

Aus dem Institut für Arbeits-, Sozial- und Umweltmedizin
der Universitätsmedizin der Johannes Gutenberg-Universität Mainz

**Weiterentwicklung und Validierung des AIGScreen, ein Instrument zur
Messung des Individuellen Arbeits- und Gesundheitsempfindens**

Inauguraldissertation
zur Erlangung des Doktorgrades der
physiologischen Wissenschaften
der Universitätsmedizin
der Johannes Gutenberg-Universität Mainz

Vorgelegt von

Annika Göktaş
aus Zeitz

Mainz, 2023

Wissenschaftlicher Vorstand: Univ.-Prof. Dr. U. Förstermann

1. Gutachter: Univ.-Prof. Dr. D.-M. Rose

2. Gutachter:

3. Gutachter:

Tag der Promotion: 04. September 2024

Inhaltsverzeichnis

Tabellenverzeichnis	i
Abbildungsverzeichnis	ii
Anlagenverzeichnis.....	iii
Abkürzungsverzeichnis	iv
1 Hintergrund.....	1
1.1 Entwicklung des AIGScreen	3
1.1.1 Analyseinstrument zur Beurteilung von Gefährdungen am Arbeitsplatz Schule	3
1.1.2 Weiterentwicklung zum BGM-Tool	5
1.2 Ziel der Arbeit.....	6
2 Theoretische Grundlagen und rechtlicher Kontext.....	10
2.1 Gesundheit und das Belastungs-Beanspruchungs-Modell.....	11
2.1.1 Der Gesundheitsbegriff nach Antonovsky.....	11
2.1.2 Das Belastungs-Beanspruchungs-Modell.....	13
2.2 Gesetzliche Pflicht zur Gefährdungsbeurteilung psychischer Belastung und Herausforderungen bei der Durchführung.....	16
2.3 Wissenschaftliche Gütekriterien	21
2.3.1 Objektivität	21
2.3.2 Reliabilität	22
2.3.3 Validität	25
2.3.3.1 Inhaltsvalidität.....	27
2.3.3.2 Konstruktvalidität und Kriteriumsvalidität.....	28
2.3.4 Resümee Haupt- und Nebengütekriterien	33
3 Methode	37
3.1 Fragebogenkonstruktion des AIGScreen.....	37
3.1.1 Fragebogenentwicklung	38
3.1.2 Erfassung psychischer Belastungen.....	42
3.1.3 Erhebungsdesign, Frageformat und Antwortskalen	47
3.2 Datenerhebung im Rahmen der Mitarbeitendenbefragung	51

3.3	Expertenbefragung.....	53
3.3.1	Vor- und Nachteile von Expertenbefragungen.....	54
3.3.2	Definition und Auswahl von Experten.....	56
3.3.3	Zielsetzung und Fragestellung.....	62
3.3.3.1	Ziel des Expertenreviews.....	63
3.3.3.2	Ziel der teilstandardisierten Expertenbefragung.....	65
3.3.4	Internes Expertenreview zur Fragebogenqualität in zwei Phasen.....	66
3.3.5	Durchführung der teilstandardisierten Expertenbefragung zur Inhaltsvalidität..	67
3.3.6	Zusammenfassung zum methodischen Vorgehen bei den expertenbeteiligten Verfahren.....	75
3.4	Reliabilitäts- und Validitätsanalyse.....	77
3.4.1	Itemanalyse.....	78
3.4.2	Multiple Imputation.....	80
3.4.3	Explorative Faktorenanalyse.....	83
3.4.4	Konfirmatorische Faktorenanalyse.....	85
4	Ergebnisse.....	88
4.1	Ergebnisse des internen Expertenreviews.....	88
4.2	Ergebnisse der teilstandardisierten Online-Befragung.....	96
4.2.1	Quantitative Auswertungen zur Inhaltsvalidität.....	97
4.2.2	Ergebnisse der quantitativen Inhaltsanalyse.....	104
4.2.3	Vergleich von Expertenmeinungen und GDA-Empfehlung.....	112
4.3	Itemanalyse, Reliabilitätsanalyse und Konstruktvalidierung.....	126
4.3.1	Beschreibung der Stichprobe.....	126
4.3.2	Itemanalyse und multiple Imputation.....	127
4.3.3	Explorative Faktorenanalyse.....	136
4.3.4	Konfirmatorische Faktorenanalyse.....	142
5	Diskussion der Ergebnisse und Implikationen für das weitere Vorgehen	145
5.1	Fragebogenreview.....	146
5.2	Expertenbefragung in der Zielgruppe.....	148

5.2.1	Vergleich der Expertenmeinungen mit Empfehlungen der GDA und wissenschaftlichen Erkenntnissen	152
5.3	Itemqualität, Reliabilität und Konstruktvalidität des AIGScreen.....	161
5.3.1	Qualität der Items.....	161
5.3.2	Resümee aus den Ergebnissen der Faktorenanalyse	166
5.4	Schlussfolgerungen für die wissenschaftliche Güte des AIGScreen	171
5.5	Notwendige weitere Anpassungen	175
5.5.1	Anpassungen im Fragebogendesign	175
5.5.2	Inhaltliche Anpassungen im Fragebogen.....	179
5.6	Limitationen.....	191
6	Zusammenfassung und Ausblick.....	196
	Literaturverzeichnis	204
	Anlagen	225
	Danksagung.....	288
	Lebenslauf	289

Hinweis: Wo im Text nur die männliche Form verwendet wurde, dient dies lediglich der sprachlichen Vereinfachung und impliziert gleichermaßen die weibliche Form.

Tabellenverzeichnis

Tabelle 1: Merkmalsbereiche und Inhalte der Gefährdungsbeurteilung gemäß der „Leitlinie Beratung und Überwachung bei psychischer Belastung am Arbeitsplatz“ (50) der GDA.....	20
Tabelle 2: Ergebnisse des Fragebogenreviews	93
Tabelle 3: Bewertung der Relevanz der Items im Rahmen der Expertenbefragung	98
Tabelle 4: Antwortverteilung auf die Frage „Würden Sie weitere Aspekte (Items) zum Merkmalsbereich erfassen?“ (n = 116)	103
Tabelle 5: Antwortverteilung auf die Frage „Ist die Definition des Merkmalsbereichs Ihrer Meinung nach vollständig?“ (n = 96).....	104
Tabelle 6: Im Kodierleitfaden definierte Arbeitsschritte	105
Tabelle 7: Kommentare zu den Arbeitsmerkmalen im Rahmen der Expertenbefragung (n = 174).....	108
Tabelle 8: Zusammenfassung der Kommentare (n = 289) in der Expertenbefragung zu einzelnen Aspekten	114
Tabelle 9: Soziodemographische Verteilung in der Stichprobe (n = 16.192) im Vergleich zur Grundgesamtheit (n = 44.642).....	127
Tabelle 10: Itemstatistiken	131
Tabelle 11: Items mit hohen Missingquoten, wenn „unzutreffend“-Antwort als fehlender Wert behandelt wird	135
Tabelle 12: Ergebnisse der explorativen Faktorenanalyse – Teststatistiken nach Faktorenzahl für die explorative Faktorenanalyse (n = 8.134)	137
Tabelle 13: Faktorladungen nach Extraktionen von 8 Faktoren (EFA, ML-Methode, Oblimin, n = 8.134)	141
Tabelle 14 - Skalenbildung bei einer 8-Faktoren-Lösung.....	142
Tabelle 15: Modellvergleich der mit CFA geschätzten Modelle für 1, 4 und 8 Faktoren Teil I (n = 8.058).....	144
Tabelle 16: Modellvergleich der mit CFA geschätzten Modelle für 1, 4 und 8 Faktoren Teil II (n = 8.058).....	144
Tabelle 17: Neuentwurf auf Basis der Fragebogenrevision (AIGScreen 3.0)	186

Abbildungsverzeichnis

Abbildung 1: Belastungs-Beanspruchungs-Modell (Eigene Darstellung)	14
Abbildung 2: Formatives Messmodell – Bsp. Sozio-Ökonomischer Status (Eigene Darstellung).....	31
Abbildung 3: Reflektives Messmodell – Bsp. Arbeitszufriedenheit (Eigene Darstellung)	32
Abbildung 4: Phasen der Entwicklung des AIGScreen.....	42
Abbildung 5: Fragen zur psychischen Belastung im AIGScreenBw 2.0	45
Abbildung 6: Ablaufmodell der quantitativen Inhaltsanalyse	75
Abbildung 7: Berufsgruppenzugehörigkeit der Teilnehmer an der Expertenbefragung (n = 92)	96
Abbildung 8: Einsatzbereich der Teilnehmenden an der Expertenbefragung (n = 93).....	97
Abbildung 9: Bewertung der Vollständigkeit der Subskala Arbeitsorganisation.....	101
Abbildung 10: Bewertung der Vollständigkeit der Subskala Arbeitsinhalt und -aufgabe	101
Abbildung 11: Bewertung der Vollständigkeit der Subskala Soziale Aspekte.....	102
Abbildung 12: Bewertung der Vollständigkeit der Subskala Arbeitsumgebung	102
Abbildung 13: Ergebnisse der explorativen Faktorenanalyse – Scree-Plot (n = 8.134)	136
Abbildung 14: Ergebnisse der explorativen Faktorenanalyse – Parallel-Analyse (n = 8.134)	137
Abbildung 15: Faktorladungen bei 4 Faktoren (EFA, Oblimin, n = 8.134)	138
Abbildung 16: Faktorladungen bei 8 Faktoren (EFA, Oblimin, n = 8.134)	139

Anlagenverzeichnis

Anlage 1: AIGScreen 2.0	225
Anlage 2 – Checkliste für Fragebogenreview	236
Anlage 3 – Fragebogen für die Expertenbefragung	239
Anlage 4: Auswertung der Freitextantworten der Expertenbefragung auf die Frage „Bitte notieren Sie, welche Aspekte Ihrer Meinung nach bezüglich der beschriebenen Anforderungen im Fragebogen enthalten sein müssen“ (n = 376)	250
Anlage 5: Ergebnisse der Itemanalyse – Häufigkeitsverteilungen der Antworten auf alle 34 Items als Histogramm	254
Anlage 6: Itemkorrelationen nach Pearson und Spearman	271
Anlage 7: Ergebnisse der explorativen Faktorenanalyse IV: Faktorladungen bei 4 extrahierten Faktoren nach Varimax-Rotation (n = 8.134)	285
Anlage 8: Ergebnisse der explorativen Faktorenanalyse V: Faktorladungen bei 6 extrahierten Faktoren nach Varimax-Rotation (n = 8.134)	286
Anlage 9: Ergebnisse der explorativen Faktorenanalyse VI: Faktorladungen bei 8 extrahierten Faktoren nach Oblimin-Rotation (n = 8.134)	287

Abkürzungsverzeichnis

ABC-L	Arbeits-Bewertungs-Check für Lehrkräfte
AIC	Akaike Information Criterion
AIGScreen	Fragebogen zur Analyse der Arbeitssituation und dem individuellen Gesundheitsempfinden
AIGScreenBw	Fragebogen zur Analyse der Arbeitssituation und dem individuellen Gesundheitsempfinden in der Bundeswehr und dem gesamten Geschäftsbereich BMVg
ASU	Institut für Arbeits-, Sozial- und Umweltmedizin der Universitätsmedizin der Johannes Gutenberg-Universität Mainz
ArbSchG	Arbeitsschutzgesetz
AVEM	Arbeitsbezogene Erlebens und Verhaltensmuster
BGF	Betriebliche Gesundheitsförderung
BGM	Betriebliches Gesundheitsmanagement
BIC	Bayesian Information Criterion
BMVg	Bundesministeriums der Verteidigung
CFA	Confirmatory Factor Analysis, Konfirmatorische Faktorenanalyse
CFI	Comparative Fit Index
COPSOQ	Copenhagen Psychosocial Questionnaire
CVI	Content Validity Index
CVR	Content Validity Ratio
df	Degrees of freedom, Freiheitsgrade
EFA	Explorative Faktorenanalyse
FASS	Fragebogen zur Arbeitssituation an Schulen
FIML	Full Information Maximum Likelihood
GDA	Gemeinsame Deutsche Arbeitsschutzstrategie
ICR	Intercoderreliabilität
KMO	Kaiser-Meyer-Olkin-Kriterium
MAP	Minimum Average Partial Correlation nach Valicer
MAR	Missing at Random
MCAR	Missing Completely at Random

MI	Multiple Imputation
MICE	Multiple Imputation with Chained Equations
ML	Maximum Likelihood
MNAR	Missing Not at Random
MSA	Measure of sampling adequacy
PCA	Principal Components Analysis, Hauptkomponentenanalyse
PHQ	Patient Health Questionnaire
RMSEA	Root Mean Square Error of Approximation
SABIC	Sample Adjusted Bayesian Information Criterion
SRMR	Standardized Root Mean Square Residual
TLI	Tucker-Lewis-Index
UM	Universitätsmedizin der Johannes Gutenberg-Universität Mainz
WHO	Weltgesundheitsorganisation (World Health Organization)

1 Hintergrund

Aus dem Arbeitsschutzgesetz (ArbSchG) geht eine umfassende Vorsorgepflicht des Arbeitgebers gegenüber den Arbeitnehmenden bei der Verhütung von Gesundheitsschäden und –gefahren hervor. Einerseits hat der Arbeitgeber alle erforderlichen Maßnahmen zum Schutz und Erhalt der Gesundheit seiner Angestellten zu ergreifen, andererseits ist er für die Ermittlung und Beurteilung von Gefährdungen am Arbeitsplatz seiner Beschäftigten verantwortlich. Bezüglich der Schutzmaßnahmen besteht darüber hinaus eine Pflicht zur Überprüfung der Wirksamkeit und – nach Möglichkeit – der Optimierung sofern die Gefährdungsbeurteilung ergibt, dass noch Gesundheitsrisiken bestehen (2).

Neben der gesetzlichen Pflicht bestehen für Unternehmen zusätzliche Anreize, Gesundheitsförderung und -management in die Unternehmensstrategie zu integrieren. Eine Vielzahl an Studien belegt den ökonomischen Nutzen von betrieblicher Gesundheitsförderung (BGF) und betrieblichem Gesundheitsmanagement (BGM) sowohl aus volkswirtschaftlicher Perspektive, zum Beispiel in Form geringerer Kosten für das Gesundheitssystem, als auch für Unternehmen, welche von einer höheren Produktivität und weniger Ausfallzeiten ihrer Mitarbeitenden profitieren (3–6). Der durchschnittliche Produktionsausfall eines Arbeitsunfähigkeitstages deutscher Arbeitnehmer beträgt etwa 102€ bis 149€. Der Gesamtverlust an Bruttowertschöpfung beträgt bis zu 266€ pro Arbeitsunfähigkeitstag. Im Jahr 2018 war der durchschnittliche Arbeitnehmer 17 Tage arbeitsunfähig. Die häufigsten Gründe für Arbeitsunfähigkeit waren Muskel-Skelett-Erkrankungen und psychische Erkrankungen (7). Während die gesetzlichen Krankenkassen leichte Rückgänge der Arbeitsunfähigkeitstage aufgrund von Muskel-Skelett-Erkrankungen melden, stieg der Anteil der psychischen Ursachen in den letzten Jahren kontinuierlich an (8–10).

Hinzu kommen Befunde aus einer Meta-Analyse, wonach die durchschnittliche Einsparung durch Ausfallzeiten die Ausgaben für Gesundheitsförderung um den Faktor 2,7 übersteigt (11). Hochart et al. (2011) zeigten in einer Längsschnittanalyse für einen Zeitraum von 3 Jahren, dass die Gesundheitsausgaben für Teilnehmer eines Gesundheitsprogramms im Erhebungszeitraum um weniger als 1% stiegen, während die Kosten für Nichtteilnehmer im gleichen Zeitraum um rund 22% zunahmen (12). Chapman (2012) konnte in einer Meta-Analyse zeigen, dass die Kosten der Berufsunfall- und Invaliditätsversicherung durch betriebliche Gesundheitsförderung um durchschnittlich 25 % gesenkt werden (3).

Gleichzeitig ergeben sich durch die Veränderung von Anforderungsstrukturen des Arbeitens und der Entstehung neuer Arbeitsformen neue Belastungsformen, welche überwiegend psychischer Natur sind (13). Die Europäische Agentur für Sicherheit und Gesundheitsschutz am Arbeitsplatz nennt beispielsweise neue Formen von Arbeitsverträgen und unsichere

Arbeitsplätze, Arbeitsintensivierung, eine schlechte Work-Life-Balance und hohe emotionale Anforderungen am Arbeitsplatz als die wichtigsten neuen Risiken im beruflichen Kontext (14).

Aufgrund von Veränderungen der Arbeitsstruktur und der Arbeitsanforderungen ist auch international eine Zunahme von Arbeitsstress und psychosozialen Gesundheitsrisiken am Arbeitsplatz zu beobachten (13–15). Unter den Gesundheitsgefahren am Arbeitsplatz finden psychische Risiken in allen EU-Mitgliedstaaten besondere Beachtung und wurden von der Europäischen Agentur für Sicherheit und Gesundheitsschutz am Arbeitsplatz hinsichtlich des Forschungsbedarfs mit der höchsten Priorität versehen (16).

Vor dem Hintergrund einer alternden Belegschaftsstruktur und dem sich abzeichnenden Fachkräftemangel in einigen Branchen, wird zudem betriebliches Gesundheitsmanagement (BGM) ein zunehmend wichtigeres Thema für Unternehmen. Diese profitieren einerseits davon, dass sie die Arbeitsfähigkeit ihrer Mitarbeitenden möglichst lange erhalten und andererseits als Arbeitgeber im Wettbewerb um potenzielle Beschäftigte und Fachkräfte attraktiv bleiben.

Wenngleich BGM die arbeitsmedizinische und sicherheitstechnische Betreuung von Unternehmen und Beschäftigten nicht ersetzen kann, könnte sie zumindest im Sinne eines ressourcenstärkenden Ansatzes helfen, die unzureichenden arbeitsmedizinischen Versorgungsstrukturen in Deutschland in Teilen abzuf puffern. Ein Bericht der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (17) stellte bereits für das Jahr 2012 eine erhebliche Diskrepanz zwischen arbeitsmedizinischem Betreuungsbedarf und den tatsächlich verfügbaren Kapazitäten in Deutschland fest. Zudem wurde ein weiteres Auseinanderdriften von Bedarf und Angebot prognostiziert. Um der Vergrößerung dieser Betreuungslücke entgegenzuwirken, wird eine bedarfsorientierte Steuerung und Planung sowie die Nutzung bestehender Ressourcen im Unternehmen empfohlen. Die Bundesanstalt für Arbeitsschutz und Arbeitsmedizin schlägt einen multidisziplinären Ansatz als eine Möglichkeit vor, den Einsatz von Ressourcen im Arbeitsschutz zu optimieren. Durch die Einbindung interner und externer Akteure wie Sicherheitsingenieure, Ergonomen, Psychologen oder Statistiker oder Institutionen wie Universitäten können Arbeitsmediziner entlastet werden und sich verstärkt auf ihre Kernaufgabe konzentrieren. Durch die Nutzung des Expertenwissens der verschiedenen Disziplinen können Synergieeffekte genutzt werden, die ein umfassenderes Wissen über Handlungsbedarfe und ein maßgeschneidertes und ganzheitliches Gesundheitsmanagement ermöglichen.

Vor dem Hintergrund dieser Problematik – einer steigenden Zahl psychischer Belastungen und Beanspruchungen am Arbeitsplatz auf der einen Seite und sinkenden Beratungs- und Betreuungskapazitäten im Arbeitsschutz und in der Arbeitsmedizin auf der anderen Seite, sollte ein Fragebogen entwickelt werden, der die folgenden Kriterien erfüllt: Erstens sollte er

in verschiedenen Branchen und Tätigkeitsbereichen leicht anwendbar sein. Zweitens sollten die Ergebnisse einen hohen Abstraktionsgrad aufweisen, um eine Vergleichbarkeit zwischen verschiedenen Erhebungseinheiten zu ermöglichen. In diesem Zusammenhang ist es wünschenswert, ein System von Indikatoren bzw. Kennzahlen abzuleiten, die einen Vergleich zwischen Organisationseinheiten sowie eine Bewertung des Erfolgs von Maßnahmen des Gesundheitsmanagements ermöglichen. Drittens sollte der Fragebogen alle relevanten Themenbereiche in Bezug auf psychische Gefährdungen bei der Arbeit abdecken.

1.1 Entwicklung des AIGScreen

Der AIGScreen entstammt einem Fragebogen zur Analyse der Arbeitssituation und dem Gesundheitsempfinden von Lehrkräften und Pädagogischen Fachkräften. Im Rahmen eines Pilotprojekts zur Einführung von Betrieblichen Gesundheitsmanagement (BGM) im Geschäftsbereich des Bundesministeriums der Verteidigung (BMVg) erfolgte eine Weiterentwicklung und Anpassung an das dortige spezifische Arbeitssetting. In Folge der Ausrollung des BGM auf den gesamten Geschäftsbereich des BMVg wurden neue Anwendungsfelder wie beispielsweise die längsschnittliche Evaluation von Maßnahmen an einzelnen Standorten ergänzt. Dadurch ergeben sich auch neue Anforderungen an den Fragebogen, insbesondere mit Blick auf die wissenschaftlichen Gütekriterien. In diesem Abschnitt soll kurz die Entwicklung des Fragebogens erläutert werden. Auf die Chronologie der jeweiligen Rahmenbedingungen und Anforderungen einzugehen, ist notwendig, um aufzuzeigen, warum bisher dem Kriterium der Praktikabilität bzw. des Praxisbezugs hohes Gewicht beigemessen wurde. Dies bedeutet jedoch nicht, dass das Instrument einer theoretischen Fundierung entbehrt. Diese wird im darauffolgenden Kapitel ausführlich beschrieben.

1.1.1 Analyseinstrument zur Beurteilung von Gefährdungen am Arbeitsplatz Schule

Im Kontext der arbeitsmedizinischen Betreuung von Lehrkräften und pädagogischen Fachkräften in Rheinland-Pfalz durch das Institut für Lehrergesundheit am Institut für Arbeits-, Sozial- und Umweltmedizin der Universitätsmedizin Mainz (ASU) wurde ein Fragebogen zur Analyse der Arbeitssituation und dem individuellen Gesundheitsempfinden (AIGScreen) entwickelt (18). Dieser sollte vor allem als Grobscreening die Gefährdungsbeurteilung psychischer Belastung an den Schulen abdecken und als Grundlage für die weitere Beratung und Durchführung von Angeboten der Gesundheitsförderung dienen (19). Ein Feinanalyse spezifischer Tätigkeiten ist nicht Gegenstand bzw. Ziel des Screeninginstrumentes. Der Wortlaut „*individuelle[s]* Gesundheitsempfinden“ bezieht sich auf die Tatsache, dass mit dem

Fragebogen ein rein subjektives Empfinden der eigenen Gesundheit der Probanden erfasst wird.

Der Inhalt des Fragebogens orientiert sich am aktuellen Stand der Forschung und an bereits entwickelten Instrumenten. Die theoretische Grundlage bildet das Belastungs-Beanspruchungs-Modell (siehe Abschnitt 2.1.2 „Das Belastungs-Beanspruchungs-Modell“). Fragen, die sich auf die Arbeitssituation beziehen, orientieren sich zum Teil an den Instrumenten „AVEM“ (Arbeitsbezogene Erlebens und Verhaltensmuster) (20), „ABC-L“ (Arbeits-Bewertungs-Check für Lehrkräfte) (21), „FASS“ (Fragebogen zur Arbeitssituation an Schulen) (22), „COPSOQ“ (Copenhagen Psychosocial Questionnaire) (22), wurden jedoch vollständig in eigene Formulierungen überführt. Die genannten Instrumente sind allesamt inhaltlich gut konzipierte Fragebögen zur Analyse von Arbeitssituationen, Arbeitsbelastungen, Stressreaktionen und / oder der wahrgenommenen Gesundheitssituation. Jedoch konnte keiner von ihnen alle interessierenden Themen im Spektrum psychischer Belastungen am Arbeitsplatz abdecken oder war nicht für ein allgemeines Screening geeignet, da die Anzahl der Fragen zu hoch war. Im Rahmen von Expertenreviews wurden die wesentlichsten Aspekte identifiziert, die mit dem AIGScreen in Bezug auf die Arbeitssituation erhoben werden sollten. In Anlehnung an bestehende Fragebögen wurden Items generiert und so ausgewählt, dass sie alle relevanten Arbeitsaspekte mit Blick auf die psychische Belastung repräsentieren. Im ersten Teil des Fragebogens werden Arbeitsmerkmale (Belastungen) erfasst, während der zweite Teil auf das physische und psychische Gesundheitsempfinden und mögliche Beanspruchungsreaktionen fokussiert. Dabei sollen insbesondere bei den Belastungen möglichst wenige Items die jeweiligen Konstrukte wie z. B. Arbeitsmenge oder Arbeitsumgebung repräsentieren. Die Anzahl der Items, die zur Erfassung der Arbeitssituation benötigt werden, konnte auf 28 reduziert werden. Diese verkürzte Version kann als praktikabler angesehen werden als z. B. der COPSOQ-Fragebogen mit mindestens 87 und bis zu 157 Items (23). Schließlich ist hinsichtlich der Bearbeitungsdauer des Fragebogens nicht nur die (Arbeits-) Zeit zu berücksichtigen, die die Teilnehmenden zur vollständigen Beantwortung aufwenden müssen, sondern auch die Akzeptanz bzw. Teilnahmebereitschaft, die mit zunehmender Länge eines Fragebogens sinkt (1, 24, 25). Eine hohe Rücklaufquote wurde jedoch als notwendige Bedingung für verlässliche Schlüsse über die Gesamtheit der Beschäftigten einer Stichprobe angesehen.

Bis 2015 wurde der AIGScreen in 156 Schulen eingesetzt. Insgesamt nahmen 4.156 Beschäftigte an der Befragung teil. Anhand des Antwortverhaltens und Rückmeldungen seitens der Probanden wurde der Fragebogen einer jährlichen Überprüfung durch eine interdisziplinär besetzte Expertengruppe unterzogen. Anhand der Review-Ergebnisse erfolgte bei erkanntem Verbesserungsbedarf eine Überarbeitung des Fragebogens in Bezug auf die

Verständlichkeit der Fragen und die Vollständigkeit hinsichtlich der für die Befragten relevanten Arbeitsbedingungen.

Im Jahr 2015 erfolgte Rahmen eines Pilotprojekts zur Einführung von BGM im Geschäftsbereich des BMVg eine Weiterentwicklung und Anpassung des Fragebogens für das neue Befragungssetting.

1.1.2 Weiterentwicklung zum BGM-Tool

In einem Erprobungsprojekt zur Implementierung von betrieblichem Gesundheitsmanagement (BGM) in elf Pilotdienststellen des BMVg wurde der für Lehrkräfte entwickelte Fragebogen an die Rahmenbedingungen des BMVg angepasst. Einerseits waren bestimmte Formulierungen anzupassen (So wurde z. B. „Kollegium“ durch „Kameraden“ ersetzt) und andererseits waren bestimmte Frageinhalte (wie z. B. zum Verhalten der Schülerinnen und Schüler) für dieses neue Arbeitssetting nicht relevant. Gleichzeitig wurden auch neue Inhalte wie beispielsweise das individuelle Gesundheitsverhalten als eine Ressource im Sinne des Belastungs-Beanspruchungs-Modells (eine Erläuterung dieses Modells findet sich in Abschnitt 2.1.2 „Das Belastungs-Beanspruchungs-Modell“) in den Fragebogen aufgenommen. Es sollte auch möglich sein, anhand der Befragungsergebnisse Aussagen über den Erfolg der angebotenen Maßnahmen treffen zu können (26). Somit wurde der AIGScreen entsprechend der eingangs skizzierten Kriterien – Indikatorenbildung, hoher Abstraktionsgrad, Niedrigschwelligkeit – umgestaltet und als Steuerungsinstrument im BGM etabliert. Denn während betriebliche Gesundheitsförderung (BGF) lediglich alle Maßnahmen innerhalb einer Organisation zur Optimierung der Arbeitsbedingungen und zur Stärkung der individuellen Ressourcen beschreibt, erweitert das Konzept von BGM die Gesamtheit dieser Maßnahmen um eine strategische Komponente, welche langfristige Planung und Zielorientierung, Steuerung und Evaluation umfasst. BGF beschreibt demnach Maßnahmen auf der Ebene der Verhaltens- und Verhältnisprävention und ist Teil des BGM, welches wiederum als ganzheitliches Managementsystem Prozesse in die Unternehmensstrategie integriert und auf ihre Wirksamkeit überprüft (27). Hinsichtlich der Bewertung des Nutzens und der Wirksamkeit von BGM existiert bisher kein einheitliches methodisches Vorgehen, wenngleich die Befunde weitestgehend übereinstimmend auf einen ökonomischen Nutzen von BGM deuten. So wurden beispielsweise der Krankenstand im Unternehmen, Versicherungskosten, Ausgleichszahlungen oder Präsentismus nachweislich durch BGM reduziert (3).

