage, inwieweit sich etablierte Modelle der Technologieakzeptanz zur Vorhersage der Nutzungsintention bzw. des Nutzungsverhaltens von statistischer Analysesoftware bei Studierenden der Psychologie bzw. Schulpsychologie eignen, kann dementsprechend ausschließlich für das TAM und nur in Hinblick auf die Nutzungsintention bejaht werden, eine Vorhersage der tatsächlichen Nutzungshäufigkeit gelingt jedoch nicht. Dieser Umstand wird in Abschnitt 1.4.4 weiter vertieft.

### 1.4.2 TAM: Zusammenhänge mit Persönlichkeit und Modellerweiterung

In Hinblick auf mögliche Zusammenhänge zwischen den Persönlichkeitseigenschaften der Big Five und den zentralen Konstrukten des TAM wurde gemäß Annahme 1a auf Basis der meta-analytisch aggregierten Vorbefunde ein solcher insbesondere zwischen Offenheit (bzw. den Facetten dieser Dimension) und der Nutzungsintention erwartet. Die Betrachtung der in Tabelle 1.6 berichteten Korrelationskoeffizienten offenbart, dass die entsprechende Beziehung am stärksten ausgeprägt scheint und sich auch in allen Facetten von Offenheit abzeichnet. Es erscheint plausibel, dass ein vielfältiges Interesse und die Aufgeschlossenheit gegenüber neuen Lernerfahrungen auch zu einer erhöhten Bereitschaft und Intention zur Nutzung neuer Technologien führt. Auch die im Vergleich am niedrigsten korrespondierende Facette — Offenheit für Gefühle (O3), d. h. das Ausmaß, in dem sich Personen als feinfühlig und sensibel beurteilen — fügt sich stimmig in diese Interpretation, da diese Eigenschaften wenig Bezug zur Nutzung neuer technologischer Entwicklungen (und hier insbesondere statistischer Anwendungssoftware) aufzuweisen scheinen. Zusätzlich weist Offenheit aber auf Dimensions- und Facettenebene auch Zusammenhänge mit der Leichtigkeit der Nutzung (PEOU) sowie der empfundenen Nützlichkeit (PU) auf. PEOU steht hierbei offenbar vornehmlich im Zusammenhang mit der Offenheit des Werte- und Normensystems (O6), wobei inhaltlich zu klären bleibt, in welcher Weise die Beurteilung des Bedienkomforts einer Technologie mit Weltoffenheit und Toleranz korrespondiert. Augenscheinlich schlüssiger wirken dahingegen die Zusammenhänge zwischen Offenheit für Handlungen (O4) und Ideen (O5) mit PU: Personen, die sich offen für neue Erfahrungen und Tätigkeiten sehen bzw. allgemein interessiert und lernbereit sind, bewerten die Nützlichkeit einer neuen Technologie tendenziell höher als Personen mit niedrigeren Ausprägungen in diesen Eigenschaften.

Als einzige weitere Persönlichkeitsdimension weist Gewissenhaftigkeit Zusammenhänge mit PU und BI auf, die tendenziell etwas niedriger ausfallen, als diejenigen von Offenheit. Die Betrachtung der Zusammenhänge auf Ebene ihrer Facetten offenbart, dass mit Ausnahme von Bedachtsamkeit (C6) alle spezifischen Persönlichkeitsbereiche der Gewis-

senhaftigkeit mit der wahrgenommenen Nützlichkeit der statistischen Anwendungssoftware PU korrespondieren. Variierende Ausprägungen der Achtsamkeit und Umsichtigkeit scheinen dementsprechend eine nachrangige Rolle für die Beurteilung von PU zu spielen.<sup>6</sup> Die beobachteten Beziehungen sind aber offenbar sowohl auf Unterschiede in der vorgenommenen Selbsteinschätzung von Kompetenz und Organisation, dem Ehrgeiz sowie dem Grad des Pflichtbewusstseins und der Disziplin zurückzuführen. Gerade letzteres könnte mit der zuvor diskutierten Wahrnehmung systemischer, externer Vorgaben und Zwänge in Beziehung stehen. Für Gewissenhaftigkeit und ihre Facetten zeichnet sich zwar kein direkter Zusammenhang mit der Nutzungshäufigkeit ab, eine zukünftige Überprüfung einer möglicherweise moderierenden Funktion in diesem Kontext könnte aber weitere Aufschlüsse über die Rolle dieser Persönlichkeitseigenschaft bieten.

In ähnlicher Weise werden auch für die Handlungsabsicht (BI) Zusammenhänge mit mehreren der Facetten von Gewissenhaftigkeit deutlich: Ordnungsliebe (C2), Pflichtbewusstsein (C3), Disziplin (C5) und Bedachtsamkeit (C6) korrespondieren jeweils positiv mit diesem Aspekt der Technologieakzeptanz. Kompetenz (C1) und Ehrgeiz (C4), die jeweils vergleichsweise hohe Zusammenhänge mit der empfundenen Nützlichkeit PU aufzeigen, sind offenbar bedeutend schwächer mit der Nutzungsintention BI verknüpft.

Unter den Facetten der Extraversion weist einzig Abenteuerlust (E5) einen positiven Zusammenhang mit PU auf, was auf die dadurch erfasste Experimentierfreudigkeit zurückgeführt werden könnte. Warum diese jedoch mit der empfundenen Nützlichkeit und nicht einem der anderen TAM-Konstrukte in Beziehung steht, bleibt ungeklärt. Dies ist ebenso bei einzelnen Facetten der Verträglichkeit der Fall: Hilfsbereitschaft (A3) korrespondiert positiv mit PEOU (Hilfsbereitschaft und Selbstlosigkeit gehen mit einem als höher empfundenen Bedienkomfort einher), Bescheidenheit (A5) zeigt einen Zusammenhang mit der Nutzungshäufigkeit (Genügsamkeit und Anspruchslosigkeit scheinen offenbar die Nutzung zu fördern). Einen vergleichbaren Zusammenhang mit der Nutzungshäufigkeit weist außerdem die Facette Selbstbeherrschung (N5) auf, die zusätzlich sowohl mit PEOU und BI korrespondiert und sich dadurch von allen anderen Facetten des Persönlichkeitsbereichs der Emotionalen Stabilität abhebt. Da dieses Merkmal, ähnlich wie die zuvor thematisierten Facetten der Gewissenhaftigkeit, Aspekte der Selbstkontrolle, Besonnenheit und Disziplin umfasst, könnten hier ebenfalls vergleichbare Mechanismen auftreten.

Inwieweit alle diese — insbesondere auch die zuletzt angeführten, über die auf Basis der Vorbefunde erwarteten Beziehungen hinausgehenden — Zusammenhänge als stabil und weithin generalisierbar erachtet werden können, bleibt (nicht zuletzt aufgrund der untergeordneten Berücksichtigung einer möglichen  $\alpha$ -Fehler-Inflation) jedoch weitgehend offen.

Vor diesem Hintergrund sind auch die in Hinblick auf die in Annahme 1b geäußerten Erwartungen erfolgten Betrachtungen kritisch zu beurteilen. Wenngleich eine diesbezügliche Erweiterung des TAM um die relevant erscheinenden Persönlichkeitsmerkmale auf Dimensions- und Facettenebene des BFSI in grundsätzlich weitgehend akzeptablen Model-

---

<sup>6</sup>Es offenbart sich jedoch ein unerwarteter Zusammenhang alleinig dieser Facette mit der Benutzerfreundlichkeit PEOU, dessen inhaltliche Interpretation zu klären bleibt.

len resultierte, konnten diese jedoch nicht ausschlaggebend zur verbesserten Erklärung der Nutzungsintention und — noch weniger — des Nutzungsverhaltens beitragen. Vor allem im Falle der Modellerweiterung auf Facettenebene ist durch die erfolgte Hinzunahme der zahlreichen spezifischen Persönlichkeitsmerkmale außerdem die Stabilität der geschätzten Modellparameter zu hinterfragen. Die mangelnde Güte dieser Modelle wurde an den sehr niedrigen Pfadkoeffizienten der Persönlichkeitsfacetten und den ausnahmslos höheren Werten in den Informationskriterien im Vergleich zum ursprünglichen TAM deutlich, welche zweifelsfrei die Präferenz der sparsameren Modelle (ohne die zusätzlichen Persönlichkeits-eigenschaften des BFSI) nahelegen.

### 1.4.3 Technikbereitschaft und Persönlichkeit

Auch die gemäß der zweiten Forschungsfrage vorgenommenen Betrachtungen unterliegen den soeben diskutierten Einschränkungen. Die überprüfte Eignung des integrativen Modells der Technikbereitschaft zur Vorhersage des Nutzungsverhaltens kann deshalb ebenfalls nur als bedingt erfolgreich beurteilt werden. Während das Modell auf Basis seiner generellen Passung zwar nicht unmittelbar abgelehnt werden konnte, erwies es sich in der Vorhersage der erfassten Nutzungshäufigkeit als völlig ungeeignet. Der Zusammenhang zwischen der Technikbereitschaft und der Nutzungshäufigkeit fiel nur gering, in seiner Tendenz sogar negativ aus. Dem Anspruch, die konkrete Techniknutzung vorherzusagen, kann das Konstrukt — zumindest in der hier betrachteten, studentischen Stichprobe — dementsprechend nicht gerecht werden (vgl. Neyer et al., 2012).

Die berichteten Zusammenhänge der Technikbereitschaft mit Persönlichkeitsmerkmalen fallen im Rahmen der erfolgten, explorativen Annäherung auf dimensionaler Ebene dagegen erwartungsgemäß aus. Emotionale Stabilität steht (als Gegenpol von Neurotizismus) in positiver Beziehung zur allgemeinen Technikbereitschaft. Für Offenheit liegt ebenfalls ein positiver, allerdings etwas niedrigerer Zusammenhang (unterhalb des Grenzwerts statistischer Signifikanz) vor, der sich in seiner Höhe jedoch deutlich von denen der anderen Persönlichkeitseigenschaften abhebt. Diese Beobachtungen korrespondieren jeweils mit den von Neyer et al. (2012) berichteten Effekten. Die Betrachtung spezifischer Persönlichkeitsfacetten legt nahe, dass der beschriebene Zusammenhang im Falle der Emotionalen Stabilität maßgeblich nur auf den der Facette Emotionale Robustheit (N6) zurückführbar scheint. Offenbar weisen Personen, die sich nur schwer aus der Ruhe bringen lassen und sich als widerstandsfähig sowie belastbar beschreiben eine höhere Technikbereitschaft auf, als Personen, bei denen diese Eigenschaften geringer ausgeprägt sind. Der Zusammenhang von Technikbereitschaft mit Offenheit wird ebenfalls in zwei einzelnen Facetten besonders deutlich: Offenheit für Fantasie (O1) und Offenheit des Werte- und Normensystems (O6). Technikbereitschaft geht tendenziell scheinbar mit besonderer Kreativität und ausgeprägtem Einfallsreichtum sowie Weltoffenheit und Aufgeschlossenheit einher.

Als weitere Facette scheint außerdem Kompetenz (C1), d. h. eine ausgeprägte Leistungsfähigkeit und Entscheidungssicherheit, mit einer erhöhten Technikbereitschaft zu korrespondieren. Allerdings handelt es sich hierbei um ein Merkmal, für dessen übergeordnete Dimension Gewissenhaftigkeit in ihrer Gesamtheit kein nennenswerter Zusammenhang mit

Technikbereitschaft beobachtet werden konnte. Die inhaltliche Bedeutung und Verallgemeinerbarkeit dieser Beziehung bleibt dementsprechend kritisch zu prüfen.

Diese Beobachtungen auf Facettenebene stehen in einem gewissen Gegensatz zu den Zusammenhängen, welche für die unterschiedlichen Persönlichkeitseigenschaften mit den Konstrukten des TAM berichtet wurden: während sich dort die Beziehungen der Technologieakzeptanz und insbesondere der Nutzungsintention mit Offenheit und Gewissenhaftigkeit auf dimensionaler Ebene ebenfalls in einer Vielzahl von ähnlichen Zusammenhängen in ihren Facetten abzeichneten, scheinen im Falle der Technikbereitschaft jeweils nur die hier angeführten, einzelnen Facetten ausschlaggebend für die Effekte der übergeordneten Dimension zu sein.

Bei spezifischer Analyse der drei Facetten von Technikbereitschaft finden sich außerdem keinerlei bedeutsame Zusammenhänge der Technikakzeptanz mit den Dimensionen oder Facetten des BFSI. Dies ist insofern verwunderlich, da es sich nach Neyer et al. (2012) hierbei um das Konstrukt handelt, das im Sinne des ursprünglichen TAM-Modells konzipiert wurde und für dieses — auch in der vorliegenden Untersuchung — teils deutlichere Zusammenhänge mit den Persönlichkeitseigenschaften beobachtet werden konnten. Im integrativen Modell der Technikbereitschaft liegen dem Anschein nach auch darüber hinaus nur vereinzelte Zusammenhänge zwischen den Technikkompetenz- und Technikkontrollüberzeugungen sowie einzelnen Persönlichkeitsfacetten vor.

Diese Unterschiede könnten den variierenden Einflüssen konkreter technologischer Entwicklungen geschuldet sein und die Ausgangsfrage nach der zeitlichen und situativen Stabilität aller hier berichteten Befunde adressieren. Während TAM, UTAUT und UTAUT2 jeweils die Einstellungen zu und die Akzeptanz einer explizit benannten Technologie erfassen, ist Technikbereitschaft (inklusive ihrer Facette Technikakzeptanz) als grundlegende, nicht an einzelne Technologien geknüpfte Haltung anzusehen. Weshalb dann jedoch die Zusammenhänge basaler, weitgehend als invariant angesehener Persönlichkeitseigenschaften mit den ebenfalls als grundlegend erachteten Persönlichkeitsmerkmalen der Technikbereitschaft niedriger und, auf Facettenebene, weniger prägnant ausfallen als beim (technikspezifischen) TAM, bleibt offen.

#### 1.4.4 Limitationen

Die vorliegenden Befunde unterliegen methodischen wie inhaltlichen Einschränkungen, welche im Folgenden diskutiert werden. Eine Stichprobe von  $N = 96$  Personen (bzw. von  $n > 80$  aufgrund einzelner fehlender Werte in den unterschiedlichen Analysen) ist für die verwendete Methodik der Strukturgleichungsmodellierung als absolutes Minimum anzusehen. Übliche Empfehlungen legen einen doppelten bis dreifachen Stichprobenumfang nahe, um genauere Parameterschätzungen zu erhalten (vgl. Bühner, 2011). Die Präzision der vorliegenden Modellparameter-Schätzungen kann demnach kritisch hinterfragt werden.

Wie in Abschnitt 1.2.3 dargestellt, wurde das Nutzungsverhalten nicht nur über die (schlußendlich betrachtete) Nutzungshäufigkeit, sondern zunächst auch über eine potentiell genauere, tägliche Selbstauskunft der Nutzungsdauer erhoben. Das verwendete Antwortformat sah hierbei eine Angabe in Stunden mit der Genauigkeit von zwei Nachkommastellen vor.

tellen vor. Dieses Format erwies sich nicht als verlässlich. Einerseits führte es zu mutmaßlichen Fehleingaben von (Stunden-)Werten im zweistelligen Bereich. Gleichzeitig war eine Häufung von ganzzahligen Angaben zu beobachten. Dies lässt darauf schließen, dass die Möglichkeit zur Präzisierung anhand von Dezimalstellen womöglich nicht von allen Teilnehmenden wahrgenommen wurde und stattdessen die Erfassung von Minuten als Bezugsgröße zu bevorzugen gewesen wären. Womöglich war einzelnen Personen eine entsprechende Selbstauskunft im erhofften Detailgrad auch tagesaktuell gar nicht möglich. Für zukünftige Erhebungen des Nutzungsverhaltens technischer Entwicklungen sollte dementsprechend nach Möglichkeit vermehrt die Erfassung objektiver, digital erfasster Nutzungsmaße in Erwägung gezogen werden.

Überraschend erscheint zunächst außerdem der Befund, dass die Nutzungsintention BI aus keinem der hier betrachteten Modelle als valider Prädiktor der berichteten Nutzungshäufigkeit hervorgeht. Die diskutierten Modelle der Technologieakzeptanz erweisen sich (auch in der Ergänzung mit spezifischen Persönlichkeitseigenschaften) als unzureichende Erklärungsansätze für die vorliegenden Nutzungsunterschiede. Diese Kluft zwischen der Verhaltensabsicht und evidenten Verhaltensweisen wird im Kontext der TAM-Forschung erst in jüngerer Zeit offen diskutiert (vgl. Nistor, 2014). Aus aktuellen Meta-Analysen geht allerdings hervor, dass überhaupt nur bei einem Anteil von etwas mehr als 20% der veröffentlichten Studien zur Technologieakzeptanz konkretes Nutzungsverhalten erhoben und der entsprechende Zusammenhang mit der Verhaltensabsicht BI empirisch ermittelt wurde (Wu & Du, 2012). Turner, Kitchenham, Brereton, Charters und Budgen (2010) berichten weiterhin von teils beträchtlichen Variationen der prädiktiven Bedeutsamkeit dieser Modelle bzw. der postulierten Verhaltensabsicht hinsichtlich des tatsächlichen Nutzungsverhaltens, insbesondere auch im Vergleich objektiver Maße und durch Selbstauskünfte erfasster Nutzungsangaben. Diese Autoren sprechen sich in der Folge ebenfalls für die vermehrte Berücksichtigung objektiver Verhaltensindikatoren aus (vgl. oben). Sie verweisen außerdem auf die teils schwierige Vergleichbarkeit unterschiedlicher Arbeiten, in denen oftmals spezifische Hypothesen an teils stark modifizierten Modellen getestet werden, ohne ausreichende Informationen zum jeweiligen Kontext, den betrachteten Technologien und/oder charakteristischen Merkmalen der erhobenen Stichprobe zu berichten. Diese Problematik wurde auch im Rahmen der zuvor präsentierten, meta-analytischen Aufbereitung deutlich. Die vorliegende Berichtlegung versucht diesem Informationsanspruch zwar angemessen gerecht zu werden, die geschilderten Befunde können aber angesichts der behandelten, spezifischen Technologie und der studentischen Stichprobe sicherlich nur einen weiteren Anhaltspunkt bei der Beantwortung der eingangs formulierten Fragestellungen bieten.

### 1.4.5 Fazit und Ausblick

Ogleich der genannten Einschränkungen können die hier berichteten Befunde einen wertvollen, differentialpsychologischen Beitrag zum verbesserten Verständnis der Rolle spezifischer Persönlichkeitseigenschaften bei der Akzeptanz und Nutzung neuer technologischer Entwicklungen leisten. Im betrachteten Kontext der Anwendung statistischer Analysesoftware unter Studierenden der Psychologie (teils mit schulpсихологischem Schwerpunkt)

können einzelne Modelle der Technologieakzeptanz und Technikbereitschaft angemessene Beschreibungsansätze unterschiedlicher Einstellungen und Handlungsabsichten bieten. Zentrale Konstrukte dieser Modelle weisen außerdem systematische Zusammenhänge mit einzelnen Persönlichkeitsdimensionen und -facetten der Big Five auf. Insbesondere die aus den aggregierten Vorbefunden meta-analytisch abgeleitete Annahme über die Bedeutung der Persönlichkeitsdimension Offenheit und ihrer Facetten wird bekräftigt. Das hierfür charakteristische vielfältige Interesse und eine stark ausgeprägte Wissbegierde werden im Technologieumfeld interessanterweise schon seit langem als zentrales Motiv im Selbstverständnis der sog. Hacker-Kultur angeführt, deren Mitglieder sich typischerweise durch eine besonders intensive Auseinandersetzung mit neuen technologischen Entwicklungen auszeichnen (vgl. Raymond, 2004; The Mentor, 1986).

An den präsentierten Befunden wird gleichwohl deutlich, dass der Erklärungs- und Vorhersagewert der diskutierten Modelle angezweifelt werden kann. Dies gilt einerseits vermutlich für die Verallgemeinerbarkeit der berichteten Zusammenhänge der Konstrukte von Technologieakzeptanz und Technikbereitschaft mit Persönlichkeitseigenschaften auf Facetenebene, die hier erstmalig berichtet wurden. Andererseits wird diese Einschränkung durch das Versagen der Modelle, die tatsächliche Nutzungshäufigkeit vorhersagen zu können, besonders evident. In dieser Hinsicht mag es zukünftig lohnenswert sein, dem entsprechenden Appell von Yarkoni und Westfall (o.D.) folgend, den (vermeintlichen) Erklärungswert derartiger Modelle weitgehend aufzugeben und stattdessen mithilfe von Techniken, die der Tradition des maschinellen Lernens zuzuschreiben sind, den Fokus auf eine präzisierte und generalisierbare Vorhersage relevanter Kriterien zu richten. In einem solchen Lichte könnte auch die Rolle differenzierter Persönlichkeitseigenschaften neu evaluiert werden. Eine beispielhafte Anwendung derartiger Prinzipien und Methoden wird in der nachfolgend vorgestellten Studie im Kontext eignungsdiagnostischer Fragestellungen demonstriert.



# Kapitel 2

## Computerintensive Methoden, Persönlichkeit und Absentismus während der Berufsausbildung

Die folgende Abhandlung beschreibt den Versuch, mithilfe von differenzierten Persönlichkeitseigenschaften die während der beruflichen Ausbildungsphase erfassten Fehlzeiten vorherzusagen. Dazu werden computerintensive Analysemethoden des maschinellen Lernens verwendet. Zusätzlich finden weitere Indikatoren der beruflichen Leistungsfähigkeit Berücksichtigung, um eine vergleichende Beurteilung der herangezogenen Methoden vornehmen zu können.

### 2.1 Einleitung

Die berufsbezogene, psychologische Eignungsdiagnostik mit dem Ziel der Vorhersage von Leistungsparametern während und nach der beruflichen Ausbildungsphase ist mit ihrer etwa hundertjährigen Tradition (vgl. Viteles, 1932) von unbestrittener Bedeutung für die differentielle, bzw. Wirtschafts- und Organisationspsychologie (Landy & Conte, 2010). In diesem Kontext haben sich insbesondere kognitive Fähigkeiten als bedeutende Prädiktoren beruflicher Leistungsfähigkeit erwiesen (Schmitt, 2014). Eine kurze Darstellung einschlägiger Befunde erfolgt in Abschnitt 2.1.1.

Das Forschungsfeld wurde zwischenzeitlich aber durch die Berücksichtigung mindestens zweier zusätzlicher Aspekte erweitert: einerseits wurde die Bedeutung von weiteren Unterschieden in der menschlichen Persönlichkeit (neben den kognitiven Fähigkeiten) als ursächliche Quelle zur Erklärung variierender beruflicher Erfolgs herangezogen. Mit ausschlaggebend war hierfür die Entwicklung des bereits vorgestellten *Fünf-Faktoren-Modells (FFM)* der Persönlichkeit (Barrick, Mount & Judge, 2001). Relevante Befunde zu dessen Zusammenhänge mit beruflicher Leistung werden nachfolgend in Kapitel 2.1.2 vorgestellt.

Auch das Verständnis einer erfolgreichen beruflichen Tätigkeit wurde über das Leistungskonzept hinaus zunehmend weiter gefasst und die traditionellen Indikatoren (wie

Abschlussnoten oder Vorgesetztenbeurteilungen) um vielfältige zusätzliche Gesichtspunkte ergänzt. In einer groben Klassifizierung sind hier neben einem Engagement, das über die Erwartungen des eigentlichen Arbeitsauftrag hinausgeht — sog. *Organizational Citizenship Behavior (OCB)* (Organ, 1997; Smith, Organ & Near, 1983) — insbesondere auch problematische, potentiell unternehmensschädigende Verhaltensweisen der Mitarbeiterinnen und Mitarbeiter — *Counterproductive Work Behavior (CWB)* (Campbell, 1990) — zu nennen. In Abschnitt 2.1.3 erfolgt eine Vorstellung dieses Konzepts.

Absentismus, d. h. das Fernbleiben vom Arbeitsplatz während vorgegebener Arbeitszeiten (Furnham, 1992), wird letzteren Verhaltensweisen zugerechnet (Sackett & DeVore, 2001). Seinen Vorbedingungen wurde im Rahmen der vorausgehend skizzierten Eignungsdiagnostik, welche kognitive Fähigkeiten und weitere Aspekte der menschlichen Persönlichkeit zu erfassen sucht, bisher allerdings nur eine vergleichsweise geringe Aufmerksamkeit zu teil. Diesbezügliche Vorbefunde werden in Abschnitt 2.1.4 zusammengefasst.

Die prädiktiven Ansätze der berufsbezogenen Eignungsdiagnostik erfolgen vor dem Hintergrund einer ausgedehnten Diskussion des sog. bandwidth-fidelity Dilemmas (Cronbach & Gleser, 1957). Eine knappe Zusammenfassung dieser Debatte wird unter Punkt 2.1.5 vorgenommen. In einer ihrer zentralen Fragen — inwiefern breite oder engere Prädiktoren besser zur Vorhersage beruflichen Erfolgs (ebenfalls in unterschiedlich spezifischer Konzeptualisierung) geeignet sind — wird zunehmend davon ausgegangen, dass eine Berücksichtigung spezifischer Persönlichkeitsmerkmale einen Mehrwert gegenüber breiteren Dimensionen bietet (Christiansen & Robie, 2011; Dudley, Orvis, Lebiecki & Cortina, 2006; Rothstein & Goffin, 2006) und dieser insbesondere auch bei der Betrachtung spezifischer Kriterien deutlich wird (Rojon, McDowall & Saunders, 2015).