Aufgrund der positiven Bilanz des Pilotprojekts, sollte eine Ausrollung des BGM auf den gesamten Gesamtgeschäftsbereich des BMVg erfolgen. Damit ergaben sich neue Anforderungen an den Fragebogen. Primär sollte er weiterhin als Instrument der Bedarfsanalyse bei der Planung und Entwicklung adäquater BGM-Angebote an den einzelnen

Standorten dienen. Gleichzeitig sollte die Befragung regelmäßig wiederholt werden, sodass sich aus einer längsschnittlichen Analyse Kennzahlen zur Beurteilung des Erfolgs des BGM und der standortspezifischen Maßnahmen ermitteln lassen. Hierfür bedarf es eines verlässlichen Messinstruments, welches alle wissenschaftlichen Gütekriterien (siehe Kapitel 2.3 „Wissenschaftliche Gütekriterien“) erfüllt. Damit lag eine Überprüfung der wissenschaftlichen Güte hinsichtlich der Kriterien Objektivität, Reliabilität und Validität des Instruments nahe. Erste Analysen im Pilotprojekt konnten die Objektivität des Instruments bestätigen, deuteten jedoch bei einigen Subskalen zur Erfassung der Arbeitsmerkmale auf die Notwendigkeit zur Verbesserung der Reliabilität hin (Cronbachs Alpha < .80). Die Validität des Instruments wurde bisher nicht überprüft, da es jedoch aufbauend auf aktuellen wissenschaftlichen Erkenntnissen und durch Experten der Arbeitswissenschaft entwickelt wurde, ist eine mindestens akzeptable Inhaltsvalidität zu erwarten.

1.2 Ziel der Arbeit

Aus den vorausgegangenen Erläuterungen lassen sich zwei wesentliche Schlüsse ziehen. Erstens, es besteht ein hoher Bedarf an einem Screeninginstrument zur ressourceneffizienten Beurteilung von psychischen Gefährdungen am Arbeitsplatz. Dieses kann und soll die Expertise von geschulten Arbeitssicherheitsfachkräften und Arbeitsmedizinern nicht ersetzen. Aber es kann als erstes Grobscreening wichtige Erkenntnisse liefern, welchen Bereichen die entsprechenden Experten eine erhöhte Aufmerksamkeit schenken sollten. Damit kann es einen Beitrag im Arbeitsschutz leisten, indem es die durch den Fachkräftemangel verursachten Versorgungsdefizite ein wenig abmildert. Dafür muss es jedoch – und das ist die zweite wichtige Schlussfolgerung – verlässlich sein. Verlässlichkeit meint hier, dass es zuverlässig alle relevanten Aspekte analysiert und messgenaue Ergebnisse zum Ausmaß etwaiger Belastungen und Risiken liefert. Erfüllt es diese Bedingung, können das Screeninginstrument und die damit gewonnenen Ergebnisse auch als Vergleichs- und Steuerungsinstrument, beispielsweise im Rahmen von BGM, eingesetzt werden. Diese Verlässlichkeit kann anhand der drei wissenschaftlichen Hauptgütekriterien Objektivität, Reliabilität und Validität bemessen werden.

Mit dem AIGScreen ist ein Grobscreening entwickelt worden, welches teilweise bereits für den genannten Zweck zum Einsatz kommt. Bisher steht jedoch eine umfassende Überprüfung der wissenschaftlichen Güte aus. Dies stellt den Gegenstand der vorliegenden Arbeit dar. Eine große Herausforderung besteht dabei darin, sowohl den wissenschaftlichen Standards gerecht zu werden als auch ein praxistaugliches Instrument zu entwickeln. Häufig finden sich in der Literatur neben den sogenannten Hauptgütekriterien, also Objektivität, Reliabilität und Validität

auch sogenannte Nebengütekriterien, welche auch als Kriterien der Praxistauglichkeit verstanden werden können.

Hierzu finden sich verschiedene Empfehlungen in der Literatur. So benennt Leonhart (28) die vier Nebengütekriterien Ökonomie, Nützlichkeit, Vergleichbarkeit und Normierung. Moosbrugger und Kelava (29) empfehlen darüber hinaus Zumutbarkeit, Fairness und Unverfälschbarkeit als Kriterien eines wissenschaftlich fundierten Tests zu berücksichtigen. Welche Nebengütekriterien zusätzlich zu den Hauptgütekriterien als bedeutsam für die Bewertung eines Instruments erachtet werden, kann auch vom Zweck, dem das Instrument dienen soll, abhängig gemacht werden. So ist Fairness, verstanden als die Abwesenheit von Nachteilen für bestimmte Gruppen von Merkmalsträgern wie z. B. Frauen oder ältere Menschen, fraglos ein zentrales Kriterium bei Leistungstests. Unverfälschbarkeit, d. h. dass Probanden durch ihr Verhalten nicht die Testergebnisse beeinflussen können, ist insbesondere dort ein relevantes Kriterium, wo Probanden die Testintention erkennen und aufgrund sozial erwünschten Antwortverhaltens steuern wollen. Da der Zielgruppe des AIGScreen ein Interesse an einem möglichst wahrheitsgemäßen Ergebnis unterstellt wird und Fragen nach dem individuellen Empfinden nicht richtig oder falsch beantwortet werden können, weshalb Probanden auch nicht besser oder schlechter in einem Test abschneiden können, sind die Kriterien der Fairness und der Unverfälschbarkeit als nachrangig zu betrachten.

Hingegen soll das Kriterium der Zumutbarkeit insbesondere deshalb berücksichtigt werden, weil eine hohe Teilnahme wünschenswert ist, um ein vollständiges und repräsentatives Bild der Arbeitssituation zu erhalten. Ein Test gilt als zumutbar, wenn der „aus seiner Anwendung resultierende Nutzen die zu testende Person in zeitlicher, psychischer sowie körperlicher Hinsicht nicht über Gebühr belastet“ (30). Vor allem auf den zeitlichen Aspekt muss hierbei Rücksicht genommen werden, da die Beantwortung des gesamten Fragebogens bereits in der aktuellen Version des AIGScreen 15 bis 20 Minuten in Anspruch nimmt und eine durchschnittliche Bearbeitungsdauer von 30 Minuten nicht überschritten werden sollte. Dies beeinflusst auch ein weiteres Nebengütekriterium, nämlich die Testökonomie. Darunter versteht man den Anspruch mit möglichst geringem Ressourcenaufwand einen möglichst hohen Nutzen zu erzielen. Neben der Begrenzung des zeitlichen Aufwands, soll auch der Materialaufwand durch den Einsatz von Online-Befragungen anstelle von Papierfragebögen auf ein notwendiges Minimum reduziert werden.

Die Nützlichkeit des AIGScreens ergibt sich aus seiner Praxisrelevanz. Erfüllt das Instrument die Hauptgütekriterien, kann es zur Gefährdungsbeurteilung psychischer Belastungen am Arbeitsplatz, zur Entwicklung von gesundheitsfördernden Maßnahmen im Betrieb und zu deren Evaluation eingesetzt werden.

Um die Testergebnisse einer Stichprobe in Relation zu anderen Populationen setzen und bewerten zu können, ist ein universell einsetzbares Instrument wünschenswert. Das bedeutet, der Fragebogen soll in verschiedenen Branchen und Tätigkeitsbereichen einsetzbar sein. Die Dimensionen des AIGScreen sollen zudem zu Evaluationszwecken in wenigen aussagekräftigen Kennzahlen zusammengefasst werden können. Da mit dem Einsatz des AIGScreen auch dem gesetzlichen Auftrag der Gefährdungsbeurteilung arbeitsbedingter psychischer Belastungen begegnet werden soll, müssen diese Dimensionen den Empfehlungen der Gemeinsamen Deutschen Arbeitsschutzstrategie (GDA) zur Durchführung einer Gefährdungsbeurteilung psychischer Belastung entsprechen.

Zusammenfassend lassen sich somit folgende Anforderungen an das Screeninginstrument benennen:

- Es soll **ökonomisch** sein. Das Erhebungsverfahren soll mit einem möglichst geringen Ressourcenaufwand insbesondere bzgl. Zeit und Personal verbunden sein. Das bedeutet in diesem Kontext, es soll die Form eines **Grobscreenings** annehmen, welches auch online eingesetzt werden kann und eine Bearbeitungsdauer von 30 Minuten nicht überschreitet.
- Es soll weiterhin den **Nebengütekriterien** Nützlichkeit, Vergleichbarkeit und Zumutbarkeit gerecht werden.
- Er soll auf Organisationsebene Handlungsschwerpunkte, d. h. besonders häufig auftretende Belastungsfaktoren, identifizieren. Dafür ist eine **Quantifizierung** der erhobenen Informationen notwendig.
- Er soll möglichst **universell** einsetzbar sein, d. h. er soll in gleicher Form für alle Bereiche innerhalb des Geschäftsbereichs des BMVg (z. B. Verwaltung, Krankenhaus, Marine, IT, Universität) eingesetzt werden können. Er soll gleichzeitig auf möglichst viele andere Branchen und Tätigkeitsbereiche übertragbar sein.
- Er soll eine standardisierte Berichterstattung ermöglichen, d. h. für alle Organisationseinheiten sollen die gleichen Merkmale in gleicher Form erhoben werden, sodass beispielsweise ein Vergleich mittels **Kennwerten** möglich ist.
- Er soll der Orientierung dienen, d. h. es soll ein grober Überblick über die Belastungssituation ermöglicht werden und gleichzeitig eine **differenzierte Bewertung der einzelnen Aspekte** durch die Mitarbeitenden erfolgen.
- Es muss inhaltlich alle Aspekte abdecken, welche von der Gemeinsamen Deutschen Arbeitsschutzstrategie (GDA) als relevant für die **Gefährdungsbeurteilung psychischer Belastung** erachtet werden (31).

- Da der Schwerpunkt der Befragung bei psychischen Belastungen liegt, welche nicht anhand manifester Variablen, sondern durch die subjektive Wahrnehmung der Betroffenen erfasst werden können (13), muss der Fragebogen demnach die **individuell empfundenen Belastungen** erfassen.

Diesen Anforderungen übergeordnet steht das Ziel der wissenschaftlichen Güte des Instruments. Wird der AIGScreen den **Hauptgütekriterien Objektivität, Reliabilität und Validität** gerecht, werden damit auch gleichzeitig viele weitere Anforderungen erfüllt. Umgekehrt gilt, dass ohne die Erfüllung der wissenschaftlichen Güte viele Anforderungen nicht erfüllt werden können. So ist beispielsweise ein Aspekt der Validität die Inhaltsvalidität, welche gewährleistet, dass ein bestimmter Aspekt oder ein latentes Konstrukt vollständig, umfassend und repräsentativ erhoben wird (32, 1). Die Reliabilität wiederum garantiert Messgenauigkeit und Wiederholbarkeit von Erhebungen, weshalb ein reliables Instrument auch zur Kennzahlenermittlung und zum längsschnittlichen Vergleich eingesetzt werden kann (33).

Auf das genaue methodische Vorgehen wird im entsprechenden Kapitel eingegangen. Hier sei nur ein kurzer Überblick der Analysen gegeben. Für die Überprüfung der meisten Gütekriterien existieren lange erprobte, empirische Methoden. So kann die Reliabilität unter anderem anhand faktoranalytischer Verfahren beurteilt werden (34). Solche Verfahren liefern auch Erkenntnisse über die interne Konsistenz und Homogenitätsstrukturen von Messinstrumenten latenter Konstrukte, weshalb ihre Ergebnisse auch Schlüsse über die Konstruktvalidität zulassen. Für andere Kriterien wie Objektivität und Inhaltsvalidität existieren keine bzw. nur rudimentäre quantitativen Verfahren, weshalb hier vor allem auf das Wissen von Experten zurückgegriffen werden muss (30).

Die genannten Verfahren bieten eine Möglichkeit, Aussagen über die wissenschaftliche Güte eines Instruments zu treffen. Gegenstand der Analyse sind, zumindest bei den quantitativen Verfahren, vor allem die mit dem Instrument erhobenen Daten. Die Analysen beeinflussen somit nicht die wissenschaftliche Güte eines Instruments. Zielsetzung dieser Arbeit ist es demnach auch nicht, am Ende ein Instrument zu präsentieren, welches zwangsläufig allen Gütekriterien vollständig gerecht wird, sondern Aussagen über die wissenschaftliche Güte, mögliche Einschränkungen bei den Anwendungen und bei der Ergebnisinterpretation der damit gewonnenen Daten aufzuzeigen sowie ggf. Verbesserungsvorschläge für eine erhöhte wissenschaftliche Güte machen zu können. Dafür wird ein Vorschlag für einen Neuentwurf des AIGScreen erarbeitet.

2 Theoretische Grundlagen und rechtlicher Kontext

Im folgenden Kapitel sollen zunächst die theoretischen Vorüberlegungen zum Inhalt des Fragebogens erläutert werden. Dabei soll genauer darauf eingegangen werden, was unter einer Gefährdungsbeurteilung psychischer Belastung zu verstehen ist und welche Inhalte diesbezüglich erfasst werden müssen. Es wird außerdem das Belastungs-Beanspruchungs-Modell erläutert, aus welchem sich das dem AIGScreen zugrundeliegende Konzept der Arbeitsbelastung ableitet. Zwar kann und soll diese Arbeit keinen umfassenden Theorienvergleich bezüglich Arbeitssituation und Arbeitsbelastung vornehmen, allerdings ist eine genaue Eingrenzung des Begriffs der Arbeitsbelastung notwendig, da dieser nach theoretischer Orientierung unterschiedlich verwendet wird. Beispielsweise existieren bezüglich der Wertung, die mit der Verwendung dieses Begriffs einhergeht, große Unterschiede. Wohingegen Belastung in der allgemeinen Sprachverwendung in der Regel als etwas Negatives verstanden wird, ist dieser Begriff im Belastungs-Beanspruchungs-Modell vollkommen wertneutral gemeint (35). Gleiche Unschärfen ergeben sich für die Verwendung des Begriffs Gesundheit, welcher insbesondere im alltäglichen Sprachgebrauch häufig ein dichotomes Konstrukt in Abgrenzung zu Krankheit darstellt. Hingegen finden sich in der Fachliteratur häufiger Konzepte, die von einem Gesundheits-Krankheits-Kontinuum ausgehen (36–38). Vor diesem Hintergrund ist es notwendig, im Sinne einer wissenschaftlich korrekten und eindeutigen Sprache, den theoretischen Hintergrund des Fragebogens zu erläutern.

Da die Gefährdungsbeurteilung psychischer Belastung nicht nur ein Instrument des BGM oder eine Maßnahme zur Erhaltung der Gesundheit und Arbeitszufriedenheit von Unternehmen sein kann, sondern auch die gesetzliche Pflicht zur Durchführung einer solchen besteht, wird im Abschnitt 2.2 auf die rechtlichen Grundlagen und offiziellen Empfehlungen zur Durchführung einer solchen Beurteilung eingegangen. Dabei sind die von der Gemeinsamen Deutschen Arbeitsschutzstrategie (GDA), einem Gremium aus Bund, Ländern und Unfallversicherungsträgern, veröffentlichten Empfehlungen besonders relevant, da sie sowohl wichtige Eckpfeiler für die methodische Umsetzung setzen als auch den inhaltlichen Rahmen bzw. die zu erfassenden Gefährdungskategorien eindeutig festlegen.

Zudem sollen die Begrifflichkeiten der Hauptgütekriterien für den Rahmen dieser Arbeit noch einmal konkret definiert werden. Das ist notwendig, da sie in der einschlägigen Literatur nicht immer einheitlich verwendet werden und häufig auch zwischen den Wissenschaftsdisziplinen unterschiedliche Konnotationen der jeweiligen Aspekte gegeben sind (30, 28). Im Sinne einer eindeutigen Sprachverwendung scheint es daher notwendig, noch einmal zusammenzutragen, wie in der Literatur die Begriffe Objektivität, Reliabilität und Validität verstanden werden und wie sie im Rahmen dieser Arbeit zu verstehen sind.

2.1 Gesundheit und das Belastungs-Beanspruchungs-Modell

Im folgenden Abschnitt sollen der Gesundheitsbegriff und das dem AIGScreen zugrundeliegende theoretische Modell erläutert werden. Es existieren vielzählige Theorien zum Zusammenhang von Umwelteinflüssen und der Entstehung von Krankheit. In Bezug auf den Arbeitskontext werden neben dem Belastungs-Beanspruchungs-Modell häufig das Gratifikationskrisenmodell nach Siegrist, das Job Demand Control Modell nach Karasek und das transaktionale Stressmodell in der einschlägigen Literatur aufgeführt (13, 39–41). Ein Theorienvergleich oder die Diskussion der jeweiligen Vor- und Nachteile einzelner theoretischer Ansätze wird hier nicht vorgenommen, da das nicht Thema dieser Arbeit ist bzw. eine wissenschaftliche Abhandlung für sich darstellen würde. Es ist jedoch wichtig, die theoretischen Vorüberlegungen des Instruments und damit verbunden auch das, was gemessen werden soll, klar einzugrenzen, in seinen Begrifflichkeiten und Dimensionen zu definieren und von anderen Konzepten und Begriffen abzugrenzen. Denn nur, wenn klar definiert ist, was wie gemessen werden soll, kann auch überprüft werden, ob diese Messung den wissenschaftlichen Hauptgütekriterien entspricht.

Im Folgenden wird daher einmal das dieser Arbeit zugrundeliegende Verständnis von Gesundheit nach Antonovsky (42) und zum anderen das Belastungs-Beanspruchungsmodell nach Rohmert (35) erläutert. Beide Ansätze eignen sich im Rahmen Betrieblicher Gesundheitsförderung besonders gut, da sie einerseits die Komplexität verschiedener Faktoren und deren Wechselwirkung bei der Entstehung von Krankheit und Beanspruchungsreaktionen berücksichtigen und andererseits auch einbeziehen, wie Umwelteinflüsse gesundheitsförderliche wirken können.

2.1.1 Der Gesundheitsbegriff nach Antonovsky

Während eine Vielzahl konkurrierender bzw. sich gegenseitig ergänzender Modelle bezüglich der Wirkung von Arbeit auf Gesundheit existieren, beziehen sich Gesundheitsförderungskonzepte in der Regel auf die Definition von Gesundheit, wie sie in der Salutogenese beschrieben und von der WHO verwendet wird. Demnach umfasst Gesundheit neben einer körperlichen Dimension auch eine psychische, eine soziale und eine existenzielle. Gesundheit ist somit mehr als die Abwesenheit manifester physischer (Krankheits-)Symptome. Gesundheit und Krankheit stellen im Rahmen dieser Theorie Pole des gleichen Kontinuums dar. Wo sich ein Individuum auf der Achse zwischen Gesundheit und Krankheit bewegt, hängt von den Umwelteinflüssen und individuellen Ressourcen ab (37).

Dem Konzept der Salutogenese nach Antonovsky folgend, ist die Verortung auf diesem Gesundheits-Krankheits-Kontinuum von einem mehrdimensionalen Zusammenspiel

verschiedener Faktoren, Risiken wie auch Ressourcen, und den jeweiligen Kontexten, in denen sich ein Individuum bewegt, abhängig. Es gibt demnach keine einfachen Ursache-Wirkungs-Beziehungen, die die Entstehung von Krankheit oder deren Abwesenheit erklären können. Das liegt einmal daran, dass Krankheit und Gesundheit nicht nur an- oder abwesend sein können, sondern beide in unterschiedlichen Ausprägungen gleichzeitig gegeben sein können. Zum Zweiten beeinflusst ein sehr komplexes und interdependentes Geflecht aus Ressourcen, Risiken und den Rahmenbedingungen verschiedener Ebenen bzw. Lebenswelten den Gesundheitszustand eines Menschen (43). Damit ergeben sich einerseits größere Herausforderungen für die Gesundheitsförderung und das betriebliche Gesundheitsmanagement, da das potenziell gesundheitsgefährdende Geflecht verschiedenere Einflussgrößen schwer in seiner gesamten Komplexität zu erfassen und entsprechend zu beeinflussen ist (zumal im Handlungsrahmen der Arbeitgeber ohnehin nur das Arbeitssetting beeinflusst werden kann). Andererseits werden damit auch neue Chancen aufgezeigt. Zum Beispiel kann im beruflichen Kontext nicht nur in puncto Gesundheitsgefahren am Arbeitsplatz und Unfallverhütung die Gesundheit der Beschäftigten geschützt und erhalten werden. Gleichzeitig wird mit diesem Konzept aufgezeigt, dass es auch sinnvoll im Sinne der Gesundheitsförderung ist, Rahmenbedingungen zu gestalten und Ressourcen zu fördern. Für den Kontext der psychischen Belastungen am Arbeitsplatz hieße das beispielsweise, dass es für Beschäftigte, die sich wegen einer zu hohen Arbeitsmenge gestresst fühlen, nicht nur die Möglichkeit gibt, die Arbeitsmenge zu reduzieren. Es kann auf Verhaltensebene auch geprüft werden, ob ihre Ressourcen gestärkt werden können, indem man sie beispielsweise in Zeitmanagement schult. Und es kann überprüft werden, ob die Rahmenbedingungen im Sinne der Verhältnisprävention optimiert werden können. Zum Beispiel könnte Stress entstehen, weil Arbeitsprozesse nicht gut aufeinander abgestimmt sind und deshalb unnötige Mehrarbeit, Frustration und Zeitverlust verursachen. Auch ein ungünstiges Verhältnis zwischen Anforderungen und Qualifikation ist denkbar, was durch Weiterbildungs- und Schulungsmaßnahmen verbessert werden kann.

Gesundheitsförderung im Unternehmen bezieht sich somit auf zwei Ebenen, die Verhältnisse am Arbeitsplatz und das individuelle Verhalten. Verhältnisprävention setzt bei den Arbeitsbedingungen an und soll die Belastungssituation der Beschäftigten im Betrieb bzw. am Arbeitsplatz so gestalten, dass diese nicht gesundheitsschädigend und bestenfalls sogar gesundheitsfördernd wirken sowie die Arbeitsleistung und -motivation fördern. Verhaltensprävention setzt bei den individuellen Ressourcen und dem Gesundheitsverhalten von Individuen an und soll Beanspruchungen minimieren. Dabei sollte sowohl bei der Verhaltens- als auch bei der Verhältnisprävention nach Möglichkeit sowohl ressourcen- als belastungsorientiert vorgegangen werden, d. h. Ziel sollte es einerseits sein, negative Faktoren zu verringern oder zu beseitigen und andererseits positive, also gesundheitsförderliche

Faktoren auszubauen und zu stärken. Da das individuelle Verhalten maßgeblich von den Rahmenbedingungen bestimmt ist, sollte verhaltenspräventiven Maßnahmen stets eine Analyse der Verhältnisse und ggf. die Ableitung adäquater Maßnahmen auf Ebene der Verhältnisprävention vorangestellt sein (44).

2.1.2 Das Belastungs-Beanspruchungs-Modell

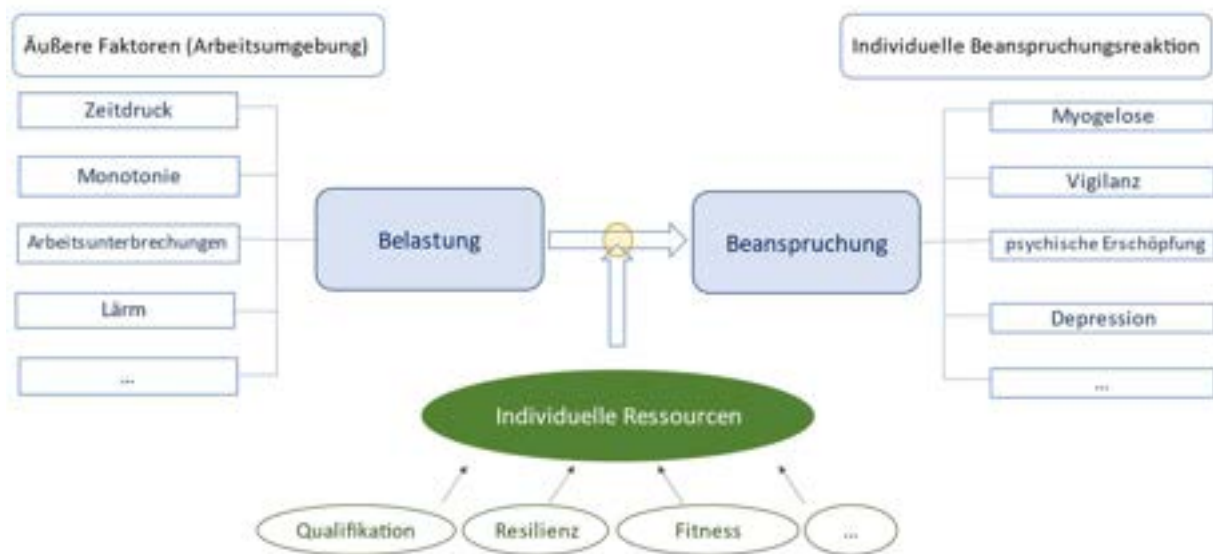
So wie der Gesundheitsbegriff von Antonovsky als internationaler Standard für das Verständnis von Gesundheit gilt, so kann das Belastungs-Beanspruchungs-Modell als internationaler Standard für ein theoretisches Modell zur Wirkung von (arbeitsbedingten) Belastungen auf das Individuum und die daraus resultierende Beanspruchung betrachtet werden. Denn dieses Modell findet nicht nur in der Arbeitsmedizin weite Verbreitung, es ist auch Grundlage der internationalen Norm DIN EN ISO 10075 „Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung“, welche die Grundlage für die Ermittlung psychischer Belastungen am Arbeitsplatz und für die entsprechenden Gestaltungsmaßnahmen bildet (40).

Entsprechend dient das Belastungs-Beanspruchungs-Modell auch als theoretische Grundlage für das Screening-Instrument AIGScreen. Grundsätzlich wird in diesem Konzept in drei Parameter unterschieden: Belastungen, Beanspruchungen und individuelle Ressourcen. Im Gegensatz zu vielen anderen Theorien und Modellen zur Arbeitsbelastung werden diese drei Parameter dabei vollkommen wertneutral verstanden (45). Das bedeutet, dass weder die Einwirkungen von Arbeitsbelastungen oder anderen Umweltfaktoren auf das Individuum noch Beanspruchungen als Reaktion auf die Arbeitsbelastung als allgemein positiv oder negativ bewertet werden. Die Beziehung zwischen Belastungen und Beanspruchungen wird zudem als von den individuellen Ressourcen beeinflusst angesehen (41). Auch darin unterscheidet sich das Belastungs-Beanspruchungs-Modell von anderen Stressmodellen, die Stressoren als Faktoren verstehen, die sowohl in der Umwelt als auch in der Person auftreten können. Stress wird in den meisten theoretischen Ansätzen als eine negative subjektive Wahrnehmung der individuellen Konstitution verstanden (42).

Für die Beurteilung des Arbeitsplatzes eignet sich das Belastungs-Beanspruchungs-Modell als Grundmodell deshalb besonders gut, weil es erlaubt, Faktoren, die auf die objektiven Arbeitsbedingungen zurückzuführen sind, klar von individuellen Faktoren zu unterscheiden und letztere ebenfalls zum Gegenstand der Situationsanalyse zu machen (46). Denn das Modell ermöglicht, das Zusammenspiel beider Aspekte (individuelle Ressourcen und Belastungen) hinsichtlich ihrer Wirkung auf die Entstehung von Beanspruchungen zu beschreiben. Abbildung 1 stellt die Wechselwirkung der drei Faktoren dar. Hier werden exemplarisch ein paar wenige Beispiele für arbeitsbedingte Belastungen wie häufige Arbeitsunterbrechungen sowie Beanspruchung, also die Reaktion von Arbeitnehmenden (z. B.

in Form von herabgesetzter Wachsamkeit bzw. Vigilanz) in Abhängigkeit der individuell verfügbaren Ressourcen dargestellt. Wie eine Belastung auf ein Individuum in Form einer Beanspruchung wirkt, hängt demnach stark von den individuellen Ressourcen ab. Diese beeinflussen die Wirkung externer Faktoren auf die Gesundheit und das Wohlbefinden einer Arbeitsperson (40).

Abbildung 1: Belastungs-Beanspruchungs-Modell (Eigene Darstellung)



Somit kann eine Belastungsquelle bei Arbeitsperson A zu einer negativen Beanspruchungsreaktion führen, während Arbeitsperson B positiv auf diese Belastung reagiert. So könnte beispielsweise eine hochkomplexe Arbeitsaufgabe, welche hohe Anforderungen an die kognitive Leistungsfähigkeit einer Person stellt, für Mitarbeitende mit entsprechend hohen Ressourcen wie zum Beispiel einer hohen Qualifikation und einem unterstützenden sozialem Umfeld, stimulierend wirken und die Arbeitsmotivation steigern, weil sich diese Person angemessen gefordert fühlt. Eine zweite Person könnte auf die gleiche Aufgabe mit Überforderung reagieren, weil ihr beispielsweise die entsprechende Qualifikation fehlt oder sie krankheitsbedingte Einschränkungen der Konzentrationsfähigkeit aufweist. Diese Sichtweise auf die Beeinflussbarkeit der durch die Arbeitssituation hervorgerufenen Beanspruchungen betont, dass sowohl die Arbeitsbedingungen als auch die arbeitende Person selbst veränderliche Variablen darstellen (46). Genauer genommen, beeinflussen bzw. können die Arbeitenden selbst die Wirkung der Arbeitssituation aktiv beeinflussen (47). Zwar haben sie keine oder wenig Einflussmöglichkeiten, was die Rahmenbedingungen des Arbeitens, also die Ebene der Verhältnisprävention, betrifft. Jedoch können sie ihre individuellen Ressourcen und somit ihre Beanspruchungsreaktion mit beeinflussen. Damit soll

nicht gesagt sein, dass der Arbeitgeber aus seiner Verantwortung zu entlassen sei oder dass Verhaltensprävention Verhältnisprävention ersetzen könne. Die aktiv gestalterische Rolle des Individuums ist deshalb so relevant für die Gefährdungsbeurteilung psychischer Belastung, als dass sie verdeutlicht, warum das subjektive Empfinden von zentraler Bedeutung für die Beurteilung der Situation ist und warum entsprechend Mitarbeitendenbefragungen oder andere Formen der direkten Beteiligung der Betroffenen an der Gefährdungsbeurteilung notwendig sind. Denn die individuelle Auswirkung von Arbeitsbelastungen ist immer subjektiv beeinflusst. Die gleiche Belastung führt in Abhängigkeit von individuellen Eigenschaften und Fähigkeiten zu unterschiedlichen Beanspruchungen und muss daher immer aus der Perspektive der Betroffenen analysiert werden (45)(45). Das schließt eine objektive Beurteilung der Arbeitsbedingungen nicht aus. Sie allein genügt jedoch nicht, um ein vollständiges Bild etwaiger psychischer Gefährdungen zu zeichnen. Wenn Kriterien für gute Arbeitsbedingungen definiert werden sollen, ist es notwendig, die Belastungen, die durch die Arbeitsbedingungen entstehen, beschreiben und erklären zu können. Um Maßnahmen zur Stärkung der individuellen Ressourcen oder zur Verbesserung des individuellen Gesundheitsverhaltens ableiten zu können, ist es außerdem notwendig, Belastungen zu identifizieren, die auf individuelle Merkmale zurückzuführen sind. Die Analyse der Arbeitsbedingungen und der individuellen Ressourcen ist daher der erste Schritt zu einem bedarfsorientierten Gesundheitsmanagement.

Letztlich ist die Berücksichtigung der individuellen Reaktion auf Belastungen auch deshalb relevant, weil sie eben interindividuell unterschiedlich ausfällt und häufig eine pauschale Aussage darüber, wie bestimmte Arbeitsmerkmale gestaltet sein sollten, gar nicht möglich ist. Diese müssen an das jeweilige Leistungsvermögen der Beschäftigten angepasst werden, denn es gilt die „Faustregel: Überforderung schädigt, Unterforderung schwächt und menschengerechte Anforderung stärkt die Gesundheit und das Leistungsvermögen“ (45).

Insofern stimmen Antonovskys Modell der Salutogenese und Rohmerts Belastungs-Beanspruchungs-Modell überein. Wann man gesund ist oder sich gesund fühlt, was krank macht und das Wohlbefinden einschränkt, hängt von einer Vielzahl von Faktoren und deren wechselseitiger Beeinflussung auf Ebene des Individuums und in der Umwelt ab. Für die Erfassung psychischer Belastungen am Arbeitsplatz bedeutet dies, dass monokausale Erklärungsansätze niemals hinreichend die Entstehung bestimmter Beanspruchungsreaktionen erklären können. Die verschiedenen Faktoren, die das subjektive Gesundheitsempfinden beeinflussen, müssen gleichzeitig und möglichst vollständig erfasst werden.

Zusammenfassend seien nochmals folgende Begriffe kurz definiert:

- **Gesundheit** wird im Rahmen dieser Arbeit verstanden als ein „Zustand des vollständigen körperlichen, geistigen und sozialen Wohlbefindens“ (48). Gemäß der Ottawa-Charta der WHO zur Gesundheitsförderung werden dabei nicht nur die körperliche Verfasstheit und die bloße Abwesenheit von Krankheit berücksichtigt, sondern auch der wesentliche Einfluss sozialer Faktoren und individueller Ressourcen betont (49).
- **Belastung** umfasst die Gesamtheit aller Umwelteinflüsse, welche auf den Menschen einwirken. Der Begriff ist wertneutral zu verstehen (13).
- **Beanspruchungen** sind „individuellen Folgen bzw. Auswirkungen“ (40) einer Belastung, beeinflusst durch die individuell verfügbaren Ressourcen.
- **Beanspruchungsreaktionen oder Beanspruchungsfolgen** sind mittel- bis langfristige individuelle physische und / oder psychische Reaktionen auf Belastungen (41, 13).
- Individuelle **Ressourcen** wirken als Puffer zwischen Belastungen und Beanspruchungen bzw. moderieren den Effekt (die Beanspruchung), den eine Belastung im Individuum auslöst (13).

2.2 Gesetzliche Pflicht zur Gefährdungsbeurteilung psychischer Belastung und Herausforderungen bei der Durchführung

Der Gesetzgeber verpflichtet den Arbeitgeber gemäß § 5 Arbeitsschutzgesetz zur „Beurteilung der für die Beschäftigten mit ihrer Arbeit verbundenen Gefährdung“ (2). In diesem Paragraphen werden neben der Arbeitsplatzgestaltung, den Arbeitsmitteln, der Qualifikation, den Arbeitsabläufen und der Arbeitszeit sowie physikalischen, chemischen und biologischen Einwirkungen auch die psychischen Belastungen als gefährdende Aspekte, die zu berücksichtigen sind, aufgeführt. Dieses umfassende Spektrum der im Rahmen der Gefährdungsbeurteilung zu berücksichtigenden Faktoren ist auf den ganzheitlichen Gesundheitsbegriff gemäß WHO-Definition wie er in Abschnitt 2.1.1 erläutert wurde, zurückzuführen (13). Die Gefährdungsbeurteilung psychischer Belastung ist somit ein Teil der gesetzlich vorgeschriebenen Gefährdungsbeurteilung. Eine rein psychische Gefährdungsbeurteilung als solche existiert nicht in den rechtlichen Vorgaben. Umgekehrt gilt jedoch, dass psychische Belastungen immer ein Gegenstand der Gefährdungsbeurteilung sein sollten.

Da unter dem Begriff psychische Belastung eine Vielzahl von Einwirkungen subsumiert wird, hat die Gemeinsame Deutsche Arbeitsschutzstrategie (GDA) eine Leitlinie „Beratung und

Überwachung bei psychischer Belastung am Arbeitsplatz“ (50) veröffentlicht. In der GDA sind der Bund als gesetzgebender Akteur und Bindeglied zum europäischen und internationalen Arbeitsschutz und zur Arbeitsschutzforschung, die Länder in Form der jeweiligen 16 staatlichen Arbeitsschutzbehörden und die Unfallversicherungsträger (Unfallkassen und Berufsgenossenschaften) vertreten (51).

Die Leitlinie soll gewährleisten, dass die Akteure im Arbeitsschutz auf einem einheitlichen Grundverständnis basierend agieren. Im Einleitungstext der Leitlinie wird die zunehmende Bedeutung psychischer Belastungen vor dem Hintergrund des Wandels der Arbeitswelt betont. Trotz des nationalen Konsenses aller Akteure im Arbeits- und Gesundheitsschutz über die Bedeutsamkeit einer Gefährdungsbeurteilung psychischer Belastung, zeigen Studien dass eben diese nicht vollständig und flächendeckend umgesetzt wird (52, 13). Häufige Gründe, welche seitens der Arbeitgeber gegen die Durchführung einer Gefährdungsbeurteilung psychischer Belastung angeführt werden, sind unter anderem Unwissenheit oder Unsicherheit bezüglich der Thematik, mangelndes Problembewusstsein auf der einen Seite und Scheu vor dem Thema aufgrund seiner Brisanz auf der anderen Seite (53). Des Weiteren sind auch fehlende Handlungsmöglichkeiten bezüglich der Umsetzung ein Hemmnis. So gab mehr als die Hälfte der Führungskräfte (54%) in einer Befragung an, dass dafür im Betrieb weder Zeit noch Personal zur Verfügung stünde. Auch seitens der Betriebsräte wird ein hohes Maß an Orientierungslosigkeit bezüglich der Durchführung berichtet. Demnach gaben über zwei Drittel an, dass im Betrieb nahezu kein Wissen darüber bestünde, wie eine Gefährdungsbeurteilung psychischer Belastungen durchzuführen sei und 89% sagten weiterhin aus, dass das Thema insgesamt „schwierig zu handhaben sei“ (53). Es ist demnach davon auszugehen, dass die Durchführung einer Gefährdungsbeurteilung psychischer Belastung zumindest nicht ausschließlich eine Haltungs-, sondern auch eine Ressourcenfrage ist.

In der Diskussion um Faktoren, die die Durchführung einer Gefährdungsbeurteilung psychischer Belastung begünstigen oder hemmen, entsteht jedoch schnell der Eindruck, dass diese ein losgelöster Prozess von der Gefährdungsbeurteilung der Arbeitssituation ist (Vgl. dazu u. a. (40, 52, 53, 45, 54–56). Mit Blick auf das Arbeitsschutzgesetz muss jedoch eindeutig festgehalten werden, dass der Aspekt Psyche explizit eine der potenziellen arbeitsbedingten Gefährdungen ist, den es im Rahmen der Gefährdungsbeurteilung zu berücksichtigen gilt. Dass der Eindruck entsteht, es handele sich beim Thema psychische Gefährdungen um einen eigenen, abgegrenzten Themenbereich kann zwei Gründe haben. Einerseits ist die Thematik im Vergleich zu anderen Gefährdungskategorien wie z. B. chemischen oder biologischen Gefährdungen noch relativ neu bzw. wurde erst spät als ein relevanter Aspekt anerkannt und muss sich, wie die Studienlage zeigt, noch immer in vielen Bereichen erst etablieren und Anerkennung als Form der arbeitsbedingten Gefährdung finden. Andererseits ist das Thema Psyche vielschichtiger und weniger gut ab- und eingrenzbar als es beispielsweise die

Handhabung schwerer Lasten ist. Entsprechend der internationalen Norm DIN EN ISO 10075 müssen im Rahmen der Beurteilung psychischer Arbeitsbelastungen gemäß dem Belastungs-Beanspruchungs-Modell sowohl Belastungen, als auch Beanspruchungen und Beanspruchungsfolgen berücksichtigt werden (47). Denn der Unterschied zu physischen Gefährdungen ist, dass die gleiche Belastung ausschließlich in Abhängigkeit der individuellen Ressourcen und subjektiven Wahrnehmung positiv oder negativ wirkt. Eine bestimmte elektromagnetische Strahlung in einer bestimmten Frequenz und Wellenlänge ist eindeutig gesundheitsgefährdend oder ist es eindeutig nicht. Zwar gibt es auch im Bereich der physischen Gefährdungen bestimmte Belastungen, die in ihrer Wirkung von den individuellen Ressourcen abhängig sind. Beispielsweise können trainierte Personen schwere Lasten mit weniger Mühe bewegen als Untrainierte. Aber auch hier sind die Grenzen dessen, was als gesundheitsschädlich definiert ist, schnell erreicht und die sich daraus ableitenden Maßnahmen sind eindeutig. Hingegen wirken psychische Belastungen interindividuell sehr unterschiedlich. Eine bestimmte Arbeitsintensität kann von einer Person als motivierend und anregend, von einer anderen als überfordernd und erdrückend empfunden werden. Dies ist einerseits eine Frage der individuellen Ressourcen und andererseits eine Frage der subjektiven Wahrnehmungen. Beide Einflussgrößen stehen wiederum in Wechselwirkung miteinander. Dabei erfolgt die Bewertung einer Situation bzw. einer Belastung in der Regel unbewusst. Die individuelle Wahrnehmung, ob eine Situation positiv oder negativ zu bewerten ist, aktiviert unterschiedliche Bereiche im Gehirn und löst unterschiedliche physiologische Reaktionen aus (40). So hängt es beispielsweise maßgeblich von der individuellen Wahrnehmung ab, ob eine große Anzahl kognitiv anspruchsvoller Arbeitsaufgaben, die einer Arbeitsperson übertragen werden, als interessant, herausfordernd, stimulierend etc. bewertet werden oder ob diese Person die Situation als belastend, überfordernd, stressig etc. empfindet. Wie in Abschnitt 2.1.2 erläutert, wird diese Bewertung stark von den individuellen Ressourcen abhängen. Da diese Bewertung jedoch zumeist unbewusst erfolgt, ist sie kaum steuerbar und kann zumindest nicht unmittelbar beeinflusst werden (40).

Der subjektiven Bewertung kommt somit ein hoher Stellenwert bei der Gefährdungsbeurteilung psychischer Belastung zu. Sie kann, wenngleich nicht als einziger, so doch als wichtigster Indikator psychischer Belastung angesehen werden (47). Aus diesem Grund gibt die GDA vor, dass die betroffenen Arbeitspersonen an der Gefährdungsbeurteilung zu beteiligen sind (50). Anders als bei anderen Gefährdungskategorien können keine allgemeinen Belastungsgrenzen formuliert werden. Berücksichtigt werden muss immer die individuelle Verarbeitung einer Belastung zu einer Beanspruchung. Diese Komplexität erklärt, warum viele Betriebe angeben, sie seien mit der Gefährdungsbeurteilung psychischer Belastung überfordert. Der Bedarf nach einem einfach anzuwendenden, niedrighschwelligen und möglichst ressourcensparsamen Instrument ist offenkundig. Oder wie es im Projektbericht der BAuA Umsetzung der

Gefährdungsbeurteilung psychischer Belastung heißt: „Es zeigte sich, dass psychosoziale Risiken komplexe Beurteilungs- und Gestaltungsprobleme sind, die mit den im betrieblichen Arbeitsschutz vertrauten Verfahrensweisen nicht ohne weiteres lösbar sind“ (57).

Die GDA-Leitlinie „Beratung und Überwachung bei psychischer Belastung am Arbeitsplatz“ formuliert Rahmenkriterien für das methodische Vorgehen und den Inhalt der Gefährdungsbeurteilung psychischer Belastung. Sie richtet sich jedoch nicht an Unternehmen, sondern an die Arbeitsschutzbehörden und Unfallversicherungsträger (50). Unternehmen finden Orientierung und Hilfestellung in der ebenfalls von der GDA veröffentlichten „Empfehlung zur Umsetzung der Gefährdungsbeurteilung psychischer Belastung“ (58).

Für die Erstellung von Instrumenten, welche zur Gefährdungsbeurteilung psychischer Belastung eingesetzt werden, hat die GDA Qualitätsgrundsätze formuliert. Hierzu zählen die Eignung für die geplanten Einsatzbereich, die Beschreibung von Anwendungsvoraussetzungen, die methodische Qualität, die Erfassung aller Tätigkeiten und Belastungsfaktoren und die Partizipation der Beschäftigten. Zudem soll das Verfahren eine Beurteilung erlauben, ob Maßnahmen zur Reduktion von Gefährdungen notwendig sind. Welche Methode und ob eine oder mehrere in einem mehrstufigen Verfahren angewandt werden sollen, gibt die Leitlinie nicht vor, da dies letztlich vom Anwendungsfall abhängt. Grobscreenings werden als gut geeignet eingestuft, um eine flächendeckende Orientierung zu gewinnen und Schwerpunkte zu identifizieren, die einer differenzierteren Analyse bedürfen (50).

Folgt man der Definition psychischer Arbeitsbelastung nach der DIN EN ISO 10075 ist diese ein komplexes, mehrdimensionales Konstrukt, in dem verschiedene Aspekte und Faktoren miteinander interagieren. Schon allein vor dem Hintergrund dieser Komplexität empfiehlt sich die Kombination mehrere Messinstrumente (47). Bedenkt man zudem die Gründe, welche von den qua Gesetz zuständigen Akteuren angeführt werden, warum diese bisher keine Gefährdungsbeurteilung psychischer Belastung durchgeführt haben, spricht dies zusätzlich für ein möglichst einfach umzusetzendes und ressourceneffizientes Grobscreening als ersten Schritt eines mehrstufigen Vorgehens. So sieht es auch die GDA, welche schriftliche standardisierte Mitarbeitendenbefragungen als effektive Methode beschreibt, die den oben genannten Qualitätskriterien gerecht wird:

„Standardisierte schriftliche Mitarbeiterbefragungen sind besonders gut geeignet, um sich einen Überblick darüber zu verschaffen, ob und welche Belastungsschwerpunkte es gibt. Sie ermöglichen es zudem, alle Beschäftigten in die Gefährdungsbeurteilung einzubeziehen.“ (58)

Demnach kann das im AIGScreen gewählte methodische Vorgehen nicht nur als praxistauglich und niedrighschwellig für die Betriebe im Rahmen der Bedarfsplanung und Evaluation von BGM-Maßnahmen erachtet werden. Sondern es ist insbesondere auch zur Erfüllung der gesetzlichen Pflicht eines jeden Arbeitgebers zur Durchführung und Dokumentation der Gefährdungsbeurteilung psychischer Belastung geeignet.

Neben methodischen Aspekten stellt sich zudem die Frage nach dem Beobachtungsgegenstand. Auch hier gibt die GDA-Leitlinie (50) Aufschluss. Sie formuliert vier Merkmalsbereiche, denen differenziertere Faktoren zugeordnet sind. In Tabelle 1 sind die Merkmalsbereiche und die zugehörigen Faktoren dargestellt.

Tabelle 1: Merkmalsbereiche und Inhalte der Gefährdungsbeurteilung gemäß der „Leitlinie Beratung und Überwachung bei psychischer Belastung am Arbeitsplatz“ (50) der GDA

Merkmalsbereich	Im Rahmen der Gefährdungsbeurteilung zu berücksichtigende Faktoren
Arbeitsinhalt/Arbeitsaufgabe	• Vollständigkeit der Aufgabe • Handlungsspielraum • Variabilität / Abwechslungsreichtum • Information(sangebot) • Verantwortlichkeiten • Qualifikation • Emotionale Inanspruchnahme
Arbeitsorganisation	• Arbeitszeit, Schichtarbeit und Pausenregime • Arbeitsablauf • Kommunikations- und Kooperationsstrukturen
Soziale Beziehung	• Zu Kolleginnen und Kollegen • Zu Vorgesetzten
Arbeitsumgebung	• Physikalische und chemische Faktoren • Physische Faktoren • Arbeitsplatz- und Informationsgestaltung • Arbeitsmittel
Neue Arbeitsformen	In Abhängigkeit vom betrieblichen Setting betrifft das bspw. • Räumliche Mobilität • Entgrenzung von Arbeit und Privatleben • Zeitliche Flexibilisierung

2.3 Wissenschaftliche Gütekriterien

Auf die Nebengütekriterien wurde bereits im Kapitel 1 eingegangen. Diese sind während etwaiger Modifikationen des Fragebogens stets zu berücksichtigen und werden teilweise im Rahmen der Überprüfung der Hauptgütekriterien implizit miterfasst. Sie stellen jedoch keinen Gegenstand der empirischen Analysen dar, weshalb im Folgenden zunächst auf die Hauptgütekriterien vertieft eingegangen wird. Eine klare Begriffsbestimmung ist hier der erste wichtige Schritt der Untersuchung, da somit auch der Analysegegenstand definiert wird. Dabei beziehen sich die wissenschaftlichen Hauptgütekriterien Objektivität, Reliabilität und Validität auf den Messvorgang im Forschungsprozess mit quantitativen Daten. Da die angewandten Methoden in der quantitativen Forschung selbst mathematisch logisch ableitbar und begründbar und somit auch frei von Zweifel und Interpretation sind, muss insbesondere das Zustandekommen der verwendeten Datenbasis auf seine Richtigkeit überprüft werden. Denn anders als die mathematische Überprüfung von Hypothesen sind das Erheben bzw. Messen von Daten und die Operationalisierung nicht frei von Störgrößen, welche die Datenbasis beeinflussen, schlimmstenfalls sogar verzerren können. Sind die Gütekriterien nicht erfüllt, sind die aus den Daten gezogenen Schlüsse nicht verlässlich: „Die strategische Bedeutung [...] besteht also darin, dass sie im wahrsten Sinne des Wortes das Fundament legen für die Glaubwürdigkeit wissenschaftlicher Forschung“ (59).

2.3.1 Objektivität

Objektivität beschreibt die Unabhängigkeit eines Erhebungsinstruments von externen Einflüssen wie Ort, Zeit oder Person. Objektivität ist dann gegeben, wenn sich das Messergebnis allein aus der Merkmalausprägung der Merkmalsträger ergibt und nicht während der Erhebung, der Interpretation oder der Auswertung vom Testleitenden, Testauswertenden oder der Person, die die Testergebnisse interpretiert und anwendet, abhängig ist. Entsprechend unterscheidet man auch zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität (29). Für die Beurteilung der Objektivität eines Instruments bestehen keine Messverfahren¹. Sie kann vielmehr anhand des Kriteriums der Standardisierung festgestellt werden. So kann die Durchführungsobjektivität z. B. durch genaue Instruktionen der Testleitenden, durch Kontrolle der Durchführungsbedingungen oder eine computerbasierte Testdurchführung, die Verzerrungseffekte durch zwischenmenschliche Beeinflussungen wie z. B. Suggestivfragen oder soziale Erwünschtheit ausschließt, gewährleistet werden. Eine Standardisierung der Erhebung z. B. durch geschlossene

¹ Die Objektivität eines Erhebungsinstruments selbst kann nicht bemessen werden. Quantitativ überprüfbar sind jedoch die erhobenen Daten und Ergebnisse hinsichtlich ihrer Objektivität. So kann bspw. die Auswertungsobjektivität anhand Kendalls Konkordanzkoeffizienten bemessen werden (29).

Antwortformate erhöht die Auswertungsobjektivität. Umfassende und eindeutige Regeln für die Testinterpretation, welche dazu führen, dass verschiedene Personen bei der Testinterpretation zu denselben Schlüssen kommen, ermöglichen zudem eine Interpretationsobjektivität (30, 28). Ist eine Standardisierung nicht möglich, müssen alle potenziellen Störvariablen kontrolliert und konstant gehalten werden. Dies kann beispielsweise durch genaue Testinstruktion der Testleiter, die alle Schritte der Erhebung detailliert festlegt, durch Auswertungsregeln, deren Anwendung empirisch überprüft werden kann (z. B. mittels Konkordanzkoeffizienten) oder durch unterstützende Anweisungen und Hilfsmittel während der Ergebnisinterpretation wie z. B. ein Auswertungsmanual und Normentabellen erfolgen (29).

Einige Autoren gehen soweit, Objektivität ausschließlich als Gütekriterium qualitativer Erhebungen zu nennen (60). Demnach ist Objektivität bei hinreichender Standardisierung der erhobenen Merkmale implizit gegeben, da keine Abweichungen bei der Operationalisierung möglich sind (61). In anderen Veröffentlichungen findet sich zudem ein Verständnis von Objektivität, welches auch Ansprüche an die kritische Selbstreflexion der Forschenden stellt. Demnach muss ihm Rahmen eines Forschungsvorhabens immer wieder das eigene Vorgehen, die Fragestellung und die entsprechende Methodenwahl kritisch hinterfragt werden. Insbesondere dort, wo Objektivität nicht oder nicht vollständig durch eine Standardisierung der Datenerhebung, -auswertung und -interpretation ermöglicht werden kann, müssen Forschende größtmögliche Transparenz bezüglich ihres Vorgehens schaffen und Entscheidungen im Forschungsprozess dokumentieren (62). Insofern geht es bei der Feststellung, ob das Kriterium der Objektivität erfüllt ist, vor allem um eine sachlogische Beurteilung der Datengewinnung und –auswertung hinsichtlich potenzieller Verzerrungseffekte. Objektivität kann dann als gegeben betrachtet werden, wenn die sich Variabilität der Daten ausschließlich aus der Unterschiedlichkeit der Probanden bezüglich der zu erhebenden Merkmale ergibt.

2.3.2 Reliabilität

Die Reliabilität bezieht sich auf die Messgenauigkeit eines Instruments. Sie kann als gegeben erachtet werden, wenn Ergebnisse unter gleichen Bedingungen zeitlich unabhängig replizierbar sind, eine nachweislich hohe interne Konsistenz besteht und hohe Korrelationen mit Instrumenten bestehen, die dasselbe Merkmal oder Konstrukt erheben sollen (63). Selbst bei Messverfahren, die das Kriterium der Objektivität erfüllen, muss davon ausgegangen werden, dass ein Teil der Varianz auf Messfehler zurückzuführen ist. Ist eine Messung reliabel, ist dieser Messfehler jedoch nicht systematisch und möglichst gering (61).

Zur Bestimmung der Reliabilität existieren verschiedene Methoden, deren Einsatz vom Messinstrument und –verfahren abhängig ist. So sind Konsistenzanalysen sinnvoll, wenn zur Messung eines latenten Merkmals eine Skala mit mehreren Items verwendet wird, da

Konsistenzanalysen überprüfen, ob alle Items dieser Skala ausreichend hoch miteinander korreliert sind. Für die Überprüfung der Wiederholbarkeit einer Messung kommen sogenannte Retests zum Einsatz, während Paralleltests oder Split-Half-Tests sich eignen, um die Vergleichbarkeit mehrerer Instrumente oder Items zu überprüfen (63).

Wesentliche Beurteilungskriterien der Reliabilität sind die Genauigkeit einer Messung und deren Replizierbarkeit. Genauigkeit meint, dass das Verfahren präzise und zuverlässig ist. Replizierbarkeit bedeutet, dass die Wiederholung einer Messung unter gleichen Bedingungen zu gleichen Ergebnissen kommt (34). Hier wird die erste Schwierigkeit bei der Bestimmung der Reliabilität deutlich. Denn viele Merkmale sind nicht dauerhaft konstant in ihrer Ausprägung. Insbesondere latente Merkmale, also Variablen, welche nicht manifest und somit nicht direkt messbar sind, können sich im Zeitverlauf ändern. Häufig ist eine Veränderung sogar erwünscht oder Teil des Forschungsvorhabens. Im Kontext dieser Arbeit wäre es zum Beispiel kein präferables Ergebnis, wenn die Beurteilung der Arbeitssituation und die subjektiv empfundenen Belastungen durch die Arbeit trotz umfangreicher Maßnahmen im Rahmen des Betrieblichen Gesundheitsmanagements gleichbleiben. Vielmehr würden Forschungshypothesen geradezu unterschiedliche Messergebnisse zu unterschiedlichen Messzeitpunkten postulieren, da man ja begründet erwarten kann, dass Maßnahmen der Gesundheitsförderung zu einer Verbesserung der individuell wahrgenommenen Arbeitsbelastung führen. Selbstverständlich wären auch andere Ergebnisse möglich. Ist das BGM nicht erfolgreich oder verändern sich andere Rahmenbedingungen, ist auch eine Verschlechterung der Belastungssituation denkbar. In jedem Fall würde man über einen längeren Zeitraum kaum erwarten, bei den gleichen Personen wiederholt zu exakt gleichen Messergebnissen zu kommen. Dieses Problem stellt sich mitunter auch bei manifesten Variablen wie z. B. dem BMI oder Zigarettenkonsum, welche ebenfalls nicht unbedingt zeitlich stabil sind bzw. auf deren Veränderung man sogar aktiv hinwirkt. Aus diesem Grund ist die Überprüfung der sogenannten Retest-Reliabilität im Rahmen von Interventionsstudien häufig nicht oder nur sehr eingeschränkt möglich. Ihre Überprüfung ist ausschließlich bei der Messung zeitstabiler Merkmale sinnvoll (63). Um dennoch die Replizierbarkeit der Messergebnisse zu überprüfen, können Paralleltests eingesetzt werden (29). Hier stehen Forschende jedoch häufig vor einem weiteren Problem. Zwar ist die der Methodik der Paralleltest zugrundeliegende Idee, dass Messergebnisse zweier inhaltlich sehr ähnlicher Tests hohe Korrelationen beim selben Messobjekt aufweisen sollten, naheliegend für die Überprüfung der Replizierbarkeit. Allerdings setzen Paralleltests voraus, dass überhaupt mehrere Instrumente existieren, die parallel eingesetzt werden können (60). Auch hier ist die Überprüfung für latente Merkmale wieder schwieriger als für manifeste. Um die Reliabilität der Messung des BMIs zu überprüfen, benötigte man schließlich nur eine zweite Personenwaage und ein zweites Maßband oder ähnlich adäquate Messinstrumente zur Erfassung von

Körpergewicht und Körpergröße. Derartige, inhaltlich gleiche Testinstrumente sind in der Regel leicht zu finden oder zumindest existent. Insbesondere bei latenten Merkmalen wie zum Beispiel psychologischen Aspekten, Einstellungen oder Persönlichkeitseigenschaften gibt es häufig kein zweites inhaltlich vergleichbares Messinstrument, welches zur Bestimmung der Reliabilität des ersten Messinstruments herangezogen werden könnte. Die Entwicklung eines zweiten Instruments wäre in der Regel sehr schwierig und zu ressourcenaufwändig, weshalb in einem solchen Fall auf Äquivalenzmessungen zurückgegriffen werden muss (34). Diesen liegt die Annahme zugrunde, dass verschiedene Indikatoren gleichwertig (äquivalent) für die Messung eines Merkmals sind, da sie Elemente der gleichen Dimension sind. Zur Überprüfung einer solchen Eindimensionalität eignen sich faktoranalytische Verfahren (64). So wurde auch im Rahmen dieser Arbeit vorgegangen. Eine genaue Beschreibung der angewandten Methodik findet sich in Kapitel 3.4 „Reliabilitäts- und Validitätsanalyse“. Eine Möglichkeit zur Beurteilung der Äquivalenz ist die Testhalbierungs- oder Split-Half-Methode, welche den Vorteil birgt, dass sie keines zusätzlichen Untersuchungsaufwands bedarf, da sie den Test zufällig in zwei Hälften teilt und die Korrelation der beiden Testhälften bemisst. Ein Nachteil ist, dass sie aufgrund der Halbierung der Itemzahl zu einer Unterschätzung der Reliabilität tendiert. Zwar existiert dafür eine Korrekturformel (Spearman-Brown-Prophecy-Formula), diese kann jedoch nur unter strengen Voraussetzungen wie der identischen Streuung der Testwerte angewandt werden (60). Diesem Problem kann wiederum durch den paarweisen Vergleich begegnet werden, was das Vorgehen bei mehr als zwei Items pro Skala jedoch schnell unübersichtlich werden lässt und auch die Urteilsfindung bei steigender Zahl zu beurteilender Korrelationen erschwert. Aus diesem Grund wird häufig der Koeffizient Cronbachs Alpha verwendet, welcher die mittlere Korrelation aller Items wiedergibt (34). Dieses Vorgehen eignet sich unter anderem auch besser zur Beurteilung der Items des AIGScreen, da jedes Item als eigenständiger Test behandelt wird (30). Dies ist insofern passend für die Reliabilitätsanalyse des AIGScreen, als dass die latenten Konstrukte, wie in Kapitel „1.2 Ziel der Arbeit“ erläutert, im Sinne der Testökonomie durch möglichst wenige Items repräsentiert werden sollen und sich deshalb nicht gegenseitig repräsentieren bzw. als Indikator für die jeweils anderen herangezogen werden können.