Anknüpfend an diese Befundlage werden in der vorliegenden Studie die Eigenschaften vielzähliger, differenzierter Persönlichkeitsmerkmale in ihrer Funktion als Prädiktoren der krankheitsbedingten und unentschuldigten Fehltag (sowie weiterer beruflicher Leistungsparameter) untersucht. Um dahingehend eine präzise, dem Verständnis der untersuchten Zusammenhänge dienliche Modellbildung zu ermöglichen, kommen computerintensive Analysemethoden zum Einsatz, die der prädiktiven Modellierung, bzw. dem maschinellen Lernen zugeordnet werden können. Diese modernen Verfahren der Datenanalyse ermöglichen eine realistische Abschätzung darüber, wie gut ein basierend auf vorliegenden Daten ermitteltes Vorhersagemodell auf neue, unbeobachtete Daten übertragbar ist. Weiterhin erlauben es einzelne der ausgewählten Methoden, nicht nur gleichzeitig eine Vielzahl von Prädiktoren für die Modellierung zu berücksichtigen, sondern auch eine (automatische) Variablenselektion vorzunehmen, d. h. Hinweise auf eine reduzierte Auswahl geeigneter Prädiktoren aus einer größeren Ausgangsmenge zu geben. Auf diese Weise wird eine sparsame Modellbildung gefördert. Diese Methodik wird in den Kapiteln 2.1.6 und 2.2.3 erläutert. Vor dem hier angedeuteten und im Weiteren ausführlich diskutierten theoretischen und empirischen Hintergrund erscheint es aussichtsreich, anhand derartiger Analysen neue Erkenntnisse über das Zusammenwirken differenzierter Persönlichkeitseigenschaften und dem Fernbleiben vom Arbeitsplatz zu ergründen. Berücksichtigung soll hierbei auch ein mögliches Zusammenspiel mit kognitiven Fähigkeiten als etablierten Prädiktoren beruflicher Leistungsfähigkeit finden.

Die Untersuchung widmet sich folglich dem Absentismus als spezifischer Form kontraproduktiven Verhaltens im Rahmen der beruflichen Ausbildungsphase. Primär soll mithilfe moderner Methoden des maschinellen Lernens überprüft werden, ob und inwieweit krankheitsbedingte und unentschuldigte Fehltage durch Unterschiede in kognitiven Fähigkeiten und differenzierten Persönlichkeitseigenschaften von Auszubildenden erklärt und vorhergesagt werden können. Ergänzend soll eine vergleichende Einschätzung der ausgewählten Modellierungsverfahren ermöglicht werden. Zu diesem Zweck wird zusätzlich auch eine Betrachtung weiterer Indikatoren beruflicher Leistungsfähigkeit vorgenommen.

Die berichteten Betrachtungen sind von explorativem Charakter, die in Abschnitt 2.1.7 formulierten und abschließend diskutierten Vorannahmen sollen dementsprechend als untersuchungsleitend verstanden werden und bilden den Ausgangspunkt aller berichteten Analysen.

### 2.1.1 Kognitive Fähigkeiten und berufliche Leistung

Die Erfassung kognitiver Fähigkeiten im Rahmen der berufsbezogenen Eignungsdiagnostik erfolgt im Regelfall über standardisierte psychologische Testverfahren. Diese Instrumente basieren teils auf unterschiedlichen Theorien der menschlichen Intelligenz. Die Mehrzahl dieser Theorien kann jedoch, wie von Carroll (1993) vorgeschlagen, in einem hierarchischen Intelligenzmodell zusammengeführt werden. Zentraler Aspekt dieses Modells ist die Annahme eines allgemeinen Intelligenzfaktors auf den letztlich alle variierenden Leistungen in untergeordneten, spezifischen Facetten verschiedener kognitiver Anforderungen (z. B. Merkfähigkeit, Verarbeitungsgeschwindigkeit) oder unterschiedlicher inhaltlicher Aufgabenstellungen (z. B. verbal, numerisch, figural) zurückgeführt werden können. Dieses Maß allgemeiner Intelligenz wird als g-Faktor der Intelligenz — engl. *General Mental Ability*; *Allgemeiner Faktor der Intelligenz (GMA)* — bezeichnet und kann, i. d. R. erfasst in Kombination der genannten spezifischen kognitiven Fähigkeiten, als Indikator für das individuelle Intelligenzniveau angesehen werden.

Die Bedeutung kognitiver Fähigkeiten (zunächst im Sinne dieses Generalfaktors der Intelligenz) für die Vorhersage beruflicher Leistungsfähigkeit wurde auf Basis vielzähliger Untersuchungen etabliert und mehrfach meta-analytisch untermauert. In wegweisenden Arbeiten konnte Ende des letzten Jahrtausends eindrücklich aufgezeigt werden, dass GMA im Vergleich unterschiedlicher Auswahlverfahren als bedeutsamster Prädiktor für die berufliche Leistungsfähigkeit in unterschiedlichen Berufsgruppen anzusehen ist (J. E. Hunter, 1986; J. E. Hunter & Hunter, 1984; Schmidt & Hunter, 1998). Vergleichbare Befunde wurden in den darauf folgenden Jahren auch in zahlreichen internationalen Untersuchungen (z. B. Bertua, Anderson & Salgado, 2005; Salgado & Anderson, 2002, 2003; Salgado, Anderson, Moscoso, Bertua & De Fruyt, 2003; Salgado, Anderson, Moscoso, Bertua, De Fruyt & Rolland, 2003) und, insbesondere auch für Ausbildungserfolg, im deutschen Sprachraum vorgelegt (Hülshager, Maier & Stumpp, 2007; Hülshager, Maier, Stumpp & Muck, 2006). Auf Basis dieser Befunde lässt sich zusammenfassend festhalten, dass GMA als valider, über unterschiedliche Situationen und Berufsgruppen generalisierbarer Prädiktor beruflicher Leistung (auch während der Ausbildungsphase) mit einer Kriteriumskorrelation von

mindestens  $\hat{\rho}_{unkorrigiert} = .20$  angesehen werden kann. Wird im Rahmen meta-analytischer Studien für Varianzeinschränkungen bzw. Messungenauigkeit korrigiert, liegen die berichteten Zusammenhänge im Bereich von  $\hat{\rho}_{korrigiert} = .40$  oder sogar darüber (Schmitt, 2014). In neueren Arbeiten wurde zunehmend auf die Bedeutung spezifischer kognitiver Fähigkeiten zur Vorhersage der beruflichen Leistungsfähigkeit und einer erfolgreichen beruflichen Ausbildungsphase hingewiesen (Goertz, Hülshager & Maier, 2014). Nicht nur wird deshalb im Abschnitt 2.1.5 die diesbezügliche inhaltliche und methodische Diskussion aufgegriffen, sondern der mögliche Einfluss spezifischer kognitiver Fähigkeiten in der vorliegenden Untersuchung auch explizit berücksichtigt. Im Fokus soll zunächst aber die Rolle weiterer Persönlichkeitseigenschaften neben den kognitiven Fähigkeiten stehen. Die diesbezüglich vorliegenden Erkenntnisse werden im folgenden Abschnitt vorgestellt.

### 2.1.2 Persönlichkeit und berufliche Leistung

In der berufsbezogenen Eignungsdiagnostik fanden in den zurückliegenden Jahren neben den kognitiven Fähigkeiten weitere Faktoren der menschlichen Persönlichkeit zunehmend an Beachtung. Dieses Interesse ist, wie angedeutet, maßgeblich in der Entwicklung des FFM begründet (Barrick & Mount, 2005, vgl. Tabelle 1.1).

Wo sich zuvor keine systematischen Zusammenhänge von Persönlichkeitsmerkmalen mit Kriterien beruflicher Leistungsfähigkeit bzw. des Ausbildungserfolgs abgezeichnet hatten und konsequenterweise von ihrer Verwendung in der Personalauswahl abgeraten wurde (Guion & Gottier, 1965), öffnete die zunehmende Akzeptanz Big Five-basierter Instrumente und ihr vermehrter Einsatz den Weg zu systematischen, meta-analytischen Betrachtungen (Barrick & Mount, 1991). Hierbei erwies sich vor allem Gewissenhaftigkeit als weitgehend universeller Prädiktor unterschiedlicher Leistungskriterien im beruflichen Kontext (Barrick et al., 2001; Salgado, 1997). Zusammenhänge zwischen den anderen Dimensionen des FFM und beruflicher Leistung wurden in einzelnen, spezifischen Betrachtungen ebenfalls sichtbar, ließen sich aber nicht in vergleichbarer Weise generalisieren.

Die beobachteten Korrelationen zwischen den Dimensionen des FFM und der beruflichen Leistungsfähigkeit fallen in aller Regel niedriger aus, als für kognitive Fähigkeiten. Sie liegen für die Persönlichkeitsfaktoren nach Schmitt (2014) bei etwa  $\hat{\rho}_{unkorrigiert} = .20$ , bzw.  $\hat{\rho}_{korrigiert} = .25$ . Obwohl diese Befunde zunächst zu einer kritischen Auseinandersetzung darüber führten, ob Persönlichkeitseigenschaften bei der Personalauswahl überhaupt zum Einsatz gebracht werden sollten (Morgeson et al., 2007; Murphy & Dzieweczynski, 2005; Ones, Dilchert, Viswesvaran & Judge, 2007; Tett & Christiansen, 2007), wurde in diesem Punkt mittlerweile weitgehende, für ihre Verwendung plädierende Einigkeit erzielt (Schmitt, 2014).

Gleichwohl wird nach wie vor die Frage rege diskutiert, in welchem Detailgrad Persönlichkeitseigenschaften im Rahmen des FFM bestmöglich erfasst und zur Vorhersage beruflichen Erfolgs zum Einsatz gebracht werden sollen (vgl. Kapitel 2.1.5). Im folgenden Abschnitt wird zunächst die in der vorliegenden Studie betrachtete Kriteriumsvariable thematisiert.

### 2.1.3 Kontraproduktives Verhalten

Der hier behandelte Absentismus in Kombination aktenkundiger krankheitsbedingter und unentschuldigter Fehltage als erfolgsrelevantes Kriterium der Berufsausbildung wird dem sog. Kontraproduktiven Verhalten am Arbeitsplatz (CWB) zugerechnet. Es handelt sich dabei um ein Konstrukt, das in unterschiedlicher Operationalisierung derartige Verhaltensweisen beschreibt, die von Mitarbeiterinnen und Mitarbeitern zwar willentlich gezeigt werden, aber den Zielen ihrer Organisation und dem eigentlichen Arbeitsauftrag entgegen stehen.

In einer wegweisenden Systematisierung des Konzeptes wurde CWB von Robinson und Bennett (1995, S. 556) dementsprechend als „*voluntary behavior that violates significant organizational norms and in so doing, threatens the well-being of the organization, its members, or both*“ umschrieben. Mit dieser Definition ging die Entwicklung einer Taxonomie zur Beschreibung kontraproduktiven Verhaltens einher, die grundsätzlich zwei Gruppen von Verhaltensweisen unterscheidet: solche, die sich gegen die Werte, Normen und Strukturen der Organisation richten und jenen, die sich gegen andere Individuen, also Vorgesetzte und Kollegen richten (Robinson & Bennett, 1995, 1997). Diese beiden Arten des Fehlverhaltens werden entsprechend als organisatorisches (engl. *organizational deviance*) und interpersonelles Fehlverhalten (engl. *interpersonal deviance*) bezeichnet. Weiterhin wird angenommen, dass sich verschiedene Ausprägungs- oder Stärkegrade im Fehlverhalten auf diesen beiden Dimensionen beobachten lassen. Auf Basis dieser Überlegungen stellten Bennett und Robinson (2000) mit der *Workplace Deviance Scale* einen Kurzfragebogen mit insgesamt 19 Items zur Erfassung der dargestellten Konstrukte per Selbstauskunft vor. Bei der Konstruktion des Instruments folgten sie einem traditionellen Testkonstruktionsansatz, der einer aktuellen, kritischen Betrachtung stellenweise nicht mehr standhalten kann (Bühner, 2011, S. 248 f.). Die Autoren äußerten bspw. das Ziel, die beiden Dimensionen von organisatorischem und interpersonellem Fehlverhalten möglichst umfassend und breit erfassen zu wollen. Gleichzeitig entschieden sie sich dazu, Trennschärfen und Faktorladungen als Grundlage einer empirischen Itemauswahl heranzuziehen. Dadurch wurde eine Steigerung der Homogenität der verbleibenden Items (und damit eine inhaltliche Beschränkung der erfassten Konstrukte) bewirkt, die dazu führte, dass etwa Aspekte des Fernbleibens vom Arbeitsplatz vom genannten Instrument inhaltlich nicht mehr erfasst werden. Laut den Autoren ähneln sich aber alle kontraproduktiven Verhaltensweisen innerhalb der beiden unterschiedlichen Bereiche derart, dass Sie stellvertretend füreinander angesehen werden können und deshalb nicht in ihrer Gesamtheit erfasst werden müssten. Dies steht allerdings im Widerspruch zur vorausgehenden Itemauswahl, bei der einzelne Inhalte gerade aufgrund ihrer ungenügenden Übereinstimmung mit anderen Aspekten aus dem Verfahren entfernt wurden. Die daraus resultierenden Bedenken hinsichtlich der inhaltlichen Einschränkungen des Verfahrens können demzufolge nicht ausgeräumt werden.

Sackett und DeVore (2001) schlugen auf Basis ihrer Definition („*[A]ny intentional behavior on the part of an organization member viewed by the organization as contrary to its legitimate interest.*“; also jegliches intentionales Verhalten eines Organisationsmitglieds, das nicht im Einklang mit den Interessen seiner Organisation steht) ein erweitertes, hier-

archisches Modell für CWB vor. Dieses sieht eine weitere Unterteilung des von Robinson und Bennett (1995, 1997) eingeführten organisatorischen Fehlverhaltens in zwei zusätzliche Kategorien vor. Fehlverhalten, das sich gegen Sachmittel der Organisation richtet, muss demzufolge unterschieden werden von solchem, das sich gegen den Produktionsprozess richtet. Letzterem ordnen Sackett und DeVore (2001) Absentismus explizit zu.

Trotz der zuvor diskutierten Einschränkungen wurde das von Bennett und Robinson (2000) entwickelte, in Hinblick auf die Erfassung von Absentismus aber unzureichende Instrument — teils auch in adaptierter, meist jedoch sehr ähnlicher Form — in der CWB-Forschung am häufigsten eingesetzt (Berry, Ones & Sackett, 2007; Gonzalez-Mulé, Mount & Oh, 2014). Die dementsprechend im eignungsdiagnostischen Kontext vorliegenden Befunde zu CWB können demzufolge nur eingeschränkt auf die Vorhersage von tatsächlichem Absentismus übertragen werden. In der vorliegenden Arbeit konnte auf die Aufzeichnungen krankheitsbedingter sowie unentschuldigter Fehlzeiten während der beruflichen Ausbildungszeit zurückgegriffen werden. Einzelne Untersuchungen, welche die Vorhersage eines vergleichbar konkreten Maßes für Absentismus zum Inhalt hatten, werden im nächsten Abschnitt vorgestellt.

### 2.1.4 Persönlichkeit und Absentismus

In frühen Untersuchungen von Zusammenhängen zwischen der menschlichen Persönlichkeit und Absentismus im Berufsleben wurde noch nicht auf das FFM als weitgehend einheitliches und umfassendes Konstrukt der Persönlichkeit zurückgegriffen. Allerdings wurden Hinweise auf Zusammenhänge einzelner Persönlichkeitseigenschaften und berufliche Fehlzeiten gegeben, wie sie nun auch auf Ebene der Dimensionen bzw. Facetten im FFM enthalten sind.

So wurden etwa für die Persönlichkeitseigenschaft Neurotizismus schon frühzeitig einfache, positive Korrelationen mit Absentismus aufgezeigt (Cooper & Payne, 1967). In einer zunehmend differenzierteren Betrachtung argumentierten Porter und Steers (1973), dass berufliche Fehlzeiten maßgeblich durch impulsive bzw. spontane Verhaltensweisen bedingt seien — die Neigung zu solchen wird im FFM zumeist als Facette Impulsivität ebenfalls dem Neurotizismus untergeordnet. In ähnlicher Weise wurden ängstliche Persönlichkeitszüge — als Unterskala Ängstlichkeit der Dimension Neurotizismus auch im FFM enthalten — mit Absentismus in Zusammenhang gebracht (Bernardin, 1977; Sinha, 1963). Cooper und Payne (1967) berichteten ferner von einem positiven Zusammenhang zwischen Extraversion und beruflichen Fehlzeiten. Eine gewisse inhaltliche Nähe zur entsprechenden Facette weist weiterhin der negative Zusammenhang auf, der von J. Hogan und Hogan (1989) für ein Geselligkeitsmaß und die Zuverlässigkeit von Mitarbeiterinnen und Mitarbeitern aufgezeigt wurde. Vereinzelt Beachtung erhielt außerdem der Zusammenhang zwischen dem Fernbleiben vom Arbeitsplatz und diversen Persönlichkeitseigenschaften, die dem Big Five-Faktor der Gewissenhaftigkeit nahe kommen oder diesem konzeptuell untergeordnet werden können. So berichtete etwa Bernardin (1977) einen bedeutsamen Zusammenhang zwischen der Stärke des Über-Ichs (eine psychoanalytische Konzeption, die der inhaltlichen Ausgestaltung der Gewissenhaftigkeits-Facette Selbstdisziplin ähnelt) und den beruflichen

Fehlzeiten.

In einer ersten Betrachtung der Zusammenhänge zwischen den Faktoren und Facetten des FFM mit beruflichen Fehlzeiten konnten Judge, Martocchio und Thoresen (1997) Effekte einzelner Persönlichkeitseigenschaften an einer Stichprobe von  $N = 89$  nichtakademischen Universitätsangestellten aufzeigen. Die Autoren berichteten von bedeutsamen Zusammenhängen auf dimensionaler, wie auch auf Facettenebene der Big Five. Von den fünf Persönlichkeitsdimensionen leisteten Extraversion und Gewissenhaftigkeit einen bedeutsamen Beitrag in der Vorhersage von Fehlzeiten, welche zeitlich nach der Erhebung der Persönlichkeitseigenschaften erfasst wurden. Die entsprechenden Effektstärken lagen bei bivariater Betrachtung nach J. Cohen (1992) im mittleren Bereich ( $r_E = .26$ ;  $r_G = .23$ ). Eine genauere Analyse auf der Facettenebene dieser beiden relevanten Dimensionen legte nahe, dass diese Zusammenhänge maßgeblich von einzelnen, differenzierteren Persönlichkeitsaspekten ausgemacht wurden. Im Falle von Extraversion waren es die beiden Facetten Erlebnishunger ( $r = .21$ ) und Geselligkeit ( $r = .28$ ), die als relevante Eigenschaften auf Facettenebene erkennbar wurden. In ähnlicher Weise wiesen die Facetten Besonnenheit ( $r = .24$ ), Pflichtbewusstsein ( $r = .27$ ) und Selbstdisziplin ( $r = .18$ ) der Dimension Gewissenhaftigkeit relevante Zusammenhänge mit beruflichen Fehlzeiten auf. Die verbleibenden Facetten dieser beiden Dimensionen wiesen geringere oder keine Zusammenhänge mit dem genannten Kriterium auf, gleichwohl fielen nach Judge et al. (1997) die in der Gesamtbetrachtung von Dimensions- vs. Facettenebene erzielten Varianzaufklärungen vergleichbar aus. Insgesamt betrug diese für alle fünf Persönlichkeitsdimensionen (und nach vorausgehender Kontrolle für die insgesamt gearbeiteten Stunden, die subjektive Einschätzung der Gesundheit, dem Alter und der Anzahl der Kinder)  $R^2 = .18$ , wobei — wie dargestellt — Extraversion und Gewissenhaftigkeit die größten und bedeutsamen Beiträge leisteten. Wurde zusätzlich für vorausgehende Fehlzeiten (im Sinne einer Neigung zu Absentismus) kontrolliert (welche in der bivariaten Betrachtung einen mittleren bis starken Zusammenhang mit den späteren Fehlzeiten aufwiesen,  $r = .41$ ), sank der Beitrag der beiden relevanten Persönlichkeitsdimensionen Extraversion und Gewissenhaftigkeit zur Varianzaufklärung späterer Fehlzeiten auf 5%, der aber an inkrementeller Bedeutung behielt, d. h. auch über die Kontrollvariablen hinausgehend zu verzeichnen war.

Ferner wurden von Darviri und Woods (2006) korrelative Zusammenhänge zwischen Verträglichkeit ( $r = -.33$ ), Extraversion ( $r = .21$ ), Offenheit ( $r = .19$ ) und einer Beurteilung der Wahrscheinlichkeit des Eintretens einer Reihe von willentlich beeinflussbaren Gründen für etwaiges zukünftiges Fernbleiben vom Arbeitsplatz bei  $N = 128$  Bauschaffenden berichtet. Weitere Zusammenhänge zwischen den Faktoren des Big Five-Modells und einer Einschätzung möglicher Gründe für das potentielle Auftreten zukünftiger Fehlzeiten, welche als unbeeinflussbar angesehen wurden sowie den objektiv erfassten Fehltagen konnten nicht festgestellt werden.

Furnham und Bramwell (2006) konnten schließlich anhand einer Erhebung von  $N = 54$  Angestellten eines Fotowerkes einen positiven Zusammenhang im Rahmen einer Regressionsanalyse zwischen der Dimension Extraversion des FFM und krankheitsbedingten Fehlzeiten aufzeigen. Im betrachteten Modell waren allerdings — neben Geschlecht und Alter als Kontrollvariablen — alle fünf Dimensionen des FFM enthalten und es konnte insgesamt

nur 14% der Varianz der erfassten Fehltagel erklären.

Zusammenfassend erscheinen auf Basis dieser Einzelstudien und den berichteten Voruntersuchungen diverse Zusammenhänge zwischen spezifischen Persönlichkeitseigenschaften und Fehlzeiten im beruflichen Kontext zwar plausibel, aufgrund beschränkter Stichprobenumfänge und teils variierender Taxonomie lassen sich daraus aber kaum Verallgemeinerungen ableiten.

Eine meta-analytische Zusammenfassung der im Zeitraum 1990–1999 veröffentlichten und einzelnen, nicht publizierten Arbeiten zum Zusammenhang des FFM mit Absentismus (erfasst durch Aufzeichnungen von Fehltagen und unangemessenen Verspätungen beim Arbeitsbeginn) konnte deshalb auch für keine der fünf Persönlichkeitsdimensionen bedeutende Korrelationen mit dem genannten Kriterium aufzeigen (Salgado, 2002). Grundlage waren hierbei insgesamt 13 Studien mit jeweils 8 bis 12 unabhängigen Stichproben bei der Betrachtung der fünf Persönlichkeitsdimensionen. Hieraus resultierten Stichprobengrößen zwischen  $N = 1339$  (Offenheit für Erfahrungen, Verträglichkeit) und  $N = 2491$  (Emotionale Stabilität). Allerdings musste der Autor auch hier einschränkend anmerken, dass nicht in allen berücksichtigten Studien solche Verfahren zur Erfassung der Persönlichkeitsstruktur zum Einsatz gekommen waren, die explizit dem FFM folgten (in diesen Fällen wurde über zwei unabhängige Beurteilungen und ggf. klärende Diskussionen eine Zuordnung zu einer korrespondierenden Big Five-Dimension vorgenommen). Vor dem Hintergrund dieser Limitation konnte auch keine weitere Auswertung auf Facettenebene erfolgen, was der Autor für zukünftige Betrachtungen mit Nachdruck anrät. Ein besonderes Augenmerk sollte laut Salgado (2002) hierbei auch auf die mögliche Kombination geeigneter Subfacetten der Big Five zur Vorhersage von Absentismus gelegt werden. Mit dieser Empfehlung beteiligte er sich zumindest implizit an der laufenden Diskussion um das sog. Bandwidth-Fidelity Dilemma, die im folgenden Abschnitt zusammengefasst wird.

### 2.1.5 Das Bandwidth-Fidelity Dilemma

Das Konzept der bandwidth-fidelity wurde von Cronbach und Gleser (1957) im Kontext psychologischer Eignungsdiagnostik eingeführt. Die *Bandbreite* (engl. *bandwidth*) bezieht sich hierbei auf den Komplexitätsgrad der Informationen, die über ein bestimmtes Testverfahren ermittelt werden. Die *Genauigkeit* (engl. *fidelity*) dieser Informationen beschreibt dagegen im Kern deren Qualität (J. Hogan & Roberts, 1996). Das benannte Dilemma besteht nun darin, dass ein Messinstrument hoher Bandbreite zwar ein breites Spektrum an Informationen verfügbar macht, diese mutmaßlich aber von vergleichsweise geringer Qualität sind. Umgekehrt liefert die Verwendung eines psychologischen Testverfahrens niedriger Bandbreite vergleichsweise weniger komplexe Information, diese aber von vermeintlich höherer Qualität. Im vorliegenden Kontext breiter Persönlichkeitsdimensionen und schmalerer Eigenschaftsfacetten besteht folglich die Annahme, auf dimensionaler Ebene Einzelmaße vergleichsweise hoher Bandbreite (mit entsprechend niedriger Qualität) und auf Facettenebene Maße reduzierten Informationsumfangs (von höherer Qualität) zu erfassen.