Zusammenfassend kann festgehalten werden, dass ein Messverfahren dann als reliabel gilt, wenn es mit möglichst kleinem Messfehler und ohne systematische Verzerrungen misst. Die Methode zur Beurteilung der Reliabilität hängt vom jeweiligen Instrument, der zugrunde liegenden Theorie, dem Untersuchungsziel des zu überprüfenden Instruments sowie den verfügbaren Instrumenten zur Reliabilitätsbestimmung ab (65, 63). Die im Rahmen dieser Arbeit angewandten Methoden, eine Konsistenzsprüfung anhand des Koeffizienten Cronbach's Alpha und eine Dimensionalitätsprüfung mittels faktoranalytischer Verfahren werden ausführlich im Kapitel 3.4 „Reliabilitäts- und Validitätsanalyse“ erläutert.

2.3.3 Validität

Der Begriff der Validität wird in Bezug auf wissenschaftliches Arbeiten unterschiedlich verwendet. Häufig nimmt er auf verschiedene Konzepte Bezug, welche teilweise nicht nur unterschiedliche Eigenschaften der wissenschaftlichen Güte beschreiben, sondern sich auch auf unterschiedliche Aspekte beziehen. So weisen Döring & Bortz (60) darauf hin, dass Validität im eigentlichen Sinn sich auf die Aussagen bzw. auf die Schlüsse wissenschaftlicher Arbeiten und die Interpretation der Daten bezieht. Jedoch wird Validität auch häufig als Überbegriff der methodischen Strenge und somit als Gütekriterium für die Methode selbst verwendet. Validität im eigentlichen Sinne meint also den „Grad der Gültigkeit wissenschaftlicher Aussagen, die aus einer quantitativ-empirischen Studie abgeleitet werden“ (60), wird jedoch auch in Bezug auf die Messinstrumente verwendet und bezeichnet somit gleichsam, ob ein Instrument auch das misst, was es messen soll bzw. was es vorgibt zu messen (62, 64).

Nicht selten findet auch eine Vermischung mit den Testgütekriterien Objektivität und insbesondere der Reliabilität statt, was auch durch die missverständliche Verwendung des Validitätsbegriffs in Bezug auf die Qualität des Messinstruments selbst begünstigt wird. Da unter den Begriff sowohl Gütekriterien bezüglich des Instruments als auch bezüglich der Qualität wissenschaftlicher Schlussfolgerungen subsumiert werden, verwundert es auch nicht, dass sich in der Literatur eine Vielzahl von Unterkategorien finden. Üblich ist die Unterscheidung zwischen Inhaltsvalidität, Konstruktvalidität und Kriteriumsvalidität. Häufig findet man auch die externe und die interne Validität, gelegentlich die prognostische Validität, die diskriminante und die konvergente Validität und einige Autoren führen zudem die Augenscheinvalidität an (64, 62, 66, 29, 28, 67, 34, 68, 63, 65, 60, 69–71).

Zur besseren Übersicht und Verständlichkeit werden all diese Konzepte an dieser Stelle kurz definiert:

- **Konvergierende oder konvergente Validität** und diskriminante Validität beschreiben das Verhältnis zu anderen Messinstrumenten. Konvergente Validität setzt voraus, dass das Messinstrument sehr hohe Korrelationen mit anderen Messinstrumenten aufweist, die dasselbe Konstrukt erfassen. Umgekehrt müssen sich für die **diskriminante oder auch divergente Validität** eines Messinstruments niedrige oder negative Korrelationen feststellen lassen, wenn sich theoretisch begründen lässt, warum zwischen den gemessenen Konstrukten kein oder ein negativer Zusammenhang besteht. Die Methoden zur Beurteilung diskriminanter und konvergenter Validität werden häufig zum Nachweis von Konstruktvalidität angewandt (62, 66).

- Somit beantwortet die **Konstruktvalidität** die Frage, ob ein Messinstrument das misst, was es messen soll und beschreibt damit den Aspekt von Validität, welcher sich auf das Messinstrument selbst bezieht (60). Sie setzt voraus, dass alle Aspekte eines theoretischen Konstrukts mit dem Messinstrument erfasst und dessen Dimensionalität adäquat abgebildet wird (65). Eine genauere Erläuterung findet sich in Abschnitt 2.3.3.2.
- Sowohl die interne als auch die externe Validität sind von der Qualität des Untersuchungsdesigns abhängig. Während die **interne Validität** die Belastbarkeit der Kausalschlüsse auf Basis der gewonnenen Daten beschreibt, bezieht sich die **externe Validität** auf die Verallgemeinerbarkeit der Ergebnisse auf andere räumliche, zeitliche und soziale Kontexte (60). Interne und externe Validität bemessen demnach die Validität im eigentlichen Sinne, nämlich die Gültigkeit und Verlässlichkeit der Interpretation der Daten.
- Die **Augenscheinvalidität** hat das Kriterium der Akzeptanz zum Gegenstand. Sie ist an keine weiteren inhaltlichen oder empirischen Bedingungen geknüpft, sondern beschreibt vielmehr eine Laienwahrnehmung des Tests (29). Obwohl sie daher aus wissenschaftlicher Sicht kein zufriedenstellendes Kriterium für die Güte eines Tests sein kann, ist sie nicht gänzlich zu vernachlässigen, da sie letztlich berücksichtigt, wie hoch die Akzeptanz bei der Zielgruppe der Messung ist.
- Um eine Aussage über die **Kriteriumsvalidität** treffen zu können, bedarf es eines sogenannten Außenkriteriums, mit dem die Messergebnisse hoch korrelieren (28). Unterschieden wird dabei zwischen konkurrenter und prädiktiver oder auch prognostischer Validität, welche sich nach dem Zeitpunkt des Eintretens des Außenkriteriums unterscheiden (65). **Konkurrente Validität** ist gegeben, wenn die Messung und das Außenkriterium mehr oder weniger zeitgleich auftreten. Eine **prognostische Validität** besteht, wenn zukünftig eintretende Kriterien anhand der Messergebnisse vorhergesagt werden können.
- **Inhaltsvalidität** schließlich kann nicht statistisch nachgewiesen werden, da diese die „Bestätigung der Angemessenheit eines diagnostischen Verfahrens durch Expertenurteile“ (69) beschreibt. Sie erfordert, dass das zu erfassende Merkmal vollständig erfasst wird und durch die verwendeten Items eine ausgewogene Repräsentation erfährt (28). Zwar bedarf die Feststellung der Inhaltsvalidität immer einer Bewertung durch Experten und kann nicht anhand objektiver Kriterien und Methoden gemessen werden, jedoch lässt sich zumindest die Intersubjektivität der Expertenurteile anhand des CVR-Index (68) quantifizieren. Eine ausführliche Erläuterung dazu erfolgt in Abschnitt 2.3.3.1.

Zunächst lässt sich feststellen, dass sich das Kriterium Validität sowohl auf die Datenerhebung (Messung) als auch auf deren Verarbeitung (Interpretation) bezieht. Einige Aspekte von Validität bedürfen gründlicher Vorüberlegungen und theoretischer Fundierung, weshalb Validität auch nicht vollständig durch Messtechniken und statistische Analysen nachgewiesen werden kann, sondern auch einer sachlogischen Überprüfung durch Experten bedarf. Aus diesem Grund ist ein sinnvoller erster Schritt zur Beurteilung der Validität eines Messverfahrens die Bewertung der Inhaltsvalidität durch fachkundige Experten.

2.3.3.1 Inhaltsvalidität

Wie für die Objektivität bestehen auch für die Inhaltsvalidität keine direkten (statistischen) Messverfahren. Lediglich für einzelne Aspekte der Inhaltsvalidität existieren quantitative Methoden wie beispielsweise der von Lawshe (1975) vorgeschlagene CVR (content validity ratio) zur Beurteilung der Relevanz eines Items (68). Vielmehr verlangt die Bewertung der Inhaltsvalidität jedoch ein Expertenurteil über die Angemessenheit des Verfahrens, das heißt, ob ein Instrument das zu erfassende Merkmal vollständig misst bzw. ob die verwendeten Items das theoretische Konstrukt vollständig repräsentieren (28). Erste notwendige Bedingung dafür ist, dass eine Definition des zu erfassenden Merkmals vorliegt, wobei zwischen einer theoretischen und einer operationalen Herleitung unterschieden wird. Die Inhaltsvalidität theoretischer Konstrukte erfordert eine argumentative Begründung der systematischen Itemselektion, welche anhand einer Theorie über den Zusammenhang zwischen Antwortverhalten auf Itemebene und der Merkmalsausprägung im zu erfassenden Konstrukt erfolgte. Die theoretisch postulierten Zusammenhänge sollten sich auch später in strukturprüfenden Verfahren abbilden lassen. Operationale Definitionen beschreiben das zu erfassende Merkmal als Summe einer Gesamtheit von verschiedenen Aspekten derselben Dimension bzw. als eine Liste von Items. Um beurteilen zu können, inwiefern ein operational definiertes Merkmal durch die verwendeten Items ausreichend repräsentiert ist, muss eine Beschreibung der Gesamtheit aller möglichen Items, das sog. Itemuniversum, vorliegen (69). Sofern keine theoretische Fundierung bezüglich der Zusammenhänge zwischen den Items besteht, sondern deren Zusammenstellung einer Indexbildung entspricht, welche letztlich das zu messende Merkmal definiert und nicht umgekehrt, liegt ein formatives Messmodell vor, welches im Rahmen der Konstruktvalidierung zu berücksichtigen ist (72).

Für den AIGScreen findet sich die theoretische Herleitung der Itemskalen und eine erste Eingrenzung des Itemuniversums einerseits unmittelbar in der Vorgaben der Gemeinsamen deutschen Arbeitsschutzstrategie (58, 50). Andererseits lässt sich die Auswahl der zu erfassenden Konstrukte sehr gut mit dem Belastungs-Beanspruchungs-Modell begründen (Vgl. Abschnitt 2.1.2 „Das Belastungs-Beanspruchungs-Modell“).

Im Rahmen der Inhaltsvalidität ist allerdings nicht nur die Auswahl der Aspekte, sondern auch die Angemessenheit von deren Erhebung bzw. Messung zu beurteilen. Auch hier ist das Wissen von Experten gefragt, welche nicht nur fachliche Kenntnisse, sondern auch ein Verständnis der Zielgruppe aufweisen sollten. Damit ein Fragebogen bzw. eine Online-Befragung den wissenschaftlichen Gütekriterien gerecht werden kann, müssen neben der inhaltlichen Vollständigkeit weitere Kriterien erfüllt sein, welche sich auf die Beantwortbarkeit und Möglichkeiten des Antwortverhaltens seitens der Probanden beziehen. So sollte im Rahmen von Expertenreviews beispielsweise auch berücksichtigt werden, ob die Fragen von der Zielgruppe beantwortet werden können oder ob soziale Erwünschtheit das Antwortverhalten beeinflussen könnte (73). Wenn z. B. Mitarbeitende befragt werden, ob sie mit der eingesetzten Software am Arbeitsplatz zufrieden sind, dann setzt das voraus, dass diese auch am Computer bzw. mit bestimmten Softwareprogrammen arbeiten. Wenn Mitarbeitenden andererseits das Gefühl haben, ihre Antworten sind nicht anonym und sie Sanktionen bei einem zu kritischen Antwortverhalten bezüglich ihrer Arbeitssituation fürchten, dann wird das Antwortverhalten durch soziale Erwünschtheit sehr wahrscheinlich positiv verzerrt sein.

Erfüllt ein Messinstrument nach subjektiver Beurteilung von Experten das Kriterium der Inhaltsvalidität, ist die Grundlage für die statistische Überprüfung der theoretisch postulierten Konstrukte geschaffen. Denn die Inhaltsvalidität ist nicht eine Form von Validität oder gar die niedrigste Stufe. Sie ist vielmehr notwendige Bedingung für alle sich anschließenden quantitativen Beurteilungsformen. Rein statistische Kennzahlen wären ohne theoretische Fundierung und sachlogische Begründung der Zusammenhänge inhaltsleer und schlicht nicht zu interpretieren (74).

2.3.3.2 Konstruktvalidität und Kriteriumsvalidität

Anders als die Inhaltsvalidität können Konstrukt- und Kriteriumsvalidität statistisch erfasst und anhand bestimmter Kennzahlen beurteilt werden. In der Regel eignen sich dafür in einem ersten Schritt Korrelationsanalysen, denen sich komplexere und aufwändigere Verfahren anschließen können (63).

Die Kriteriumsvalidität befasst sich, wie in Abschnitt 2.3.3 erläutert, mit der Vorhersagekraft eines Instruments bezüglich eines Außenkriteriums. Dieses Außenkriterium muss in seiner Ausprägung unabhängig und unbeeinflusst vom Messinstrument und den damit erhobenen Merkmalen sein (75) und gleichzeitig soll ein Zusammenhang zwischen den Testwerten und dem ausgewählten Kriterium theoretisch oder sachlogisch gut begründbar sein (33). Entsprechend schwierig ist es, geeignete Kriterien zu identifizieren, welche nicht nur die genannten Anforderungen erfüllen, sondern auch noch verfügbar sind. Wenn man

beispielsweise ein Instrument zur Erfassung des individuellen Interesses an Kunst und Kultur bezüglich seiner Kriteriumsvalidität beurteilen möchte, läge es zunächst nahe, als Kriterien die Häufigkeit von Museen- oder Theaterbesuchen heranzuziehen. Allerdings kann dieser Zusammenhang erheblich von Drittvariablen beeinflusst sein. Menschen, die sich für Kunst und Kultur interessieren, aber nur über ein geringes Einkommen verfügen, können sich derartige Besuche nicht so häufig leisten. Menschen mit schweren gesundheitlichen Einschränkungen hingegen können sowohl über das Interesse als auch die finanziellen Mittel verfügen, aber dennoch aufgrund ihrer Verfassung nicht in der Lage zur Teilhabe an kulturellen Veranstaltungen sein. In beiden Fällen wäre das Außenkriterium „Museen- und Theaterbesuche“ zur Beurteilung des Interesses an Kunst und Kultur nur eingeschränkt geeignet. Werden solche Störgrößen im Rahmen der Kriteriumsvalidierung nicht ausreichend bedacht, kann ein Instrument schnell fälschlicherweise als nicht ausreichend valide hinsichtlich seiner Vorhersage- bzw. Übereinstimmungsvalidität bezüglich der interessierenden Außenkriterien eingeschätzt werden. Umgekehrt kann es zu einer Überschätzung der Kriteriumsvalidität kommen, wenn ein Außenkriterium zur Beurteilung herangezogen wird, welches in einer direkten oder indirekten Beziehung zu den erhobenen Merkmalen steht.

Noch schwieriger als die umfassende Berücksichtigung aller potenziellen Störvariablen ist die Identifikation eines geeigneten Außenkriteriums bei vielen latenten, also nicht direkt messbaren Merkmalen. In diesen Fällen kommt dem Expertenurteil eine (noch) höhere Bedeutung zu, da diese in Fällen, in denen kein geeignetes Außenkriterium herangezogen werden kann, dieses gewissermaßen ersetzen (76). Der Messgegenstand des AIGScreen stellt hierfür ein gutes Beispiel dar. Gemessen werden soll die psychische Belastung im Arbeitskontext. Wie im Kapitel 2.1 erläutert, handelt es sich dabei nicht um ein direkt messbares Merkmal, sondern vielmehr um eine individuelle Empfindung bzw. Verarbeitung externer Einflüsse, welche zudem bei gleichen Umweltfaktoren nicht für jede Person gleich, sondern in Abhängigkeit ihrer individuellen Ressourcen (man könnte hier auch sagen „Störvariablen“) ausgeprägt ist. Weder der komplexe Verarbeitungsprozess noch dessen Ergebnis in Form von Beanspruchungen ist direkt messbar. Hierzu könnten weitere Instrumente zur Erfassung psychischer Erkrankungen wie zum Beispiel der PHQ zur Messung von Depression (77) herangezogen werden. Allerdings handelt es sich dabei wiederum erneut um eine Messung latenter Eigenschaften und eine gewisse inhaltliche Überschneidung zwischen dem zu beurteilenden Instrument und dem Kriterium ist nicht auszuschließen. In solchen Fällen spricht man in der Fachliteratur auch gelegentlich von innerer Validierung. Hierbei werden höhere Anforderungen an die Korrelation zwischen Messinstrument und Kriterium gestellt, als es bei der Verwendung von Außenkriterien der Fall ist. Zudem ist eine notwendige Bedingung, dass das verwendete Instrument selbst bereits hinsichtlich seiner psychometrischen Güte überprüft und als hinreichend valide eingestuft wurde (33).

Da Kriteriumsvalidität als „Zusammenhang zwischen Testergebnis und konkreten Leistungen oder Verhaltensweisen außerhalb der Testsituation“ (71) definiert ist, ist es auch denkbar sogenannte Outcome-Variablen als Kriterien heranzuziehen. Insbesondere bei psychologischen Tests ist es ein probates Vorgehen, Kriterien zu verwenden, welche sich sowohl aus der Theorie als auch aus der einschlägigen Forschung ableiten lassen (76). Im Fall des AIGScreen sind denkbare Outcome-Variablen die Arbeitsunfähigkeits- bzw. Ausfalltage im Beruf, bestimmte medizinische Atteste, welche potenziell aus der arbeitsbedingten Belastungssituation heraus hervorgegangene Beanspruchungsreaktionen attestieren oder auch Merkmale des Arbeitsverhaltens und der Einstellung wie Engagement, Commitment oder Arbeitsmotivation.

Bevor jedoch derartige Prognosen gemacht oder Zusammenhänge mit bestimmten Outcomes postuliert werden können, muss garantiert werden können, dass die Konstruktvalidität gegeben ist, also dass das Instrument misst, was es vorgibt zu messen (71). Die Konstruktvalidität hat gegenüber der Inhalts- und der Kriteriumsvalidität einen wesentlichen Vorteil: Sie erlaubt eine „eindeutige, unzweifelhafte Aussage darüber [...], was ein Test tatsächlich misst“ (33), da sie nicht auf zum Teil subjektiv beeinflusste Expertenurteile oder mehr oder weniger gut geeignete Außenkriterien angewiesen ist, sondern einzig auf der statistischen Hypothesenprüfung basiert.

Die Konstruktvalidität kann auf zwei Wegen überprüft werden. Einerseits stehen Methoden der Dimensionalitäts- und Konsistenzüberprüfung zur Verfügung. Andererseits kann die Überschneidung mit inhaltsgleichen oder inhaltsähnlichen Instrumenten getestet werden (63). Bei letzterem Vorgehen ist es gut möglich, dass die Korrelation mit Kriteriumswerten getestet wird, weshalb die Konstruktvalidierung unter Umständen die Kriteriumsvalidierung mit abdecken kann (71). Im Rahmen der Kriteriumsvalidierung wird also sowohl überprüft wie ein Instrument misst als auch, was es misst.

Für das weitere Verständnis soll abschließend der Begriff des Konstrukts genauer erläutert werden. Mit Konstrukt ist das zu erfassende Merkmal gemeint (63). Diese Definition trifft jedoch auf jede Form von Erhebung zu und beschreibt auch manifeste, eindimensionale Merkmale wie zum Beispiel das Körpergewicht oder ob eine Person regelmäßig Zigaretten raucht. Bei einem Konstrukt handelt es sich jedoch, wie der Begriff schon sagt, um etwas Konstruiertes, also etwas Menschengemachtes: „We create constructs to explain consistencies in complex human behaviors. In much educational and psychological research, constructs are latent. In other words, they are unseen and must be inferred from observable behaviors—either naturally occurring behaviors or responses to standardized stimuli“ (76). Konstrukte beschreiben also immer ein nicht mit wissenschaftlichen Methoden direkt und objektiv messbares, ein nicht unmittelbar sinnlich wahrnehmbares, sondern ein sogenanntes latentes Merkmal dar. Meist

handelt es sich dabei um komplexere Sachverhalte, welche mehrere Dimensionen oder zumindest mehrere Aspekte beinhalten und erst durch Benennung als solches zu einem Gegenstand zusammengefasst werden.

Psychische Belastungen im Arbeitskontext stellen exakt so ein komplexes Konstrukt dar, welches nur anhand von mehreren standardisierten Indikatorvariablen erhoben werden kann. Bei derartigen Konstrukten unterscheidet man grundsätzlich zwischen formativen und reflexiven Messmodellen. Diese unterscheiden sich im Wesentlichen anhand der postulierten Wirkungsrichtung. Während bei formativen Modellen davon ausgegangen wird, dass sie das Konstrukt kausal beeinflussen – man könnte auch sagen „verursachen“ –, wird bei reflektiven Modellen umgekehrt davon ausgegangen, dass die erhobenen Indikatoren das Produkt eines bestimmten Konstruktes sind (78). Ein typisches Beispiel für ein formatives Messmodell ist der sozioökonomische Status. Dieser ergibt sich aus verschiedenen Indikatoren wie Einkommen, Bildung, Beruflicher Stellung und je nach gewähltem Indikatorsystem noch weiteren Ressourcen wie Prestige oder sozialen Netzwerken. Beispiele für ein reflektives Messmodell sind psychologische Konstrukte wie die Arbeitszufriedenheit (79). Diese würde sich unter anderem in der Arbeitsleistung, der empfundenen Verbundenheit mit dem Unternehmen, der Freude an der Arbeit, der Bereitschaft zu Überstunden, häufigen bzw. wenigen Fehltagen usw. niederschlagen. Beide Beispiele sind zur Veranschaulichung noch einmal grafisch in Abbildung 2 und Abbildung 3 dargestellt.

Abbildung 2: Formatives Messmodell – Bsp. Sozio-Ökonomischer Status (Eigene Darstellung)

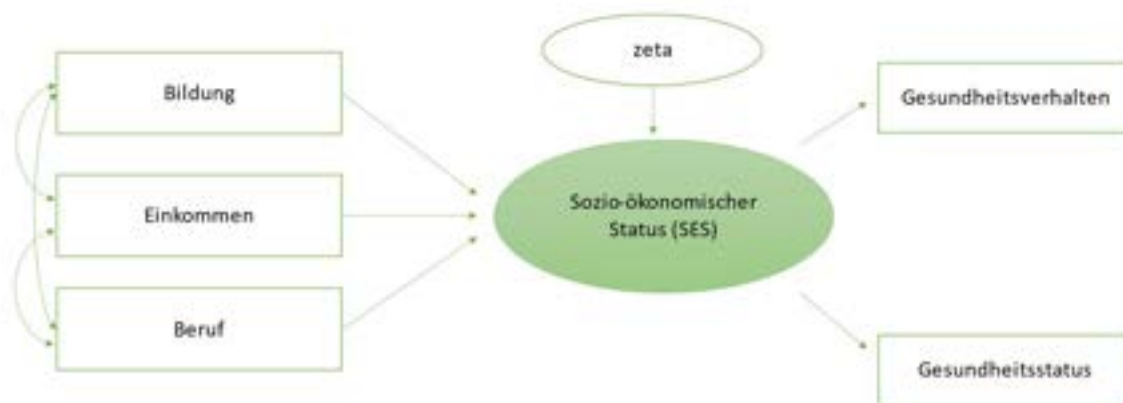
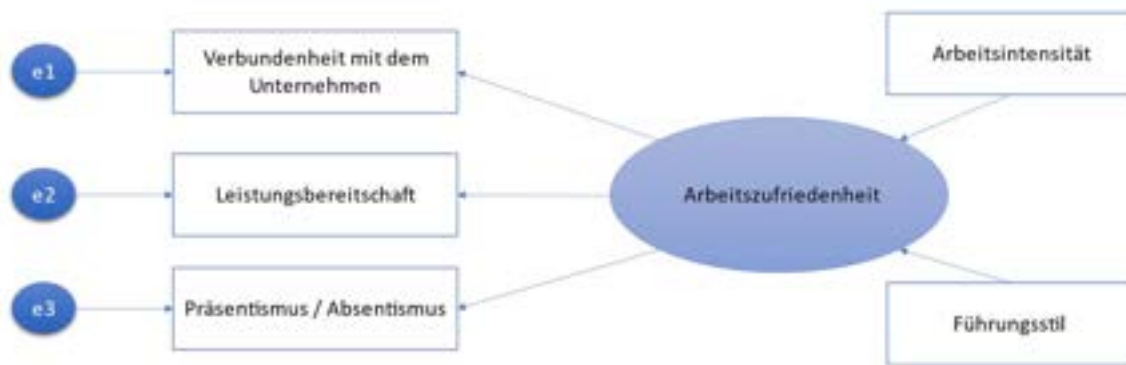


Abbildung 3: Reflektives Messmodell – Bsp. Arbeitszufriedenheit (Eigene Darstellung)



Anhand der Pfeilrichtungen wird ersichtlich, dass die Kausalrichtung zwischen beiden Messmodellen unterschiedlich ist. Bei dem formativen Messmodell geht die Kausalrichtung von den zu messenden Indikatoren zum Konstrukt. Bei den reflektiven Modellen verläuft die Kausalrichtung umgekehrt vom Konstrukt, das gemessen werden soll zu den Indikatoren. Das bedeutet jedoch nicht, dass reflektive Messmodelle nicht von anderen externen Faktoren beeinflusst werden. Im Falle der Arbeitszufriedenheit hängt diese beispielsweise von der Arbeitsmenge bzw. Arbeitsintensität und dem Führungsstil ab. Während bei formativen Messmodellen meist davon ausgegangen werden kann, dass die einzelnen Indikatoren auch untereinander korreliert sind bzw. sich gegenseitig beeinflussen, ist dies bei reflektiven Messmodellen nicht der Fall (80). Ein weiterer wichtiger Unterschied ist, dass die Indikatoren bei formativen Messmodellen nicht austauschbar sind, da sie das Konstrukt definieren. Für die reflektiven Modelle gilt hingegen, dass die Indikatoren austauschbar sind und dass sich durch Hinzunahme und Auslassen einer oder mehrerer Indikatoren die Bedeutung des Konstrukts nicht verändert, da dieses ja die Indikatoren verursacht und nicht umgekehrt (72). In Abbildung 2 und Abbildung 3 wird außerdem eine unterschiedliche Verortung der Messfehler (in den Abbildungen als „zeta“ bzw. „e“ dargestellt) deutlich. Während bei formativen Modellen wie dem sozio-ökonomischen Status die Messfehler nicht auf Indikatoren- bzw. Itemebene, sondern auf Konstruktebene zu verorten sind, werden diese bei reflektiven Modellen auf Ebene der Indikatoren bzw. Items verortet (80). Daraus ergibt sich ein wesentlicher Vorteil hinsichtlich der empirischen Implikationen für die reflektiven Messmodelle. Denn da die Konstrukte selbst nicht gemessen werden, sondern wie oben beschrieben immer nur über Indikatoren ermittelt werden können, ist es bei formativen Modellen weder möglich den Messfehler zu identifizieren noch ihn entsprechend zu beeinflussen. Hingegen können im Rahmen reflektiver Modelle die Items, die das Messergebnis verzerren mittels Faktorenanalyse identifiziert und ggf. aus der Messung des Konstrukts ausgeschlossen werden (72). Ein Ausschluss von Indikatoren aus der Messung wäre unproblematisch, da

reflektiven Messmodellen Eindimensionalität unterstellt wird. Das bedeutet, es wird davon ausgegangen, dass alle zur Messung des Konstrukts verwendeten Indikatoren dasselbe zugrundeliegende Konstrukt beschreiben und somit auch jeder Indikator das Konstrukt erfasst bzw. misst (78). Die Annahme der Ein- bzw. Mehrdimensionalität (bei formativen Modellen) wirkt sich auch auf die interne Konsistenz aus. Die Korrelationen der Items untereinander sollten bei reflektiven Modellen erwartungsgemäß hoch sein und mit steigender Faktorladung in Bezug auf das zu messende Konstrukt zunehmen (78). Bei der Messung sowohl reflektiver als auch formativer Modelle ist darauf zu achten, dass alle Items gleich gepolt sind, d. h. entweder sollen alle Items positiv oder negativ mit dem Konstrukt korrelieren. Allerdings kann die Stärke des Zusammenhangs bei formativen Modellen variieren, wohingegen die Items zur Messung reflektiver Modelle eine ähnlich starke Beziehung zum Konstrukt im Sinne der Austauschbarkeit haben sollen (72).

Gemeinsam ist beiden Messmodellen, dass sie etwas beschreiben, dass nicht unmittelbar messbar und somit sehr wahrscheinlich auch nicht in Gänze durch – egal wie viele und wie valide – Indikatoren erfasst und empirisch beschrieben werden kann. Weder einzelne Messwerte auf Itemebene noch Summenscores repräsentieren latente Konstrukte vollständig (80). Diesen Sachverhalt gilt es bei der Ergebnisinterpretation genauso zu berücksichtigen wie beim Verwenden von latenten Konstrukten als erklärende Variablen von weiteren Konstrukten.

2.3.4 Resümee Haupt- und Nebengütekriterien

Den Gegenstand der nachfolgend erläuterten Erhebungen und Analysen stellen die wissenschaftlichen (Haupt-) Gütekriterien Objektivität, Reliabilität und Validität dar. Objektivität erfordert die Freiheit von Verzerrungseffekten im Rahmen der Operationalisierung und kann bei ausreichender Standardisierung der Erhebung als weitgehend gesichert betrachtet werden (61). Reliabilität beschreibt die Genauigkeit und Replizierbarkeit der Messungen. Die Validität umfasst eine Vielzahl von Subkonzepten und Kriterien. Im Kontext dieser Arbeit liegt der Fokus auf der Fragestellung, ob das Instrument das misst, was es messen soll.

Wenngleich die Nebengütekriterien nicht Gegenstand der quantitativen und qualitativen Analysen im Rahmen dieser Arbeit sind, ist die Fragebogenentwicklung darüber hinaus nicht frei von Rahmenbedingungen, die die Erfüllung bestimmter Nebengütekriterien erforderlich machen. Im Falle des AIGScreen betrifft dieses insbesondere die Aspekte Ökonomie, Zumutbarkeit, Nützlichkeit und in Bezug auf das Ziel der Erhebung und den damit weiterverfolgten Nutzen, nämlich die strategische Implementierung und Steuerung von Maßnahmen des Betrieblichen Gesundheitsmanagements, auch das Nebengütekriterium Normierung, also die Möglichkeit des Vergleichs von Individual- oder Gruppenergebnissen mit

Normdaten. Die Anforderung, dass eine Erhebung in einem angemessenen Kosten-Nutzen-Verhältnis steht, findet sich wahrscheinlich bei nahezu jeder Befragung. Da im Fall des AIGScreen eine Befragung im Rahmen der Arbeitszeit geplant ist, ist ein besonders hohes Interesse an einer möglichst ressourceneffizienten Datenerhebung nur nachvollziehbar. Da eine hohe Beteiligung für aussagekräftige Ergebnisse notwendig ist, ist auch das Kriterium der Zumutbarkeit – wie in den meisten anderen Befragungen auch – eines, dem der Fragebogen unbedingt gerecht werden sollte. Die Nützlichkeit ergibt sich aus dem Ziel der Befragung selbst.

Durch die Überprüfung der psychometrischen Güte des Instruments werden Teilaspekte dieser Kriterien implizit miterfasst und überprüft. So schreiben Moosbrugger & Kelava (2012), dass das Kriterium der Nützlichkeit erfüllt ist, wenn „für das gemessene Merkmal praktische Relevanz besteht und die auf seiner Grundlage getroffenen Entscheidungen (Maßnahmen) mehr Nutzen als Schaden erwarten lassen“ (30). Man kann davon ausgehen, dass bei einer verlässlichen Messung der interessierenden Merkmale genau das der Fall ist. Werden psychische Belastungen möglichst genau erfasst, können auf den Erhebungsergebnissen aufbauend bedarfsgerechte und zielführende Maßnahmen zur Verbesserung des Wohlbefindens der Beschäftigten abgeleitet werden. Umgekehrt gilt, dass bei fehlender psychometrischer Güte des AIGScreen die Nützlichkeit durchaus in Frage stünde, da fehlerhafte Messungen zu falschen Schlüssen und somit im schlimmsten Fall tatsächlich zu schädlichen Maßnahmen führen könnten. Auch für die Normierung bedarf es zunächst eines Instruments, welches den wissenschaftlichen Gütekriterien entspricht. Denn die Entwicklung von Kennwerten und Vergleichszahlen bzw. der Vergleich von Befragungsergebnissen zwischen Gruppen setzt voraus, dass in allen (Sub-)Stichproben verlässlich, d. h. gleich und die gleichen Merkmale, gemessen wurden (70). Sind die Kriterien der Reliabilität und Validität erfüllt, spricht nichts dagegen, beispielsweise die Ergebnisse von einer Teilstichprobe (eine bestimmte Berufsgruppe, eine Dienststelle, ein Dienort, Frauen vs. Männer, Jüngere vs. Ältere etc.) mit der Gesamtstichprobe zu vergleichen und Aussagen über ein Besser- oder Schlechterabschneiden dieser einzelnen Gruppen zu treffen.

Die Zumutbarkeit eines Fragebogens hängt nicht nur von objektiven Kriterien wie der Umfragedauer ab, sondern auch von der Verständlichkeit der Fragen, dem Umstand, ob die Befragten die Fragen aufgrund ihres Vorwissens oder ihrer Vorerfahrung beantworten können und der Akzeptanz bzw. der Augenscheinvalidität (76, 29, 28, 81). Die Augenscheinvalidität wird, auch wenn es nicht explizit Gegenstand der Erhebung ist, im Rahmen der Expertenbefragung mit erhoben, da daran Vertreter der Zielgruppe des AIGScreen beteiligt sein werden. Sie sollen einschätzen ob einerseits alle relevanten Gegenstandsbereiche erfasst sind bzw. ob bestimmte Frageinhalte aufgrund mangelnder Relevanz für die Zielgruppe

zu streichen sind und ob die vorgestellten Items aus Probandensicht verständlich und beantwortbar sind. Die Befragungsökonomie wird ebenfalls nicht direkt erhoben. Allerdings ist auch hier zu erwarten, dass durch den regelmäßigen Austausch mit der Zielgruppe ein gutes Verhältnis von Wissenschaftlichkeit und Wirtschaftlichkeit des Fragebogens umgesetzt werden kann. Zudem ist eine hohe Messgenauigkeit (Reliabilität) einzelner Items hilfreich, um den Fragebogen möglichst kurz bzw. die Anzahl der verwendeten Items möglichst gering halten zu können.

Die Genauigkeit einer Messung kann über eine Vielzahl und Vielfalt an Methoden der Reliabilitätsanalysen beurteilt werden. Dabei handelt es sich bei den verfügbaren Methoden ausschließlich um solcher quantitativer Art. Die Replizierbarkeit der Messergebnisse ist kein maßgebliches Kriterium für den AIGScreen, da mit diesem Fragebogen zeitlich instabile Merkmale erfasst werden. Jedoch sollen, um die wissenschaftliche Güte des Instruments beurteilen zu können, Äquivalenzmessungen zur Bewertung der Eindimensionalität der latenten Konstrukte vorgenommen werden. Die damit erfasste interne Konsistenz bemisst gleichzeitig die Konstruktvalidität. Der Begriff der Konstruktvalidität wird in der Literatur nicht einheitlich verwendet und umfasst mehrere empirische Befunde und Vorgehensweisen (30, 28). Konstruktvalidität kann als gegeben betrachtet werden, wenn sich das zugrundeliegende theoretische Konstrukt empirisch messbar in seiner internen Struktur sowie in seinen externen Beziehungen abbilden lässt (69). Für die Bewertung der inneren Struktur eignet sich der Koeffizient Cronbachs Alpha, welcher anhand der durchschnittlichen Korrelation der Items einer Skala bemessen wird. Latente Konstrukte, deren Itemauswahl vor allem theoriebasiert ist und für die Vorannahmen bezüglich ihrer Interkorrelation und Dimensionalität bestehen, können mittels faktoranalytischer Verfahren überprüft werden (63). Die im Rahmen dieser Arbeit untersuchten Skalen zur Beurteilung der psychischen Belastung stellen reflektive Messmodelle dar. Das bedeutet, sie sollen eindimensional mit einer hohen internen Konsistenz sein. Einzelne Items können ausgetauscht werden, wenn sie sich als ungeeignet für die Repräsentation des zu messenden Konstrukts erweisen.

Voraussetzung für Konstruktvalidität ist, dass das Instrument objektiv und reliabel ist und Inhaltsvalidität als gegeben betrachtet werden kann (76). Denn grundsätzlich gilt das die drei Hauptgütekriterien Objektivität, Reliabilität und Validität in genannter Reihenfolge aufeinander aufbauen. Ist ein Instrument nicht objektiv, kann es auch nicht reliabel messen. Und eine schlechte Reliabilität wiederum wirkt sich negativ auf die Validität eines Instruments aus. Umgekehrt gilt jedoch nicht, dass ein reliables Instrument zwangsläufig auch valide ist. Diese Nachweise müssen stets separat und aufeinander aufbauend erbracht werden (28, 76). Letztlich definieren alle drei Gütekriterien den Standard einer guten wissenschaftlichen

Messung, auch wenn die Validität im methodischen Vorgehen die letzte Stufe und somit sinnbildlich die Königsdisziplin darstellt (71, 63).

Der Nachweis von Validität bedarf somit einer umfassenden und mehrstufigen Analyse, welche statistische Verfahren und logisch-fachliche Überprüfungen kombiniert. Ob das zu messende Merkmal vollständig erfasst wird (Inhaltsvalidität), kann letztlich nur durch Experten des jeweiligen Gegenstands- bzw. Fachbereichs beurteilt werden. Die postulierte Struktur der erhobenen Merkmale wird zwar mittels quantitativer Verfahren überprüft, setzt jedoch zunächst eine theoretische Modellierung der Dimensionen und Zusammenhänge voraus (Konstruktvalidität). Gleiches gilt für die Überprüfung der Prognosekraft hinsichtlich eines Außenkriteriums (Kriteriumsvalidität). Auch die Bewertung der prognostischen Eignung eines Messinstruments ist nur dann sinnvoll, wenn sich dieses aus theoriegeleiteten Hypothesen ableiten lässt.

3 Methode

Im folgenden Kapitel sollen zunächst die Vorüberlegungen und Rahmenbedingungen, die für die Entwicklung des AIGScreen wesentlich waren, genauer erläutert werden. Dabei wird deutlich, dass es sich bereits jetzt um ein berufsgruppenübergreifend einsetzbares Instrument handelt, auch wenn die später präsentierten Ergebnisse ausschließlich aus Daten einer Mitarbeitendenbefragung im Geschäftsbereich des BMVg stammen.

Die Überprüfung der psychometrischen Güte eines Instruments ist in der Regel ein mehrstufiger Prozess, welcher nicht nur, wie oben im Zitat angedeutet, mehrerer Erhebungen und Revisionen bedarf, sondern auch die Anwendung verschiedener Methoden und Erhebungsdesigns beinhaltet. So wurde auch im Rahmen dieser Arbeit vorgegangen. Um sowohl das tiefe Detailwissen von Experten nutzen zu können als auch objektivierbare statistische Analysen zur Beurteilung der psychometrischen Güte durchführen zu können, wurde ein Mixed-Method-Ansatz gewählt. Je nach zu beurteilendem Aspekt der wissenschaftlichen Güte, eignen sich quantitative oder qualitative Verfahren besser. So kamen Expertenbefragungen bzw. Fragebogenreviews zum Einsatz, um die Inhaltsvalidität zu bewerten. Hier standen Fragen der Aktualität und Angemessenheit des Fragebogeninhalts bezüglich des aktuellen Stands der Forschung sowie Optimierungsmöglichkeiten hinsichtlich des Fragebogendesigns im Vordergrund. Durch die Beteiligung der Zielgruppe des Fragebogens wurde zudem die Augenscheinvalidität in den Erhebungen berücksichtigt.

Zur Beurteilung von Reliabilität und Konstruktvalidität eignen sich quantitative Verfahren. Hier kamen insbesondere Itemanalysen und Faktorenanalysen zur Bewertung von Verteilungsstruktur, potenziellen Verzerrungseffekten sowie Skalenhomogenität und Eindimensionalität zum Einsatz. Der für die Faktorenanalysen verwendete Datensatz wurde mittels multipler Imputation für die fehlenden Werte ergänzt.

Sämtliche Erhebungen fanden im Rahmen eines Forschungsprojektes zum Betrieblichen Gesundheitsmanagement im Geschäftsbereich des BMVg statt und wurden entsprechend durch die Ethikkommission der Landesärztekammer Rheinland-Pfalz, den Datenschutzbeauftragten der Universitätsmedizin Mainz sowie dem Datenschutzbeauftragten des Kommando Sanitätsdienst der Bundeswehr genehmigt.

3.1 Fragebogenkonstruktion des AIGScreen

Eine ausführliche Erläuterung zum Hintergrund und der Entstehungsgeschichte des Fragebogens AIGScreen wurde bereits im Kapitel 1.1 gegeben. Dabei ging es primär um die Rahmenbedingungen der Entstehung, die Begründung der Notwendigkeit der Entwicklung des AIGScreen sowie die damit verbundenen Ziele und Anforderungen. Diese Aspekte sollen an

dieser Stelle nicht wiederholt erörtert werden. Vielmehr sollen aufbauend auf den beschriebenen Anforderungen und auf den in Kapitel 2 erläuterten theoretischen Grundlagen, die Überlegungen und Abwägungen, die zum gewählten Erhebungsdesign geführt haben, dargestellt werden.

Zum Anwendungszweck kann zusammenfassend gesagt werden, dass der AIGScreen für Mitarbeitendenbefragungen im Geschäftsbereich des BMVg zur Bedarfsanalyse und zur Bewertung des Erfolgs von Betrieblichen Gesundheitsmanagement (BGM) eingesetzt werden soll. Das Erhebungsinstrument soll in Form eines universell einsetzbaren, standardisierten Grobscreenings zur Ist-Zustandserhebung individueller und arbeitsbezogener Belastungen und Beanspruchungen angewendet werden, aus deren Ergebnissen sich Maßnahmen im Rahmen von BGM und Kennzahlen für eine Längsschnittanalyse ableiten lassen. Da insbesondere zur Erhebung psychischer Belastungen die individuelle Wahrnehmung entscheidend ist, soll die subjektive Bewertung der Situation durch die Mitarbeitenden erfasst werden. Für eine Mitarbeitendenbefragung eignen sich mit Blick auf die Ressourceneffizienz schriftliche bzw. Online-Befragungen mittels standardisiertem Fragebogen. Die theoretische Basis für den AIGScreen bildet das Belastungs-Beanspruchungs-Modell (Vgl. Kapitel 2.1). Die zu erfassenden Merkmale leiten sich aus der Leitlinie der Gemeinsamen Deutschen Arbeitsschutzstrategie ab (Vgl. Kapitel 2.2).

3.1.1 Fragebogenentwicklung

Obwohl es bereits einige Instrumente für den Anwendungsbereich der Beurteilung arbeitsbedingter psychischer Belastungen gibt, konnte keines identifiziert werden, das alle Anforderungen umfassend erfüllt und einer vollständigen Gefährdungsbeurteilung psychischer Belastungen gerecht würde (31). Die Bundesanstalt für Arbeitsschutz und Arbeitsmedizin gibt im Rahmen ihrer „Toolbox“ einen Überblick über 97 Instrumente zur Beurteilung psychischer Gefährdungsfaktoren (82). Die Mehrzahl dieser Instrumente wurde für den beschriebenen Zweck als ungeeignet angesehen, da sie entweder auf die Einbeziehung von Experten während der unmittelbaren Datenerhebung angewiesen sind, was den angestrebten Zielen der Niedrigschwelligkeit und Wirtschaftlichkeit widerspricht, oder weil sie qualitative bzw. nicht standardisierte Daten erheben und damit eine indikatorenbasierte Bewertung von BGM erschweren bis unmöglich machen. Andere Instrumente wie COPSQ – The Copenhagen Psychosocial Questionnaire (83) oder SPA – Screening psychischer Arbeitsbelastung (47) sind im Hinblick auf wissenschaftliche Qualitätskriterien gut entwickelt, aber sehr umfangreich und entsprechend zeitaufwändig. Zudem wurden bereits bestehende Instrumente wie ISTA – Instrument zur stressbezogenen Tätigkeitsanalyse (84) oder WDQ – Work Design Questionnaire (85) nicht oder nicht vordergründig mit dem Ziel entwickelt, alle Inhalte der gesetzlich geforderten Risikobewertung abzudecken. Trotz ihrer guten theoretischen

Fundierung fehlen ihnen daher entscheidende Aspekte zur Erfüllung des gesetzlichen Auftrags wie beispielsweise Aspekte der Arbeitsorganisation im Falle des WDQ oder soziale Aspekte und bestimmte Arbeitsumgebungsfaktoren bei ISTA. Der KFZA – Kurz-Fragebogen zur Arbeitsanalyse (86), der mit dem Ziel entwickelt wurde, eine Kurzversion etablierter Instrumente zu generieren, ist noch nicht validiert, was ein Kriterium für anwendbare Gefährdungsbeurteilungsinstrumente in den GDA-Empfehlungen ist (87). Andere Instrumente wie der Fragebogen zur Arbeitssituation an Schulen (FASS) (88) oder der von der Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege entwickelte Fragebogen zur psychischen Belastung in der stationären Krankenpflege „miab“ (89) sind oft an sehr spezifische Berufsgruppen angepasst, aber nicht allgemein anwendbar.

Um den in Kapitel 1.2 genannten Anforderungen gerecht zu werden, wurde daher ein eigener Fragebogen zur Beurteilung der Arbeitssituation und des individuellen Gesundheitsempfindens (AIGScreen) entwickelt (90). Der Inhalt des Fragebogens orientiert sich am aktuellen Stand der Forschung und an bereits entwickelten Instrumenten (18). Die Fragen, die sich auf die Arbeitssituation beziehen, orientieren sich zum Teil an dem AVEM, ein persönlichkeitsdiagnostisches Instrument für die berufsbezogene Rehabilitation (20), dem Arbeits-Bewertungs-Check für Lehrkräfte (ABC-L) (21) und den bereits erwähnten Instrumenten FASS (22) und COPSOQ (91). Dies sind gut konzipierte Fragebögen für die Analyse von Arbeitssituationen, Arbeitsbelastungen, Stressreaktionen und / oder der wahrgenommenen Gesundheitssituation. Leider konnte keiner von ihnen alle interessierenden Themen abdecken oder war für ein allgemeines Screening ungeeignet, da die Anzahl der Fragen jeweils zu hoch war. Im Rahmen von Expertenreviews wurden für die Entwicklung des Grobscreenings für die rheinland-pfälzischen Lehrkräfte und Pädagogischen Fachkräfte die wichtigsten Aspekte in Bezug auf psychische Belastungen bei der Arbeit ermittelt und in Anlehnung an bestehende Fragebögen wurden Items erstellt und ausgewählt. Die Anzahl der Items, die zur Erfassung der Arbeitssituation benötigt werden, konnte auf 28 reduziert werden. Die aktuell vorliegende Version des AIGScreen umfasst 34 Items, da im Rahmen der Weiterentwicklung und unter Berücksichtigung wissenschaftlicher Erkenntnisse 6 Aspekte ergänzt wurden. Damit handelt es sich immer noch um einen sehr ressourceneffizienten Fragebogen. Zum Vergleich: Der COPSOQ-Fragebogen arbeitet mit mindestens 87 und bis zu 157 Items (23).

An der Weiterentwicklung des Fragebogens für die Verwendung im Rahmen von BGM beteiligt war eine multidisziplinär besetzte Arbeitsgemeinschaft aus Arbeitsmedizinern, Psychologen und Militärpsychologen, Beschäftigtenvertretern, BGM-Beauftragten und Forschungspartnern aus acht Universitäten und Forschungsinstituten mit den Schwerpunkten Epidemiologie, Sozialmedizin, Ernährungswissenschaft, Psychologie, Arbeitsschutz, Sportwissenschaft und Erhebungsmethoden (26). Das Pilotprojekt startete im Januar 2015 an elf Dienststellen, die

als repräsentativ für die Organisationsstruktur aller Einheiten der Bundeswehr angesehen wurden. Dazu gehören die Verwaltung, vertreten durch das Bundesministerium, das Bundesamt für das Personalwesen, das Bundesamt für Ausrüstung, Informationstechnik und Nutzung der Bundeswehr und das Ernährungsamt, sowie Einheiten des Heeres, vertreten durch ein Korvettengeschwader, ein Panzergrenadierbataillon, ein Einsatzführungskommando und ein Bataillon für elektronische Kampfführung sowie die medizinische Versorgung, vertreten durch das Bundeswehrkrankenhaus Ulm.

Aufbauend aus den Erfahrungen der Pilotstudie und Erkenntnissen aus weiteren Expertendiskussionen wurde der Fragebogen zu Beginn des Folgeprojekts zur bundesweiten Ausrollung des BGM und somit zur flächendeckenden Befragung aller Beschäftigten im Geschäftsbereich BMVg angepasst. Einige Überarbeitungen betrafen ausschließlich den Sprachgebrauch, insbesondere mit Blick auf eine geschlechterneutrale Sprache. Missverständliche bzw. schwer verständliche Formulierungen wurden ebenfalls korrigiert. Zum Beispiel wurde die Aussage: „Ich empfinde den Lärmpegel an meinem Hauptarbeitsplatz als nicht störend“ umformuliert zu: „Ich empfinde den Geräuschpegel an meinem Hauptarbeitsplatz als angenehm“ um Schwierigkeiten bei der Beantwortung der Frage aufgrund der doppelten Verneinung auszuschließen.

Die Erhöhung der Fragezahl um 6 zusätzliche Items ist zum Teil auch darauf zurückzuführen, dass im Expertenreview auffiel, dass einige Aussagen nicht das Kriterium der Eindeutigkeit erfüllten und somit eine verlässliche Interpretationsleistung der Teilnehmenden im Sinne der Inhaltsvalidität nicht als gewährleistet betrachtet werden konnte (1). Dies betraf beispielsweise die Aussage „Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet bzw. habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit“. Hier sind zwei inhaltlich unterschiedliche Aspekte (Vorbereitung / Einarbeitung sowie Weiterbildungsmöglichkeiten) enthalten, die die Befragten nur getrennt beantworten können. Aus diesem Grund wurde das Item in zwei neue Aussagen aufgeteilt, welche lauten „Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet“ und „Ich habe ausreichend Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit“. Ein Item („Ich werde den an mich gestellten Anforderungen gerecht“) wurde im Sinne der Validität zudem gestrichen, da zu erwarten war, dass es aus Sicht der Probanden mehr auf die subjektive Einschätzung der eigenen individuellen Leistung abzielt als auf die von außen vorgegebenen Arbeitsanforderungen, welche eigentlich damit erfasst werden sollten. Der für das Folgeprojekt entwickelte Fragebogen ist in Anlage 1 dargestellt. Die einzelnen Schritte bzw. Forschungsprojekte, die der Entwicklung und Weiterentwicklung des AIGScreen gedient haben und die damit verbundenen Rahmenbedingungen sind für einen besseren Überblick in Abbildung 4 zusammengefasst.

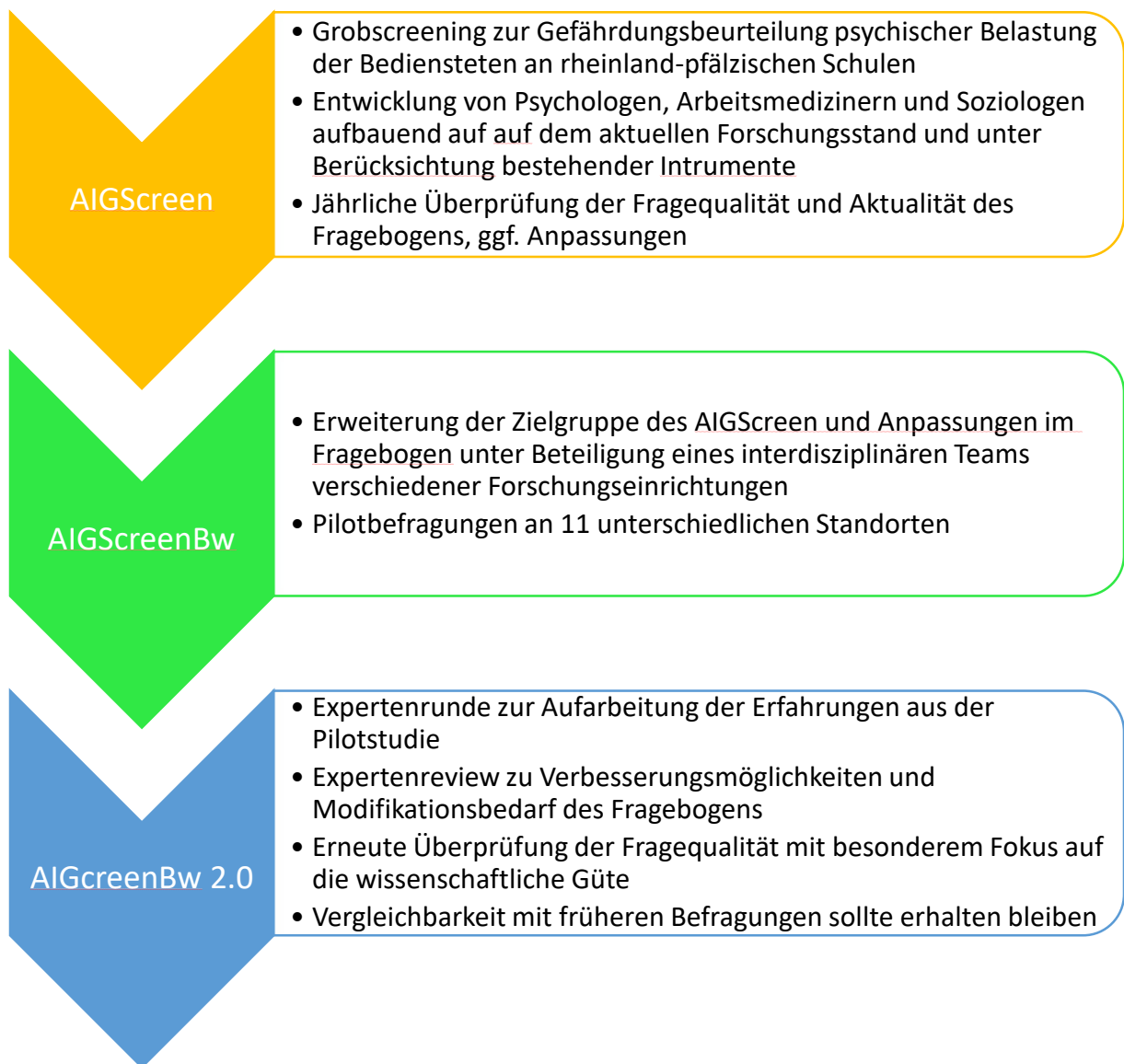
Wie bereits erläutert, entstammt das im Rahmen dieser Arbeit hinsichtlich seiner psychometrischen Güte zu überprüfende Instrument ursprünglich einem Fragebogen zur Analyse der Arbeitssituation und dem Gesundheitsempfinden von Lehrkräften und Pädagogischen Fachkräften in Rheinland-Pfalz und wurde für die wissenschaftliche Begleitung der Implementierung von BGM im Geschäftsbereich des BMVg weiterentwickelt. Dabei handelt es sich um sehr unterschiedliche Anwendungskontexte. Dass der AIGScreen dennoch in so verschiedenen Bereichen zur Beurteilung psychischer Belastungen eingesetzt werden kann², ist ein großer Vorteil des Instrumentes. Zwar wurden die hier verwendeten Daten ausschließlich im Rahmen von Erhebungen im Geschäftsbereich des BMVg, also mit dem AIGScreenBw 2.0, erhoben, da es sich jedoch weitgehend um dasselbe Instrument handelt, wird im Rahmen dieser Arbeit in der Regel etwas kürzer vom AIGScreen gesprochen.

Neben dem mit dem BGM verbundenen Zielen, war es den Forschenden, die an der Fragebogenentwicklung beteiligt waren, ein Anliegen, möglichst wenige Veränderungen am Instrument vorzunehmen. Denn je ähnlicher die Instrumente sind, desto besser ist eine Vergleichbarkeit der damit bereits gewonnenen Daten möglich. Eine breite, berufsgruppen- und tätigkeitsbereichsübergreifende Datenbasis eröffnet mehr Möglichkeiten für wissenschaftliche Untersuchungen und Fragestellungen. Dieses zusätzlich gewonnene Erkenntnispotenzial wurde vor allem dann als Argument berücksichtigt, wenn bestimmte Veränderungen im Fragebogen umstritten waren, von den beteiligten Akteuren unterschiedlich eingeschätzt wurden und anhand des aktuellen Forschungsstands nicht eindeutig beantwortet werden konnten.

Erste Analysen im Pilotprojekt konnten die Objektivität des Instruments bestätigen, deuteten jedoch bei einigen Subskalen zur Erfassung der Arbeitsmerkmale auf die Notwendigkeit zur Verbesserung der Reliabilität hin (Cronbachs Alpha < .80).