Hinsichtlich der Vorhersage beruflicher Leistungsparameter wird vor dem Hintergrund dieser Überlegungen diskutiert, wie diese gegenläufigen Tendenzen der Informationsbe-



schaffenheit optimiert werden können. So wurde argumentiert, berufliche Leistung im Sinne eines komplexen und mehrdimensionalen Konstruktes durch (vergleichbar) breite Persönlichkeitseigenschaften vorherzusagen (Ones & Viswesvaran, 1996). Andere Autoren hielten dagegen, auch im Falle einer solchen Vorhersage auf spezifische Persönlichkeitsmerkmale zurückzugreifen (Ashton, 1998; Ashton, Jackson, Paunonen, Helmes & Rothstein, 1995; Paunonen & Ashton, 2001). Große Zustimmung erhält die These, dass die höchste Kriteriumsvalidität dann erzielt wird, wenn die Prädiktoren in ihrer Bandbreite an den Spezifitätsgrad der jeweiligen Kriterien angepasst werden (sog. *bandwidth-matching*) (z. B. Barrick & Mount, 2005; Dudley et al., 2006; J. Hogan & Holland, 2003; Jenkins & Griffith, 2004; Schneider, Hough & Dunnette, 1996; Stewart, 1999). Vor diesem Hintergrund wird angenommen, dass bei einer Kombination aus spezifischen Prädiktoren und spezifischen Kriterien die Kriteriumsvalidität am höchsten ausfällt (J. Hogan & Roberts, 1996; Rojon et al., 2015; Tett, Steele & Beaugard, 2003), unspezifischere Prädiktoren aber bei breiteren Kriterien ihre Bedeutung behalten.

Diese Diskussion basiert jedoch maßgeblich auf — wie von Christiansen und Robie (2011) herausgestellt, meist impliziten — Überlegungen zur inhaltlichen Breite der erfassten Konstrukte nach bestimmten Vorgaben. Die Wahl der Formulierung einer gegebenen Anzahl von Items in einem Fragebogen kann zweifelsfrei zu inhaltlichen Unterschieden in den Konstrukten führen, zu welchen die erhaltenen Antworten dieser Fragen nach den Prinzipien traditioneller Testkonstruktion meist aufsummiert werden. Bei einer Gegenüberstellung spezifischer und breiterer Konstrukte, die auf diese Weise erfasst wurden, ist dann auch in der beschriebenen Weise davon auszugehen, dass es Unterschiede in der Validität bei der Vorhersage ebenfalls in ihrer Breite variierender Kriterien gibt.

Wenn aber ein Testverfahren wie der NEO-PI-R mit angenommener, hierarchischer Struktur (oder ein dahingehend vergleichbarer Intelligenztest) Anwendung findet, geht bei einer Aggregation der erfassten Daten potentiell immer Information verloren, d. h. alle in den breiten Persönlichkeitsfaktoren vorliegenden interindividuellen Unterschiede sind vollumfänglich auch in den untergeordneten, spezifischen Facetten enthalten. Dies ist unabhängig von der ursprünglichen inhaltlichen Breite der erfassten Konstrukte oder der eines vorherzusagenden Kriteriums. Eine Aufnahme der spezifischeren Facetten in ein Regressionsmodell sollte in einem solchen Fall immer mindestens zur gleichen (oder einer größeren) Varianzaufklärung des Kriteriums führen, wie ein Modell mit einer geringeren Anzahl an Dimensionen, die aus den ursprünglichen Facetten gebildet wurden (Christiansen & Robie, 2011; Nunnally, 1978; Paunonen & Nicol, 2001).

Die vorausgegangenen Abschnitte deuten an, welchen Herausforderungen sich eine angemessene Modellierung im Kontext eignungsdiagnostischer Vorhersagen stellen muss. Die Zusammenhänge relevanter Prädiktor- und Kriteriumsvariablen müssen in adäquater Form abgebildet werden, wobei potentiell eine Vielzahl von Prädiktoren Berücksichtigung finden kann. In Hinblick auf eine möglichst hohe Präzision der Vorhersage müssen diese u. U. auf eine Auswahl besonders geeigneter Variablen reduziert und ggf. ein Vergleich mehrerer konkurrierender Modellierungsansätze vorgenommen werden. Im folgenden Abschnitt werden die methodischen Strategien zur Begegnung dieser Herausforderungen vorgestellt.

### 2.1.6 Methodische Abwägungen und maschinelles Lernen

Die Anpassungsgüte eines statistischen Modells kann mithilfe verschiedener Maße beurteilt werden. In diesen kommt der Anspruch an adäquate Modelle zum Ausdruck, die Bedingungen der Realität möglichst exakt abzubilden, d. h. dass ihre Vorhersagen den tatsächlichen Beobachtungen so ähnlich wie möglich sein sollten. Dementsprechend wird etwa ein Vergleich der vom Modell geschätzten —  $\hat{Y} = \hat{f}(X)$  — mit den bekannten, beobachteten Kriteriumswerten —  $Y$  — vorgenommen, üblicherweise in Form einer Quadratsumme der Differenzen dieser Werte unter Berücksichtigung aller Beobachtungen. Dieses Maß wird als *Residual Sum of Squares; Residuenquadratsumme (RSS)* bezeichnet:

$$RSS = \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.1)$$

Die Schätzung der Modellparameter  $\beta_0, \beta_1, \dots, \beta_p$  mithilfe der Methode der kleinsten Quadrate im Rahmen der linearen Regression erfolgt derart, dass diese Größe minimiert wird:

$$RSS = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \quad (2.2)$$

#### Determinationskoeffizient, MSE und RMSE

Im Kontext psychologischer Studien erhält üblicherweise der auf den vorausgegangenen Prinzipien aufbauende Determinationskoeffizient  $R^2$  an Beachtung, da dieser bei der Anwendung (multipler) linearer Regressionsmodelle eingängig als Quadrat des multiplen Korrelationskoeffizienten  $R$  interpretiert werden kann. Er ist definiert als der Varianzanteil (an der insgesamt beobachteten Variation), welcher durch das Modell erklärt werden kann:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.3)$$

Der im maschinellen Lernen übliche Ansatz, die Anpassung eines Vorhersagemodells zu beurteilen, ist die Berechnung des sog. *Mean Square Error; Mittlere quadratische Abweichung (MSE)*. Dieser ergibt sich aus der durchschnittlichen quadrierten Differenz zwischen den beobachteten Werten und den durch das Modell vorhergesagten Werten in einem gegebenen Datensatz.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (2.4)$$

Eine der vereinfachten Interpretation dienende Modifikation dieses Maßes stellt der sog. *Root Mean Square Error (RMSE)* dar. Es handelt sich um die Wurzel des MSE:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2} \quad (2.5)$$

Dieser Wert drückt die durchschnittliche Abweichung der durch das Modell geschätzten von den tatsächlich beobachteten Werten in der ursprünglichen Einheit der Kriteriumsvariable aus.

### Überanpassung und Bias-Varianz-Dilemma

In der Tradition des maschinellen Lernens wird jedoch nachdrücklich betont, dass eine Modellanpassung über die Minimierung der vorgestellten Maße ausschließlich an bekannten Daten nicht zur Schätzung von Modellen führt, die auch angemessen auf neue, unbeobachtete Daten (aus vergleichbaren Situationen) übertragbar sind. Vielmehr wird davor gewarnt, vermeintlich gut passende Modelle ungeprüft auf solche Daten zu übertragen (vgl. z. B. Hastie, Tibshirani & Friedman, 2011, S. 221). Dieses Phänomen wird als *Überanpassung* (engl. *overfitting*) bezeichnet.

Hierzu kann gezeigt werden, dass sich der für ein gegebenes Modell in neuen Daten erwartete MSE — unter der Annahme (unendlich) vieler Modellanpassungen und -überprüfungen — aus einer etwaigen systematischen Verzerrung des Modellschätzers (dem sog. *Bias*), einer gewissen Flexibilität des Modells (die sich in einer Variabilität der vorhergesagten Werte äußert) sowie der Varianz eines unsystematischen Fehlerterms zusammensetzt (vgl. Hastie et al., 2011, S. 223 ff.). Der Bias der geschätzten Werte spiegelt wider, inwieweit die erhaltene Schätzung modelleigenen Einschränkungen unterworfen ist und dadurch ggf. systematisch von dem funktionalen Prozess abweicht, der den wahren Zusammenhängen in der Population zugrunde liegt und dessen möglichst exakte Abbildung man anstrebt. Die Flexibilität des Modells, bzw. die Variabilität seiner Vorhersagen (die womöglich zu einer besonders guten Beschreibung der gegebenen Daten geführt hat, aber nicht auf neue Daten generalisierbar ist) schlägt sich in der Varianz der Vorhersagewerte nieder. Zusammen mit dem quadrierten Bias und der Varianz des unsystematischen Fehlerterms ergibt sich in der Summe der mittlere Vorhersagefehler eines Modells für neue Daten:

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \text{Var}\left(\hat{f}(x_0)\right) + \left(\text{Bias}\left(\hat{f}(x_0)\right)\right)^2 + \text{Var}(\epsilon) \quad (2.6)$$

In der Regel haben Modelle niedriger Komplexität (etwa mit wenigen Prädiktoren und einfachen, z. B. linearen Funktionsverläufen) einen vergleichsweise hohen Bias. Sie nehmen Vereinfachungen vor, die der Realität unter Umständen nicht gerecht werden. Ihre Vorhersagen können dementsprechend verzerrt sein und systematisch von den wahren Gegebenheiten abweichen. Modelle, die hingegen auf komplexeren, beispielsweise nicht-linearen Zusammenhängen basieren und zusätzliche Modellparameter enthalten, führen i. d. R. zu Vorhersagen mit einem geringeren Bias. Gleichzeitig erhöht sich ihre Flexibilität und damit die Variabilität ihrer Vorhersagen, da für die Schätzung ihrer Parameter eine größere Vielfalt an möglicherweise stichprobenspezifischen (Fehl-)Informationen bei der Modellbildung berücksichtigt werden. Modelle hoher Flexibilität laufen demnach Gefahr, zu stark an die Spezifika beobachteter Daten angepasst zu werden. Es kommt dann zur genannten Überanpassung und das erhaltene Modell kann infolgedessen nicht adäquat auf weitere Beobachtungen übertragen werden.

Auch in der psychologischen Methodenliteratur finden sich seit langem diesbezügliche Hinweise. So konnte etwa schon Larson (1931) im Kontext der psychologischen Diagnostik bestätigen, dass in gegebenen Datensätzen ermittelte, multiple Regressionskoeffizienten  $R^2$  häufig „täuschend groß“ erscheinen und bei der Übertragung der zugrunde liegenden Modelle auf neue, vergleichbare Datensätze deutlich geringer ausfallen. Der entsprechende Unterschied fiel mit zunehmender Prädiktorenanzahl und Höhe des Koeffizienten immer stärker aus. Der generalisierende Populationsparameter  $\rho^2$  wird in derartigen Fällen dementsprechend durch  $R^2$  systematisch überschätzt (J. Cohen & Cohen, 1983; Stevens, 2012; Wherry, 1931).

Teilweise wird versucht, diesen Überschätzungen mithilfe von analytisch abgeleiteten Korrekturtermen zu begegnen. Hierbei gilt es allerdings zu beachten, dass zusätzlich eine Unterscheidung zwischen  $\rho$  und einem weiteren Parameter  $\rho_c$  vorgenommen werden müsste (Yin & Fan, 2001). Ersterer bringt die Schätzung des wahren (multiplen) Korrelationskoeffizienten in der Population auf Basis der beobachteten Stichprobenverhältnisse zum Ausdruck. Letzterer bezieht sich auf die Übertragung eines in einer beobachteten Stichprobe geschätzten Modells zur Vorhersage in einer neuen Stichprobe aus der gleichen Population. Es kann gezeigt werden, dass bei der zweiten Interpretation die Überschätzung des Parameters durch  $R$  stärker ausfällt, als im ersten Fall. Der Zusammenhang zwischen diesen beiden Populationswerten und dem Erwartungswert von  $R$  nimmt dementsprechend die Form  $E(R) > \rho > \rho_c$  an. Wird hingegen bei der Übertragung eines an einer Stichprobe angepassten Modells auf eine neue Stichprobe der gleichen Population das Bestimmtheitsmaß empirisch bestimmt (dann  $R_c$ ), stellt dessen Erwartungswert bei Wiederholung dieses Vorgehens an (unendlich) vielen Stichproben eine Annäherung an  $\rho_c$  dar:  $E(R_c) \approx \rho_c$  (Claudy, 1978; Cummings, 1982; Herzberg, 1969). In der psychologischen Forschung wird jedoch  $R^2$  oftmals als unkorrigierter Schätzer für  $\rho^2$ , respektive  $\rho_c^2$  herangezogen.

Zwar konnten Yin und Fan (2001) in einer umfangreichen Literaturrecherche insgesamt 15 Korrekturformeln ausmachen, die genutzt werden könnten, um der resultierenden Überschätzung entgegenzuwirken, allerdings verwiesen die Autoren darauf, dass offenbar nicht nur eine gewisse Unsicherheit hinsichtlich der Bezeichnung und Verwendung dieser Adjustierungs-Formeln herrscht, sondern dass außerdem a) von gängiger Statistiksoftware i. d. R. nur eine einzelne, zur korrigierten Schätzung von  $\rho^2$  entwickelte Formel implementiert und b) sich deren Korrektur in einer berichteten Simulationsstudie als unzureichend erwies (zumeist als *Wherry-Formel* bezeichnet, vgl. Yin & Fan, 2001). Es muss deshalb davon ausgegangen werden, dass die in der psychologischen Forschung als Effektstärkemaße berichteten Determinationskoeffizienten selbst bei Angabe eines korrigierten Maßes als Schätzer der entsprechenden Populationswerte ungeeignet sind und insbesondere im Falle der angestrebten Übertragung eines Modells auf neue Daten der gleichen Population, d. h. bei einer Schätzung von  $\rho_c^2$ , eine deutliche Überschätzung erfolgt (vgl. auch Stelzl, 2005, S. 129 ff.).

In zunehmendem Umfang verfügbare Rechenkapazitäten erlauben es mittlerweile jedoch, Schätzungen dieser Größen in vertretbarer Zeit über eine empirische Kreuzvalidierung oder andere Resampling-Verfahren, d. h. Stichprobenwiederholungen vorzunehmen. Das entsprechende Vorgehen wird im folgenden Abschnitt skizziert.

## Resampling: Kreuzvalidierung

Eine Abschätzung der Übertragbarkeit erstellter Modelle auf neue, zukünftige Datensätze sollte aufgrund der beschriebenen Eigenschaften des Vorhersagefehlers idealerweise an ebensolchen demonstriert werden. Dies hat gegenüber korrigierten Anpassungsmaßen den Vorteil, dass ein unverfälschter, direkter Schätzer für den Vorhersagefehler resultiert und keine zusätzlichen Vorannahmen in mögliche Korrekturen einfließen müssen. Außerdem ist der entsprechende Ansatz flexibel und in unterschiedlichen Anwendungsfällen der Modellauswahl einsetzbar, selbst wenn beispielsweise Modelle betrachtet werden, deren exakte Prädiktorenanzahl nicht bestimmt oder Fehlervarianzen nur schwer geschätzt werden können (James, Witten, Hastie & Tibshirani, 2013, S. 213).

Im Rahmen dieses Vorgehens sollte (mindestens) ein Datensatz zur Modellierung, d. h. zur Schätzung der Modellparameter (engl. *training set*, *Trainingsstichprobe*) und ein zweiter, unabhängiger Datensatz zur Abschätzung des Vorhersagefehlers (engl. *test set*, *Teststichprobe*) herangezogen werden. In Fällen, in denen eine Auswahl zwischen mehreren konkurrierenden Modellen getroffen werden soll, wird empfohlen, einen zusätzlichen Datensatz vorzuhalten (engl. *validation set*, *Validierungsstichprobe*). Zunächst werden dann die Modellparameter an der Trainingsstichprobe geschätzt, durch eine Vorhersage an der Validierungsstichprobe werden die Vorhersagefehler der unterschiedlichen Modelle ermittelt und verglichen. Schließlich wird mithilfe der Teststichprobe eine Abschätzung der Vorhersagegenauigkeit des favorisierten Modells an neuen, bis dato unbekanntem Daten vorgenommen (Hastie et al., 2011, S. 219 ff.). Dieses Vorgehen ist (bisher) in der psychologischen Praxis aufgrund der meist beschränkten Verfügbarkeit von Daten aber nicht immer verlässlich zu realisieren.<sup>1</sup>

Mit dem Aufkommen zunehmend leistungsfähiger Computertechnologie wurden deshalb Verfahren entwickelt, die einen vergleichbaren Ansatz der Modellprüfung auch an begrenzten Datensätzen erlauben, die sog. *Resampling*-Verfahren. Der Begriff Resampling beschreibt hierbei die gemeinsame Eigenschaft unterschiedlicher Methoden, die darin besteht, dass aus den verfügbaren Daten viele Datensätze generiert werden, die zwar (in unterschiedlicher Form) auf den vorhandenen Daten basieren, aber eine generalisierende Betrachtungsweise ermöglichen, welche mit der Erhebung neuer Datensätze vergleichbar ist. Ursprünglich kamen diese Verfahren zur Ermittlung der Genauigkeit einzelner Parameterschätzer zum Einsatz (z. B. Bootstrap-Konfidenzintervalle), im maschinellen Lernen stellen sie außerdem den de facto Standard zur Überprüfung von Vorhersagemodellen dar.

Die gängigsten Verfahren sind der soeben genannte *Bootstrap* und die *Kreuzvalidierung*, wobei sich hinter diesen beiden Begriffen wiederum eine Vielzahl unterschiedlicher

---

<sup>1</sup>In entsprechenden Voranalysen wurden auch für die vorliegende Arbeit wiederholt zufällige Aufteilungen des verfügbaren Datensatzes in Trainings- und Teststichproben im Verhältnis  $3/4$  zu  $1/4$  vorgenommen — also  $n_{\text{train}} \approx 527$ , bzw.  $n_{\text{test}} \approx 175$ . Nach einer kreuzvalidierten Parameterschätzung unterschiedlicher Modelle in der Trainingsstichprobe wurde die Genauigkeit deren Vorhersage anhand ihrer Passung in der Teststichprobe ermittelt. Selbst bei diesen Stichprobengrößen ergaben sich in Folge unterschiedlicher, zufallsbedingter Stichprobenaufteilungen teils nicht vernachlässigbare Schwankungen, die unterschiedliche Interpretationen sowohl in Hinblick auf die Modellauswahl, als auch die jeweilige Präzision der Vorhersage nahelegten.

Ausgestaltungen der jeweiligen Methode verbergen. Das Prinzip des Bootstrap basiert auf wiederholten Ziehungen aus den ursprünglichen Daten, wobei die gezogenen Beobachtungen jeweils wieder „zurückgelegt“ werden. Dieses Vorgehen wird solange wiederholt, bis ein „neuer“ Datensatz in der gleichen Größe vorliegt wie der ursprüngliche Datensatz. Dieser kann nun einzelne Beobachtungen mehrfach, andere gar nicht enthalten. Die Schätzung der Modellparameter erfolgt dann an dieser Bootstrap-Stichprobe (engl. *bootstrap sample*), die also als Trainingsstichprobe fungiert. Der Vorhersagefehler derart angepasster Modelle wird anschließend für die Beobachtungen ermittelt, die nicht in diesem neuen Datensatz enthalten waren (und dementsprechend als Teststichprobe dienen, sog. *out of bag sample*).

Im Falle einer Kreuzvalidierung werden die verfügbaren Daten in eine vorgegebene Anzahl an Teildatensätzen aufgeteilt. Folgt man einer üblichen Unterteilung, werden zehn Teildatensätze gebildet (man spricht dann von einer *10-fold cross validation* — allgemein *k-fold cross validation*, also *k-fach Kreuzvalidierung*). Die Validierung eines zu prüfenden Modells erfolgt, indem an jeweils zusammengenommen neun dieser Datensätzen eine Schätzung der Modellparameter vorgenommen und damit eine Vorhersage der Beobachtungen des zehnten Teildatensatzes getroffen wird, bei der dann der Vorhersagefehler (etwa in Form der mittleren quadratischen Abweichung) ermittelt wird. Dieses Vorgehen wird derart wiederholt, dass jeder der zehn Teildatensätze zur Ermittlung des Vorhersagefehlers herangezogen wird (sie dienen in diesem Sinne alle einmalig als Teststichprobe), während in den jeweils neun verbleibenden Teildatensätzen die Parameterschätzung erfolgt (sie bilden dann also jeweils die Trainingsstichprobe). Die so in den zehn Durchgängen erhaltenen Vorhersagefehler werden gemittelt, um eine Gesamtaussage über die Vorhersagepräzision des geprüften Modellierungsansatzes zu erhalten.

Eine Erweiterung dieses Prozederes stellt die sog. *wiederholte k-fach Kreuzvalidierung* (engl. *repeated k-fold cross validation*) dar, für welche der beschriebene Vorgang mehrfach wiederholt wird, also beispielsweise zehn mal eine Aufteilung des Gesamtdatensatzes in mehrere Teildatensätze sowie die beschriebene Schätzung der Parameter, bzw. des Vorhersagefehlers erfolgt (Bischof, Mersmann, Trautmann & Weihs, 2012; Simon, 2007). Würde beispielsweise eine 10-fach Kreuzvalidierung zehnmal wiederholt werden, wären es letztlich 100 unterschiedliche Datenkonstellationen, an denen eine Modellschätzung und -überprüfung erfolgen würde — allesamt basierend auf den ursprünglich vorliegenden Daten. Mithilfe dieses Vorgehens können unterschiedliche Modelle angepasst und hinsichtlich ihrer Vorhersagegenauigkeit geprüft werden. Im folgenden Abschnitt werden solche Modelle vorgestellt, die als Erweiterungen des in der psychologischen Forschung häufig eingesetzten linearen Modells angesehen werden können.

## Shrinkage- und Regularisierungsmodelle

In der aktuellen, dem maschinellen Lernen gewidmeten Statistikkultur werden vielfältige Modellierungsansätze diskutiert (für eine Einführung siehe: James et al., 2013). Neben unterschiedlichen Ausgestaltungen und Erweiterungen linearer Modelle (*Generalisiertes Lineares Modell (GLM)*, siehe Abschnitt 2.2.3) werden im Kontext der prädiktiven Modellierung (engl. *predictive modeling*) auch Methoden eingesetzt, die grundsätzlich nicht

zwingend von einem linearen Zusammenhang zwischen Prädiktor- und Kriteriumvariablen ausgehen (z. B. Entscheidungsbäume und Random Forest, Support Vector Machine, künstliche neuronale Netzwerke). Diese Verfahren erlauben teils Vorhersagen, die in ihrer Vorhersagepräzision über die Möglichkeiten der zuvor genannten Modelle hinausgehen können, allerdings oftmals nur bedingt nachvollziehbare Rückschlüsse auf funktionale Zusammenhänge zwischen Prädiktoren und Kriterium zulassen und deshalb auch als *Black-Box-Modelle* bezeichnet werden (Breiman, 2001b).

Jedoch lassen sich bereits auch durch Modelle, die „nur“ die Komplexität (regulärer) multipler Regressionsmodelle reduzieren (und damit im Sinne des Bias-Varianz-Dilemmas die Varianz der Vorhersagen tendenziell stärker verringern, als der Bias eines solchen, weniger flexiblen Modells zunimmt), teils verbesserte Vorhersagen unter Beibehaltung der vergleichsweise einfachen Interpretationsmöglichkeiten linearer Modelle erzielen. Einzelne entsprechende Verfahren, bei denen aus einem Satz von  $p$  Prädiktoren eine Teilmenge  $m$  ausgewählt wird, haben teilweise zwar auch schon in der psychologischen Forschung Beachtung erhalten (*best subset selection*, *Vorwärts-/Rückwärtsregression* und *Schrittweise Regression*), sind aber mit unterschiedlichen Einschränkungen verbunden. Im Falle der *best subset selection* werden sämtliche Modelle mit allen möglichen Kombinationen aus den  $p$  Prädiktoren ermittelt. Dieses Vorgehen ist äußerst rechenintensiv und selbst auf leistungsfähigen Computern für nicht mehr als etwa 20 bzw. 40 Prädiktoren (nur im Falle der linearen multiplen Regression, mit optimierten Algorithmen) realisierbar. Die Vorwärts-, Rückwärts- sowie Schrittweise Regression sind computational weniger aufwändig und stehen deshalb auch in gängiger Statistiksoftware zur Verfügung (Hastie et al., 2011, S. 57 ff.). In der psychologischen Literatur werden auf diese Weise ermittelte Prädiktorteilmengen meist aber nicht an neuen Daten validiert, bzw. ihre Vorhersagegenauigkeit nicht überprüft. Es erfolgt dementsprechend dann eine Optimierung und Auswahl der Parameter an einem einzelnen Datensatz, worunter in vielen Fällen die Generalisierbarkeit der Vorhersagegüte des resultierenden Modells leidet (vgl. Überanpassung).

Die nachfolgend vorgestellten und in dieser Untersuchung angewendeten Verfahren sollten im Gegensatz dazu deshalb in das beschriebene Vorgehen eingebettet betrachtet werden, bei welchem die zur Modellierung verfügbaren Daten nicht nur zur Schätzung der Modellparameter herangezogen werden, sondern auch, um eine Ermittlung der Vorhersagegüte des jeweiligen Modells in der Verallgemeinerung auf potentiell neue Datensätze vorzunehmen.