² Mittlerweile wird das Instrument für weitere, ebenfalls sehr unterschiedliche, Berufsgruppen und Arbeitssettings wie bspw. bei Beschäftigten der evangelischen Landeskirche der Pfalz eingesetzt.

Abbildung 4: Phasen der Entwicklung des AIGScreen



3.1.2 Erfassung psychischer Belastungen

Der Begriff „Psychische Arbeitsbelastung“ ist in der DIN-Norm EN ISO 10075-1 (92) definiert. Er subsumiert alle von außen auf die Psyche des Menschen einwirkenden Einflüsse. Diese können kognitiver, informativer oder emotionaler Art sein. Da Belastung in dieser Definition einen wertneutralen Begriff darstellt, hängt es von den individuellen Ressourcen und anderen Aspekten der Arbeitssituation ab, inwiefern psychische Belastungen zu Beanspruchungen führen.

Die theoretische Grundlage dieser Definition und auch des AIGScreen ist das in Kapitel 2.1.2 ausführlich beschriebene Belastungs-Beanspruchungs-Modell, das sowohl Belastung als auch Beanspruchung als wertneutrale Begriffe versteht. Das bedeutet, dass weder die Einwirkungen von Arbeitsbelastungen auf das Individuum noch Beanspruchungen als Folge

der Arbeitsbelastung als positiv oder negativ bewertet werden. Die Beziehung zwischen Belastung und Beanspruchung wird als von individuellen Merkmalen beeinflusst angesehen (93).

Joiko et al. (2010) definieren psychische Belastungen als „[...] die Gesamtheit aller erfassbaren Einflüsse, die von außen auf den Menschen zukommen und psychisch auf ihn einwirken“ (94). Welche Reaktionen diese Einwirkungen auslösen ist hingegen interindividuell verschieden. Die Bezeichnung „psychisch“ bezieht sich gemäß DIN EN ISO 10075-1 (92) auf wahrnehmende (informative), informationsverarbeitende (kognitive oder geistig-mentale) und gefühlsmäßige (emotionale) Vorgänge im Menschen. Angesprochen sind also Sinnesorgane, Denken, Lernen, Gedächtnis, Konzentration sowie Empfindungen. Psychische Belastungen müssen nicht zwangsläufig negative Auswirkungen haben. Berufliche Anforderungen können eine Herausforderung darstellen und bei erfolgreicher Bewältigung zu Arbeitszufriedenheit führen. Die gleiche berufliche Situation kann von einer anderen Person aber auch als Überforderung erlebt werden, weil sie z. B. nicht über das erforderliche Grundlagen- oder Spezialwissen verfügt. Insbesondere wenn sich Beschäftigte stressauslösenden Faktoren hilflos ausgesetzt fühlen, kann sich eine psychische Anspannung mit einhergehenden körperlichen Beschwerden einstellen (95, 96). Psychische Belastungen führen nicht bei allen Menschen zu gleichen Beanspruchungen. Persönliche Eigenschaften, Fähigkeiten und Fertigkeiten wie Ausbildung, Alter, Lebenserfahrung, die eigene Gesundheit oder Motivation für eine Arbeit sind Faktoren, die das Erleben einer Arbeitssituation stark beeinflussen. Zusätzlich haben arbeitssituationsbedingte Umstände wie Entscheidungsfreiheit, soziale Unterstützung und Wertschätzung durch Führungskräfte oder die Unterstützung durch Kollegen ganz erheblichen Einfluss auf die resultierende Arbeitsbeanspruchung (97).

Für die Beurteilung der arbeitsbedingten psychischen Belastung eignet sich das Belastungs-Beanspruchungs-Modell besonders gut, weil es erlaubt, Faktoren, die auf die objektiven Arbeitsbedingungen zurückzuführen sind, klar von individuellen Faktoren abzugrenzen. Darüber hinaus ist es möglich, das Zusammenwirken beider Aspekte im Hinblick auf ihre Wirkung bei der Entstehung von Beanspruchungen zu beschreiben (93). Wenn Kriterien für gute Arbeitsbedingungen definiert werden sollen, ist es notwendig, die Beanspruchungen, die durch die Arbeitsbelastung entstehen, beschreiben und erklären zu können. Um Maßnahmen abzuleiten, die die individuellen Ressourcen stärken oder das individuelle Gesundheitsverhalten verbessern, ist es notwendig, Belastungen trennscharf zu erfassen (39).

Eine große Herausforderung besteht darin, geeignete Indikatoren für die psychischen Belastungs- und Beanspruchungsdimensionen zu verwenden, da es sich um latente Variablen handelt. Der aktuelle Kernfragebogen des AIGScreenBw enthält 34 Items zur Erfassung von Belastungen am Arbeitsplatz. Die Auswahl der Konstrukte und ihrer Indikatoren orientiert sich

an den Empfehlungen und Leitlinien der für den Arbeitsschutz und die Arbeitsmedizin zuständigen Institutionen: der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) (98), der Deutschen Gesetzlichen Unfallversicherung (DGUV) (99) und der Gemeinsamen Deutschen Arbeitsschutzstrategie (GDA) (31).

In Anlehnung an die GDA-Empfehlungen (31, 58) besteht der Fragebogen aus den vier Hauptdimensionen Arbeitsorganisation, soziale Aspekte, Arbeitsumfeld sowie Arbeitsinhalt und Arbeitsaufgabe. Diese Dimensionen umfassen jeweils sehr heterogene Aspekte. Die *Arbeitsumgebung* umfasst die physikalischen, chemischen und ergonomischen Arbeitsbedingungen sowie die Gestaltung des Arbeitsplatzes, die Informationen und die Arbeitsmittel. *Soziale Aspekte* beziehen sich sowohl auf die Beziehung zum Vorgesetzten als auch auf die Beziehung zwischen Kollegen. Die sozialen Beziehungen zum Vorgesetzten ergeben sich aus der organisatorischen Einbindung der eigenen Arbeit in das Arbeitssystem, der notwendigen arbeitsbezogenen Interaktion und dem Kontakt mit dem Vorgesetzten. Zu letzteren zählen auch Aspekte wie Unterstützung, Anerkennung bzw. Rückmeldung zu den Arbeitsleistungen oder Konflikte. Die sozialen Beziehungen zu den Kollegen entstehen durch die arbeitsbezogene Interaktion und den Kontakt zu den Kollegen. Auch hier sind Unterstützung und Konflikte relevante Aspekte. Zudem kann die Interaktionsdichte sowohl bei besonders hoher als auch bei besonders niedriger Kontaktzahl eine Gefährdung darstellen. Die Dimension *Arbeitsorganisation* umfasst die Arbeitszeit, die Arbeitsabläufe und die Zusammenarbeit mit Kollegen. Der Faktor Arbeitszeit ist im Sinne des Arbeitsschutzgesetzes definiert als die Zeit vom Beginn bis zum Ende der Arbeit sowie die Lage und Verteilung der Arbeitszeiten und der Pausen und Ruhezeiten. Der Arbeitsablauf konzentriert sich auf Störungen, die die planmäßige Ausübung von Tätigkeiten in vorhersehbarer oder unvorhersehbarer Weise behindern. Die GDA subsumiert auch die Arbeitsintensität, also die quantitativen Anforderungen einer Tätigkeit im Sinne des Verhältnisses von erforderlicher Arbeitsmenge zu verfügbarer Zeit, sowie das Verhältnis von Aufgabenkomplexität zu verfügbarer Zeit unter Arbeitsorganisation. *Arbeitsinhalte und Arbeitsaufgaben* umfassen Handlungsspielraum, Vollständigkeit, Variabilität, Verantwortung, Information, Qualifikation und emotionale Anforderungen. Die Vollständigkeit umfasst die Kontrolle, Organisation, Vorbereitung und Ausführung eines in sich abgeschlossenen Produkts oder einer in sich abgeschlossenen Dienstleistung. Der Handlungsspielraum ist dadurch definiert, inwieweit der zeitliche Ablauf, die Zeit und die Dauer sowie die Wahl der Arbeitsmittel und -methoden und die Definition der Arbeitsziele selbst bestimmt werden. Variabilität liegt vor, wenn die Aufgaben und Anforderungen variieren und unterschiedliche Fähigkeiten und Fertigkeiten zur Erledigung der Aufgaben und zur Entscheidungsfindung erforderlich sind. Die Verantwortung beschreibt die Pflichten gegenüber Personen und Objekten, die ein Stelleninhaber bei der Ausübung einer bestimmten Tätigkeit hat. Die Qualität der Informationen ist definiert durch die

Vollständigkeit, Relevanz, Aktualität, Verständlichkeit, Zugänglichkeit und Eindeutigkeit der angebotenen Informationen zu Arbeitsaufgaben, Arbeitsgegenstand, gewünschten Ergebnissen oder Arbeitsablauf. Qualifizierung umfasst die quantitative und qualitative Übereinstimmung von tatsächlichen und geforderten beruflichen Kompetenzen (Fachwissen, Arbeitstechniken, Fähigkeiten und Fertigkeiten). Sie umfasst auch soziale und kommunikative Kompetenzen, die den Anforderungen einer bestimmten Tätigkeit entsprechen, sowie die Fähigkeit, auf Probleme zu reagieren und sie angemessen zu lösen. Die emotionalen Anforderungen beziehen sich hauptsächlich auf den Dienstleistungssektor. Einerseits können die Bedürfnisse und Interessen sowie die verbalen oder nonverbalen Äußerungen von Kunden, Klienten oder Patienten emotional belastend sein. Andererseits kann die Arbeitsaufgabe das Zeigen von erwünschten Emotionen erfordern, die zum Teil im Widerspruch zu den eigenen Gefühlen stehen.

Da es sich im Falle arbeitsbedingter psychischer Belastungen um einen sehr umfangreichen und komplexen Sachverhalt handelt, ist die Formulierung eines Itemuniversums kaum möglich. Im hier vorliegenden Fall würde dies beispielsweise erfordern, dass sämtliche potenzielle Belastungsfaktoren erfasst werden müssten. Darüber hinaus gäbe es eine Vielzahl an Formulierungsmöglichkeiten, um den jeweiligen Faktor zu beschreiben. Ein Mensch oder auch eine Gruppe von Menschen wäre nicht in der Lage diese potentiell unendliche Menge an Items zu überblicken, angemessen zu beurteilen und eine adäquate Auswahl zu treffen. In solchen Fällen ist es weder ökonomisch, noch zielführend, vollständige Itemuniversen formulieren zu wollen (100, 63). Dann ist es ausreichend und zielführender eine Gruppe von Experten zu beteiligen, welche in ihrem Urteil hinsichtlich inhaltlicher Vollständigkeit und Repräsentation übereinstimmen sollten (33). So wurde auch bei der Itemauswahl für den AIGScreen vorgegangen. Involviert war ein interdisziplinäres Team aus Arbeitsmedizinern, Psychologen und Soziologen.

Abbildung 5 zeigt den Kernfragebogen zur Erfassung arbeitsbedingter psychischer Belastungen in der letzten Version des AIGScreenBw 2.0 (Version Paper-Pencil-Befragung)³.

Abbildung 5: Fragen zur psychischen Belastung im AIGScreenBw 2.0

Bitte schätzen Sie ein, inwiefern die folgenden Arbeitsmerkmale an Ihrem Hauptarbeitsplatz vorhanden sind.		nein	eher nein	eher ja	ja	Unzu- tref- fend
Kategorie: Arbeitsmenge, zeitliche Aspekte, Arbeitspausen						
10	Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit ausgewogen.	☐	☐	☐	☐	☐
11	Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen.	☐	☐	☐	☐	☐

³ Die Abbildung hier entspricht dem Layout der Paper-Pencil-Befragung. Die Online-Version des Fragebogens weicht im Layout ab, ist hinsichtlich des Inhaltes und der Fragereihenfolge jedoch vollkommen identisch.

Bitte schätzen Sie ein, inwiefern die folgenden Arbeitsmerkmale an Ihrem Hauptarbeitsplatz vorhanden sind.		nein	eher nein	eher ja	ja	Unzu- tref- fend
12	Meine Arbeitszeitverteilung ist mir angenehm.	☐	☐	☐	☐	☐
13	Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen.	☐	☐	☐	☐	☐
14	Meine Arbeitspausen sind ausreichend.	☐	☐	☐	☐	☐
15	Meine Arbeitspausen sind störungsfrei.	☐	☐	☐	☐	☐
Kategorie: Arbeitsanforderungen, Kontrolle, organisatorische Arbeitsbedingungen						
16	Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet.	☐	☐	☐	☐	☐
17	Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit.	☐	☐	☐	☐	☐
18	Meine Arbeit ist in angenehmem Maße abwechslungsreich.	☐	☐	☐	☐	☐
19	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z. B. Personalverantwortung).	☐	☐	☐	☐	☐
20	Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material).	☐	☐	☐	☐	☐
21	Ich kann erkennen, ob ich meine Arbeit gut gemacht habe.	☐	☐	☐	☐	☐
22	Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit.	☐	☐	☐	☐	☐
23	Ich habe einen angemessenen Handlungsspielraum (z. B. Entscheidungsfreiheit) bei meiner Arbeit.	☐	☐	☐	☐	☐
24	Ich werde bei meiner Arbeit selten unterbrochen.	☐	☐	☐	☐	☐
25	Die nötigen Arbeitsmittel sind vorhanden.	☐	☐	☐	☐	☐
26	Arbeitsräume sind in ausreichender Anzahl vorhanden.	☐	☐	☐	☐	☐
27	Die Einrichtung an meinem Hauptarbeitsplatz ist angemessen.	☐	☐	☐	☐	☐
Kategorie: Soziale Aspekte der Arbeit						
28	Mein Kundenkreis für interne Dienstleistungen (z. B. auch Patienten, Leistungsempfänger) verhält sich angemessen.	☐	☐	☐	☐	☐
29	Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.	☐	☐	☐	☐	☐
30	Die Anerkennung meiner Arbeit durch meinen Mitarbeiterkreis ist angemessen.	☐	☐	☐	☐	☐
31	Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis.	☐	☐	☐	☐	☐
32	Mit meinen Mitarbeitenden gibt es keine Konflikte.	☐	☐	☐	☐	☐
33	Die Anerkennung meiner Arbeit durch meine unmittelbaren Vorgesetzten ist angemessen.	☐	☐	☐	☐	☐
34	Ich erfahre genügend Unterstützung von meinen unmittelbaren Vorgesetzten.	☐	☐	☐	☐	☐

Bitte schätzen Sie ein, inwiefern die folgenden Arbeitsmerkmale an Ihrem Hauptarbeitsplatz vorhanden sind.		nein	eher nein	eher ja	ja	Unzu- tref- fend
35	Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte.	☐	☐	☐	☐	☐
36	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Beschäftigten ist angemessen.	☐	☐	☐	☐	☐
37	Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Beschäftigten.	☐	☐	☐	☐	☐
38	Mit den Untergebenen/den mir unterstellten Beschäftigten gibt es keine Konflikte.	☐	☐	☐	☐	☐
39	Konflikte werden offen besprochen.	☐	☐	☐	☐	☐
Kategorie: Umgebungsbedingungen am Hauptarbeitsplatz						
40	Den Geräuschpegel an meinen Hauptarbeitsplatz empfinde ich als angenehm.	☐	☐	☐	☐	☐
41	Die äußeren Bedingungen (Klima, Beleuchtung, Belüftung etc.) an meinem Hauptarbeitsplatz sind angenehm.	☐	☐	☐	☐	☐
42	Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen.	☐	☐	☐	☐	☐
43	Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann (z. B. Arbeitshöhe, Sitzhöhe, Beinfreiheit, Durchgangsbreiten/Sicherheitsabstände).	☐	☐	☐	☐	☐
44	Sonstiges:					

3.1.3 Erhebungsdesign, Frageformat und Antwortskalen

Die Entscheidung, die Ermittlung psychischer Belastungen am Arbeitsplatz mittels eines standardisierten Fragebogens durchzuführen, liegt in den eingangs genannten Anforderungen an den Fragebogen begründet. Besonders vordergründige Motive waren der Wunsch nach einer quantitativen Bemessung verschiedener arbeitsbedingter Belastungsfaktoren sowie der Vergleichbarkeit von verschiedenen Stichproben und Kennzahlengenerierung einerseits und einem möglichst ressourceneffizienten Vorgehen andererseits. Dies bedeutet jedoch nicht, dass das gewählte Erhebungsdesign ein Kompromiss zwischen Praktikabilität und Eignung für die Beurteilung psychischer Belastungen darstellt. Die Gemeinsame Deutsche Arbeitsschutzstrategie (GDA) empfiehlt sogar Mitarbeitendenbefragungen, sofern diese bestimmte Qualitätskriterien erfüllen (50). Die Qualitätskriterien, die vom AIGScreen – evtl. mit Ausnahme der wissenschaftlichen Güte – erfüllt werden, wurden bereits in Abschnitt 2.2

beschrieben. Mitarbeitendenbefragungen haben gegenüber anderen Erhebungsformen einige Vorteile. So ermöglichen sie in der Regel, eine vollständige Erfassung der Situation, da sowohl alle Beschäftigten einbezogen werden können als auch ein breites Spektrum an Belastungen erfragt werden kann. Mitarbeitendenbefragungen liefern einen ressourceneffizienten und objektiven Überblick und ermöglichen die Identifikation von Interventionsschwerpunkten. Ergriffene Maßnahmen können hinsichtlich ihres Erfolgs mittels Referenzwerten evaluiert werden (31). Die Kehrseite dieser Vorteile ist, dass zunächst eine ausreichend hohe Beteiligung erzielt werden muss, um aussagekräftige Schlüsse ziehen zu können. Ein Faktor, der die Teilnahmebereitschaft verringern kann, können Bedenken bezüglich des Datenschutzes sein. Diesen kann zwar durch die Zusicherung von Anonymität begegnet werden. Das wiederum kann allerdings die Möglichkeiten der Auswertung einschränken. Wenn beispielsweise Organisationseinheiten mit kleiner Beschäftigtenzahl analysiert werden sollen, ist dies nur möglich, wenn der Rücklauf ausreichend hoch ist, da andernfalls aus den Antworten Rückschlüsse auf einzelne Personen gezogen werden könnten. Aus Sicht des Expertenteams überwogen jedoch die Vorteile. Neben den genannten Vorteilen ist außerdem darauf hinzuweisen, dass eine Beteiligung der Beschäftigten an der Beurteilung psychischer Gefährdungen in jedem Fall von der GDA gefordert wird (50) und dass mittels umfangreichen standardisierten Datensätzen mehr wissenschaftliche Fragestellungen bearbeitet werden können als dies beispielsweise mittels (unstandardisierten) Beobachtungsprotokollen einer Arbeitsplatzbegehung möglich wäre. Standardisierte Erhebungen haben außerdem den Vorteil, wie in Abschnitt 2.3.1 erläutert, dass sie einen guten, wenn auch nicht hinreichenden Indikator für die Erfüllung des Gütekriteriums der Objektivität darstellen.

Um den Rücklauf weiter zu erhöhen, wurden zusätzlich Paper-Pencil-Befragungen für die Dienststellen angeboten, an denen nicht alle Beschäftigten einen Computerarbeitsplatz haben. Damit konnten auch Personen ohne unmittelbaren Internetzugang ohne zusätzliche Hürden an der Befragung teilnehmen. Da Erfahrungen aus dem Pilotprojekt im Geschäftsbereich des BMVg zeigten, dass die Datenqualität bei den Online-Befragungen insgesamt besser ist und die digitale Erfassung auch das weniger ressourcenaufwendige Verfahren darstellt, wurde zudem eine dritte Erhebungsmöglichkeit in Form von Tablet-PC entwickelt. Die Tablets waren so konfiguriert, dass sie ausschließlich für den Zweck der Umfrageteilnahme benutzt werden konnten. Die Umfrage wurde direkt auf den Geräten installiert, sodass sie auch in Bereichen ohne Internetzugang eingesetzt werden können (z. B. auf hoher See oder in besonderen, abgeschirmten Sicherheitsbereichen). Die Befragung war mit einem automatischen Time-out versehen und einmal abgespeicherte Angaben konnten nicht erneut aufgerufen werden, sodass mehrere Personen nacheinander ein Tablet für die Befragungsdurchführung nutzen konnten, ohne die Anonymität der Angaben der einzelnen Probanden zu gefährden.

Um Ermüdungseffekten vorzubeugen, war es wichtig, die durch die alleinige Darstellungsform der Fragen entstehende kognitive Beanspruchung so wenig anspruchsvoll wie möglich zu gestalten. Dies umfasst das Seitenlayout, die Anwenderfreundlichkeit der Seite bzw. die sogenannte Softwareergonomie sowie die Menge der Fragen pro Seite (101). Letzteres war nicht einfach festzulegen, da mit unterschiedlichen Endgeräten auch die Darstellungsform und die Seitenlänge variiert. Eine dynamische Seite, die sich den Rahmenbedingungen des jeweiligen Endgeräts automatisch anpasst, war jedoch gewünscht, um die Zugriffsbarrieren möglichst niedrig zu halten. So konnten interessierte Personen bei Bedarf beispielsweise auch über ihr privates Mobiltelefon an der Befragung teilnehmen. Da eine Befragungsteilnahme vom Arbeitgeber jedoch ausdrücklich erwünscht war und daher im Rahmen der Arbeitszeit durchgeführt werden konnte, wurde angenommen, dass die überwiegende Mehrheit der Teilnehmenden an der Online-Befragung dafür ihren Arbeitscomputer nutzt. Eine Verifizierung dieser Annahme konnte im Projektverlauf nicht vorgenommen werden, da die dafür notwendigen Informationen im Sinne des Datenschutzes nicht erhoben wurden. Die Anzahl der Fragen bzw. die Seitenlänge bemisst sich daher an den Darstellungsmöglichkeiten eines Desktop-Computers. Die Anzahl der Fragen pro Umfrageseite wurden so gewählt, dass die Teilnehmenden nicht zu lange scrollen mussten und inhaltlich sinnvolle Blöcke entstanden ohne dabei Kontexteffekte zu produzieren (1).

Für alle 34 Items wurde eine positive Formulierung gewählt (z. B. "Meine Pausen sind ungestört", "Es gibt keine Konflikte zwischen mir und meinen Mitarbeitern"). Dadurch müssen Probanden weniger häufig umdenken bzw. sich bei der Beantwortung der Fragen nicht ständig neu auf der Antwortskala orientieren, was das Risiko versehentlich falsch gegebener Antworten und Ermüdungseffekte verringert. Außerdem erfolgte die Polung gemäß der gewohnten Richtung für den europäischen Raum von links (negativ) nach rechts (zunehmend positiv) (102). Die vollverbalisierten Antwortoptionen können auf einer 4-Punkte-Likert-Antwortskala mit einer zusätzlichen Antwortmöglichkeit "unzutreffend" bewertet werden. Laut Wolter & Junkermann (2018) sollte stets eine „Weiß nicht/Unzutreffend“-Optionen angeboten werden, um zu vermeiden, dass ahnungs- oder meinungslose Probanden sich gedrängt fühlen, ein Votum abzugeben (103). Diese kann die Ergebnisse unter Umständen erheblich verzerren. Da zudem davon auszugehen war, dass bestimmte Aspekte tatsächlich nicht von allen Befragten beantwortet werden können (z. B. können Fragen zur sozialen Beziehung mit unterstellten Mitarbeitenden nur von Führungspersonen beantwortet werden), musste schon deshalb eine Antwortoption „unzutreffend“ angeboten werden, um die Augenscheinvalidität aus Sicht der Teilnehmenden nicht zu gefährden. Da sowohl die Teilnahme an der Befragung als auch die Beantwortung einzelner Fragen freiwillig war, bestand bei allen Items die Möglichkeit, diese nicht zu beantworten bzw. im Fragebogen zu überspringen. Voll verbalisierte Antwortskalen wurden verwendet, da sie für die Probanden eine schnellere und

leichtere Beantwortung ermöglichen und eine intersubjektiv einheitlichere Interpretation gewährleisten, was sich positiv auf die Reliabilität und Validität der gewonnenen Daten auswirkt (104, 105).

Auf eine Mittelkategorie wurde verzichtet. In der Literatur können hierzu widersprüchliche Positionen und Forschungsergebnisse gefunden werden. So wird einerseits argumentiert, dass Probanden eine Mittelkategorie als Teil des möglichen Meinungsspektrums angeboten werden sollte. Nur so, so die Argumentation, können auch alle Personen ihre – möglicherweise neutrale – Position in den Antwortoptionen wiederfinden (106). Andere argumentieren, dass die Mittelkategorien ein „easy way out“ (105) sein können für Probanden, die nicht lange über die Antwort nachdenken wollen oder auch für solche, die sich keine Meinung bilden wollen oder können oder diese nicht kundtun möchten (107). Die Argumente gegen eine Mittelkategorie finden sich in der Satisficing-Theorie (108), wonach neutrale Antwortoption insbesondere bei Menschen mit geringer ausgeprägten kognitiven Fähigkeiten oder Desinteresse am Befragungsthema eine Tendenz zur Mitte, sprich ein neutrales Antwortverhalten, fördern würden (109). Diese Theorie wird bestärkt durch Erhebungen zur Verwendung einer „Weiß nicht“-Kategorie. Wurde diese Option nicht zur Verfügung gestellt, zeigte sich, dass mehr Personen eine Meinung zur jeweiligen Frage äußerten als wenn sie den „leichten Ausweg“ in die neutrale Antwortoption nehmen konnten (105). Da zudem eine „unzutreffend“-Antwortoption von den Befragungsteilnehmenden gewählt werden konnte für die Items, die diese tatsächlich aufgrund ihrer beruflichen Position oder Tätigkeit nicht beantworten können, wurde auf eine Mittelkategorie verzichtet.

Bei einer Stichprobe von Pädagogischen Fachkräften und Lehrkräften in Rheinland-Pfalz im Jahr 2014 / 2015 ($n = 964$) betrug Cronbachs Alpha für die jeweiligen Subskalen $\alpha = 0,85$ für Arbeitsinhalte und Arbeitsaufgaben, $\alpha = 0,87$ für Soziale Aspekte, $\alpha = 0,81$ für Arbeitsorganisation und $\alpha = 0,82$ für Arbeitsumgebung. Die interne Konsistenz der vier Skalen konnte jedoch im Rahmen der Pilotstudie und Adaption des AIGScreen für den Geschäftsbereich des BMVg nur noch eingeschränkt bestätigt werden. Folgende Werte wurden für die vier Skalen für Cronbachs Alpha bei einer Stichprobengrößen von $n = 2.076$ ermittelt: $\alpha = 0,83$ für Arbeitsinhalte und Arbeitsaufgaben, $\alpha = 0,80$ für Soziale Aspekte, $\alpha = 0,79$ für Arbeitsorganisation und $\alpha = 0,72$ für die Arbeitsumgebungsfaktoren. In der einschlägigen Fachliteratur wird für Cronbachs Alpha allerdings erst ein Wert ab $\geq 0,80$ als akzeptabel angegeben (64). Vor dem Hintergrund des Anwendungszwecks, insbesondere die geplanten Gruppen- und Längsschnittvergleiche, ergibt sich sowohl mit Blick auf die Qualitätskriterien der GDA als auch hinsichtlich der mit der Befragung verbundenen wissenschaftlichen Einsatzziele die Notwendigkeit einer Überprüfung der wissenschaftlichen Güte des Instruments sowie ggf. der Überarbeitung, wenn die Gütekriterien nicht zufriedenstellend erfüllt werden.

3.2 Datenerhebung im Rahmen der Mitarbeitendenbefragung

Der Feldzugang wurde im Rahmen des wissenschaftlichen Folgeprojekts zur bundesweiten Ausrollung von Betrieblichen Gesundheitsmanagement (BGM) im Geschäftsbereich des Bundesministeriums der Verteidigung (BMVg) ermöglicht. Die Projektlaufzeit begann im Januar 2018 und endete im Juli 2021. Da die Analyse der psychometrischen Eigenschaften des Fragebogens einen Teil der zu leistenden Arbeitspakete im Rahmen des Forschungsprojektes darstellte, erstreckte sich die Phase der Datenerhebung für die Überprüfung von Reliabilität und Validität des AIGScreen nicht über den gesamten Projektzeitraum, sondern endete bereits im Oktober 2019.

Da sich das Folgeprojekt auf den gesamten Geschäftsbereich bezog, sollten auch alle Mitarbeitenden befragt werden bzw. zur Teilnahme an der Befragung im Projektzeitraum eingeladen werden. Die Befragung erfolgte dabei zeitlich gestaffelt für jeweils vier bis acht Organisationseinheiten in einem Zeitraum von circa vier Wochen. Zudem wurde mit Abschluss des Pilotprojektes beschlossen, dass für die Längsschnittanalysen und Maßnahmenevaluation eine Wiederholungsbefragung nach circa drei Jahren anzusetzen sei.

Die Koordination der Befragung, also die Entscheidung welche Organisationseinheiten zu welchem Zeitpunkt und in welchem Format (ausschließlich Online oder zusätzlich mit Papierfragebögen⁴) befragt werden soll, oblag ausschließlich der Bundeswehr. Aufgrund des langen Befragungszeitraums, dem hohen Interesse und der hohen Teilnahmebereitschaft insbesondere zu Beginn des Projektes, dem grundsätzlichen Ziel alle Dienststellen im Projektzeitraum mindestens einmal befragt zu haben und der Vielzahl und Heterogenität, der bis Oktober 2019 befragten Dienststellen bzw. Organisationseinheiten ist jedoch davon auszugehen, dass die Stichprobe mindestens repräsentativ für den Geschäftsbereich BMVg ist. Genauere Informationen zur Stichprobe finden sich in Abschnitt 4.3.1 „Beschreibung der Stichprobe“.

Der Zugang zur Online-Befragung erfolgte über eine von der Universitätsmedizin Mainz (UM) betreuten Website. Die UM war auch im gesamten Projektzeitraum der Host für das eingesetzte Online-Befragungstool „Limesurvey“. Sämtliche Daten wurden ausschließlich auf UM-eigenen Servern gespeichert.

Für interessierte Personen, die nicht an der Online-Befragung teilnehmen konnten oder wollten, wurde ein Papierfragebogen zur Verfügung gestellt. Diese wurden an die jeweiligen Dienstorte zu Händen des zuständigen BGM-Koordinierenden vor Ort kurz vor dem geplanten

⁴ Die vorher erläuterte Möglichkeit der Tablet-Befragung wurde in diesem Befragungszeitraum noch entwickelt und erst zu einem späteren Zeitpunkt eingesetzt.

Befragungszeitraum versandt. Der BGM-Koordinierende oder eine andere benannte Person trug vor Ort Sorge, dass die Anonymität der Befragungsteilnehmenden und ihrer Angaben gewahrt blieb. Nach Abschluss des Befragungszeitraums wurden die Fragebögen zur weiteren elektronischen Eingabe an das Institut für Arbeits-, Sozial- und Umweltmedizin (ASU) zurückgeschickt. Die Erfassung der Daten erfolgte überwiegend automatisiert mittels einer Scannersoftware. Fragebögen, die nicht eingelesen werden konnten, wurde manuell von geschulten Mitarbeitenden des ASU erfasst.

Inhaltlich und in der Reihenfolge der Fragen sind beide Erhebungsformate identisch. Insgesamt umfasst der Fragebogen 134 Fragen. Davon erfassen 34 standardisierte Items in geschlossenem Antwortformat die individuell empfundene Arbeitssituation in den vier von der Gemeinsamen Deutschen Arbeitsschutzstrategie empfohlenen Dimensionen Arbeitsumgebung, Arbeitsorganisation, Arbeitsinhalt und -aufgabe sowie arbeitsbedingte soziale Beziehungen. Diese Skalen gilt es im Rahmen der Arbeit hinsichtlich ihrer wissenschaftlichen Güte zu überprüfen. Darüber hinaus wurden im Rahmen der Befragung gemäß dem Belastungs-Beanspruchungs-Modell individuelles Gesundheitsverhalten und individuelle Ressourcen, etwaige Belastungsfolgen in Form gesundheitlicher Beschwerden, soziodemographische Informationen, die persönliche Einstellung zum BGM-Angebot⁵ sowie Informationen zum Beschäftigungsverhältnis und Auslandseinsätzen erhoben. Der vollständige Fragebogen ist in Anlage 1 zu finden.

Unabhängig davon, ob die Befragung online oder in Papierform angeboten und durchgeführt wurde, wurde allen Mitarbeitenden der befragten Dienststelle zugesichert, dass ihnen weder durch die Teilnahme noch durch die Nicht-Teilnahme Vor- oder Nachteile entstünden. Dies wurde u. a. durch eine Zusicherung der vollständigen Anonymität der Angaben gewährleistet. Zudem war die Teilnahme an der Befragung freiwillig und es wurde auch nicht erhoben, welche Beschäftigten teil- bzw. nicht teilgenommen haben.

Lediglich die Dienststellen- und Abteilungszugehörigkeit der Befragungsteilnehmenden wurde erfasst. Dies war für die spätere Zuordnung und dienststellenspezifische Auswertung der Befragungsergebnisse notwendig. Auswertungen für Gruppen und Organisationseinheiten von weniger als 15 Teilnehmenden wurden und werden nicht vorgenommen, sodass auch anhand der Gruppenergebnisse Rückschlüsse auf einzelne Personen ausgeschlossen werden konnten.

Um dennoch längsschnittliche Analysen der Entwicklung des individuellen Arbeits- und Gesundheitsempfindens zu ermöglichen, wurden die Probanden gebeten einen individuellen, nur ihnen bekannten Code zu generieren. Dafür wurde ein Muster aus Zahlen und Buchstaben

⁵ Die drei betreffenden Fragen wurden erst nachträglich im April 2018 im Fragebogen ergänzt.

vorgeschlagen, die nur den Teilnehmenden bekannt sind und im Rahmen der Befragung oder des Dienstverhältnisses nicht erfasst wurden (z. B. Buchstaben aus dem Vornamen der Mutter).

Für das gesamte Vorgehen zur Mitarbeitendenbefragung liegt sowohl vom Datenschutzbeauftragten der UM als auch vom Datenschutzbeauftragten des Kommando Sanitätsdienst der Bundeswehr ein positives Votum vor. Darüber hinaus wurde ein positives Votum der Ethik-Kommission der Landesärztekammer Rheinland-Pfalz für das Forschungsvorhaben eingeholt.

3.3 Expertenbefragung

Da die Inhaltsvalidität notwendigerweise gegeben sein muss, wenn auch die darauf aufbauenden Gütekriterien Reliabilität und Konstruktvalidität erfüllt werden sollen (64), sollte in einem ersten Analyseschritt der bereits bestehende Fragebogen AIGScreen von Experten evaluiert werden. Bewertet werden sollte dabei, ob sich die verwendeten Konstrukte gemäß dem derzeitigen Forschungsstand und der Praxis der Gefährdungsbeurteilung begründen lassen und in ihrer Definition plausibel sind. Zudem sollte geprüft werden, ob mit den ausgewählten Konstrukten und Items alle relevanten Aspekte für ein Grobscreening abgedeckt sind und die verwendeten Items die ihnen zugrundeliegenden Konstrukte ausreichend repräsentieren. Darüber hinaus galt es den Fragebogen auf grundlegende Qualitätsstandards der Fragebogenentwicklung und auf das Gütekriterium der Objektivität zu prüfen.

Die Frage, wer ein Experte ist, ist ein vieldiskutiertes und nicht abschließend oder zumindest nicht einheitlich beantwortetes Thema in der Fachliteratur (110–114). Klar ist, dass sich die Auswahl der Experten wenigstens teilweise aus der Fragestellung ergibt, auf die die jeweiligen Experten eine Antwort finden sollen (115). Da es hier mehrere Fragestellungen zu unterschiedlichen Aspekten zu klären gab, war es naheliegend, auch mehrere Expertengruppen zu beteiligen. Während die Expertise für Aspekte wie Praxistauglichkeit, Augenscheinvalidität oder Nützlichkeit eindeutig bei der Zielgruppe, den Beschäftigten im Geschäftsbereich des BMVg gesehen wurde, wurde das Expertenwissen für Aspekte der Fragebogengestaltung, Objektivität oder Vergleichbarkeit unter den wissenschaftlichen Mitarbeitenden des ASU gesehen.

Aufgrund der unterschiedlichen Verfügbarkeit der Expertengruppen und der inhaltlich gut trennbaren Erkenntnisinteressen wurde der Reviewprozess zweigeteilt. Dies barg den Vorteil, dass sich auch der Aufwand für die befragten Experten deutlich reduzierte, weil diese sich nur mit einem Teilaspekt befassen mussten, welcher zudem in ihrem Wissensbereich lag und von

ihnen keine zusätzliche Einarbeitungszeit verlangte. Beide Aspekte wurden als vorteilhaft hinsichtlich Mitwirkungsbereitschaft, Motivation und Rücklaufquote erachtet.

3.3.1 Vor- und Nachteile von Expertenbefragungen

Obwohl die Methode Expertenbefragung selbst hinsichtlich ihrer wissenschaftlichen Güte kritisiert werden kann, findet sie breite Anwendung in der Forschung, insbesondere in der Fragebogenentwicklung (70, 116). Die Kritik leitet sich aus dem Umstand ab, dass es sich bei Expertenbefragungen um keine vollständig standardisierten, messbaren Verfahren handelt. Erfasst werden subjektive Einschätzungen, welche nicht anhand eines objektiven Standards in ihrer Richtigkeit und Gültigkeit bewertet werden können (112). Häufig geäußerte Kritikpunkte betreffen insbesondere die Objektivität des Verfahrens. So kann bereits die Auswahl der Experten an sich verzerrt sein, da es sich um eine soziale Attribution handelt, die von vielen Faktoren beeinflusst wird, die sich nicht auf die Eignung für den Befragungszweck beziehen (z. B. Prestige, Macht). Zweitens kann dem Expertenurteil selbst ein hoher Grad an Subjektivität unterstellt werden. Manche Autoren argumentieren, dass Experten bezüglich ihrer Wissensdomäne aufgrund ihres tiefen Wissens und persönlicher Erfahrungen in dem Bereich wesentlich weniger wertfrei und neutral, sondern mit vorgefertigten Meinungen Sachverhalte bearbeiten (117).

Damit müssten ebenfalls die Reliabilität und die Validität der Ergebnisse in Frage gestellt werden. Allerdings besteht „überwältigende wissenschaftliche Evidenz [eigene Übers.]“ (117), dass Expertenurteile sehr wohl der Überprüfung wissenschaftlicher Kriterien standhalten. So konnten Presser & Blair (1994) nachweisen, dass Expertenreviews von Fragebögen im Vergleich zu anderen Methoden (kognitiven Pretests, konventionelles Testen, Behavior Coding / strukturierte Verhaltensbeobachtung) die meisten Probleme identifizieren konnten (118). Gemeinsam mit der strukturierten Verhaltensbeobachtung (Behavior Coding) stellten sie in der Untersuchung die reliabelste Methode dar. Zudem waren sie im Vergleich zu den anderen drei Methoden die kostengünstigste. Auch Olson (2010) überprüfte die Eignung von Expertenreviews für die Fragebogenevaluation und stellte fest, dass auch bei geringer Reliabilität die Summe der Expertenurteile erfolgreich die Items identifizierte, welche sich in der Erhebung auch als problematisch (z. B. in Form von item-non-response) erwiesen (116).

Forschende sollten sich bei der Durchführung von Expertenbefragungen der Möglichkeit der Verzerrung durch subjektive Meinungen und individuelle Präferenzen bewusst sein und Expertenurteile daher nicht unreflektiert als endgültige Ergebnisse übernehmen. Vielmehr können und sollen Expertenmeinungen als Orientierung bei der Identifikation von potenziellen Problemen einzelner Fragen eingesetzt werden (70). Dies trifft in der Regel jedoch nicht auf das Gesamtkonzept des Erhebungsinstrumentes zu: „Whereas structured expert reviews are

very concrete with regard to each question, they do not provide evidence on the interplay of questions“ (70). Dies ist aus Sicht der Autoren des *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System* jedoch kein Argument gegen die Nutzung von Expertenbefragungen, sondern muss lediglich bei der Ergebnisinterpretation berücksichtigt werden. Vielmehr empfehlen die Autoren den Einsatz von Expertengruppen und verweisen darauf, dass dies nach eigenen Recherchen die zweithäufigste Methode (nach Interviewerschulungen) der Instrumentenentwicklung und -evaluation in den meisten nationalen europäischen Statistikbehörden und internationalen Forschungseinrichtungen darstellt (70). Die Beteiligung von Experten bei der Fragebogenentwicklung wird dringend empfohlen, um einerseits sicherzustellen, dass die für die Fragestellung notwendigen Informationen vollständig erhoben werden und dass andererseits das Fragedesign dem Standard guter wissenschaftlicher Praxis entspricht.

Auch Faulbaum et al. (2009) verweisen auf die Notwendigkeit, den Fragebogen wiederholt systematisch auf Qualitätsgefährdungen zu überprüfen. Zur Identifikation solcher Gefährdungen werden Fragebewertungssysteme und Expertenrunden als anerkannte und gut bewährte Verfahren empfohlen (1). Klein et al. (2017) sprechen sich in einem Artikel dafür aus, dass Expertenbeteiligung, entgegen aller Einwände aus verschiedenen Disziplinen, einen unersetzlichen Beitrag im Forschungsprozess leisten (117). Sie sind demnach nicht nur effektiver als andere Methoden hinsichtlich ihrer Kosten-Nutzen-Bilanz, sondern sie liefern auch tiefes, umfassendes und spezifisches Detailwissen, das mit anderen Methoden nicht genutzt werden könnte. Den Autoren zufolge finden sich in der Forschung und Wissenschaftsliteratur genügend Belege, dass Expertenurteile als zuverlässig und präzise hinsichtlich ihres wissenschaftlichen Werts angesehen werden können. Darüber hinaus bringen Experten einer bestimmten Wissensdomäne ein hohes Maß an intrinsischer Motivation mit in den Forschungsprozess und sind aufgrund ihrer persönlichen Involviertheit bereit, intensiver an Fragestellungen zu arbeiten.

In Anbetracht der Vorteile, die Expertenbefragungen mit sich bringen und die durch andere Verfahren nicht substituiert werden könnten, findet sich in der Fachliteratur auch Konsens darüber, dass sie im Rahmen von Fragebogenentwicklungen Anwendung finden sollten (70, 116, 1, 30, 119, 120, 71). Dennoch müssen potenzielle Verzerrungseffekte sowohl bei der Auswahl von Experten als auch bezüglich der Expertenurteile berücksichtigt werden. Entsprechende Vorsicht muss bei der Ergebnisauswertung und Interpretation von Daten walten. Auch wenn es unmöglich ist, vollständig unverzerrte Expertenpanels zu generieren, sollte im Studiendesign dieses Risiko besondere Beachtung finden. Potenziellen Verzerrungseffekten kann nicht nur durch eine vorsichtige Ergebnisinterpretation, sondern vielmehr auch durch eine Vielfalt mehrerer kombinierter Auswahlkriterien für die

Expertenrekrutierung begegnet werden (111). Wer als Experte gilt und wie Experten identifiziert werden können, soll im nachfolgenden Kapitel erörtert werden.

3.3.2 Definition und Auswahl von Experten

Die Auswahl von Experten bzw. die Bestimmung von Auswahlkriterien für Experten zu einer bestimmten Fragestellung kann, je nach Methode und Denkschule unterschiedlich erfolgen. So verfolgt der **soziologische Ansatz** ein konstruktivistisches Begriffsverständnis, wonach Expertise ein Zuschreibungskriterium der Umwelt darstellt. Eine allgemeine oder auch kontextspezifische Anerkennung als Experte ist demnach das wesentliche Kriterium, das eine Person zum Experten macht (111). Anders sieht es die **behavioristische Schule**, wonach sich der individuelle Grad von Expertise an der eigenen Performanz im Verhältnis zu anderen Experten, aber auch an der eigenen Selbsteinschätzung bemisst. Als Experte gilt hier, wer einen hohen Genauigkeitsgrad bei bisherigen Einschätzungen an den Tag legte und sich nahe am Mittelwert vieler verschiedener Experteneinschätzungen bewegt. Die Genauigkeit von Einschätzungen zu einem bestimmten Sachverhalt steigt diesem Ansatz nach mit der Diversität der beteiligten Experten. So sieht es auch der **kognitive Ansatz**: „Scholars should not generally view experts as "all-knowing single answer decision makers" but as knowledgeable consultants who disagree because they see alternative paths" (ebd.). Anders als die behavioristische Sichtweise bemisst dieser Ansatz Expertise jedoch nicht primär an der individuellen Performanz, sondern vor allem daran, wie das Wissen erlangt wurde und wie Fragestellungen aufgegriffen werden. Der Experte unterscheidet sich dabei vom Nicht-Experten durch eine besser, d. h. genauere und detaillierte Problembeschreibung. Er weist ein vernetzteres Denken auf und bezieht eine Vielzahl an Informations- und Erfahrungsquellen in die Problemlösung mit ein. Seine Fähigkeiten hat er durch ein intensives und domänenspezifisches Training erlangt, dass es ihm ermöglicht überdurchschnittlich tiefes Wissen in die Bewertung einfließen zu lassen und in detaillierteren Kategorien zu denken, als Nicht-Experten es könnten (ebd.).

Bei der Auswahl von Experten im Rahmen dieser Arbeit wurden alle drei Ansätze berücksichtigt. Einerseits können Experten anhand der Zuschreibung von Expertise in Form eines bestimmten Titels, Abschlusses oder der beruflichen Position identifiziert werden. Andererseits war tiefes domänenspezifisches Wissen unabdingbar, welches sich in der Fähigkeit ausdrücken sollte, Fragestellungen in sehr detaillierten Kategorien bearbeiten zu können. Darüber hinaus wurde anerkannt, dass es bei Fragen, die die Inhaltsvalidität eines Messinstruments für latente Konstrukte betreffen, nicht die eine richtige Antwort gibt. Vielmehr sollten die besten Empfehlungen und Hinweise aus dem Mittel vieler verschiedener

Expertenmeinungen ermittelt werden. Dafür sollte nicht nur eine personelle, sondern auch eine fachliche Vielfalt in der Gruppe der Experten erreicht werden.

Neben der fachlichen Disziplin kann auch das jeweilige Erkenntnisinteresse bzw. der Forschungsgegenstand definieren, wer als Experte gilt. So unterscheiden Weiss & Shanteau (2003) in vier Kategorien von Experten: Solche, die Urteile fällen (z. B. Diagnosen erstellen oder Noten vergeben), solche, die Vorhersagen treffen (z. B. bezüglich des Wetter oder der Börsenentwicklung), solche, die anleiten (z. B. als Trainer oder bei der Entwicklung von Testkriterien) und solche, die man ihrer überdurchschnittlichen Leistung bemessen kann (z. B. Sportler oder Musiker) (110). Einen Konsens aller verschiedenen Ansätze, wer als Experte gilt, kann in der Definition von Bogner et al. (2014) gefunden werden:

„Im Begriff des Experten steckt die lateinische Sprachwurzel „expertus: erprobt, bewährt. Dieses „expertus“ leitet sich von einem Verb her, das es nur in der Passivform gibt, nämlich „experiri: prüfen, ausprobieren. In Lexika werden Experten gewöhnlich als Sachverständige, Fachleute, Kenner charakterisiert. Der Experte ist jemand, so heißt es in der Encyclopedia Britannica, der sachkundig ist und über Spezialwissen verfügt“ (114).

Das gefragte Spezialwissen für die Expertenbefragungen im Rahmen dieser Arbeit bezieht sich einerseits auf methodische Aspekte im Bereich Fragebogenentwicklung und Erhebungsdesign, andererseits auf inhaltlich-fachliche Kompetenzen im Bereich arbeitsbedingte psychische Belastung sowie drittens auf Kenntnisse bezüglich der Zielgruppe der Befragung. Die beteiligten Experten sollen also vor allem in der Lage sein, Urteile zu fällen. Da es schwer möglich schien, eine Vielzahl an Experten zu identifizieren, welche tiefer gehendes Wissen in allen drei Bereichen aufweisen, wurden die zu bearbeitenden Fragestellungen und die dafür benötigten Gruppen von Experten in zwei Subgruppen aufgeteilt. Eine Gruppe sollte für das Fragebogenreview eingesetzt werden und benötigte somit umfangreiches Wissen in den Bereichen Fragebogenkonstruktion und empirische Forschungsmethoden. Die zweite Gruppe sollte beurteilen, inwiefern sich der Fragebogen für die Zielgruppe eignet und ob er in Bezug auf deren Bedürfnisse und die bestehenden Rahmenbedingungen vollständig ist (Augenscheinvalidität). Hierfür wurde Erfahrungswissen aus dem Arbeitssetting als notwendiges Kriterium angesehen, weshalb ausschließlich Beschäftigte im Geschäftsbereich BMVg dafür in Frage kamen. Experten beider Gruppen mussten vertiefte Kenntnisse im Bereich arbeitsbedingte psychische Belastungen und Gefährdungsbeurteilung aufweisen, da die Inhaltsvalidität eines Instruments letztlich nur von Experten beurteilt werden kann, die mit dem inhaltlich zu erfassenden Konstrukt umfassend vertraut sind (121).

Nachdem definiert ist, wer als Experte gilt, stellt sich die Frage, wie Experten identifiziert und für den Forschungsprozess ausgewählt werden können. Auch hier unterscheiden sich die

Auswahlkriterien je nach Disziplin. So wurde bereits erläutert, dass im soziologischen Ansatz die **Zuschreibung** das wesentliche Kriterium darstellt. Dabei hat die Einschätzung von Personen, die in dem jeweiligen Feld tätig sind, in dem Expertise gefragt ist, besonderes Gewicht. Wenn sich mehrere Personen einer bestimmten Profession darin einig sind, dass eine Person ein Experte ist, dann – so die Logik bei diesem Ansatz – ist diese Person auch ein Experte (122). Davon nicht ganz unabhängig ist das Kriterium des politischen Einflusses bzw. von **Einflussmöglichkeiten**. Personen, die in einem bestimmten Bereich potenziell Macht ausüben können, können aus Sicht des soziologischen Ansatzes ebenfalls als Experten angesehen werden (111). Allerdings ist dies eher ein schwaches Auswahlkriterium, da mächtige oder einflussreiche Personen nicht zwangsläufig über eine hohe Kompetenz oder Spezialwissen verfügen. Zudem wurden im Rahmen dieser Arbeit keine Fragestellungen verfolgt, für die politische Macht eine hilfreiche Kompetenz darstellt. Dieses Kriterium kam somit für die Auswahl von Experten nicht infrage. Ein anderes, ebenfalls dem soziologischen, aber auch dem kognitiven Ansatz zuzuordnendes Kriterium stellt **persönliche Involviertheit** dar (111). Die Auswahl orientiert sich demnach am Interesse einer Person an dem Thema. Ein großer Vorteil dieses Kriteriums ist es, dass die Personen aufgrund des Interesses an der Thematik ein hohes Maß an intrinsischer Motivation aufweisen und wahrscheinlich eine umfangreiche Rückmeldung erzielt wird. Andererseits müsste das persönliche Interesse für eine objektive Auswahl zunächst messbar gemacht werden. Darüber hinaus besteht die Gefahr der Selbstselektion. Dennoch wurde das Kriterium insbesondere für die zweite Gruppe von Experten, die sich aus der Zielgruppe rekrutiert, als sinnvoll erachtet. Persönliche Involviertheit ist hier aufgrund des Forschungskontextes a priori gegeben. Möglichen Verzerrungseffekten durch Selbstselektion sollte durch eine für Expertenbefragungen überdurchschnittlich große Stichprobengröße begegnet werden.

Weitere Auswahlkriterien im Rahmen des kognitiven Ansatzes lassen sich anhand von **Tests zu individuellen Kompetenzen oder verifizierbarem Wissen** ermitteln (111). Das grundsätzliche Problem von Wissenstest zur Ermittlung von Experten ist, dass man Experten ja vor allem dort braucht, wo es noch keinen „Gold Standard“ (122), also kein ausreichend fundiertes Wissen, bezüglich eines bestimmten Sachverhaltes gibt. Das besondere Problem im Kontext dieser Arbeit ist, dass es sich bei den verfolgten Fragestellungen eben nicht um verifizierbares Wissen handelt. Es können in dem Sinne keine richtigen oder falschen Antworten gegeben werden, sondern aus der Vielzahl von Empfehlungen soll ein möglichst guter Konsens erarbeitet werden. Tests zu individuellen kognitiven Fähigkeiten können sich auf unterschiedliche Aspekte beziehen. Dabei werden in der Literatur häufig zwei Kriterien genannt. Das erste Kriterium bezieht sich auf die sogenannte **innere Konsistenz**, häufig auch als within-reliability oder intra-person reliability bezeichnet. Das zweite Kriterium beschreibt die Fähigkeit detaillierte Unterscheidungen zu machen, die sog. „discrimination ability“ (110–113,

122). Ersteres Kriterium verlangt eine zeitliche Konsistenz von Einschätzungen und Urteilen desselben Experten. D. h. in Wiederholungsbefragungen sollte zum selben Sachverhalt von derselben Person dieselbe Aussage getroffen werden. Daher wird in diesem Zusammenhang auch gern von intrapersonaler Reliabilität gesprochen. Diese Form der Konsistenz ist zwar in Bezug auf Messergebnisse ein wichtiges Kriterium, zur Beurteilung von Expertise jedoch unzureichend. Schließlich kann eine hohe Aussagenkonsistenz auch dann erreicht werden, wenn konstant falsche Einschätzung nach demselben Muster abgegeben werden (122). Dann sind zwar die Aussagen einer Person in sich konsistent. Sie sind jedoch auch konsistent falsch. Hinzu kommt ein Problem bezüglich der Praktikabilität dieses Kriteriums. Ob potenzielle Experten konsistente Einschätzungen abgeben, kann nur anhand von Wiederholungsbefragungen festgestellt werden. Dies würde letztlich zu einer Erhebung in der Erhebung führen, den Fokus vom eigentlichen Forschungsgegenstand ablenken und forschungsökonomisch schwer begründbar sein. Letzteres insbesondere deshalb, weil andere kognitive Fähigkeiten wie „**discrimination ability**“ als Kriterium herangezogen werden können. Wie bereits erläutert, erfordert dies eine genauere, feinere und detaillierte Problembeschreibung im Vergleich zum Laien (111). Experten erkennen und benennen Unterschiede demnach, die Nicht-Experten verborgen sind. Zudem ist es substantiell, dass sie die unterschiedlichen Aspekte auch unterschiedlich behandeln, beispielsweise die jeweils geeigneten Lösungsvorschläge machen. Dinge, die ähnlich, aber eben doch nicht gleich sind, können nur von Experten erkannt werden und es zeichnet einen Experten aus, dass er weiß, wie man diese feinen Unterschiede erkennt und weiß, wie sie zu behandeln sind: „Knowing *how* [Hervorhebung im Text] to combine information is of no value without knowing *what* [Hervorhebung im Text] information to combine“ (122). Da die hier verfolgten Fragestellungen eine zum Teil stark differenzierte Analyse einzelner Aspekte verlangt, stellen die an die Experten gerichteten Fragen gewissermaßen selbst ein Auswahlkriterium dar. Denn nur, wer sich imstande sieht, adäquate Antworten auf die Fragen zu geben, kann an der Expertenbefragung teilnehmen. Hier kommt ein weiteres, wenn auch nicht hinreichendes Kriterium, hinzu, nämlich die **Selbsteinschätzung**. Der Vorteil ist, dass dieses Kriterium ohne größeren Aufwand und ohne größere Kosten angewandt werden kann. Ein offensichtlicher Nachteil dieses Kriteriums ist, dass die Bewertung der eigenen Fähigkeiten nicht zwangsläufig mit den tatsächlichen Fähigkeiten korreliert. Einige Studien weisen jedoch darauf hin, dass selbsternannte Experten eine höhere Kompetenz aufweisen, als Personen, die sich selbst nicht als Experten sehen (111). Dennoch muss diese Form der Selbsteinschätzung mit großer Vorsicht behandelt werden und genügt allein nicht als Auswahlkriterium (110, 111).

Weniger subjektive oder auch weniger von Persönlichkeitseigenschaften abhängige Kriterien stellen **externale Eigenschaften** wie beispielsweise Erfahrung, berufliche Position und Akkreditierung dar. Diese können allesamt an weitestgehend objektiven Tatsachen bemessen

werden wie z. B. bestimmten Zertifikate und Abschlüssen, der Berufserfahrung gemessen in Jahren oder beruflichen Stationen, der Qualifikation bemessen an Titeln oder der Stellenbeschreibung, der wissenschaftliche Befähigung bemessen an der Zahl der Publikationen, der Mitgliedschaft in bestimmten Gremien und Expertenkreisen usw. (111, 122, 112). Shanteau et al. (2002) warnen jedoch davor, diese Kriterien überzubewerten. Zum einen sind sie häufig interkorreliert – so steigt in der Regel die Anzahl von Zertifikaten mit den Jahren der Berufserfahrung – und zum anderen sind die Zusammenhänge weder linear noch zwingend kausal. Mit den Jahren an Erfahrung steigt nicht unbedingt die Kompetenz und Inhaber des gleichen Zertifikats können sich erheblich in ihren Fähigkeiten unterscheiden (122). Der Vorteil solcher Kriterien ist, dass sie häufig leicht und für alle potenziellen Experten gleichermaßen verfügbar sind. Sie erleichtern somit die Identifikation von Experten und können den Auswahlprozess fairer und objektiver machen. Schließlich sollten die Kritikpunkte bezüglich ihrer Aussagekraft zwar berücksichtigt werden, allerdings kann davon ausgegangen werden, dass sie zumindest ein Mindestmaß an fachlichem Wissen garantieren. Denn einen bestimmten Abschluss erreicht man nicht ohne ein bestimmtes, meist sehr umfangreiches Maß an Wissen und in einer beruflichen Position kann man nur dann dauerhaft verbleiben, wenn die Qualifikationen eine ausreichend zufriedenstellende Leistung ermöglichen. Daher eignen sich externale Eigenschaften besonders dort, wo sich das Wissen und die Fähigkeiten von Experten nicht anderweitig bemessen und beurteilen lässt. Der oben erwähnte, fehlende Goldstandard kann damit zwar nicht ersetzt werden, aber Erfahrung und Akkreditierung garantieren zumindest ein Mindestmaß an Eignung und sollen daher in das Set von Auswahlkriterien aufgenommen werden. Für den Kontext dieser Arbeit bedeutet das, dass Personen, die als Experte für die erste Gruppe, Expertenreview zur Evaluation des Fragebogens, ausgewählt werden, mindestens das Grundstudium in einem Fach abgeschlossen haben müssen, in dem empirische Methoden Bestandteil der Ausbildung sind, und selbst bereits praktische Erfahrungen bei der Durchführung von standardisierten Befragungen gesammelt haben müssen. Für die zweite Gruppe ist eine notwendige Bedingung für die Teilnahme an der Expertenbefragung, dass sie Berufserfahrung im Geschäftsbereich BMVg vorweisen und eine fachliche Qualifikation im Bereich Arbeits- und Gesundheitsschutz, Arbeitsmedizin, Betriebliches Gesundheitsmanagement, Psychologie oder eine vergleichbare Befähigung bezüglich Gesundheitsschutz und arbeitsbedingten psychischen Belastungen aufweisen.