Grundsätzlich gilt, dass der Vorhersagefehler linearer Modelle i. d. R. einen relativ geringen Varianzanteil aufweist und dann einem geringen Bias unterliegt, wenn tatsächlich ein linearer Zusammenhang zwischen den Prädiktoren sowie dem Kriterium besteht, d. h. wenn eine Veränderung in den Prädiktoren immer zu einer proportionalen Veränderung im Kriterium führt und alle weiteren Modellannahmen erfüllt sind (vgl. Kuhn & Johnson, 2013, S. 122 ff.). Hinsichtlich der Stabilität der Parameterschätzer muss aber darüber hinaus das Verhältnis der verfügbaren Information zur Anzahl der Modellparameter berücksichtigt werden. Das heißt, die vorhandenen Beobachtungen müssen zur Anzahl der Prädiktoren des Modells, bzw. deren Parametern in Bezug gesetzt werden. Übersteigt die Anzahl der Beobachtungen  $n$  deutlich die Anzahl der im Modell aufgenommenen Variablen  $p$  (wie in

der vorliegenden Studie) und weisen diese keine hohen Interkorrelationen auf (Multikollinearität), führt der Schätzer der kleinsten Quadrate i. d. R. auch zu stabilen Parameterschätzungen. Insbesondere wenn die Anzahl der Beobachtungen  $n$  jedoch nicht deutlich größer ist als die Anzahl an Modellparametern  $p$ , weisen die Parameterschätzungen eine hohe Variabilität auf (und das resultierende, womöglich überangepasste Modell ist ggf. nur schlecht, d. h. mit hohem Vorhersagefehler auf neue Datensätze übertragbar). Im Falle von mehr Modellparametern  $p$  als Beobachtungen  $n$  gibt es keine eindeutige Lösung für die Parameterschätzung durch das Prinzip der kleinsten Quadrate, d. h. eine herkömmliche Regressionsanalyse ist gar nicht durchführbar.

In diesen Situationen kann entweder eine Teilmenge an  $m < n$  Prädiktorvariablen ausgewählt und für diese mit der Methode der kleinsten Quadrate eine Parameterschätzung vorgenommen werden (vgl. Ausführungen oben) oder die Parameterschätzer bestimmten Einschränkungen unterworfen (engl. *constraints*) und in der Folge geschrumpft werden (engl. *shrinkage*), um eine Schätzung zu ermöglichen. Die Gruppe entsprechender Verfahren werden deshalb als *Shrinkage*-Verfahren bzw. *Regularisierungsmethoden* bezeichnet. Sie unterwerfen alle potentiellen Prädiktoren bestimmten Einschränkungen, wodurch diese in der Höhe ihrer Parameterschätzer reduziert und ggf. gleich Null gesetzt werden. Hierdurch kann auch die Interpretierbarkeit eines Modells durch die Entfernung redundanter oder unbedeutender Prädiktoren erhöht werden. Sie wirken auf diese Weise außerdem einer möglichen Überanpassung durch zu viele Prädiktoren und instabilen Parameterschätzungen entgegen, wie sie etwa durch Multikollinearität verursacht werden können. Wie angedeutet dienen diese Modellierungsansätze schließlich dazu dienen, den Varianzanteil des Vorhersagefehlers derart zu reduzieren, dass er die gleichermaßen bedingte, aber potentiell geringere Zunahme des Bias übersteigt und insgesamt zu einer besseren Vorhersagegenauigkeit eines auf diese Weise erhaltenen Modells führt (vgl. Hastie et al., 2011, S. 61 ff.; Hastie, Tibshirani & Wainwright, 2015; James et al., 2013, S. 214 ff.).

Das älteste dieser Verfahren ist die sog. *Ridge Regression* (Hoerl & Kennard, 1970). Hierbei wird nicht ausschließlich die Summe der Fehlerquadrate minimiert, sondern die Optimierung der Parameterschätzung wird ergänzt um einen weiteren Term, der den sog. *Komplexitäts-* oder *Tuning-*Parameter  $\lambda$  enthält. Die Schätzung der Parameter  $\beta_0^{\text{ridge}}$ ,  $\beta_1^{\text{ridge}}$ , ...,  $\beta_p^{\text{ridge}}$  im Rahmen der Ridge Regression minimiert demnach den folgenden Ausdruck:

$$PRSS_R = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (2.7)$$

Der zusätzlich Bestandteil  $\lambda \sum_j \beta_j^2$  dieser Formel hat die Funktion eines Strafterms, der die geschätzten Regressionskoeffizienten in ihrer Größe einschränken kann (engl. *shrinkage penalty*). Die Stärke dieser Einschränkung wird durch den (mithilfe eines geeigneten Resampling-Verfahrens) zu optimierenden Parameter  $\lambda$  bestimmt: ist  $\lambda$  gleich Null, entspricht die Parameterschätzung der Ridge Regression dem Regressionsschätzer nach der Methode der kleinsten Quadrate. Je höher die Werte sind, die  $\lambda$  annimmt, desto stärker wird der Einfluss des Strafterms und eine Minimierung der Formel 2.7 führt zu einer immer kleineren Schätzung der Regressionsparameter, bis sie sich schließlich für  $\lambda \rightarrow \infty$  dem Wert Null



annähern. Der Vorteil dieser Erweiterung gegenüber einer Schätzung gemäß dem Prinzip der kleinsten Quadrate liegt begründet im zuvor diskutierten Zusammenwirken von Bias und Varianz eines Modells. Eine nicht weiter eingeschränkte Schätzung der Parameter aller Prädiktorvariablen resultiert in vielen Fällen in einem Modell, das bei Bestehen eines linearen Zusammenhangs zwar als erwartungstreu betrachtet werden kann (d. h. einen niedrigen Bias besitzt), dessen Flexibilität aber nicht gänzlich vernachlässigt werden kann und dessen Vorhersagen folglich in Abhängigkeit der zur Modellierung herangezogenen Daten hohen Streuungen unterworfen sein können. Eine Einschränkung der Parameterschätzer in Form einer Reduktion im Rahmen der Ridge Regression führt tendenziell zu einer Verzerrung der Modellvorhersagen (d. h. einem erhöhtem Bias), reduziert gleichzeitig aber in stärkerem Maße die Varianz der Vorhersagewerte und verbessert dadurch insgesamt die Genauigkeit der Vorhersage. In der gegenläufigen Beziehung von Bias und Varianz gilt es demnach den Wert für den Parameter  $\lambda$  zu bestimmen, der durch die einhergehende Einschränkung in den Schätzern der Regressionsparameter zu einem Minimum im Vorhersagefehler führt.

Eine von Tibshirani (1996) vorgestellte Alternative zur Ridge Regression ist der *Least Absolute Shrinkage and Selection Operator (Lasso)*. Hierzu wird durch eine entsprechende Schätzung der Parameter  $\beta_0^L, \beta_1^L, \dots, \beta_p^L$  die folgende Größe minimiert:

$$PRSS_L = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (2.8)$$

Anstelle des Strafterms  $\lambda \sum_j \beta_j^2$  der Ridge Regression findet im Rahmen des Lasso der Ausdruck  $\lambda \sum_j |\beta_j|$  Verwendung. Dies hat zur Folge, dass die Parameterschätzer  $\beta^L$  nicht nur in ihrer Größe reduziert werden, sondern tatsächlich den Wert Null annehmen können und in diesem Fall aus dem Modell entfernt werden. Es findet ggf. also auch eine Auswahl der im Modell verbleibenden Variablen statt. Ein derart sparsameres Modell kann dem Parsimonitätsprinzip entsprechen und ist gegebenenfalls einfacher zu interpretieren als ein Modell mit einer größeren Zahl an Prädiktoren.

Eine zusätzliche Erweiterung (und Verbindung) dieser Methoden stellt das *Elastic Net* dar. Es kombiniert die Strafterme der beiden zuvor genannten Ansätze und führt dadurch sowohl zu einer allgemeinen Einschränkung, als auch zu einer Auswahl relevanter Parameter. Dieses Verfahren soll demnach auch im Falle hoch korrelierter Prädiktoren zu zuverlässig geschätzten Modellen führen (Zou & Hastie, 2005). Die Parameter  $\beta_0^E, \beta_1^E, \dots, \beta_p^E$  werden über die Minimierung des folgenden Ausdrucks geschätzt:

$$PRSS_E = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \left( \alpha \beta_j^2 + (1 - \alpha) |\beta_j| \right) \quad (2.9)$$

Neben dem bereits bekannten  $\lambda$ -Parameter enthält dieser Strafterm den Parameter  $\alpha$ , der die anteilige Gewichtung der beiden zuvor vorgestellten Terme bestimmt.  $\alpha$  stellt wie  $\lambda$  einen Tuning-Parameter dar, der gezielt ermittelt werden muss (siehe 2.2.3 für eine Beschreibung der praktischen Umsetzung). In der hier vorgestellten Parametrisierung ent-

spricht das Elastic Net für  $\alpha = 0$  der Lasso-Schätzung,  $\alpha = 1$  führt zu einer Schätzung analog zur Ridge Regression.<sup>2</sup>

### 2.1.7 Annahmen der Untersuchung

Die berichteten theoretischen Überlegungen, dargestellten Vorbefunde und erläuterten methodischen Entwicklungen führen für die nachfolgend berichteten Betrachtungen zu den folgenden untersuchungsleitenden Annahmen, die im Weiteren einer Überprüfung unterzogen werden:

#### Annahme 1

Krankheitsbedingte und unentschuldigte Fehltage während der beruflichen Ausbildungsphase können in kombinierter Betrachtung (Absentismus) durch eignungsdiagnostisch erfasste Persönlichkeitsmerkmale vorhergesagt werden. Vor dem Hintergrund des bandwidth-fidelity Dilemmas erscheint es unter Beachtung der bandwidth-matching Hypothese aussichtsreich, bei diesem Kriterium auch die Prädiktorseite vergleichbar spezifisch zu betrachten, d. h. auf der Facettenebene menschlicher Persönlichkeit im Sinne des NEO-PI-R.

#### Annahme 2

Ein Einfluss von Intelligenz als allgemein etablierter Prädiktor der beruflichen Leistungsfähigkeit wird vermutet und — im Rahmen der aktuellen Debatte zur Betrachtung allgemeiner versus differenzierter Fähigkeiten — ebenfalls in Form spezifischer kognitiver Fähigkeiten mitberücksichtigt.

#### Annahme 3

Das gewählte methodische Vorgehen erlaubt einen Vergleich der resultierenden Modelle hinsichtlich der Genauigkeit ihrer Vorhersagen. Die vorgestellten Eigenschaften legen nahe, dass Shrinkage- bzw. Regularisierungsmethoden stabilere Modellschätzungen ermöglichen, als ein multiples Regressionsmodell. In Abhängigkeit der Beziehungen zwischen den einzelnen Prädiktoren sowie ihren jeweiligen Zusammenhängen mit dem Kriterium können entweder Ridge Regression, Lasso oder Elastic Net zu präzisen Vorhersagen führen.

#### Addendum

An diese Annahmen anknüpfende Betrachtungen werden analog für zusätzliche Modelle (und weitere Indikatoren der Leistungsfähigkeit während der beruflichen Ausbildungsphase — Vorgesetztenbeurteilungen, Berufsschulnoten und die Leistung in der Abschlussprüfung) vorgenommen, um eine vergleichende Beurteilung der verwendeten Auswertungsverfahren und Modellierungstechniken zu ermöglichen.

---

<sup>2</sup>Im R-Paket `glmnet` (Friedman, Hastie, Simon & Tibshirani, 2015) wurde der Elastic Net Strafterm in der Form  $\lambda \sum_j (1 - \alpha) / 2\beta_j^2 + \alpha|\beta_j|$  implementiert. Dies macht eine entgegengesetzte Interpretation der  $\alpha$ -Werte notwendig:  $\alpha = 1$  entspricht dann dem Lasso-Strafterm,  $\alpha = 0$  der Ridge Regression.

## 2.2 Methode

In den folgenden Abschnitten werden zunächst die wesentlichen Charakteristika des vorliegenden Datensatzes (2.2.1) sowie der verwendeten Messinstrumente (2.2.2) vorgestellt. Daran anschließend werden in Kapitel 2.2.3 der vollzogene Auswertungsprozess beschrieben und einzelne methodische Aspekte vertiefend dargelegt.

### 2.2.1 Datenerhebung und Stichprobenzusammensetzung

Der betrachtete Datensatz umfasst  $N = 702$  Jugendliche und junge Erwachsene, die in den Jahren 2008 bis 2010 eine Berufsausbildung beim Industriepartner dieser Untersuchung oder einem mit diesem kooperierenden Unternehmen angetreten hatten. Es handelt sich um 247 Frauen und 455 Männer im Alter von 14 bis 31 Jahren ( $M = 17.40$ ;  $SD = 2.19$ ). Diese gingen einem von 37 Ausbildungsberufen nach. Aufgrund einer entsprechenden Branchenzugehörigkeit der beteiligten Unternehmen entstammt der Großteil der betrachteten Ausbildungsplätze der chemischen Industrie: die größte Gruppe an Auszubildenden waren Chemikantinnen und Chemikanten ( $n = 135$ ), gefolgt von Chemie- ( $n = 107$ ) und Biologie-laborantinnen und -laboranten ( $n = 71$ ). Das Geschlecht der Bewerberinnen und Bewerber sowie deren Alter zum Zeitpunkt der Einstellungsuntersuchung (siehe unten) wurden in den weiteren Analysen als Kontrollvariablen berücksichtigt (vgl. Tabelle 2.1).

Die in der Stichprobe enthaltenen Auszubildenden durchliefen einen mehrstufigen Auswahlprozess, bei dem zunächst ihre Bewerbungsunterlagen geprüft wurden und eine erste Vorauswahl der Bewerberinnen und Bewerber stattfand. Hierbei wurden der Lebenslauf und die Schulzeugnisse hinsichtlich bestimmter Fachnoten und etwaiger Fehlzeiten geprüft. Anschließend wurden geeignete Kandidatinnen und Kandidaten zu einem Einstellungstest eingeladen, bei dem neben kognitiven auch weitere spezifische Fähigkeiten erfasst wurden. Außerdem wurden zu diesem Zeitpunkt Geschlecht und Alter der Bewerberinnen und Bewerber festgehalten. Mit Beginn des Auswahlzeitraums für den Ausbildungsjahrgang 2008 wurde den Bewerberinnen und Bewerbern zusätzlich zu diesem etablierten Vorgehen ein Verfahren zur Erfassung der Big Five-Persönlichkeitsstruktur vorgegeben. Die Resultate dieser Erhebung wurden nicht (bzw. bei späteren Jahrgängen nur zum seltenen Ausschluss von einzelnen Kandidaten mit spezifischen Extremwerten, d. h. als Kriterium im Sinne einer Negativselektion) zur Auswahl herangezogen. Auf Basis der Ergebnisse in den Leistungstests fand anschließend wieder eine Auswahl an Kandidatinnen und Kandidaten statt, die in einer dritten Bewerbungsstufe zu einem teilstrukturierten Interview eingeladen wurden. Diese, von Mitarbeiterinnen und Mitarbeitern der Personalabteilungen und Ausbildungsleiterinnen und -leitern geführten Gespräche, bildeten schließlich die finale Grundlage zur Entscheidung, welchen Personen ein Ausbildungsplatz angeboten wurde. Die Erhebung der Daten dieses Auswahlprozesses begann im September 2007 und wurde bis September 2010 fortgesetzt, die Erfassung aller weiteren in der vorliegenden Untersuchung betrachteten Variablen erfolgte im Frühjahr 2014.

## 2.2.2 Verwendete Instrumente

Die im Rahmen des zuvor beschriebenen Auswahlprozesses eingesetzten psychologischen Testverfahren werden in den folgenden Abschnitten kurz vorgestellt. Danach erfolgt die Beschreibung der primären Kriteriumsvariable.

### Persönlichkeitseigenschaften

Zur Erfassung der Persönlichkeitseigenschaften der Bewerberinnen und Bewerber kam das NEO-PI-R (Ostendorf & Angleitner, 2004) zum Einsatz. Dieses Verfahren erlaubt eine umfassende, dem FFM der Persönlichkeit folgende und zugleich detaillierte Beschreibung der Persönlichkeit auf Facettenebene (vgl. Tabelle 1.1). Der Antwortbogen des NEO-PI-R wurde mit Zustimmung der Rechteinhabenden im Vergleich zur kommerziell erhältlichen Version modifiziert und in einer kompakten Form vorgelegt. Für die Beantwortung der insgesamt 240 Items des Verfahrens waren auf zwei Seiten jeweils 30 Antwortvorgaben in vier Spalten vorgesehen. Das Antwortformat stimmte mit dem Originalverfahren überein. Es handelte sich um 5-stufige Likert-Skalen, auf denen die Kandidatinnen und Kandidaten ihren Einstellungen zu den enthaltenen Aussagen in den Abstufungen von „Starke Ablehnung“ über „Neutral“ bis „Starke Zustimmung“ Ausdruck verleihen konnten. Das Verfahren wurde gemäß den Vorgaben des Manuals ausgewertet und einzelne Items dementsprechend derart umkodiert, dass die ermittelten Summen bei zunehmender Größe einer höheren Ausprägung der jeweiligen Persönlichkeitseigenschaft entsprechen. Die Summenrohwerte jeweils aller sechs Facetten der fünf Persönlichkeitsdimensionen dienen als Grundlage der nachfolgend berichteten Analysen.

### Kognitive Fähigkeiten

Als Intelligenztest wurde im dargestellten Auswahlprozess der sog. *Höchster-Intelligenztest (HIT)* eingesetzt (Heß, 1994). Es handelt sich dabei um eine hausinterne Weiterentwicklung des Intelligenz-Strukturtest 70 (IST-70) von Amthauer (1973). Konkret wurde die kognitive Leistungsfähigkeit in Form der Faktoren verbale (mit den Aufgabengruppen Satzergänzung, Wortauswahl und Analogien), numerische (Rechenaufgaben und Zahlenreihen) und figural-räumliche Intelligenz (Figurenauswahl und Schlauchfiguren) erfasst. Weiterhin wurde durch die Vorgabe entsprechender Aufgaben die Merkfähigkeit der Bewerberinnen und Bewerber ermittelt. Die individuellen Testwerte ergaben sich aus den Summen der Richtiglösungen in den unterschiedlichen Aufgabengruppen — eine stärkere Ausprägung der jeweiligen Fähigkeit findet dementsprechend in zunehmend höheren Testwerten Ausdruck.

### Absentismus

Als primäre abhängige Variable wurden die krankheitsbedingten und unentschuldigten Fehltag betrachtet. Diese wurden über den Zeitraum der Ausbildung für alle Auszubildenden erfasst und in einer zentralen Datenbank vorgehalten. Diese beiden Maße des Absentismus wurden zusammengefasst, indem sie aufaddiert und vereinzelt registrierte Halb-

tage zu vollen Tagen aufgerundet wurden. Im nachfolgenden Abschnitt 2.2.3 wird eine Schilderung des weiteren Auswertungsprozesses vorgenommen, inklusive der Beschreibung einer durchgeführten Transformation der hier berichteten Daten. Weiterhin hinzugezogene Kriteriumsvariablen werden in Abschnitt 2.3.3 vorgestellt.

### 2.2.3 Auswertungsprozess

Die zuvor formulierten Annahmen und Fragestellungen zielen auf eine möglichst präzise Vorhersage der Fehlzeiten während der beruflichen Ausbildungsphase ab. Dies spiegelt auch das Interesse der auszubildenden Unternehmen wieder, frühzeitig diejenigen Bewerberinnen und Bewerber zu identifizieren, die im weiteren Ausbildungsverlauf dazu neigen könnten, hohe Fehlzeiten zu akkumulieren. Eine Häufung von Fehlzeiten oder gar ein Abbruch der Ausbildung ist i. d. R. mit hohen Kosten für das Ausbildungsunternehmen verbunden (Wenzelmann & Lemmermann, 2012). Es wird mit den krankheitsbedingten und unentschuldigten Fehltagen deshalb die Betrachtung eines spezifischen erfolgsrelevanten Kriteriums im Rahmen des Ausbildungsverlaufs vorgenommen und dieses soll durch die im Auswahlprozess gesammelten Informationen bestmöglich vorhergesagt werden.

Eine deskriptive Analyse der Fehlzeiten legt nahe, dass ein großer Teil der betrachteten Auszubildenden keine oder nur wenige Fehlitage während des Ausbildungszeitraums ansammelte. Es wird deutlich, dass es sich bei den Fehltagen dementsprechend um ein Merkmal handelt, das eine stark linkssteile Verteilung aufweist (vgl. Tabelle 2.1 und Abbildung 2.1). Die Vorhersage eines solchen Kriteriums mithilfe eines Modellierungsansatzes, der eine Normalverteilung der abhängigen Variable (bzw. deren Residuen) annimmt, scheint in einem derartigen Fall wenig aussichtsreich (wird aber zu Vergleichszwecken ebenfalls erfolgen). Zwei Vorgehensweisen bieten erfolgversprechendere Ansätze:

- die Wahl eines Vorhersagemodells, dem eine passende Verteilungsannahme für die abhängige Variable zugrunde liegt oder
- die Transformation der (schiefen) Kriteriumsvariable mithilfe eines geeigneten Transformationsverfahrens und ihre Überführung in eine Variable, die (annäherungsweise) durch eine Normalverteilung modelliert werden kann.

Die vorliegende Untersuchung stellt beide Herangehensweisen vor. Einerseits werden die beobachteten Fehlitage unverändert als abhängige Variable herangezogen. Einen (augenscheinlich) geeigneten Modellierungsansatz stellt dann im vorliegenden Fall der Fehlitage als abzählbare Ereignisse im Zeitraum der beruflichen Ausbildungsphase ein GLM dar, für das eine (bedingte) Poissonverteilung der abhängigen Variable angenommen wird (sog. Poisson Regression). In einem alternativen Ansatz wird die abhängige Variable zunächst einer Yeo-Johnson-Transformation unterzogen und auf diese Weise (annähernd) in eine Normalverteilung überführt, bevor eine lineare Modellierung erfolgt.

In beiden Fällen wird das jeweilige multiple Regressionsmodell (mit allen potentiellen Prädiktoren) verglichen mit Modellen, in denen durch Ridge Regression, Lasso und Elastic Net die Regressionskoeffizienten beschränkt oder ganz auf Null reduziert werden. Wie

zuvor in Abschnitt 2.1.6 erläutert, erfordern die letztgenannten Modellierungsansätze eine Optimierung ihres Strafterms über die jeweiligen Tuning-Parameter  $\lambda$  und ggf.  $\alpha$ . Diese Optimierung wird im Rahmen eines geeigneten Verfahrens der Stichprobenwiederholung realisiert. Gleichzeitig soll auch, wie beschrieben, die über den vorliegenden Datensatz hinausgehende Vorhersagegenauigkeit für die unterschiedlichen Modelle ermittelt werden. Aus diesem Grund muss — ebenfalls mithilfe von Resampling — eine generalisierbare Abschätzung der Modellpassungen erfolgen. Die zur Umsetzung dieser Überlegungen notwendigen methodischen Grundlagen werden in den nachfolgenden Abschnitten vorgestellt.

### Yeo-Johnson-Transformation

Die Yeo-Johnson-Transformation (Yeo & Johnson, 2000) gehört zu den sog. *Potenz-Transformationen* (engl. *power transform*) und kann als Erweiterung in der Familie der Box-Cox-Transformationen (Box & Cox, 1964) angesehen werden. Im Gegensatz zu diesen kann die Yeo-Johnson-Transformation in den zu transformierenden Werten auch den Bereich gleich oder kleiner Null berücksichtigen. Mithilfe einer Maximum-Likelihood-Schätzung wird ein Parameter  $\lambda$  so bestimmt, dass gemäß nachfolgender Funktion eine Transformation der Ausgangswerte  $y_i$  in die näherungsweise normalverteilten Werte  $y_i^{(\lambda)}$  erfolgt:

$$y_i^{(\lambda)} = \begin{cases} \frac{(y_i+1)^\lambda - 1}{\lambda} & \text{für } y_i \geq 0 \\ \frac{(-y_i+1)^{2-\lambda} - 1}{\lambda} & \text{für } y_i < 0 \end{cases} \quad (2.10)$$

Generell dienen die Potenz-Transformationen durch die monotone Überführung schiefer Verteilungen in (annäherungsweise) Normalverteilungen der Vereinheitlichung von Varianzen und der verbesserten Interpretierbarkeit korrelativer Zusammenhänge. Sie können dadurch zur Erfüllung der Voraussetzungen (multipler) Regressionsanalysen wie der Annahme eines linearen Zusammenhangs und der Homoskedastizität der Residuen beitragen (Fox, 2015).

### Poisson Regression

Die Poisson Regression ist eingebettet in das Rahmenwerk der GLM (McCullagh & Nelder, 1989; Nelder & Baker, 2006). Hierbei handelt es sich um eine Verallgemeinerung des linearen Regressionsmodells, die Verteilungen aus der Exponentialfamilie für das Kriterium zulässt (d. h. nicht dessen Normalverteilung voraussetzt) und dieses über eine sog. Link-Funktion (d. h. nicht zwingend über eine identische Abbildung) mit der Linearkombination der Prädiktorvariablen verknüpft. Die Poisson Regression dient hierbei als Modellierungstechnik für Zähldaten, also abzählbare Ereignisse (Tutz, 2011). Es wird die Annahme getroffen, dass die (auf die Prädiktoren bedingte) abhängige Variable einer Poisson-Verteilung folgt —  $Y \sim P(\lambda)$  — und, im kanonischen Fall der üblicherweise gebräuchlichen Link-Funktion, der Logarithmus ihres Erwartungswertes  $E(Y) = \lambda$  durch eine Linearkombination der unabhängigen Variablen modelliert werden kann:

$$\log(\lambda) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2.11)$$

Es gilt zu beachten, dass die Poisson-Verteilung einzig über den Parameter  $\lambda$  definiert wird, der neben ihrem Erwartungswert auch ihre Varianz bestimmt:  $E(Y) = \text{var}(Y) = \lambda$ . Dies impliziert, dass bei einem größeren Mittelwert auch eine stärkere Streuung angenommen wird. Wenngleich dieser Zusammenhang zeitweise in realen Daten beobachtet werden kann, ist die Poisson-Verteilung in dieser Hinsicht nicht flexibel. Dies kann dazu führen, dass Fälle auftreten (insbesondere auch bei der Beobachtung besonders seltener Ereignisse), in denen zwar der Mittelwert einer beobachteten Verteilung (bzw. der entsprechende Populationsschätzer) angemessen ermittelt werden kann, gleichzeitig die dazugehörige Varianz aber (teils stark) unterschätzt wird, da tatsächlich eine größere Variabilität vorliegt. Dieses häufig auftretende Phänomen wird als *Überstreuung* (engl. *overdispersion*) bezeichnet. Eine weitere Herausforderung bei der Modellierung von Zähldaten entsteht in Situationen, in denen die abhängige Variable viele Nullen enthält, das entscheidende Ereignis also in vielen Fällen überhaupt nicht aufgetreten ist (sog. *zero-inflation*). Wenn eine oder gar beide dieser Eigenschaften auftreten, kann es hilfreich sein, auf alternative Modellierungsansätze zurückzugreifen, die teilweise als Erweiterung der Poisson Regression angesehen werden können und im Folgenden kurz vorgestellt werden.

### Weitere Modellierungstechniken

Ein Teil der nachfolgend vorgestellten Modelle dient wie angedeutet explizit dazu, im entsprechenden Bedarfsfall möglichen Schwächen der Poisson Regression zu begegnen und erlauben dann eine angemessenere Modellierung. Weiterhin wird mit dem Random Forest ein Verfahren knapp vorgestellt, das der computerwissenschaftlichen Tradition des maschinellen Lernens entstammt. Als solches hat es zunächst nicht das Ziel, eine inhaltlich erklärende Modellierung des in Form eines stochastischen Modells angenommenen Zusammenhangs zwischen Prädiktor- und Kriteriumsvariablen vorzunehmen. Vielmehr dient es dem Bestreben, algorithmisch eine möglichst präzise Vorhersage abhängiger Variablen auf Basis der in den unabhängigen Variablen enthaltenen Informationen zu erreichen (Breiman, 2001b). Es eignet sich bei entsprechender Verwendung allerdings auch zur Modellierung etwaiger nicht-linearer Zusammenhänge zwischen den Prädiktor- und Kriteriumsvariablen.

**Zero-inflated models:** Eine Möglichkeit zur Berücksichtigung gehäuft auftretender Nullen bei der Vorhersage von Zählwerten stellen die sog. *zero-inflated models* dar (Lambert, 1992; Mullahy, 1986). Es handelt sich hierbei um *mixture models*, denen die Annahme zugrunde liegt, dass die betrachteten Daten nicht aus einer homogenen Population stammen, sondern zwei Subpopulationen unterschieden werden können: eine, in der das abzählbare Ereignis grundsätzlich nicht auftritt und eine zweite, in der ein Auftreten möglich, aber nicht zwingend ist. Beide Subpopulationen tragen dementsprechend zur vermehrten Beobachtung von Nullen bei. Die resultierenden, latenten Mischverteilungen stellen die Ausgangsbasis für Modelle dieser Klasse dar.

**Hurdle models:** Die sog. *Hurdle*-Modelle (engl. für *Hürde*) erlauben eine explizite Modellierung von gehäuft auftretenden Nullen in den Daten und begegnen auf diese Weise ebenfalls dem geschilderten Problem einer möglichen Überstreuung bei der Poisson

Regression (Creel & Loomis, 1990; Mullahy, 1986). Für diese Modelle werden zwei Komponenten oder Teilmodelle kombiniert: zunächst wird anhand einer binären Entscheidung modelliert, ob eine Null oder ein anderer, positiver Wert auftritt. Im zweiten Fall wird dann bspw. eine entsprechend auf positive Fälle beschränkte Poisson- oder negative Binomialverteilung herangezogen, um eine angemessene Vorhersage zu treffen. Es handelt sich um einen schrittweisen Entscheidungsprozess, bei dem im ersten Schritt also eine metaphorische Hürde genommen werden muss, damit ein (positiv) von Null abweichender Zählwert modelliert wird.

**Negativ binomiale Regression:** Dem geschilderten Problem der Überstreuung kann außerdem begegnet werden, indem direkt auf ein Modell mit einer flexibleren Verteilungsannahme zurückgegriffen wird, als sie die Poisson-Verteilung aufweist. Oftmals wird in Hinblick auf diese Anforderung die negative Binomialverteilung gewählt, die im Gegensatz zur Poisson-Verteilung durch zwei Parameter definiert wird:  $NB(\nu, \mu)$ . Sie stellt folglich den Ausgangspunkt einer Negativen Binomialen Regression dar (Hilbe, 2011), in welcher zwar der Mittelwert  $\mu$  dem der Poisson Regression entspricht, die Varianz diejenige der Poisson Regression aber um den Faktor  $\mu^2/\nu$  übersteigt. Je kleiner  $\nu$  gewählt bzw. geschätzt wird, desto stärker unterscheidet sich die angenommene Streuung vom Mittelwert der resultierenden Verteilung. Aus diesem Grund wird  $1/\nu$  auch als Überstreuungs-Parameter bezeichnet.<sup>3</sup>

**Random Forest:** Ein *Random Forest (RF)* ist die Aggregation weitgehend unabhängiger Entscheidungsbäume (Breiman, 2001a). Im Kontext der Regression unterteilt ein einzelner Entscheidungsbaum den (mehrdimensionalen) Antwortraum auf Basis der unabhängigen Variablen in verschiedene Regionen, so dass über alle Beobachtungen (und Regionen) hinweg die RSS minimiert wird, wenn in den unterschiedlichen Regionen der Mittelwert der jeweiligen Beobachtungen zur weiteren Vorhersage genutzt wird. Zur Erstellung eines RF werden auf der Basis von Bootstrap-Stichproben mehrere tiefe Entscheidungsbäume generiert (meist mindestens 500), die der Einschränkung unterworfen sind, dass an jeder Verzweigung nicht alle ( $p$ ), sondern nur eine zufällige, beschränkte Auswahl ( $m$ ) der möglichen Prädiktoren berücksichtigt wird (bei Regressionen üblicherweise  $m = p/3$ ). Die resultierenden Entscheidungsbäume sind nur vergleichsweise gering miteinander korreliert und ihre (gemittelten) Vorhersagen sind von entsprechend reduzierter Fehlervarianz.

Der vorausgegangene Abschnitt bietet nur eine knappe, vereinfachte Übersicht der in dieser Untersuchung als Erweiterungen und Alternativen zu den vorausgehenden und in Abschnitt 2.1.6 ausführlicher besprochenen Verfahren herangezogenen Modellierungstechniken. Für detaillierte Erläuterungen dieser und der hier besprochenen Methoden sei auf

---

<sup>3</sup>Es gilt zu beachten, dass diese Eigenschaft ausschließlich zu unterschiedlichen Standardfehlern, nicht aber Punktschätzern der Modellparameter führt. Mit Ausnahme von schätzungs- bzw. implementierungsbedingten Abweichungen sollten sich in den Modellvorhersagen deshalb keine Unterschiede zur Poisson Regression ergeben.



Hastie et al. (2011), James et al. (2013), Kuhn und Johnson (2013), Tutz (2011) sowie Zeileis, Kleiber und Jackman (2008) verwiesen.

### Parameter Tuning und Nested Resampling

Dem Bedarf nach den benannten Zielen einer wiederholten Ermittlung der mittleren quadratischen Abweichung an verschiedenen Teststichproben und einer gleichzeitigen Optimierung der Strafterme einzelner Modelle wird, wie angedeutet, mit jeweiligen Stichprobenwiederholungen Rechnung getragen. Es ist hierzu notwendig, mehrere Resampling-Verfahren zu kombinieren. Dies wird in Form eines geschachtelten Resampling-Ansatzes (*nested resampling*) realisiert (Bischl et al., 2012; Simon, 2007).

Hierzu werden im computationalen Sinne zwei Instruktionsschleifen ineinander verschachtelt. Die innere Resampling-Schleife dient bei den entsprechenden Modellen (Ridge Regression, Lasso und Elastic Net) der Optimierung der Tuning-Parameter. In den vorliegenden Betrachtungen bedeutet das konkret, dass für alle einhundert (bzw. bei Elastic Net eintausend) vorgegebenen Werte des Parameterraums von  $\lambda$  (und  $\alpha$ ) durch eine 10-fach Kreuzvalidierung der jeweilige durchschnittliche mittlere quadratische Fehler der resultierenden Modelle ermittelt wird. Der/die Parameter/-kombination mit dem kleinsten Durchschnitt mittlerer quadratischer Abweichungen wird zur Modellierung an der gesamten Trainingsstichprobe und Schätzung des Vorhersagefehlers in der äußeren Resampling-Schleife übernommen.

Bei dieser handelt es sich in der hier berichteten Untersuchung um eine zehnmals wiederholte 10-fach Kreuzvalidierung. Für die auf diese Weise in der äußeren Schleife resultierenden einhundert Modellierungen wird wiederum der Durchschnitt des mittleren quadratischen Fehlers in den Teststichproben gebildet und dadurch eine Abschätzung der Vorhersagegenauigkeit des betrachteten Modellierungsansatzes gewonnen. Diese wird schließlich direkt zwischen den unterschiedlichen Modellen verglichen, um eine Entscheidung zugunsten der Modellierungstechnik mit dem kleinsten Durchschnittswert der mittleren quadratischen Abweichung zu fällen.

### Technische Umsetzung

Die Zusammenführung und Aufbereitung der über mehr als sechs Jahre hinweg gesammelten Daten erfolgte mit *IBM SPSS Statistics*, *LibreOffice Calc* (The Document Foundation, 2014) und *GNU R* (R Core Team, 2016). Die letztgenannte Software wurde weiterhin zur Durchführung aller beschriebenen Analysen verwendet. Hierbei kamen neben der Basisinstallation die nachfolgenden Pakete zum Einsatz.

Die Erstellung deskriptiv-statistischer Beschreibungen wurde durch die Verwendung von *psych* (Revelle, 2016) unterstützt. Für die Berechnung der Yeo-Johnson-Transformation wurde auf das Paket *car* (Fox & Weisberg, 2016) zurückgegriffen. *mlr* (Bischl et al., 2015) bot in Verbindung mit *parallelMap* (Bischl & Lang, 2015) das Rahmengerüst für die vorgenommenen Vergleiche unterschiedlicher Modellierungstechniken. Die Modelle wurden mithilfe der Pakete *glmnet* (Ridge Regression, Lasso und Elastic Net; Friedman et al.,

---

2015), `pascal` (Zero-inflated Modelle, Hurdle Modelle; Jackman, Tahk, Zeileis, Maimone & Fearon, 2015), `MASS` (Negativ Binomiale Regression; Ripley, 2015) sowie `randomForest` (Random Forest; Breiman, Cutler, Liaw & Wiener, 2015) angepasst. Grafische und tabellarische Ausgaben wurden mit `ggplot2` (Wickham & Chang, 2016), `GGally` (Schloerke et al., 2016) und `xtable` (Dahl, 2016) aufbereitet.

## 2.3 Ergebnisse

Die Ergebnisse der zuvor skizzierten Analysen werden im Folgenden berichtet. In Kapitel 2.3.1 wird zunächst eine Darstellung einleitender Betrachtungen vorgenommen, gefolgt von den zentralen Ergebnissen der erfolgten Modellierungen (Abschnitt 2.3.2) sowie ergänzenden Resultaten (Abschnitt 2.3.3).

### 2.3.1 Vorbetrachtungen

Eine Übersicht deskriptiv-statistischer Kennwerte aller Prädiktor- und Kriteriumsvariablen für die Vorhersage der (unveränderten) krankheitsbedingten und unentschuldigten Fehltag (im Weiteren *primäres Modell*) kann Tabelle 2.1 entnommen werden.

Die Häufigkeitsverteilung des Absentismus in Kombination der krankheitsbedingten und unentschuldigten Fehltag (als abhängige Variable) ist zusätzlich in Abbildung 2.1 grafisch dargestellt.

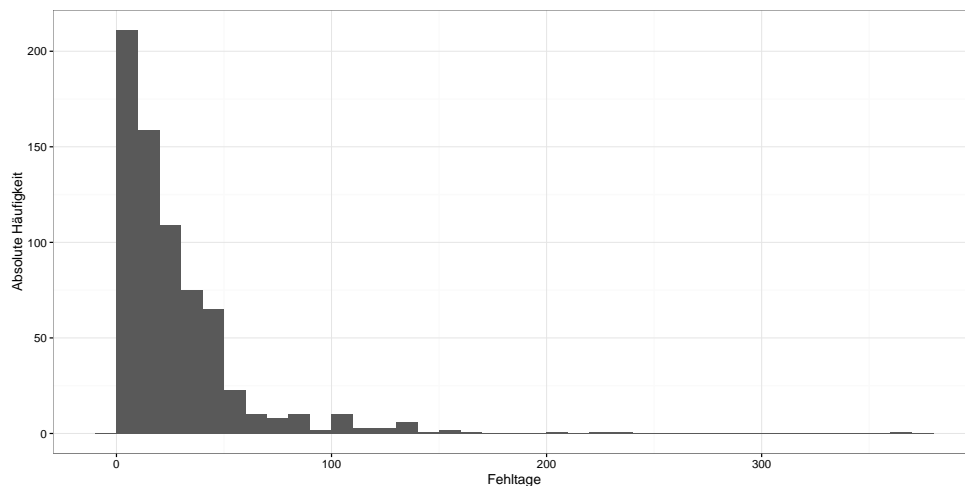


Abbildung 2.1: Häufigkeitsverteilung der krankheitsbedingten und unentschuldigten Fehltag.

Es wird deutlich, dass es sich um eine stark linkssteile Verteilung ( $M = 26.93$ , *Schiefe* = 3.76) mit großer Streuung ( $SD = 32.26$ ) und Spannweite ( $R = 369$ ) handelt. Ihre starke Streuung geht insbesondere im hohen Wertebereich auf einzelne Extremwerte zurück: insgesamt  $n = 16$  Personen blieben während ihrer Ausbildungszeit an mehr als 127 Tagen (entspricht etwa den Arbeitstagen eines halben Jahres im relevanten Zeitraum) der betrieblichen Ausbildung fern. Obwohl diese extremen Fälle als Ausreißerwerte deklariert werden könnten, wäre es unpassend, diese aus der weiteren Untersuchung auszuschließen. Es handelt sich schließlich genau um jene Fälle, die durch ihre starke Ausprägung bei einer Vorhersage der Fehlzeiten von besonderem Interesse für ein auswählendes Unternehmen sind. Derartig lange Fehlzeiten erfordern eine spezielle Aufmerksamkeit auf Unternehmensseite und angemessene Unterstützungsmaßnahmen oder sie führen gar zum Abbruch einer

Tabelle 2.1: Deskriptive Statistiken der im primären Modell enthaltenen abhängigen und unabhängigen Variablen.

Variable	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	<i>Schiefe</i>	<i>Kurtosis</i>
Fehltage	26.93	32.26	0.00	369.00	3.76	24.49
Geschlecht	0.35	0.48	0.00	1.00	0.62	-1.62
Alter	17.40	2.19	14.00	31.00	1.79	5.53
Verbal	13.44	2.08	4.67	19.00	-0.36	0.07
Numerisch	11.66	3.10	3.00	19.50	-0.02	-0.49
Figural-räuml.	12.12	2.73	4.50	19.00	-0.11	-0.33
Merkfähigkeit	13.81	3.75	3.00	20.00	-0.12	-0.72
N1: Ängstlichkeit	14.10	3.78	1.00	24.00	0.15	-0.10
N2: Reizbarkeit	11.17	3.57	2.00	27.00	0.44	0.76
N3: Depression	10.48	4.09	0.00	28.00	0.31	0.31
N4: Soziale Befangenheit	14.28	3.56	3.00	27.00	0.12	0.30
N5: Impulsivität	13.98	3.64	2.00	30.00	0.42	0.86
N6: Verletzlichkeit	9.83	3.19	0.00	24.00	0.17	0.72
E1: Herzlichkeit	23.61	3.06	11.00	32.00	-0.37	0.82
E2: Geselligkeit	21.91	4.19	8.00	32.00	-0.35	0.31
E3: Durchsetzungsfähigkeit	18.56	3.85	7.00	30.00	-0.13	-0.27
E4: Aktivität	18.44	2.78	8.00	27.00	-0.02	0.49
E5: Erlebnissuche	19.03	4.17	3.00	30.00	-0.34	0.16
E6: Positive Emotionen	22.52	3.81	6.00	32.00	-0.36	0.48
O1: Fantasie	16.54	3.88	2.00	29.00	0.25	0.08
O2: Ästhetik	18.47	5.02	4.00	31.00	-0.18	-0.29
O3: Gefühle	19.75	3.51	9.00	31.00	0.06	-0.17
O4: Handlungen	17.59	3.36	7.00	28.00	0.10	-0.19
O5: Ideen	20.95	4.18	2.00	32.00	-0.19	0.56
O6: Werte- u. Normensystem	18.20	2.76	7.00	28.00	0.32	0.67
A1: Vertrauen	19.90	3.30	5.00	31.00	-0.42	1.32
A2: Freimütigkeit	20.72	3.61	8.00	30.00	-0.38	0.10
A3: Altruismus	24.17	3.12	5.00	32.00	-0.65	2.28
A4: Entgegenkommen	19.21	3.63	5.00	30.00	-0.31	0.59
A5: Bescheidenheit	18.35	3.42	6.00	29.00	-0.24	0.50
A6: Gutherzigkeit	21.36	3.13	6.00	31.00	-0.27	1.53
C1: Kompetenz	22.26	2.98	12.00	31.00	-0.20	-0.01
C2: Ordnungsliebe	20.46	3.59	6.00	31.00	-0.26	0.23
C3: Pflichtbewusstsein	24.02	3.20	11.00	32.00	-0.23	0.01
C4: Leistungsstreben	21.42	3.40	11.00	31.00	0.01	-0.14
C5: Selbstdisziplin	23.15	3.75	10.00	32.00	-0.39	0.47
C6: Besonnenheit	19.57	3.93	8.00	32.00	-0.18	-0.04

*Anmerkung.* Krankheitsbedingte und unentschuldigte Fehltage, demographische Variablen (Geschlecht: 0 = Männer, 1 = Frauen; Alter in Jahren) sowie mit dem HIT bzw. NEO-PI-R erfasste Summenwerte der Intelligenzfaktoren und Persönlichkeitsfacetten.

Ausbildung und sind dann, wie dargestellt, für das Ausbildungsunternehmen mit erhöhten Kosten verbunden. Auch ist nicht davon auszugehen, dass derartige Fälle nicht genera-

lisierbar wären, d. h. auch für zukünftige Generationen von Auszubildenden muss davon ausgegangen werden, dass vereinzelt vergleichbar lange Fehlzeiten auftreten können und diese Fälle dementsprechend als repräsentativ angesehen werden können.

Eine Betrachtung der im Anhang B.1 angefügten grafischen Darstellungen der Zusammenhänge zwischen den Variablen im primären Modell (ab Seite 91) zeigt auf, dass die Fälle mit extremen Ausprägungen in der Kriteriumsvariable i. d. R. Ausprägungen im mittleren Bereich der unterschiedlichen unabhängigen Variablen aufweisen, sie also im Sinne der Modellierung keine hohen Hebelwerte darstellen und ihr Einfluss auf die Modellgestaltung dementsprechend begrenzt ausfällt.

Diese Einschätzung wurde durch regressionsdiagnostische Analysen (jeweils anhand des vollständigen multiplen, bzw. Poisson-Regressionsmodells) untermauert. Hierbei wurde deutlich, dass sich unter den (in überschaubarer Anzahl,  $n < 10$ ) auf Basis studentisierter Residuen potentiell als Ausreißer klassifizierbaren Fällen (vgl. Fox, 2015, S. 265 ff.) nur vereinzelt solche mit extremen Kriteriumswerten wiederfinden ließen. Ein möglicher Einfluss derartiger Fälle auf die Modellierung lässt sich dementsprechend nicht allein auf hohe Ausprägungen in der abhängigen Variable zurückführen und spricht aufgrund offensichtlich komplexerer Wechselwirkungen zusätzlich gegen einen vorschnellen Ausschluss einzelner Fälle.<sup>4</sup> Anhang B.2 (ab Seite 95) kann zusätzlich eine Korrelationstabelle mit den paarweisen Zusammenhängen zwischen den unabhängigen Variablen und allen Kriterien der nachfolgend präsentierten Modelle entnommen werden.

### 2.3.2 Hauptergebnisse

Die Passung der unterschiedlichen Modelle zur Vorhersage der primären abhängigen Variable, den krankheitsbedingten und unentschuldigtem Fehltagen, wird anhand der (über die insgesamt 100 Modellschätzungen der zehnmals wiederholten 10-fach Kreuzvalidierung im äußeren Resampling-Loop gemittelten) mittleren quadratischen Abweichung (MSE), bzw. deren Quadratwurzel (RMSE) bestimmt. Diese Kennwerte sind für die betrachteten Modelle in Tabelle 2.2 aufgeführt. In Abbildung 2.2 findet sich eine grafische Aufbereitung der mittleren quadratischen Abweichungen der Poisson Regression sowie für die drei zusätzlichen Shrinkage- bzw. Regularisierungsmethoden, die auf diesem generalisierten linearen Modell aufbauen. Daraus geht hervor, dass sich die mittleren quadratischen Fehler dieser Modelle nur unwesentlich voneinander unterscheiden.

Die Modellpassung der linearen multiplen Regression, der zuvor vorgestellten zusätzlichen Modelle, die möglichen Schwachpunkten der Poisson Regression begegnen können (vgl. 2.2.3), und des Random-Forests als Verfahren aus dem Methoden-Repertoire des maschinellen Lernens kann in vergleichbarer Weise Abbildung 2.3 entnommen werden. In einer

---

<sup>4</sup>Die Auswirkungen eines möglichen Fallausschlusses, basierend auf extremen Kriteriumswerten, wurden nichtsdestotrotz im Rahmen zusätzlicher explorativer Analyse untersucht. Bei diesen wurden ausschließlich diejenigen Personen berücksichtigt, die weniger als ein halbes Jahr an Krankheits- und unentschuldigtem Fehltagen während ihrer (i. d. R. dreijährigen) Ausbildung vorzuweisen hatten ( $n = 686$ , vgl. oben). Die resultierenden Ergebnisse waren im Wesentlichen mit den nachfolgenden Befunden vergleichbar. Auch dies spricht gegen einen etwaigen Fallausschluss.

Tabelle 2.2: Anpassungsgüte der unterschiedlichen Modellierungsansätze bei der Vorhersage krankheitsbedingter und unentschuldigter Fehltage.

Modell	$MSE$	$RMSE$	$R_c^2$
Lineare Regression	1069.98	32.71	0.01
Poisson Regression	1598.80	39.98	0.01
Ridge Poisson Regression	1597.93	39.97	0.01
Lasso Poisson Regression	1598.00	39.97	0.00
Elastic Net Poisson Regression	1598.04	39.98	0.01
Zero-inflated Poisson Regression	1066.17	32.65	0.01
Hurdle Poisson Regression	1066.17	32.65	0.01
Negative Binomiale Regression	1599.13	39.99	0.01
Random Forest	1111.74	33.34	0.00

*Anmerkung.* Teststichproben- $MSE$ ,  $-RMSE$  und  $-R_c^2$  (zehnmalig wiederholte 10-fach Kreuzvalidierung).

Nebeneinanderstellung mit Abbildung 2.2 wird deutlich, dass die in Abbildung 2.3 enthaltene Vorhersagegenauigkeit der Negativen Binomialen Regression etwa den vorausgegangenen Modellen nahe kommt, alle anderen Modelle im Durchschnitt tendenziell niedrigere mittlere quadratische Abweichungen aufweisen. Sie erlauben folglich eine etwas höhere Genauigkeit in der Vorhersage der betrachteten abhängigen Variable.

Zusätzlich ist in Tabelle 2.2 das Bestimmtheitsmaß  $R_c^2$  enthalten. Dieses wird über die quadrierte Korrelation der modellierten vorhergesagten Werte mit den beobachteten Werten in den Teststichproben ermittelt. Es kann in diesem Sinne zwar nicht (bzw. nur bei der multiplen linearen Regression) direkt als erklärter Varianzanteil an der Kriteriumsvariable, aber genereller als Maß für die Vorhersagegüte des Modells, interpretiert werden und entspricht in dieser Verwendung nach Everitt (2002) der Bedeutung eines Determinationskoeffizienten.

Weiterhin wurden die krankheitsbedingten und unentschuldigsten Fehltage vor einer erneuten Modellierung mithilfe der Yeo-Johnson-Transformation in eine (annähernd) normalverteilte Form überführt ( $\lambda = .20$ ) und anschließend eine Prädiktion auf Basis linearer Vorhersagemodelle vorgenommen. Die Häufigkeitsverteilung dieser transformierten abhängigen Variable ist in Abbildung 2.4 dargestellt. Die deskriptiven Kennwerte dieser Variable sind in Tabelle 2.3 enthalten (die Tabelle enthält außerdem weitere, in Abschnitt 2.3.3 referenzierte Variablen). Abbildung 2.5 bietet einen Vergleich der mittleren quadratischen Abweichungen der entsprechenden linearen multiplen Regression sowie der linearen Ridge Regression, den Lasso- und Elastic Net-Modellen. In der Abbildung sind zu Vergleichszwecken ebenfalls die kleinsten quadratischen Abweichungen des Random Forests enthalten. In Tabelle 2.4 sind für alle diese Modelle auch  $MSE$ ,  $RMSE$  und  $R_c^2$  abgetragen.

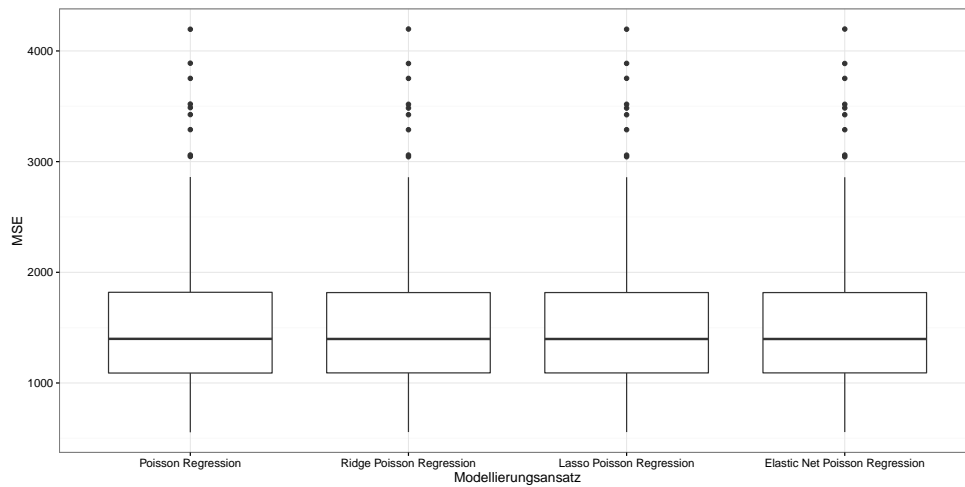


Abbildung 2.2: Boxplots der mittleren quadratischen Abweichung bei der Vorhersage krankheitsbedingter und unentschuldigter Fehltage für die Poisson Regression, Poisson Ridge Regression, Poisson Lasso und Poisson Elastic Net.

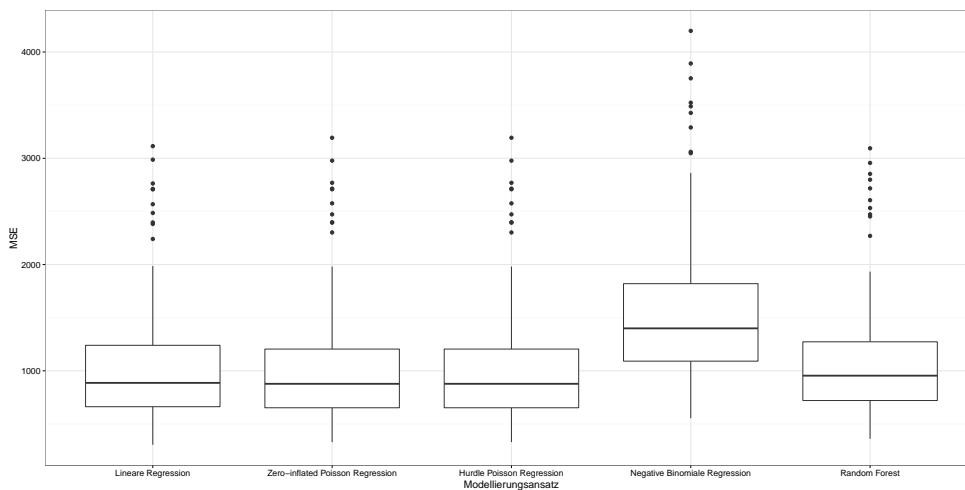


Abbildung 2.3: Boxplots der mittleren quadratischen Abweichung bei der Vorhersage krankheitsbedingter und unentschuldigter Fehltage für die Lineare Regression, Zero-inflated Poisson Regression, Hurdle Poisson Regression, Negative Binomiale Regression und Random Forest.

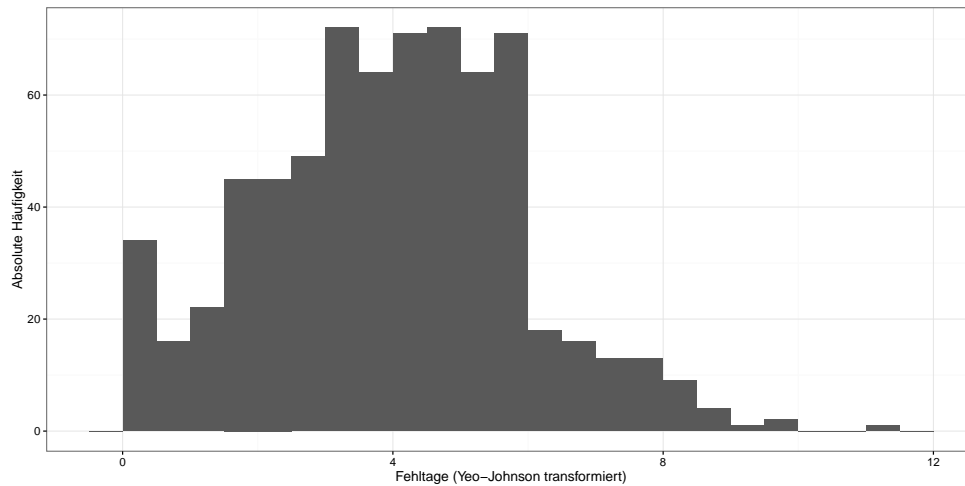


Abbildung 2.4: Häufigkeitsverteilung der Yeo-Johnson transformierten krankheitsbedingten und unentschuldigten Fehltage.

Tabelle 2.3: Deskriptive Statistiken der zusätzlichen abhängigen Variablen: Yeo-Johnson transformierte Fehltage, Leistungsbeurteilungen während der Ausbildungszeit, Berufsschulnoten und Ergebnis der Abschlussprüfung.

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	<i>Schiefe</i>	<i>Kurtosis</i>
FT (YJ)	702	3.96	1.94	0.00	11.32	0.09	0.07
LB	698	5.96	0.80	2.00	7.00	-0.63	0.97
BN	684	2.35	0.77	1.00	6.00	0.39	0.02
AP	667	78.80	8.07	45.00	96.00	-0.36	0.33

*Anmerkungen.* FT (YJ) = Yeo-Johnson transformierte Fehltage; LB = Durchschnittliche Leistungsbeurteilung; BN = Durchschnittliche Berufsschulnoten; AP = Abschlussprüfung in Prozent.



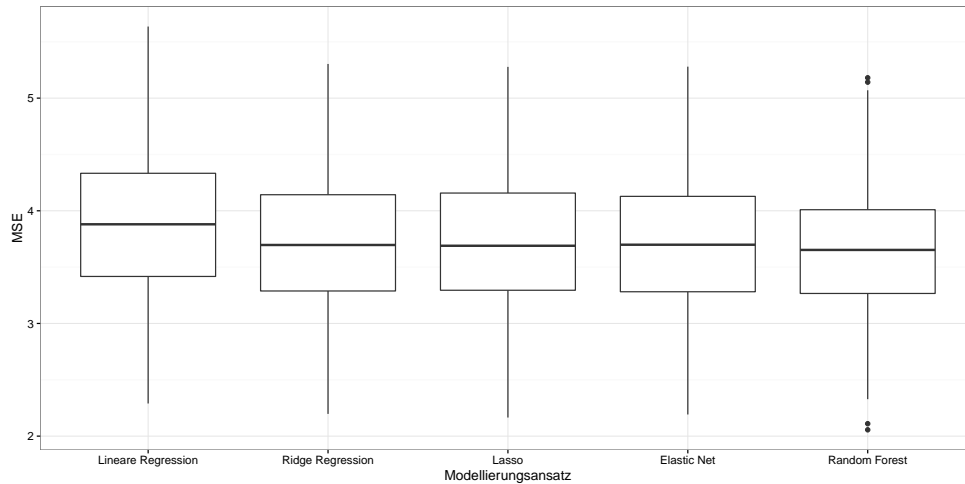


Abbildung 2.5: Boxplots der mittleren quadratischen Abweichung für unterschiedliche Modellierungsansätze bei der Vorhersage Yeo-Johnson transformierter krankheitsbedingter und unentschuldigter Fehltag.

Tabelle 2.4: Anpassungsgüte der unterschiedlichen Modellierungsansätze bei der Vorhersage der Yeo-Johnson transformierten krankheitsbedingten und unentschuldigter Fehltag.

Modell	$MSE$	$RMSE$	$R_c^2$
Lineare Regression	3.86	1.97	0.01
Ridge Regression	3.70	1.92	0.02
Lasso	3.70	1.92	0.02
Elastic Net	3.71	1.93	0.02
Random Forest	3.64	1.91	0.04

*Anmerkung.* Teststichproben- $MSE$ ,  $-RMSE$  und  $-R_c^2$  (zehnmalig wiederholte 10-fach Kreuzvalidierung).

### 2.3.3 Zusätzliche Befunde

In weiteren Untersuchungen, die analog zu den vorausgegangenen Modellvergleichen erfolgten, wurde das folgende Analyseziel angestrebt: Als referentielle Betrachtung wurden Modellvorhersagen verschiedener Leistungsmaße mithilfe der bekannten, primär linearen Modellierungsansätze (Lineare Regression, Ridge Regression, Lasso, Elastic Net und, ergänzend, RF) vorgenommen.

Hierzu wurden das Mittel der (von den Ausbildungsleiterinnen und -leitern abgegeben) Leistungsbeurteilungen während der Ausbildungszeit ( $n = 698$ ), der Notendurchschnitt der Berufsschulnoten der Azubildenden ( $n = 684$ ) und das Ergebnis in der gemäß gesetzlicher Vorgabe durchzuführenden Abschlussprüfung ( $n = 667$ ) herangezogen. Tabelle 2.3 enthält auch die entsprechende Übersicht der deskriptiven Kennwerte dieser Variablen, ihre Häufigkeitsverteilungen können Abbildung 2.6 entnommen werden.

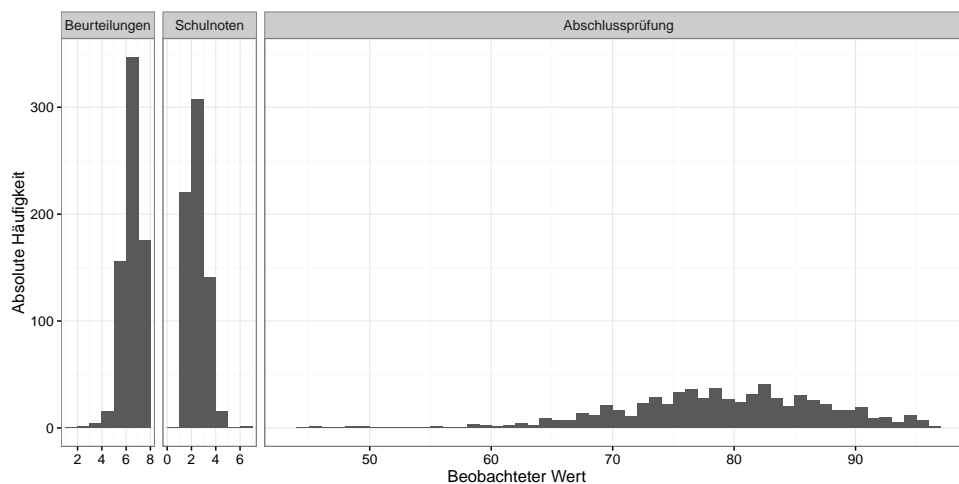


Abbildung 2.6: Häufigkeitsverteilungen der Leistungsbeurteilungen während der Ausbildungszeit, des Notendurchschnitts der Berufsschulnoten und des Ergebnisses der Abschlussprüfung.

Eine vergleichende Betrachtung der genannten Vorhersagemodelle ist schließlich in Abbildung 2.7 dargestellt. Tabelle 2.5 enthält für diese abhängigen Variablen  $MSE$ ,  $RMSE$  und  $R_c^2$  der unterschiedlichen Modellierungsansätze.

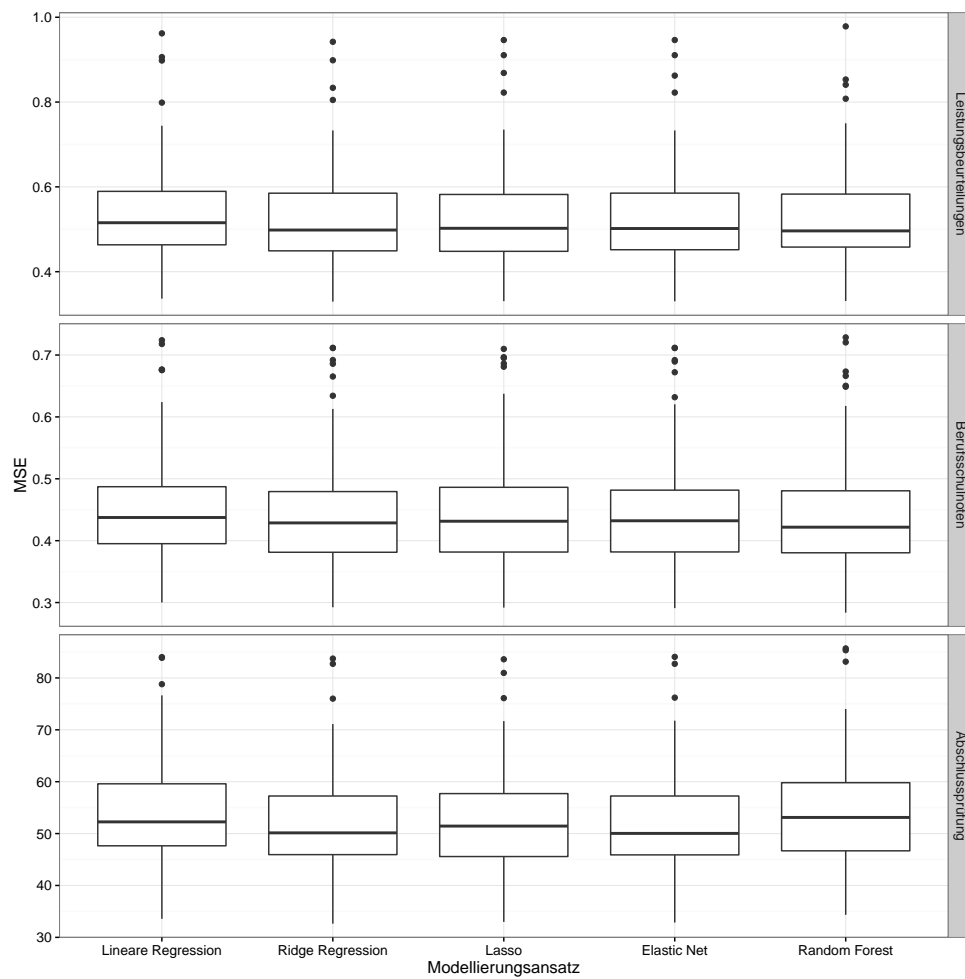


Abbildung 2.7: Boxplots der mittleren quadratischen Abweichung für unterschiedliche Modellierungsansätze bei der Vorhersage der Leistungsbeurteilungen während der Ausbildungszeit, des Notendurchschnitts der Berufsschulnoten und des Ergebnisses der Abschlussprüfung.

Tabelle 2.5: Anpassungsgüte der unterschiedlichen Modellierungsansätze bei der Vorhersage der zusätzlichen abhängigen Variablen: Yeo-Johnson transformierte Fehltag, Leistungsbeurteilungen während der Ausbildungszeit, Notendurchschnitt der Berufsschulnoten und Ergebnis der Abschlussprüfung.

Abhängige Variable	Modell	$MSE$	$RMSE$	$R_c^2$
Leistungsbeurteilungen	Lineare Regression	0.54	0.73	0.16
	Ridge Regression	0.52	0.72	0.18
	Lasso	0.52	0.72	0.17
	Elastic Net	0.52	0.72	0.17
	Random Forest	0.52	0.72	0.18
Berufsschulnoten	Lineare Regression	0.45	0.67	0.24
	Ridge Regression	0.44	0.67	0.25
	Lasso	0.45	0.67	0.25
	Elastic Net	0.45	0.67	0.25
	Random Forest	0.44	0.66	0.26
Abschlussprüfung	Lineare Regression	53.92	7.34	0.18
	Ridge Regression	52.13	7.22	0.20
	Lasso	52.76	7.26	0.19
	Elastic Net	52.25	7.23	0.20
	Random Forest	53.92	7.34	0.17

*Anmerkung.* Teststichproben- $MSE$ ,  $-RMSE$  und  $-R_c^2$  (zehnmalig wiederholte 10-fach Kreuzvalidierung).

## 2.4 Diskussion

Es folgt eine Interpretation der berichteten Ergebnisse in Hinblick auf die eingangs formulierten untersuchungsleitenden Annahmen und zusätzlich aufgeworfenen Fragestellungen (Kapitel 2.4.1 und 2.4.2). Daran anschließend werden in Abschnitt 2.4.3 Einschränkungen der vorgenommenen Analysen diskutiert, um schließlich mögliche Implikationen aus der vorliegenden Untersuchung abzuleiten (Abschnitt 2.4.4).

### 2.4.1 Überprüfung der zugrunde gelegten Annahmen

Die zentralen Annahmen der vorliegenden Studie müssen als widerlegt betrachtet werden: differenzierte Persönlichkeitsmerkmale, wie sie von HIT und NEO-PI-R als spezifische kognitive Fähigkeiten und Persönlichkeitseigenschaften auf Facettenebene erfasst werden, eignen sich nicht zur Vorhersage krankheitsbedingter und unentschuldigter Fehltag während der beruflichen Ausbildungsphase (Annahmen 1 und 2). Unterschiedliche Modellierungsansätze resultierten in kreuzvalidierten Determinationskoeffizienten  $R_c^2 \leq .01$  für die Vorhersage der unveränderten Absentismusvariable. Es wird deutlich, dass es in den Teststichproben nahezu keine Zusammenhänge zwischen den beobachteten und durch die Modelle vorhergesagten Fehltag gibt (vgl. auch Korrelationstabellen in Anhang B.2). Minimale Unterschiede, die in der mittleren quadratischen Abweichung ( $MSE$  bzw. deren Wurzel  $RMSE$ ) zwischen den Modellen sichtbar wurden, sind vor diesem Hintergrund vernachlässigbar. Diese Ergebnisse stehen in Einklang mit den in Abschnitt 2.1.4 berichteten Vorbefunden: zwar waren in vereinzelt Studien bedeutsame Zusammenhänge zwischen Persönlichkeitseigenschaften und Absentismus berichtet worden (z. B. Cooper & Payne, 1967; Furnham & Bramwell, 2006; Judge et al., 1997), aufgrund der teils uneinheitlichen Konzeption der betrachteten Konstrukte und begrenzter Stichprobenumfänge schien eine Generalisierbarkeit jedoch zweifelhaft. Diese Vermutung wurde bereits durch die von Salgado (2002) berichteten, meta-analytischen Befunde untermauert. Die hier berichteten Ergebnisse entkräften jedoch auch die weiterführende Annahme, dass eine Betrachtung auf Ebene differenzierter, ggf. kombinierter Persönlichkeitsfacetten eine präzisere Vorhersage ermöglichen könnte.

Weiterhin ist festzuhalten, dass die Verwendung von Shrinkage- bzw. Regularisierungsverfahren zu keiner Verbesserung der Vorhersage im Vergleich zu einer unbeschränkten Poisson Regression führte (Annahme 3). Auch mit weiteren Modellierungstechniken, die eine Überstreuung, bzw. das gehäufte Auftreten von Nullen in den Daten gezielt berücksichtigen, konnte die Vorhersage nicht entscheidend verbessert werden. Gleiches gilt für das beispielhaft herangezogene, computerwissenschaftliche Verfahren des maschinellen Lernens, den Random Forest. Alle Modelle führten zu einer Vorhersage, deren Anpassungsgüte bestenfalls mit der einer linearen Regression vergleichbar ist, obwohl deren Voraussetzungen in diesem Betrachtungsfall nicht ansatzweise erfüllt sind: es besteht grundsätzlich kein linearer Zusammenhang zwischen den involvierten Prädiktor- und Kriteriumsvariablen, die resultierenden Residuen eines solchen Modells sind weder normalverteilt noch homoskedastisch.

Auch eine Yeo-Johnson-Transformation der Absentismusvariable führte nicht zu einer entscheidenden Verbesserung der Anpassung. Ein Großteil der Modellschätzungen resultierte nach wie vor in einem kreuzvalidierten Determinationskoeffizienten von  $R_c^2 \leq .02$ . Zwar stellt nach dieser Umwandlung die lineare Regression ein angemesseneres Modell zur Vorhersage der transformierten Kriteriumsvariable dar, jedoch erwiesen sich die linearen Zusammenhänge zwischen dieser und den im Modell enthaltenen Prädiktoren nach wie vor als recht gering (vgl. ebenfalls Korrelationstabelle in Anhang B.2). Eine potentiell durch Ridge Regression, Lasso oder Elastic Net herbeigeführte Reduktion der Parameterschätzer — und damit einhergehende Reduzierung der Varianzkomponente des erwarteten Vorhersagefehlers der jeweiligen Modelle — ermöglicht vor diesem Hintergrund keine entscheidende Modellverbesserung. Lediglich mit dem Random Forest aus dem Repertoire des maschinellen Lernens gelang eine leicht verbesserte, insgesamt aber nach wie vor unzureichende Vorhersage ( $R_c^2 = .04$ ).

Erst mit der Betrachtung zusätzlicher Leistungsindikatoren der beruflichen Ausbildungsphase zeichnete sich tendenziell die angenommene Überlegenheit von Shrinkage- bzw. Regularisierungsmethoden im Vergleich zu linearen Regressionsmodellen ab. Zunächst lässt sich festhalten, dass die Vorhersagegüte für die drei zusätzlichen Kriterien deutlich besser ausfiel, als für die zuvor berichteten Fehlzeiten. Für die Vorhersage der Leistungsbeurteilungen lag der Determinationskoeffizient im Wertebereich  $.16 < R_c^2 < .18$ , für die Berufsschulnoten im Bereich  $.24 < R_c^2 < .26$  sowie für den erreichten Prozentwert in der Abschlussprüfung bei  $.17 < R_c^2 < .20$ . Den Korrelationstabellen in Anhang B.2 lässt sich entnehmen, dass sich für diese Kriterien erwartungsgemäß auch bivariat höhere, lineare Zusammenhänge mit unterschiedlichen Prädiktoren beobachten lassen. Insbesondere für die spezifischen kognitiven Fähigkeiten, aber auch für einzelne Persönlichkeitsfacetten werden Zusammenhänge mit den Kriterien des Ausbildungserfolgs sichtbar, die in ihren Größenordnungen den eingangs berichteten Erwartungen entsprechen (vgl. Abschnitte 2.1.1 und 2.1.2). Weiterhin wird deutlich, dass die lineare Regression, deren Parameter mithilfe der Methode der kleinsten Quadrate geschätzt wurden, jeweils am unteren Ende der Wertebereiche von  $R_c^2$  lag. Die Determinationskoeffizienten der Shrinkage- bzw. Regularisierungsmethoden fallen durchgängig etwas höher aus. Diejenigen des zusätzlich miteinbezogenen Verfahrens des maschinellen Lernens (Random Forest) fluktuierten stärker, sie fielen bei der Vorhersage der Abschlussprüfung am niedrigsten, für die Prädiktion der Berufsschulnoten am höchsten aus.

### 2.4.2 Schätzung der Determinationskoeffizienten

Anhand der drei Leistungskriterien lässt sich auch die erwartete, in Abschnitt 2.1.6 thematisierte Problematik bei der Verwendung des Determinationskoeffizienten  $R^2$  als Schätzer für die Populationswerte  $\rho^2$  bzw.  $\rho_c^2$  aufzeigen. Wie sich Tabelle 2.6 entnehmen lässt, liegt das unkorrigierte, in der Gesamtstichprobe ermittelte  $R^2$  bei allen drei Leistungsindikatoren bis zu acht Prozentpunkte über dem kreuzvalidierten  $R_c^2$ . Wie angenommen begegnet auch das korrigierte  $R^2$  dieser Überschätzung nur unzureichend, die entsprechenden Werte liegen durchgängig vier Prozentpunkte über den kreuzvalidierten Koeffizienten.

Tabelle 2.6: Determinationskoeffizient, korrigierter Determinationskoeffizient und kreuzvalidierter Determinationskoeffizient der Linearen Regression für die zusätzlichen abhängigen Variablen.

Abhängige Variable	$R^2$	Korrigiertes $R^2$	$R_c^2$
Leistungsbeurteilungen	0.24	0.20	0.16
Berufsschulnoten	0.31	0.28	0.24
Abschlussprüfung	0.26	0.22	0.18

*Anmerkungen.*  $R^2$  = Determinationskoeffizient Gesamtdatensatz, Korrigiertes  $R^2$  = Korrektur gem. *Wherry-Formel*,  $R_c^2$  = Teststichproben-Determinationskoeffizient (zehnmäßig wiederholte 10-fach Kreuzvalidierung).

### 2.4.3 Limitationen

Vorausgegangen Untersuchungen entsprechend wurde in der vorliegenden Analyse keine Trennung zwischen (möglicherweise beabsichtigten) unentschuldigtem und (potentiell unverschuldeten) krankheitsbedingtem Fehltag vorgenommen (vgl. Furnham & Bramwell, 2006; Judge et al., 1997; Salgado, 2002), obwohl eine solche Unterscheidung konzeptuell seit langem diskutiert wird (Driver & Watson, 1989; Hackett & Guion, 1985). Neben der so aufrecht erhaltenen Vergleichbarkeit der Untersuchungen wurde diese Entscheidung auch durch den Umstand begründet, dass im verfügbaren Datensatz insgesamt überhaupt nur  $n = 51$  Fälle unentschuldigter Fehltag verzeichnet waren. Es wurde also lediglich bei 7% der betrachteten Personen mindestens ein unentschuldigter Fehltag registriert. Die resultierende Verteilung wies eine niedrige Streuung bei stark ausgeprägter Schiefe auf ( $M = 0.75$ ;  $SD = 5.32$ ;  $Schiefe = 11.03$ ). Eine angemessene Modellierung eines derart selten auftretenden Merkmals erschien wenig aussichtsreich. Gleichwohl führte die isolierte Betrachtung von ausschließlich krankheitsbedingtem Fehltag in einer Voranalyse zu Ergebnissen, die mit den hier berichteten Befunden vergleichbar waren. Es ist dementsprechend davon auszugehen, dass die vorliegenden Resultate eher für die als krankheitsbedingt registrierten Fehltag aussagekräftig sind und weniger Rückschlüsse auf möglicherweise bewusste, unentschuldigte Fehltag zulassen. Es ist weiterhin denkbar, dass es insbesondere die letzteren wären, die in dispositionalen Ursachen begründet und als solche auch im Rahmen der Eignungsdiagnostik erfasst werden könnten. Gleichzeitig ist festzuhalten, dass die vorliegenden Beobachtungen vor dem Hintergrund eines kontinuierlichen Anstiegs von Fehltag aufgrund psychischer Erkrankungen in den zurückliegenden Jahren erfolgt sind (DAK-Gesundheit, 2015; Techniker Krankenkasse, 2015). Dies lässt den Schluss zu, dass auch derartige Krankheitsursachen nicht adäquat mit eignungsdiagnostisch erfassten Persönlichkeitsmerkmalen vorhergesagt werden können, oder dass die beobachteten, krankheitsbedingten Fehltag doch vermehrt auf situativ bedingte und ggf. unverschuldete Ursachen zurückgehen. Es darf jedoch nicht vernachlässigt werden, dass in motivationaler und dispositonaler Hinsicht aus Sicht des Arbeitgebers vermutlich keine strikte Trennung

zwischen diesen beiden Formen von Absentismus erfolgen kann. Es sind sowohl Fälle denkbar, in denen eine bewusste Entscheidung zum Fernbleiben vom Arbeitsplatz mit einer (in diesem Fall) ungerechtfertigten Krankmeldung einhergeht und in anderen Situationen ein eigentlich vertretbarer Krankheitsfall nicht zu einer solchen führt. Vor diesem Hintergrund wäre eine Vorhersage beider Aspekte reizvoll. Sie scheint aber — zumindest in der hier betrachteten, kombinierten Form wie auch für ausschließlich als krankheitsbedingt verzeichnete Fehltage — nicht möglich.

Eine weitere Einschränkung der hier berichteten Befunde liegt in der beschriebenen Auswahlpraxis des kooperierenden Unternehmens begründet. Wie dargestellt erfolgt eine erste Vorauswahl der Bewerberinnen und Bewerber auch anhand etwaiger (schulischer) Fehlzeiten. Vergangene Fehlzeiten haben sich in vorausgegangenen Studien vielfach als Prädiktor zukünftiger Fehlzeiten erwiesen (z. B. Breugh, 1981; A. Cohen & Golan, 2007; Ivancevich, 1985). Es muss dementsprechend davon ausgegangen werden, dass die Variabilität potentieller Fehlzeiten in der vorliegenden Stichprobe einer Einschränkung unterworfen ist, da Kandidatinnen und Kandidaten mit einer hohen Anzahl vorausgegangener Fehlzeiten nicht in die Betrachtung aufgenommen wurden. So wurde zwar (ganz im Sinne der kooperierenden Unternehmen) für vorausgegangene Fehlzeiten kontrolliert, möglicherweise wurden weiterreichende Zusammenhänge zwischen Persönlichkeitseigenschaften und Absentismus aber auch aufgrund dieses Vorgehens in den vorliegenden Daten nicht sichtbar.

In Hinblick auf die angewandten Methoden ist schließlich anzumerken, dass auch die hier vorgestellten Verfahren nur eine Auswahl potentiell geeigneter Modelle zur Vorhersageoptimierung darstellen. Es gäbe weitere Modellierungsansätze, welche beispielsweise die betrachtete Negative Binomiale Regression mit den Eigenschaften von Zero-inflated oder Hurdle-Modellen kombiniert (Zeileis et al., 2008). Zusätzlich wäre auch eine Regularisierung dieser Modelle (auch in ihrer jeweiligen, hier vorgestellten Grundform) denkbar. Möglich wäre ebenso die Anwendung Generalisierter Additiver Modell (Hastie & Tibshirani, 1986), welchen eine additive Kombination auch nicht-linearer Funktionen der einzelnen Prädiktoren zugrunde liegt. Auch für den Random Forest als Verfahren des maschinellen Lernens könnte eine weitere Optimierung einzelner Modellparameter erfolgen und ggf. zu verbesserten Vorhersagen führen.

#### 2.4.4 Fazit und Ausblick

Aus den berichteten Befunden lässt sich ableiten, dass im Rahmen traditioneller Eignungsdiagnostik erfasste Maße der menschlichen Persönlichkeit auch unter expliziter Berücksichtigung vielfältiger differenzierter Persönlichkeitseigenschaften und spezifischer kognitiver Fähigkeiten nur höchst eingeschränkt zur Vorhersage von Fehlzeiten während der beruflichen Ausbildungsphase geeignet sind. Dies gilt trotz der Verwendung erweiterter Modellierungsansätze, die durch ihre Verteilungsannahmen dem vergleichsweise seltenen Auftreten von krankheitsbedingten und unentschuldigtem Fehltagen besser gerecht werden sollten, als die lineare, multiple Regression. Auch können Shrinkage- bzw. Regularisierungsmethoden, die einer zunehmenden Flexibilität von Modellen mit größerer Prädiktoranzahl durch Einschränkungen der Parameterschätzungen begegnen, auf Basis der hier nicht vorhandenen



Zusammenhänge zu keiner Verbesserung der Vorhersagegüte führen.

Anhand der hier untersuchten großen Stichprobe an Auszubildenden verschiedener Ausbildungsberufe wird verdeutlicht, dass die angenommene und erhoffte Vorhersage von Fehlzeiten nicht in den unterschiedlichen Manifestationen des Absentismus und berufsübergreifender Form vorgenommen werden kann. Es wäre aber denkbar, dass zukünftig spezifischere, bspw. auch nach Berufsgruppen getrennte Betrachtungen (vgl. Ziegler et al., 2014) eine angemessenere Modellierung erlauben. Gleiches gilt für Untersuchungen, die nicht nur Jugendliche und junge Erwachsene während der beruflichen Ausbildungsphase, sondern auch ein größeres Altersspektrum erwerbstätiger Personen miteinbeziehen könnten.

Deutlich wird aber, dass die angenommenen Vorzüge der hier vorgestellten Shrinkage- und Regularisierungsmethoden im Falle substantieller Zusammenhänge zwischen den Prädiktor- und Kriteriumsvariablen wie erwartet zum Tragen kommen. An den Beispielen unterschiedlicher Indikatoren der beruflichen Ausbildungsleistung wird aufgezeigt, dass sie eine präzisere, weniger fehlerbehaftete Vorhersage zukünftiger Beobachtungen erlauben.

Um die Auswahl geeigneter Modelle und deren Übertragbarkeit auf zukünftige Datensätze bestmöglich abzusichern, sollten die hier vorgestellten Prinzipien der Berücksichtigung und gemeinsamen Abwägung vielfältiger, teils computerintensiver Modellierungsverfahren sowie die Überprüfung ihrer Anpassungsgüte mithilfe von Resampling-Techniken dementsprechend vermehrt Anwendung in der psychologischen Forschung und Diagnostik finden.

# Epilog

Die vorausgehend berichteten Befunde bekräftigen — neben den bereits en détail diskutierten Ergebnissen — den Eindruck, dass eine zunehmend technisch geprägte Lebensumwelt mit menschlichem Erleben und Verhalten in komplexer Weise interagiert.

In der Betrachtung des Zusammenspiels von Mensch und Maschine gelang in der vorliegenden Arbeit zwar offenkundig weder eine einfache, erschöpfende Erklärung der Unterschiede von Akzeptanz und Nutzung technischer Entwicklungen über grundlegende, spezifische Persönlichkeitseigenschaften, noch konnte mithilfe computerintensiver Methoden des maschinellen Lernens im betrachteten Kontext eine vollumfängliche Vorhersage unterschiedlicher Verhaltensindikatoren und Leistungsparameter vorgenommen werden.

Es wird jedoch deutlich, dass sich durch eine intensive Auseinandersetzung mit den beiden Themenbereichen sowie ihrer Kombination vielfältige Möglichkeiten neuer Perspektiven und potentiell bedeutungsvoller Befunde eröffnen. In einer Welt, in der Erleben und Verhalten immer mehr von technologischen Entwicklungen und algorithmischen Entscheidungen beeinflusst werden, erscheint ein dementsprechend kombiniertes Verständnis von zunehmender Bedeutung für die psychologische Forschung, Lehre und Praxis. . .

# Anhang A

## Technologieakzeptanz und -nutzung

### A.1 Korrelationstabelle

Die nachfolgende Tabelle A.1 enthält die korrelativen Zusammenhänge der Konstrukte des UTAUT und des erweiterten UTAUT2 mit den Persönlichkeitseigenschaften des FFM in Form der Dimensionen und Facetten des BFSI.

Tabelle A.1: Korrelationen der UTAUT- und UTAUT2-Konstrukte mit den Persönlichkeitsdimensionen und -facetten.

Persönlichkeitsmerkmal	PE	EE	SI	FC	HM	PV	HT	BI
N: Emotionale Stabilität	0.21*	0.16	0.16	0.14	0.13	0.10	0.05	0.02
E: Extraversion	0.15	-0.05	0.08	0.12	-0.06	0.00	0.14	0.10
O: Offenheit	0.29**	0.24*	0.07	0.27**	0.24*	0.08	0.12	0.27**
A: Verträglichkeit	0.05	0.16	0.09	0.14	0.13	-0.04	0.02	-0.01
C: Gewissenhaftigkeit	0.26**	0.17	0.09	0.13	0.10	0.08	0.17	0.22*
N1: Unbekümmertheit	0.15	0.07	0.14	0.05	0.12	0.18	0.00	0.02
N2: Gelassenheit	0.09	0.12	0.18	0.15	0.10	0.09	-0.05	-0.04
N3: Positive Grundstimmung	0.17	0.09	0.12	0.20*	0.06	0.12	0.02	0.01
N4: Soziale Selbstsicherheit	0.16	0.09	0.09	0.08	0.06	-0.03	0.16	0.02
N5: Selbstbeherrschung	0.21*	0.28**	0.10	0.03	0.26**	0.05	0.07	0.15
N6: Emotionale Robustheit	0.16	0.09	0.08	0.05	0.04	0.02	0.04	-0.02
E1: Freundlichkeit	0.09	0.03	0.11	0.14	0.01	0.00	0.19	0.07
E2: Geselligkeit	0.01	-0.08	0.01	0.06	-0.06	-0.04	0.08	0.06
E3: Durchsetzungsfähigkeit	0.12	-0.03	-0.06	0.00	-0.12	-0.12	0.08	0.06
E4: Tatendrang	0.16	-0.03	0.10	0.06	-0.06	0.03	0.11	0.08
E5: Abenteuerlust	0.26**	0.00	0.06	0.08	0.05	0.02	0.16	0.23*
E6: Fröhlichkeit	0.11	-0.06	0.12	0.18	-0.05	0.07	0.07	0.03
O1: Phantasie	0.17	0.20*	0.06	0.22*	0.16	0.08	0.06	0.19
O2: Ästhetik	0.21*	0.09	-0.01	0.20	0.13	0.09	-0.05	0.18
O3: Gefühle	0.09	0.04	0.00	0.04	0.15	0.01	0.21*	0.06
O4: Handlungen	0.34***	0.21*	0.10	0.23*	0.25*	0.09	0.15	0.28**
O5: Ideen	0.28**	0.18	0.05	0.24*	0.16	0.03	0.14	0.26*
O6: Werte- und Normensystem	0.16	0.29**	0.08	0.21*	0.25*	0.04	0.01	0.16
A1: Vertrauensbereitschaft	-0.01	-0.04	0.12	0.05	-0.07	0.07	0.00	-0.03
A2: Aufrichtigkeit	0.10	0.20	0.03	0.25*	0.11	-0.07	0.12	-0.04
A3: Hilfsbereitschaft	0.01	0.21*	0.01	0.14	0.06	-0.12	-0.01	-0.02
A4: Entgegenkommen	0.07	0.05	0.17	0.06	0.14	0.02	-0.01	0.00
A5: Bescheidenheit	0.01	0.18	0.08	0.14	0.21*	-0.06	0.03	0.09
A6: Gutmütigkeit	0.04	0.10	0.04	0.03	0.13	0.01	0.01	-0.01
C1: Kompetenz	0.23*	0.08	0.09	0.10	0.07	0.11	0.14	0.19
C2: Ordnungsliebe	0.24*	0.08	0.07	0.08	0.10	0.13	0.14	0.11
C3: Pflichtbewusstsein	0.23*	0.18	0.12	0.14	0.08	0.06	0.12	0.21*
C4: Ehrgeiz	0.28**	0.10	-0.02	0.05	0.02	0.03	0.12	0.22*
C5: Disziplin	0.20*	0.13	0.04	0.11	0.13	0.00	0.20*	0.20*
C6: Bedachtsamkeit	0.13	0.25*	0.17	0.17	0.11	0.11	0.14	0.17

Anmerkungen. PE = Erwartete Leistung; EE = Erwarteter Aufwand; SI = Sozialer Einfluss; FC = Förderliche Umgebungsbedingungen; HM = Hedonistische Motivation; PV = Preis-Leistungs-Verhältnis; HT = Gewohnheit; BI = Nutzungsintention.  
 \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

# Anhang B

## Computerintensive Methoden

### B.1 Paarweise Abbildungen

Die nachfolgenden Abbildungen bieten einen detaillierten, grafischen Überblick über die Zusammenhänge zwischen den einzelnen Prädiktorvariablen sowie dem Kriterium der unentschuldigten und krankheitsbedingten Fehltage im primären Modell.

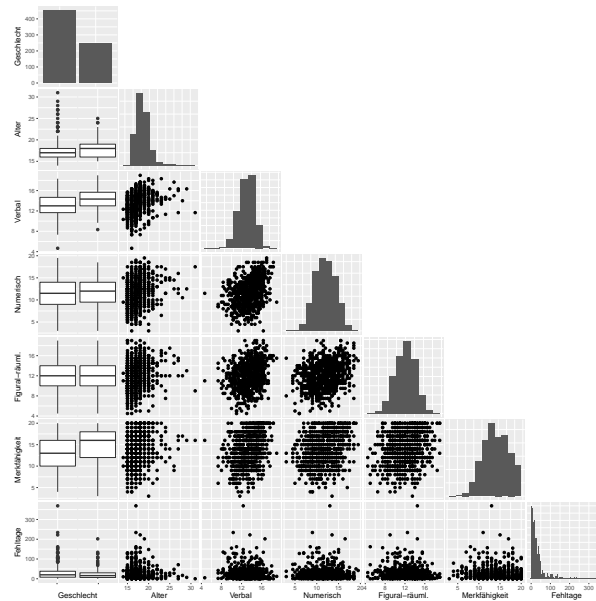


Abbildung B.1: Grafische Veranschaulichung der Zusammenhänge zwischen unentschuldig-ten und krankheitsbedingten Fehltagen, Geschlecht, Alter sowie kognitiven Fähigkeiten.

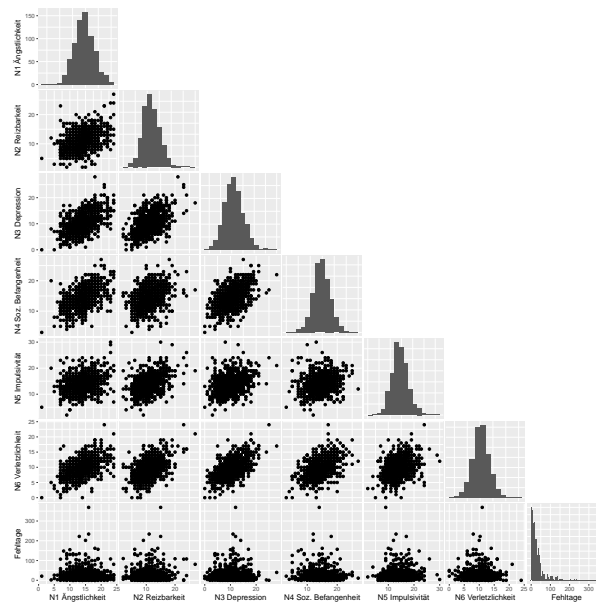


Abbildung B.2: Grafische Veranschaulichung der Zusammenhänge zwischen unentschuldig-ten und krankheitsbedingten Fehltagen sowie den Facetten der Dimension Neurotizismus aus dem NEO-PI-R.

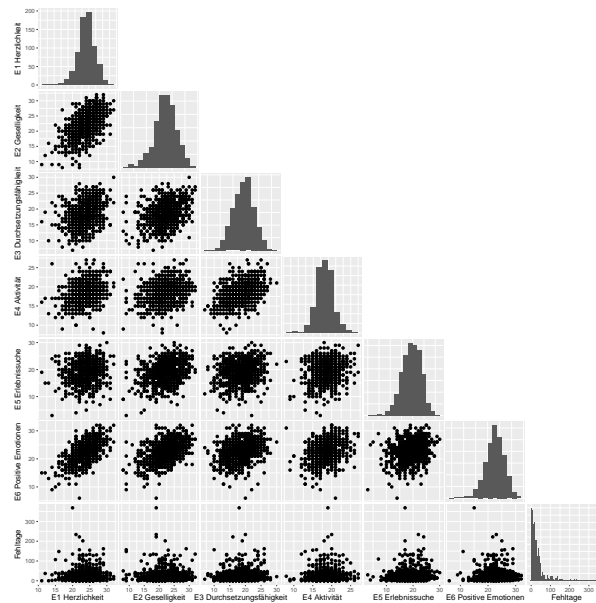


Abbildung B.3: Grafische Veranschaulichung der Zusammenhänge zwischen unentschuldig-ten und krankheitsbedingten Fehltagensowie den Facetten der Dimension Extraversion aus dem NEO-PI-R.

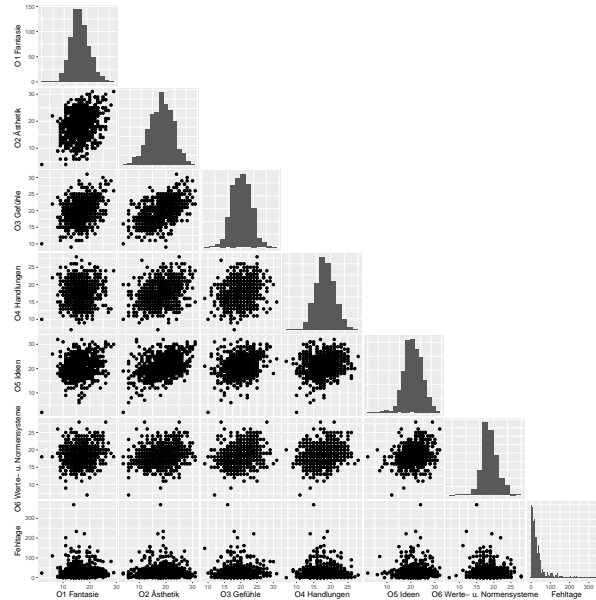


Abbildung B.4: Grafische Veranschaulichung der Zusammenhänge zwischen unentschuldig-ten und krankheitsbedingten Fehltagensowie den Facetten der Dimension Offenheit aus dem NEO-PI-R.

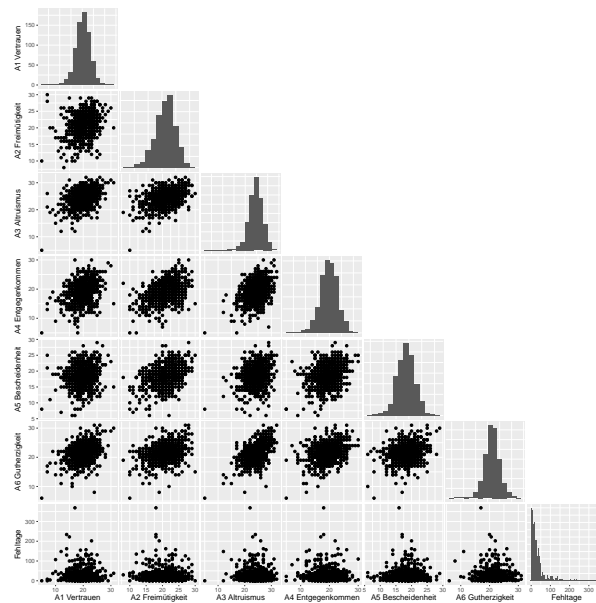


Abbildung B.5: Grafische Veranschaulichung der Zusammenhänge zwischen unentschuldig-ten und krankheitsbedingten Fehltagen sowie den Facetten der Dimension Verträglichkeit aus dem NEO-PI-R.

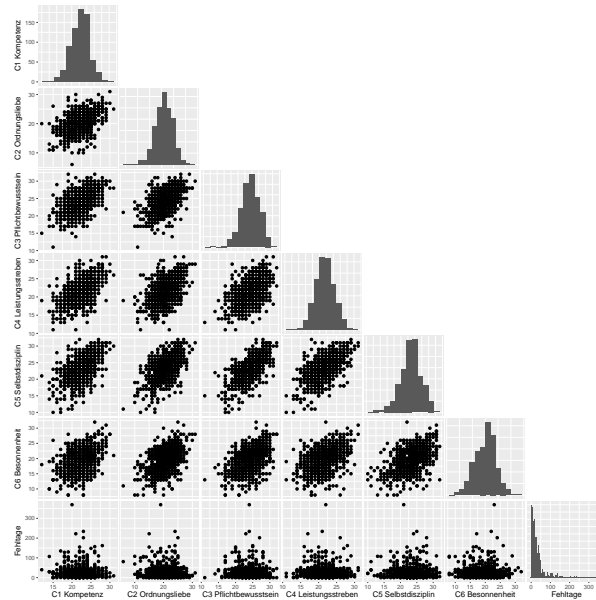


Abbildung B.6: Grafische Veranschaulichung der Zusammenhänge zwischen unentschuldig-ten und krankheitsbedingten Fehltagen sowie den Facetten der Dimension Gewissenhaftig-keit aus dem NEO-PI-R.



## B.2 Korrelationstabelle

Auf der folgenden Seite sind die numerischen Zusammenhänge zwischen den Prädiktorvariablen (Geschlecht, Alter, spezifische kognitive Fähigkeiten sowie Persönlichkeitseigenschaften auf Facettenebene des NEO-PI-R) und den abhängigen Variablen (krankheitsbedingte und unentschuldigte Fehltage, Yeo-Johnson transformierte Fehltage, durchschnittliche Berufsschulnoten sowie prozentuales Ergebnis der Abschlussprüfung) der unterschiedlichen Modelle tabellarisch zusammengefasst.

Tabelle B.1: Korrelationen von Geschlecht, Alter, den spezifischen kognitiven Fähigkeiten und Persönlichkeitsfacetten mit allen abhängigen Variablen.

Variable	FT	FT (YJ)	LB	BN	AP
Geschlecht	-0.10**	-0.08*	0.32***	-0.28***	0.26***
Alter	-0.06	-0.09*	0.29***	-0.37***	0.26***
Verbal	-0.12**	-0.12**	0.29***	-0.36***	0.36***
Numerisch	-0.02	-0.04	0.20***	-0.28***	0.29***
Figural-räuml.	-0.04	-0.05	0.12**	-0.19***	0.18***
Merkfähigkeit	-0.07	-0.11**	0.23***	-0.29***	0.22***
N1: Ängstlichkeit	0.02	0.02	0.01	0.00	0.01
N2: Reizbarkeit	0.07	0.06	-0.09*	0.09*	-0.10**
N3: Depression	0.01	-0.01	-0.04	0.05	-0.05
N4: Soziale Befangenheit	-0.05	-0.06	0.00	-0.04	0.02
N5: Impulsivität	0.04	0.06	-0.01	0.03	-0.02
N6: Verletzlichkeit	-0.02	-0.04	-0.02	0.00	-0.01
E1: Herzlichkeit	0.02	0.07	0.07	-0.03	0.07
E2: Geselligkeit	-0.01	0.05	0.05	0.00	0.04
E3: Durchsetzungsfähigkeit	0.10**	0.16***	0.02	0.02	-0.03
E4: Aktivität	0.02	0.05	0.07	0.01	0.01
E5: Erlebnissuche	0.09*	0.13***	-0.10*	0.13***	-0.17***
E6: Positive Emotionen	-0.02	0.05	0.14***	-0.10**	0.12**
O1: Fantasie	0.01	0.06	0.00	-0.03	0.03
O2: Ästhetik	-0.02	-0.02	0.15***	-0.18***	0.21***
O3: Gefühle	-0.02	0.03	0.12**	-0.14***	0.18***
O4: Handlungen	0.01	0.04	0.19***	-0.13***	0.18***
O5: Ideen	0.03	0.03	0.08*	-0.18***	0.17***
O6: Werte- u. Normensystem	-0.07*	-0.06	0.18***	-0.23***	0.25***
A1: Vertrauen	-0.06	-0.03	0.06	-0.02	0.04
A2: Freimütigkeit	-0.06	-0.06	0.10**	-0.04	0.10**
A3: Altruismus	-0.04	-0.03	0.09*	-0.08*	0.10*
A4: Entgegenkommen	-0.08*	-0.09*	0.06	-0.05	0.09*
A5: Bescheidenheit	-0.01	-0.03	0.04	-0.03	0.11**
A6: Guthertzigkeit	-0.04	0.01	0.03	-0.04	0.08*
C1: Kompetenz	-0.03	-0.03	0.15***	-0.16***	0.17***
C2: Ordnungsliebe	-0.07	-0.06	0.17***	-0.19***	0.14***
C3: Pflichtbewusstsein	-0.08*	-0.09*	0.17***	-0.19***	0.17***
C4: Leistungsstreben	0.00	0.00	0.12**	-0.11**	0.08*
C5: Selbstdisziplin	0.00	-0.02	0.15***	-0.15***	0.14***
C6: Besonnenheit	-0.04	-0.05	0.03	-0.06	0.09*

Anmerkungen. FT = Fehltag; FT (YJ) = Yeo-Johnson transformierte Fehltag; LB = Durchschnittliche Leistungsbeurteilung; BN = Durchschnittliche Berufsschulnoten; AP = Abschlussprüfung in Prozent.

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

# Literaturverzeichnis

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. doi:10.1016/0749-5978(91)90020-T
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Englewood Cliffs, NJ: Prentice-Hall.
- Allport, G. W. & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1), i–171. doi:10.1037/h0093360
- Amthauer, R. (1973). *Intelligenz-Struktur-Test (IST 70)*. Göttingen: Hogrefe.
- Arendasy, M., Hornke, L. F., Sommer, J., Häusler, M., Wagner-Menghin, M., Gittler, G., . . . Körtner, T. (2010). *INSBAT: Intelligenz-Struktur-Batterie*. Mödling: SCHUHFRIED GmbH.
- Arendasy, M., Sommer, M. & Feldhammer, M. (2009). *BFSI: Big-Five Struktur Inventar*. Mödling: SCHUHFRIED GmbH.
- Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior*, 19(3), 289–303. JSTOR: 3100173
- Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E. & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality*, 29(4), 432–442. doi:10.1006/jrpe.1995.1025
- Barnett, T., Pearson, A. W., Pearson, R. & Kellermanns, F. W. (2015). Five-factor model personality traits as predictors of perceived and actual usage of technology. *European Journal of Information Systems*, 24(4), 374–390. doi:10.1057/ejis.2014.10
- Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1–26. doi:10.1111/j.1744-6570.1991.tb00688.x
- Barrick, M. R. & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance*, 18(4), 359–372. doi:10.1207/s15327043hup1804\_3
- 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. doi:10.1111/1468-2389.00160
- Bennett, R. J. & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, 85(3), 349–360. doi:10.1037/0021-9010.85.3.349

- Bernardin, H. J. (1977). The relationship of personality variables to organizational withdrawal. *Personnel Psychology*, *30*(1), 17–27. doi:10.1111/j.1744-6570.1977.tb02317.x
- Berry, C. M., Ones, D. S. & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*(2), 410–424. doi:10.1037/0021-9010.92.2.410
- Bertua, C., Anderson, N. & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, *78*(3), 387–409. doi:10.1348/096317905X26994
- Bischl, B. & Lang, M. (2015). *parallelMap: Unified interface to parallelization back-ends*. R package version 1.3. Zugriff unter <https://CRAN.R-project.org/package=parallelMap>
- Bischl, B., Lang, M., Richter, J., Bossek, J., Judt, L., Kuehn, T., ... Kotthoff, L. (2015). *mlr: Machine learning in R*. R package version 2.5. Zugriff unter <https://github.com/mlr-org/mlr>
- Bischl, B., Mersmann, O., Trautmann, H. & Weihs, C. (2012). Resampling methods for meta-model validation with recommendations for evolutionary computation. *Evolutionary Computation*, *20*(2), 249–275. doi:10.1162/EVCO\_a\_00069
- Box, G. E. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–252. JSTOR: 2984418
- Breaugh, J. A. (1981). Predicting absenteeism from prior absenteeism and work attitudes. *Journal of Applied Psychology*, *66*(5), 555–560. doi:10.1037/0021-9010.66.5.555
- Breiman, L. (2001a). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324
- Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. doi:10.1214/ss/1009213726
- Breiman, L., Cutler, A., Liaw, A. & Wiener, M. (2015). *randomForest: Breiman and Cutler's random forests for classification and regression*. R package version 4.6-12. Zugriff unter <https://CRAN.R-project.org/package=randomForest>
- Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion*. München: Pearson Studium.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Hrsg.), *Handbook of industrial and organizational psychology* (Bd. 1, S. 687–732). Palo Alto, CA: Consulting Psychologists Press.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Cattell, R. B. (1946). *Description and measurement of personality*. New York, NY: World Book.
- Christiansen, N. D. & Robie, C. (2011). Further consideration of the use of narrow trait scales. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *43*(3), 183–194. doi:10.1037/a0023069
- Claudy, J. G. (1978). Multiple regression and validity estimation in one sample. *Applied Psychological Measurement*, *2*(4), 595–607. doi:10.1177/014662167800200414

- Cohen, A. & Golan, R. (2007). Predicting absenteeism and turnover intentions by past absenteeism and work attitudes: An empirical examination of female employees in long term nursing care facilities. *Career Development International*, 12(5), 416–432. doi:10.1108/13620430710773745
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. doi:10.1037/0033-2909.112.1.155
- Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cooper, R. & Payne, R. (1967). Extraversion and some aspects of work behavior. *Personnel Psychology*, 20(1), 45–57. doi:10.1111/j.1744-6570.1967.tb02268.x
- 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.
- Creel, M. D. & Loomis, J. B. (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. *American Journal of Agricultural Economics*, 72(2), 434–441. doi:10.2307/1242345
- Cronbach, L. J. & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Cummings, C. C. (1982). *Estimates of multiple correlation coefficient shrinkage*. Vortrag auf dem 66th Annual Meeting of the American Educational Research Association. New York, NY.
- Dahl, D. B. (2016). *xtable: Export tables to L<sup>A</sup>T<sub>E</sub>X or HTML*. R package version 1.8-2. Zugriff unter <https://CRAN.R-project.org/package=xtable>
- DAK-Gesundheit. (2015). *DAK-Gesundheitsreport 2015*. DAK-Gesundheit. Hamburg. Zugriff unter [http://www.dak.de/dak/download/Vollstaendiger\\_bundesweiter\\_Gesundheitsreport\\_2015-1585948.pdf](http://www.dak.de/dak/download/Vollstaendiger_bundesweiter_Gesundheitsreport_2015-1585948.pdf)
- Darviri, S. V. & Woods, S. A. (2006). Uncertified absence from work and the big five: An examination of absence records and future absence intentions. *Personality and Individual Differences*, 41(2), 359–369. doi:10.1016/j.paid.2006.02.003
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. doi:10.2307/249008
- Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38(3), 475–487. doi:10.1006/imms.1993.1022
- Davis, F. D. & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. *International Journal of Human-Computer Studies*, 45(1), 19–45. doi:10.1006/ijhc.1996.0040
- Devaraj, S., Easley, R. F. & Crant, J. M. (2008). How does personality matter? Relating the five-factor model to technology acceptance and use. *Information Systems Research*, 19(1), 93–105. doi:10.1287/isre.1070.0153
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1), 417–440. doi:10.1146/annurev.ps.41.020190.002221

- Driver, R. W. & Watson, C. J. (1989). Construct validity of voluntary and involuntary absenteeism. *Journal of Business and Psychology*, 4(1), 109–118. doi:10.1007/BF01023041
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E. & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91(1), 40–57. doi:10.1037/0021-9010.91.1.40
- Epskamp, S. (2014). *semPlot: Path diagrams and visual analysis of various SEM packages' output*. R package version 1.0.1. Zugriff unter <https://CRAN.R-project.org/package=semPlot>
- Everitt, B. (2002). *The Cambridge dictionary of statistics* (2. Aufl.). Cambridge University Press.
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fox, J. (2015). *Applied regression analysis and generalized linear models* (3. Aufl.). Los Angeles, CA: SAGE Publications.
- Fox, J. & Weisberg, S. (2016). *car: Companion to applied regression*. R package version 2.1-2. Zugriff unter <https://CRAN.R-project.org/package=car>
- Friedman, J., Hastie, T., Simon, N. & Tibshirani, R. (2015). *glmnet: Lasso and elastic-net regularized generalized linear models*. R package version 2.0-2. Zugriff unter <https://CRAN.R-project.org/package=glmnet>
- Furnham, A. (1992). *Personality at work*. London: Routledge.
- Furnham, A. & Bramwell, M. (2006). Personality factors predict absenteeism in the workplace. *Individual Differences Research*, 4(2), 68–77.
- Goertz, W., Hülshager, U. R. & Maier, G. W. (2014). The validity of specific cognitive abilities for the prediction of training success in Germany. *Journal of Personnel Psychology*, 13(3), 123–133. doi:10.1027/1866-5888/a000110
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Hrsg.), *Review of personality and social psychology* (Bd. 2, S. 141–165). Beverly Hills, CA: SAGE Publications.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26–42. doi:10.1037/1040-3590.4.1.26
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt & F. Ostendorf (Hrsg.), *Personality psychology in Europe* (Bd. 7, S. 7–28). Tillburg, The Netherlands: Tillburg University Press.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R. & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84–96. doi:10.1016/j.jrp.2005.08.007
- Gonzalez-Mulé, E., Mount, M. K. & Oh, I.-S. (2014). A meta-analysis of the relationship between general mental ability and nontask performance. *Journal of Applied Psychology*, 99(6), 1222–1243. doi:10.1037/a0037547

- Guion, R. M. & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18(2), 135–164. doi:10.1111/j.1744-6570.1965.tb00273.x
- Hackett, R. D. & Guion, R. M. (1985). A reevaluation of the absenteeism-job satisfaction relationship. *Organizational Behavior and Human Decision Processes*, 35(3), 340–381. doi:10.1016/0749-5978(85)90028-7
- Harrell, F. E., Jr. (2016). *Hmisc: Harrell miscellaneous*. R package version 3.17-4. Zugriff unter <https://CRAN.R-project.org/package=Hmisc>
- Hastie, T. & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–310. JSTOR: 2245459
- Hastie, T., Tibshirani, R. & Friedman, J. (2011). *The elements of statistical learning: Data mining, inference, and prediction* (2. Aufl.). New York, NY: Springer.
- Hastie, T., Tibshirani, R. & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations*. Boca Raton, FL: Taylor & Francis.
- Herzberg, P. A. (1969). The parameters of cross-validation. *Psychometrika Monograph Supplement*, 16, 1–10.
- Heß, M. (1994). *Der Hoechster Intelligenztest. Revision auf Basis des I-S-T 70 von R. Amthauer*. Interner Bericht der Hoechst AG. Frankfurt.
- Higgins, J. P. T. & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. doi:10.1002/sim.1186
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560. doi:10.1136/bmj.327.7414.557
- Hilbe, J. M. (2011). *Negative binomial regression* (2. Aufl.). Cambridge University Press.
- Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. doi:10.1080/00401706.1970.10488634
- Hogan, J. & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology*, 74(2), 273–279. doi:10.1037/0021-9010.74.2.273
- Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88(1), 100–112. doi:10.1037/0021-9010.88.1.100
- Hogan, J. & Roberts, B. W. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, 17(6), 627–637. JSTOR: 2488580
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi:10.1080/10705519909540118
- Hülsheger, U. R., Maier, G. W. & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, 15(1), 3–18. doi:10.1111/j.1468-2389.2007.00363.x
- Hülsheger, U. R., Maier, G. W., Stumpp, T. & Muck, P. M. (2006). Vergleich kriteriumsbezogener Validitäten verschiedener Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland. *Zeitschrift für Personalpsychologie*, 5(4), 145–162. doi:10.1026/1617-6391.5.4.145

- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340–62. doi:10.1016/0001-8791(86)90013-8
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. doi:10.1037/0033-2909.96.1.72
- Ivancevich, J. M. (1985). Predicting absenteeism from prior absence and work attitudes. *Academy of Management Journal*, *28*(1), 219–228. doi:10.2307/256070
- Jackman, S., Tahk, A., Zeileis, A., Maimone, C. & Fearon, J. (2015). *pscl: Political Science Computational Laboratory, Stanford University*. R package version 1.4.9. Zugriff unter <https://CRAN.R-project.org/package=pscl>
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. New York, NY: Springer.
- Jenkins, M. & Griffith, R. (2004). Using personality constructs to predict performance: Narrow or broad bandwidth. *Journal of Business and Psychology*, *19*(2), 255–269. doi:10.1007/s10869-004-0551-9
- Judge, T. A., Martocchio, J. J. & Thoresen, C. J. (1997). Five-factor model of personality and employee absence. *Journal of Applied Psychology*, *82*(5), 745–755. doi:10.1037/0021-9010.82.5.745
- Krampen, G. (1991). *Fragebogen zu Kompetenz-und Kontrollüberzeugungen (FKK)*. Göttingen: Hogrefe.
- Kuhn, M. & Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1–14. doi:10.1080/00401706.1992.10485228
- Landy, F. J. & Conte, J. M. (2010). *Work in the 21st century: An introduction to industrial and organizational psychology*. John Wiley & Sons.
- Lang, F. R., Lüdtke, O. & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica*, *47*(3), 111–121. doi:10.1026//0012-1924.47.3.111
- Larson, S. C. (1931). The shrinkage of the coefficient of multiple correlation. *Journal of Educational Psychology*, *22*(1), 45–55. doi:10.1037/h0072400
- Leiner, D. J. (2013). SoSci Survey (Version Version 2.4.00-i). München: SoSci Survey GmbH. Zugriff unter <https://www.soscisurvey.de>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. doi:10.1007/BF02296272
- McCrae, R. R. & Costa, P. T. (1985). Updating Norman’s „adequacy taxonomy“: Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*, *49*(3), 710–721. doi:10.1037/0022-3514.49.3.710
- McCrae, R. R. & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. doi:10.1037/0022-3514.52.1.81
- McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models*. Boca Raton, FL: Chapman & Hall/CRC.



- McElroy, J. C., Hendrickson, A. R., Townsend, A. M. & DeMarie, S. M. (2007). Dispositional factors in internet use: Personality versus cognitive style. *MIS Quarterly*, *31*(4), 809–820. JSTOR: 25148821
- 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*(3), 683–729. doi:10.1111/j.1744-6570.2007.00089.x
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, *33*(3), 341–365. doi:10.1016/0304-4076(86)90002-3
- Murphy, K. R. & Dziewieczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, *18*(4), 343–357. doi:10.1207/s15327043hup1804\_2
- Nelder, J. A. & Baker, R. J. (2006). Generalized linear models. *Encyclopedia of Statistical Sciences*. doi:10.1002/0471667196.ess0866.pub2
- Neyer, F. J., Felber, J. & Gebhardt, C. (2012). Entwicklung und Validierung einer Kurzskaala zur Erfassung von Technikbereitschaft. *Diagnostica*, *58*(2), 87–99. doi:10.1026/0012-1924/a000067
- Nistor, N. (2014). When technology acceptance models won't work: Non-significant intention-behavior effects. *Computers in Human Behavior*, *34*, 299–300. doi:10.1016/j.chb.2014.02.052
- Nunnally, J. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
- Ones, D. S., Dilchert, S., Viswesvaran, C. & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, *60*(4), 995–1027. doi:10.1111/j.1744-6570.2007.00099.x
- Ones, D. S. & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, *17*(6), 609–626. JSTOR: 2488579
- Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, *10*(2), 85–97. doi:10.1207/s15327043hup1002\_2
- Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae*. Göttingen: Hogrefe.
- Özbek, V., Alımaçık, Ü., Koc, F., Akkılıç, M. E. & Kaş, E. (2014). The impact of personality on technology acceptance: A study on smart phone users. *Procedia - Social and Behavioral Sciences*, *150*, 541–551. doi:10.1016/j.sbspro.2014.09.073
- Paunonen, S. V. & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, *81*(3), 524–539. doi:10.1037/0022-3514.81.3.524
- Paunonen, S. V. & Nicol, A. A. (2001). The personality hierarchy and the prediction of work behaviors. In R. Hogan & B. W. Roberts (Hrsg.), *Personality psychology in the workplace* (S. 161–191). Washington, DC: American Psychological Association.
- Porter, L. W. & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. *Psychological Bulletin*, *80*(2), 151–176. doi:10.1037/h0034829

- Punnoose, A. C. (2012). Determinants of intention to use eLearning based on the technology acceptance model. *Journal of Information Technology Education: Research*, 11(1), 301–337.
- R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Zugriff unter <https://www.R-project.org/>
- Raymond, E. S. (2004). The jargon file, Version 4.4.7. Zugriff unter <http://catb.org/jargon>
- Revelle, W. (2016). *psych: Procedures for psychological, psychometric, and personality research*. R package version 1.6.6. Zugriff unter <https://CRAN.R-project.org/package=psych>
- Ripley, B. (2015). *MASS: Support functions and datasets for Venables and Ripley's MASS*. R package version 7.3-45. Zugriff unter <https://CRAN.R-project.org/package=MASS>
- Robinson, S. L. & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal*, 38(2), 555–572. doi:10.2307/256693
- Robinson, S. L. & Bennett, R. J. (1997). Workplace deviance: Its definition, its manifestations, and its causes. In R. J. Lewicki, R. J. Bies & B. H. Sheppard (Hrsg.), *Research on negotiation in organizations* (Bd. 6, S. 3–27). US: Elsevier Science/JAI Press.
- Rojon, C., McDowall, A. & Saunders, M. N. K. (2015). The relationships between traditional selection assessments and workplace performance criteria specificity: A comparative meta-analysis. *Human Performance*, 28(1), 1–25. doi:10.1080/08959285.2014.974757
- Rosen, P. A. & Kluemper, D. H. (2008). The impact of the big five personality traits on the acceptance of social networking website. *AMCIS 2008 Proceedings*, Vortrag 274.
- Rosseel, Y. (2015). *lavaan: Latent variable analysis*. R package version 0.5-20. Zugriff unter <https://CRAN.R-project.org/package=lavaan>
- Rothstein, M. G. & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, 16(2), 155–180. doi:10.1016/j.hrmmr.2006.03.004
- Sackett, P. R. & DeVore, C. J. (2001). Counterproductive behaviors at work. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Hrsg.), *Handbook of industrial, work, and organizational psychology* (Bd. 1, S. 145–164). London: SAGE Publications.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82(1), 30–43. doi:10.1037/0021-9010.82.1.30
- Salgado, J. F. (2002). The big five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10(1-2), 117–125. doi:10.1111/1468-2389.00198
- Salgado, J. F. & Anderson, N. (2002). Cognitive and GMA testing in the European Community: Issues and evidence. *Human Performance*, 15(1-2), 75–96. doi:10.1080/08959285.2002.9668084

- Salgado, J. F. & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European Community. *European Journal of Work and Organizational Psychology, 12*(1), 1–17. doi:10.1080/13594320244000292
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C. & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European Community meta-analysis. *Personnel Psychology, 56*(3), 573–606. doi:10.1111/j.1744-6570.2003.tb00751.x
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology, 88*, 1068–1081. doi:10.1037/0021-9010.88.6.1068
- Schloerke, B., Crowley, J., Cook, D., Briatte, F., Marbach, M., Thoen, E. & Elberg, A. (2016). *Ggally: extension to ggplot2*. R package version 1.0.1. Zugriff unter <https://CRAN.R-project.org/package=GGally>
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274. doi:10.1037/0033-2909.124.2.262
- Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 45–65. doi:10.1146/annurev-orgpsych-031413-091255
- Schneider, R. J., Hough, L. M. & Dunnette, M. D. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior, 17*(6), 639–655. JSTOR: 2488581
- Schroeders, U. & Wilhelm, O. (2011). Computer usage questionnaire: Structure, correlates, and gender differences. *Computers in Human Behavior, 27*(2), 899–904. doi:10.1016/j.chb.2010.11.015
- SCHUHFRIED GmbH. (2011). Wiener Testsystem (Version 6.72.029). Mödling: SCHUHFRIED GmbH.
- Schulenberg, S. E. & Melton, A. M. (2008). The computer aversion, attitudes, and familiarity index (CAAFI): A validity study. *Computers in Human Behavior, 24*(6), 2620–2638. doi:10.1016/j.chb.2008.03.002
- Sharma, A. & Citurs, A. (2004). Incorporating personality into UTAUT: Individual differences and user acceptance of IT. *AMCIS 2004 Proceedings*, Vortrag 421.
- Simon, R. (2007). Resampling strategies for model assessment and selection. In W. Dubitzky, M. Granzow & D. Berrar (Hrsg.), *Fundamentals of data mining in genomics and proteomics* (S. 173–186). New York, NY: Springer.
- Sinha, A. K. P. (1963). Manifest anxiety affecting industrial absenteeism. *Psychological Reports, 13*(1), 258–258. doi:10.2466/pr0.1963.13.1.258
- Smith, C. A., Organ, D. W. & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology, 68*(4), 653–663. doi:10.1037/0021-9010.68.4.653
- Stelzl, I. (2005). *Fehler und Fallen der Statistik: für Psychologen, Pädagogen und Sozialwissenschaftler*. Münster: Waxmann.

- Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. New York, NY: Routledge.
- Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology, 84*(6), 959. doi:10.1037/0021-9010.84.6.959
- Svendsen, G. B., Johnsen, J.-A. K., Almås-Sørensen, L. & Vittersø, J. (2013). Personality and technology acceptance: The influence of personality factors on the core constructs of the technology acceptance model. *Behaviour & Information Technology, 32*(4), 323–334. doi:10.1080/0144929X.2011.553740
- Swanson, E. B. (1974). Management information systems: Appreciation and involvement. *Management Science, 21*(2), 178–188. doi:10.1287/mnsc.21.2.178
- Techniker Krankenkasse. (2015). *Gesundheitsreport 2015*. Techniker Krankenkasse. Hamburg. Zugriff unter <https://www.tk.de/centaurus/servlet/contentblob/718612/Datei/143830/Gesundheitsreport-2015.pdf>
- Tett, R. P. & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*(4), 967. doi:10.1111/j.1744-6570.2007.00098.x
- Tett, R. P., Steele, J. R. & Bearegard, R. S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior, 24*(3), 335–356. doi:10.1002/job.191
- The Document Foundation. (2014). LibreOffice (Version 4.2). Berlin: The Document Foundation.
- The Mentor. (1986). The conscience of a hacker („Hacker Manifesto“). *Phrack, Inc. 1*(7). Zugriff unter <http://www.phrack.org/archives/issues/7/3.txt>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288. JSTOR: 2346178
- Tishkovskaya, S. & Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education, 20*(2), 2012.
- Tupes, E. C. & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings*. USAF. Lackland Air Force Base, Texas.
- Tupes, E. C. & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality, 60*(2), 225–251. doi:10.1111/j.1467-6494.1992.tb00973.x
- Turner, M., Kitchenham, B., Brereton, P., Charters, S. & Budgen, D. (2010). Does the technology acceptance model predict actual use? A systematic literature review. *Information and Software Technology, 52*(5), 463–479. doi:10.1016/j.infsof.2009.11.005
- Tutz, G. (2011). *Regression for categorical data*. Cambridge University Press.
- Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research, 11*(4), 342–365. doi:10.1287/isre.11.4.342.11872
- Venkatesh, V. & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273–315. doi:10.1111/j.1540-5915.2008.00192.x

- Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. doi:10.1287/mnsc.46.2.186.11926
- Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478. JSTOR: 30036540
- Venkatesh, V., Thong, J. & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178.
- Viechtbauer, W. et al. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. doi:10.18637/jss.v036.i03
- Viechtbauer, W. (2015). *metafor: Meta-analysis package for R*. R package version 1.9-8. Zugriff unter <https://CRAN.R-project.org/package=metafor>
- Viteles, M. S. (1932). *Industrial psychology*. New York, NY: W. W. Norton.
- Wenzelmann, F. & Lemmermann, H. (2012). Betriebliche Kosten von Vertragslösungen. *Berufsbildung in Wissenschaft und Praxis: BWP*, *41*(5), 4–5.
- Wherry, R. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *The Annals of Mathematical Statistics*, *2*(4), 440–457. JSTOR: 2957681
- Wickham, H. & Chang, W. (2016). *ggplot2: An implementation of the grammar of graphics*. R package version 2.1.0. Zugriff unter <https://CRAN.R-project.org/package=ggplot2>
- Wu, J. & Du, H. (2012). Toward a better understanding of behavioral intention and system usage constructs. *European Journal of Information Systems*, *21*(6), 680–698. doi:10.1057/ejis.2012.15
- Yarkoni, T. & Westfall, J. (o.D.). Choosing prediction over explanation in psychology: Lessons from machine learning. Zugriff unter [http://jakewestfall.org/publications/Yarkoni\\_Westfall\\_choosing\\_prediction.pdf](http://jakewestfall.org/publications/Yarkoni_Westfall_choosing_prediction.pdf)
- Yeo, I.-K. & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959. doi:10.1093/biomet/87.4.954
- Yin, P. & Fan, X. (2001). Estimating  $R^2$  shrinkage in multiple regression: A comparison of different analytical methods. *The Journal of Experimental Education*, *69*(2), 203–224. doi:10.1080/00220970109600656
- Zeileis, A., Kleiber, C. & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25. doi:10.18637/jss.v027.i08
- Ziegler, M., Bensch, D., Maaß, U., Schult, V., Vogel, M. & Bühner, M. (2014). Big five facets as predictor of job training performance: The role of specific job demands. *Learning and Individual Differences*, *29*, 1–7. doi:10.1016/j.lindif.2013.10.008
- Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Methodological)*, *67*(2), 301–320. doi:10.1111/j.1467-9868.2005.00503.x