Auch der **Konsens zwischen Experten** wird häufig als Reliabilitätsmaß (consensus reliability, between-expert reliability) angeführt (122, 111). Dabei handelt es sich zwar um ein notwendiges, jedoch ebenfalls um kein hinreichendes Kriterium. Die zugrundeliegende Annahme ist, dass Experten einer bestimmten Wissensdomäne in ihren Meinungen und Einschätzungen übereinstimmen sollten. Ist das nicht der Fall, so ist dies ein Indikator dafür,

dass wenigstens einige der Personen keine Experten sind (122). Dieses Postulat ist dort zutreffend, wo klar zwischen Richtig und Falsch unterschieden werden kann. In den meisten Bereichen, in denen Methoden der Expertenbefragung eingesetzt werden, besteht dieses Wissen jedoch (noch) nicht, sondern soll eben durch die Beteiligung von Experten generiert werden. In solchen Fällen zeigt sich die gleiche Problemstellung wie bezüglich der inneren Konsistenz: Es ist durchaus möglich, dass Konsens zwischen allen Experten besteht, aber dass dieser Konsens dennoch falsch oder irreführend ist (110). Die Konsistenz von Aussagen, sowohl von einer Person zu mehreren Zeitpunkten als auch von mehreren Personen im Rahmen einer Befragung, sagt also nichts über die Richtigkeit oder Verlässlichkeit von diesen Aussagen aus. Umgekehrt gilt jedoch, dass Konsens eine notwendige Bedingung ist, um sicher zu sein, dass die Einschätzungen überhaupt richtig sein können. Dissens wäre ein klares Indiz dafür, dass einzelne Expertenmeinungen nicht oder nicht vollständig zutreffend sind (110). Wendet man außerdem mehrere Kriterien für die Expertenauswahl gleichzeitig an, ist davon auszugehen, dass angemessen kompetente Personen identifiziert werden und ein Konsens durchaus als Orientierung zur Beantwortung der Fragestellung herangezogen werden kann. Studien konnten zudem zeigen, dass konsensuale Gruppenergebnisse bessere Ergebnisse erbrachten als Einzelmeinungen von qualifizierten Experten (111). Zudem können konsensorientierte Verfahren Verzerrungseffekte anderer Auswahlverfahren ausbalancieren: „From a probabilistic perspective, averaging methods tend to outperform experts because they counterbalance a long list of social-influence processes and judgment biases that undermine accuracy“ (111). Im Rahmen dieser Arbeit soll daher bei Fragen die methodische Aspekte bzw. das Erhebungsdesign betreffen, ein Konsens innerhalb der Expertengruppe angestrebt werden. Bezüglich der inhaltlichen Ausgestaltung und der Relevanz einzelner Aspekte arbeitsbedingter psychischer Belastung wird ebenfalls dem Mittel aller Einschätzung mehr Gewicht als Einzelmeinungen beigemessen werden.

Zusammenfassend kann festgehalten werden, dass die Auswahl der Experten für das Fragebogenreview anhand der Kriterien Zuschreibung, persönliche Involviertheit, Selbsteinschätzung und externalen Eigenschaften erfolgte und ein konsensuales Ergebnis angestrebt wurde. Anhand dieser Kriterien wurde eine Gruppe von wissenschaftlichen Mitarbeitenden des ASU identifiziert. Anhand ihrer Ausbildung und beruflichen Erfahrung konnte von einer angemessenen Qualifikation ausgegangen werden. Die persönliche Involviertheit war durch die Nähe zum Projekt gegeben. Dies barg den zusätzlichen Vorteil, dass sie gut mit dem Anwendungszweck, dem Hintergrund des AIGScreen und der zugrundeliegenden Theorie vertraut waren und gleichzeitig ein bezüglich des Anwendungsgebiets erfahrenes und interdisziplinär aufgestelltes Team darstellten, welches kurzfristig und niedrigschwellig Rückmeldung geben konnte. Ferner erlaubten Vorgaben des

Datenschutzes keine Weitergabe von Daten an Dritte, weshalb diese unter Umständen mit unvollständigen Informationen hätten arbeiten müssen.

Als geeignete Expertengruppe für Befragung zur inhaltlichen Vollständigkeit des Fragebogens konnten Mitarbeitenden im Geschäftsbereich BMVG identifiziert werden, sofern sie über zusätzliche externale Eigenschaften verfügten, die ihre fachliche Qualifikation bestätigte oder mittels Zuschreibung durch andere Beschäftigte als solche Experten empfohlen wurden. Ein hohes Maß an persönlicher Involviertheit konnte aufgrund der unmittelbaren Betroffenheit durch das Forschungsvorhaben auch für diese Gruppe angenommen werden.

3.3.3 Zielsetzung und Fragestellung

Da Konstruktvalidität nur dann als nachweislich gegeben betrachtet werden kann, wenn zuvor die Inhaltsvalidität überprüft und in diesem Rahmen Korrelationsmuster a priori definiert wurden (64), sollte eine Inhaltsvalidierung durch Expertenbefragungen erfolgen. Ziel war es dabei, den Fragebogen AIGScreen in seiner Konstruktion zu replizieren: Lassen sich die verwendeten Konstrukte gemäß dem derzeitigen Forschungsstand und der Praxis der Gefährdungsbeurteilung begründen und definieren? Sind mit den ausgewählten Konstrukten und Items alle relevanten Aspekte für ein Grobscreening abgedeckt? Repräsentieren die verwendeten Items die ihnen zugrundeliegenden Konstrukte ausreichend? Ferner sollte die intersubjektive Verständlichkeit sowie die Sinnhaftigkeit der einzelnen Items aus Sicht der Zielgruppe überprüft werden, was die Beteiligung derselben unerlässlich machte.

Da für die Inhaltsvalidität keine quantitativen Messverfahren existieren, verlangt Ihre Überprüfung ein Expertenurteil über die Angemessenheit des Verfahrens, d. h. ob ein Instrument das zu erfassende Merkmal vollständig misst bzw. ob die verwendeten Items das theoretische Konstrukt vollständig repräsentieren (123). Erste notwendige Bedingung dafür ist, dass eine Definition des zu erfassenden Merkmals vorliegt (64). Eine ausführliche Beschreibung und Begriffsbestimmung der mit dem AIGScreen zu messenden Konstrukte findet sich in Kapitel 3.1.2, die theoretischen Grundlagen dafür wurden in Kapitel 2.1 erläutert. Die Inhaltsvalidität theoretischer Konstrukte erfordert eine argumentative Begründung der systematischen Itemselektion, welche anhand einer Theorie über den Zusammenhang zwischen Antwortverhalten auf Itemebene und der Merkmalsausprägung im zu erfassenden Konstrukt erfolgt. Die theoretisch postulierten Zusammenhänge sollten sich auch später in strukturprüfenden Verfahren abbilden lassen (69).

Die Expertenbefragung schafft somit die Grundlage für anschließende Reliabilitätsanalysen. Hierfür muss nicht nur eine inhaltliche Vollständigkeit des zu messenden Konstrukts gegeben sein, sondern es muss auch die Fragebogenqualität wissenschaftlichen Standards

entsprechen und qualitätsgefährdende Aspekte müssen aus dem Erhebungsdesign ausgeschlossen sein. Entsprechend sollte der Fragebogen in einem zweiten Verfahren mit Expertenbeteiligung auch in seiner Eignung als Testverfahren als solches anhand allgemeiner Standards für die Testkonstruktion überprüft werden.

Da mehrere Fragestellungen verfolgt wurden, für die es unterschiedliche Expertengruppen brauchte, wurde eine Zweiteilung der Erhebung vorgenommen. Zuerst wurde ein internes Expertenreview des Instrumentes vorgenommen. Hier standen Qualitätskriterien für ein wissenschaftliches Erhebungsinstrument im Fokus der Analyse. Erst in einem zweiten Schritt, einer teilstandardisierten Expertenbefragung, wurde Fragen nach der inhaltlichen Vollständigkeit und der Augenscheinvalidität nachgegangen.

3.3.3.1 Ziel des Expertenreviews

Wie bereits eingangs erläutert, empfehlen die europäischen Statistikbehörden den Einsatz von Expertengruppen zur Fragebogenevaluation, um dessen methodische und qualitative Eignung für den Einsatzzweck zu überprüfen (70). Diese Gruppen sollten neben Experten für das inhaltliche Thema und für Fragebogendesign auch Personen involvieren, die an der Datenverarbeitung und -auswertung beteiligt sind. Dies wurde erfüllt, indem alle diese Personengruppen unter den wissenschaftlichen Mitarbeitenden des ASU vertreten waren. Aufgabe der beteiligten Experten war es das Fragebogendesign hinsichtlich seiner Eignung für den geplanten Einsatzzweck zu beurteilen. Dafür sollte überprüft werden, ob es qualitätsgefährdende Aspekte bei den einzelnen Fragen oder Antwortskalen gab und ob die mit dem Instrument erhobenen Daten sich für die wissenschaftliche Auswertung eignen bzw. ob durch das Erhebungsdesign bedingte Probleme bei der Erhebung zu erwarten seien. Dafür wurde in Orientierung an den Empfehlungen von Brancato et al. (2006) sowie Faulbaum et al. (2009) eine Checkliste zur Überprüfung der Fragequalität entwickelt (70, 1). Die vollständige Liste ist in Anlage 2 zu finden. Wesentliche zu bewertende Aspekte waren:

- Wortlaut und Verständlichkeit der Fragen
- Verständlichkeit von Anweisungen und Hilfstexten
- Fragebogenstruktur und potenzielle Effekte durch die Fragenreihenfolge oder andere Kontexteffekte
- Layout: Übersichtlichkeit, Intuitivität, Ergonomie, Navigation
- Angemessenheit der kognitiven Anforderungen und Vorannahmen über die Zielgruppe
- Orthografie und Grammatik
- Antwortoptionen: Vollständigkeit, Eignung, Trennschärfe, Eindeutigkeit, Eignung der Unzutreffend-Kategorie
- Soziale Erwünschtheit und Suggestion

Bezüglich des ersten Punktes, **Wortlaut und Verständlichkeit der Fragen**, galt es kritisch zu überprüfen, ob die Formulierungen und eventuell gewählte Fachbegriffe für die Zielgruppe verständlich und eindeutig in ihrer Bedeutung sind (30). Würden Fragen unterschiedlich verstanden oder interpretiert, würde dies erheblich die Reliabilität und Validität der erhobenen Daten gefährden, da somit interindividuell unterschiedliche Aspekte erhoben werden würden. Ebenso wichtig ist in diesem Zusammenhang die **Verständlichkeit von Anweisungen und Hilfstexten**, da Probleme auch hier zu variierendem Antwortverhalten führen könnten. Darüber hinaus können schwer verständliche Anweisungen die Probanden schnell überfordern und frustrieren und somit die Abbruchquote erhöhen. Dies trifft auch zu, wenn die **Angemessenheit der kognitiven Anforderungen und Vorannahmen über die Zielgruppe** nicht adäquat sind. Es ist wichtig, dass ein Fragebogen die Befragten in ihrer Lebenswelt „abholt“. Das heißt, es sollten grundsätzlich nur Fragen gestellt werden, die die Befragten auch beantworten können, weil sie bereits Berührungspunkte mit der Thematik hatten, entsprechende Erfahrungen machen konnten oder sich eine Meinung bilden konnten (124). Gleichzeitig ist es wichtig, die Befragten nicht zu unterfordern oder zu langweilen.

Das **Layout** eines Fragebogens, egal ob als Papierversion oder im Online-Format, hat einen erheblichen Einfluss auf das Antwortverhalten von Befragten und muss daher bestimmten Standards und wissenschaftlichen Erkenntnissen zu Darstellungseffekten gerecht werden (125, 124, 1, 25). Die Experten sollten im Rahmen des Reviews insbesondere darauf achten, ob sowohl bei der Online- als auch bei der Papierversion eine Übersichtlichkeit von Inhalten, Antwortskalen und Frageblöcken gegeben ist, ob insbesondere bei der Online-Erhebung die Darstellungsform (z. B. Schriftgröße, Farbkontraste) ergonomischen Ansprüchen genügen, ob die Orientierung und Navigation in der Befragung leichtfällt und intuitiv möglich ist. Beispielsweise wurde in Abschnitt 3.1.3 „Erhebungsdesign, Frageformat und Antwortskalen“ darauf hingewiesen, dass im europäischen Raum die gewohnte Lesrichtung und Polung von Negativ zu Positiv von links nach rechts erfolgt. In diesem Zusammenhang ist es auch wichtig die **Fragebogenstruktur und potenzielle Effekte durch die Fragenreihenfolge oder andere Kontexteffekte** zu prüfen. So ist es beispielsweise möglich, dass die Beantwortung einer Frage durch vorausgegangene Fragen beeinflusst wird (1). Wird beispielsweise zunächst nach der körperlichen Aktivität gefragt und der Befragte verortet sich selbst in der niedrigsten Kategorie könnte dies dazu führen, dass er in einer Anschlussfrage zum individuellen Gesundheitsverhalten dieses negativer bewertet als er es normalerweise tun würde. Derartige Verzerrungsgefahren sollten von den Experten identifiziert werden. Auch **Soziale Erwünschtheit und Suggestion** können das Antwortverhalten beeinflussen bzw. verzerren und sollten nach Möglichkeit unterbunden werden. Während ersteres bei heiklen Frage nicht ausgeschlossen werden kann, sind Suggestivformulierungen in jedem Fall zu vermeiden (124). Wie **Antwortoptionen** idealerweise gestaltet sein sollten, wurde bereits in Abschnitt

3.1.3 „Erhebungsdesign, Frageformat und Antwortskalen“ erörtert. Die Experten sind aufgefordert zu überprüfen, ob die Qualitätsstandards eingehalten werden oder Anpassungen an den Antwortskalen notwendig sind. Schließlich soll auch berücksichtigt werden, ob **Orthografie und Grammatik** fehlerfrei sind. Fehlerhafte Fragen könnten die Glaubwürdigkeit der Befragung gefährden und die Teilnahmebereitschaft senken.

Ziel war es neben der Bewertung der einzelnen, hier aufgeführten Aspekte einen Konsens zwischen den Experten zu erarbeiten. Dabei sollte Einigkeit sowohl bei etwaigen Änderungsbedarfen als auch bezüglich der Beibehaltung von Items bestehen.

3.3.3.2 Ziel der teilstandardisierten Expertenbefragung

Im Rahmen einer teilstandardisierten Befragung von Beschäftigten im Geschäftsbereich des BMVg soll die grundsätzliche Angemessenheit des Verfahrens, d. h. ob ein Instrument das zu erfassende Merkmal vollständig misst bzw. ob die verwendeten Items das theoretische Konstrukt vollständig repräsentieren, überprüft werden.

Für die Bewertung der Inhaltsvalidität sind insbesondere zwei Aspekte von Interesse: die Relevanz der erfassten Merkmale und die vollständige Repräsentation der Skalen durch die verwendeten Items jeweils aus Sicht der Experten. Dieser Ansatz bietet die Möglichkeit nicht nur den Fragebogen, sondern auch die Empfehlungen der Institutionen im deutschen Arbeitsschutz aus Sicht der Experten zu evaluieren. Eine Diskrepanz zwischen dem Urteil der befragten Experten und offiziellen Empfehlungen, welche Aspekte bezüglich der Gefährdungsbeurteilung psychischer Belastung erfasst werden sollten, ist nicht auszuschließen. Erkenntnisse können dazu dienen, bisher bestehende Lücken zwischen offiziellen Empfehlungen und den Anforderungen bei der praktischen Umsetzung aufzuzeigen. Ziel dieser Untersuchung ist daher vor allem, der Frage nachzugehen, inwiefern bestehende Empfehlungen aus Praxissicht akzeptiert bzw. gleichermaßen als wichtig wahrgenommen werden. Anhand der Ergebnisse können Handlungsempfehlungen für die Umsetzung von Bedarfsanalysen und psychischen Gefährdungsbeurteilungen sowie ferner auch für die Zusammensetzung von Expertengruppen bzw. Gesundheitsgremien in der betrieblichen Praxis abgeleitet werden. Dafür wurden Beschäftigte im Geschäftsbereich BMVg in ihren verschiedenen Rollen im Bereich Arbeits- und Gesundheitsschutz und Gesundheitsförderung gebeten, ihre Einschätzung abzugeben, inwiefern sie ihre Situation und Bedarfe im Fragebogen abgebildet sehen. Die zugrundeliegenden Fragestellungen waren,

- ob sich der Fragebogen aus Sicht der Befragten für eine Bedarfsanalyse eignet und Handlungsschwerpunkte aus den Ergebnissen abgeleitet werden können,

- ob aus Sicht der Befragten alle wesentlichen Aspekte der arbeitsbedingten psychischen Belastungen erfasst werden und
- ob Formulierungen und Inhalt den spezifischen Rahmenbedingungen des betreffenden Arbeitssettings gerecht werden.

Da es den Befragten grundsätzlich ermöglicht werden sollte hinsichtlich aller Aspekte (sowohl inhaltlich als auch in Bezug auf Wortlaut und Formulierungen) Anmerkungen und Verbesserungsvorschläge zu äußern, war eine vollständige Standardisierung der Befragung nicht möglich. Da es gleichzeitig Ziel der Befragung war, den Durchschnitt der Expertenurteile zu ermitteln, um möglichen Verzerrungen durch Einzelmeinungen entgegenzuwirken und sich dem Wahrheitswert mehr anzunähern, wie in Abschnitt 3.3.2 erläutert, wurde die Befragung jedoch wo immer möglich in geschlossenen Antwortformaten durchgeführt. Dort, wo es möglich und angemessen war, sollten die Ergebnisse quantifizierbar sein.

Die Augenscheinvalidität stellt, wie in Abschnitt 2.3.3 erläutert, kein wissenschaftliches Gütekriterium im eigentlichen Sinn dar und wird daher nicht standardmäßig im Rahmen von Fragebogenvalidierungen überprüft. Allerdings kann davon ausgegangen werden, dass es insbesondere bei der Durchführung von freiwilligen Befragungen ein wichtiges Kriterium darstellt. Denn wenn Personen freiwillig teilnehmen sollen, müssen sie für sich einen individuellen Nutzen bzw. einen Sinn in der Teilnahme erkennen können. Insofern schien die Berücksichtigung der Augenscheinvalidität aus Sicht der Zielgruppe sinnvoll. Da die Expertengruppe für die teilstandardisierte Befragung sich ausschließlich aus Vertretern der Zielgruppe zusammensetzt und diese die Relevanz der Items bewerten sollen, wird die Augenscheinvalidität miterfasst.

3.3.4 Internes Expertenreview zur Fragebogenqualität in zwei Phasen

Im vom statistischen Arbeitsprogramm (SPC) der Europäischen Kommission herausgegebenen *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System* (70) wird empfohlen Expertenreviews in strukturierter Form durchzuführen. Im Sinne einer besseren Vergleichbarkeit der einzelnen Einschätzungen sollte demnach ein durch Checklisten standardisiertes Experteninterview, welches die sieben Bereiche Leseschwierigkeiten, Anleitungsschwierigkeiten, Probleme der Eindeutigkeit, Probleme der Logik, Erinnerungsschwierigkeiten, Probleme hinsichtlich Sensitivität oder Verzerrung und Probleme bei den Antwortkategorien abdeckt, durchgeführt werden. Die Standardisierung der Erhebung erfolgte durch die itemweise anzuwendende Checkliste in Anlage 2. Hier sollten die Experten für jedes Problem, dass ihnen bezüglich einer

Frage auffiel einen entsprechenden Eintrag machen. Diese Liste diene der Vorstrukturierung der anschließenden Gruppendiskussion zur Konsensfindung.

Dieses zweistufige Vorgehen wurde einerseits gewählt, um Ressourcen zu sparen. So konnte man beispielsweise auf die Diskussion in der Gruppe verzichten, wenn bereits im Vorfeld von allen Experten unabhängig voneinander kein Problem oder Änderungsvorschlag bzgl. eines Items gemacht wurde. Andererseits konnte durch die Alleinarbeit im ersten Schritt die gegenseitige Beeinflussung vermieden werden. Es wurde davon ausgegangen, dass die Arbeit allein mit einer Checkliste zunächst zu genaueren Ergebnissen führt, während man in der Gruppendiskussion dazu tendiert, einzelne Gedanken oder Anmerkungen als weniger wichtig und erwähnenswert zu empfinden, sodass diese eventuell nicht in den Arbeitsprozess einfließen können. Letzteres könnte zudem auch durch Ermüdungseffekte begünstigt werden. Durch die Alleinarbeit im ersten Schritt konnten die Experten sich hingegen eine unabhängige und gefestigte Meinung bilden und sich die Arbeitszeit frei einteilen bzw. bei Bedarf Pausen einlegen.

Aspekte, die von einem oder mehreren Experten als problematisch identifiziert wurden, wurden anschließend in der Gruppe diskutiert und sich auf eine Lösungsstrategie verständigt. Methodisch heikel ist dabei, dass Expertenreviews in der Regel v. a. einzelne Fragen und nicht den gesamten Kontext betrachten. Dieses Zusammenspiel, der Fragefluss und Gesamtkontext des Fragebogens müssen jedoch bei der Bewertung des Instruments ebenfalls berücksichtigt werden (70). Hierfür war das Gruppengespräch, in dem der gesamte Fragebogen evaluiert werden sollte, gut geeignet, da einzelne Kritikpunkte stets im Gesamtkontext des Instruments und der Anmerkungen aller Experten dazu diskutiert wurden.

3.3.5 Durchführung der teilstandardisierten Expertenbefragung zur Inhaltsvalidität

Die Teilnahme an der Mitarbeitendenbefragung AIGScreenBw war, wie in Kapitel 3.2 erläutert, freiwillig. Den Teilnehmenden entstand durch die Teilnahme kein Vorteil bzw. es wurden keine zusätzlichen Anreize zur Befragungsteilnahme gegeben. Mitarbeitende sollten aus Eigeninteresse an der Thematik teilnehmen. Für eine ausreichende Rücklaufquote ist somit eine hohe Akzeptanz der Erhebung bei der Zielgruppe, die sogenannte Augenscheinvalidität, notwendig (30). Diese kann nur von der Zielgruppe selbst bewertet werden. Aus diesem Grund sollte die Inhaltsvalidität des AIGScreens in Form von Experteninterviews mit Beschäftigten im Geschäftsbereich des BMVg überprüft werden. Die Datenerhebung fand im Rahmen des Forschungsprojekts zum BGM im Geschäftsbereich des BMVg im Zeitraum 22. Juni bis 16. August 2019 statt. Da es sich um einen Arbeitgeber mit sehr unterschiedlichen Tätigkeitsfeldern handelt, welche neben dem militärischen Bereich z. B. auch Verwaltung, Krankenhäuser, IT, Logistik oder Forschungseinrichtungen umfassen und zudem eine Vielzahl

von unterschiedlichen Standorten in ganz Deutschland sowie einigen Auslandsstationierungen existieren, sollte die Teilnahme an der Befragung allen Experten im Bereich Arbeits- und Gesundheitsschutz, die im Geschäftsbereich BMVg tätig sind, ermöglicht werden, um ein möglichst repräsentatives Bild nicht nur der verschiedenen Professionen sondern auch der unterschiedlichen Arbeitssettings zu erhalten. Als Experte gilt in diesem Zusammenhang jede Person, die detailliertes oder spezielles Wissen in einem spezifischen und abgrenzbaren Bereich hat (126). Dieses Kriterium kann für alle in Abschnitt 3.3.2 genannten Berufsgruppen (Arbeitsmedizin, Gesundheitsmanager, Fachkräfte für Arbeitssicherheit etc.) im Geschäftsbereich des BMVg als gegeben betrachtet werden.

Die große Anzahl von Experten im gewählten Setting machte die Wahl der Erhebungsmethode bzw. des Fragebogendesigns zu einer besonderen Herausforderung. Üblicherweise empfiehlt sich für die Gewinnung zusätzlicher Informationen und der Exploration bisher nicht oder wenig erforschter Bereiche ein offenes, d. h. unstandardisiertes Erhebungsformat (127). Allerdings erschien eine vollständig unstandardisierte Erhebung mit Blick auf eine potentielle Stichprobengröße von circa 1.000 Teilnehmenden⁶ nicht geeignet bzw. wäre das dadurch generierte Textmaterial kaum in einem angemessenen Zeitraum auswertbar gewesen. Brancato et al. (128) empfehlen im Sinne einer besseren Vergleichbarkeit ein standardisiertes Experteninterview. Darüber hinaus bestanden zum Zeitpunkt der Erhebungsplanung bereits einige Erkenntnisse und Vorüberlegungen, die sich zum Teil aus der Praxiserfahrung und zum Teil aus dem internen Expertengutachten zur Validität des Instruments ergaben und die in die Gestaltung der Experteninterviews einfließen sollten. Viertens sollte im Rahmen der Befragung geschlossene Antwortformate zur Bewertung der Relevanz der Items eingesetzt werden, um den CVI (content validity index) berechnen zu können. Aus diesen Gründen wurde ein teilstandardisiertes Vorgehen gewählt. Der Fragebogen hat bereits ein zu geschlossenes Format, um noch als Leitfadeninterview zu rangieren, wie es klassischerweise bei Experteninterviews eingesetzt wird. Aber er erfüllt durchaus das Kriterium der „geschlossenen Offenheit“ (117). Das bedeutet, dass alle Probanden zunächst in geschlossenen Antwortformaten die einzelnen Fragen beurteilen sollten (z. B. „Bitte bewerten Sie die Relevanz des jeweiligen Items für diesen Merkmalsbereich“ auf einer Antwortskala von „1 – Nicht relevant“ bis „6 – Höchst relevant“) und sich an einige geschlossene Fragen (z. B. „Ist die Definition des Merkmalsbereichs Ihrer Meinung nach vollständig?“ mit den Antwortoptionen „Ja“ und „Nein“) die Möglichkeit einer Freitextangabe anschloss (z. B. „Welche weiteren Items würden Sie erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich relevant?“), wo die Experten ihre Antwort begründen sollten. Somit wurde den Kriterien der

⁶ Dieser Richtwert entspricht Angaben der Bundeswehr. Die genaue Zahl ließ sich nach Auskunft zu keinem Stichtag aufgrund von Fluktuationen in den involvierten Gremien, Berufungen und Stellenbesetzungen ermitteln.

Offenheit und Flexibilität von Expertenbefragungen (126) Rechnung getragen und den Experten genügend Raum gegeben, ihre eigenen Gedanken und Vorschläge zu formulieren. Um eine Aussage darüber treffen zu können, inwiefern sich die Einschätzung von Experten aus der betrieblichen Praxis mit allgemeinen Empfehlungen der Arbeitsschutzinstitutionen deckt bzw. um Unterschiede zu identifizieren, wurde den Experten eine Beschreibung der vier Hauptmerkmalsbereiche vorgelegt, die laut BAuA (129, 97) und GDA (130) im Rahmen einer Gefährdungsbeurteilung psychischer Belastung zu erfassen sind. Diese sind die in Kapitel 3.1.2 beschriebenen Subdimensionen Arbeitsumgebung, soziale Aspekte, Arbeitsinhalt und Arbeitsaufgabe sowie Arbeitsorganisation. Der vollständige Fragebogen findet sich in Anlage 3. Da die Befragung ausschließlich online durchgeführt wurde, handelt es sich bei der Darstellung im Anhang ausschließlich um einen inhaltlichen Entwurf. Auf eine Paper-Pencil-Befragung wurde verzichtet, da alle Beschäftigten in der Zielgruppe einen Computerarbeitsplatz hatten.

Der eingesetzte Fragebogen unterteilte sich in drei Blöcke. Der erste Block bestand aus einer offenen Frage zu relevanten Aspekten bezüglich des Einsatzzweckes des AIGScreen aus Sicht der Befragten. Diese Frage wurde zuerst gestellt, um Priming-Effekte durch andere Frageinhalte und Kontexteffekte zu vermeiden (75). Im zweiten Block wurde die Relevanz und Vollständigkeit des AIGScreen-Kernfragebogens beurteilt. Hierbei wiederholten sich dieselben sieben Fragen für die jeweiligen vier Subdimensionen arbeitsbedingter psychischer Belastung. Dabei wurden Relevanz der einzelnen Items und die Vollständigkeit der Subskala zunächst in einem geschlossenen Antwortformat beurteilt. Daran schlossen sich vier dichotome (Ja / Nein) Fragen zu Vollständigkeit und Umfang der Subskalen und deren Definition an. Wenn die Probanden Änderungsbedarf hinsichtlich eines Aspekts sahen, konnten sie diesen genauer in einem Freitextfeld erläutern. Abgeschlossen wurden alle vier Abschnitte mit einer offenen Frage für weitere Anmerkungen.

Im dritten Block wurden Informationen zum Tätigkeitsbereich und der Qualifikation der Befragten erhoben. Alle Angaben sowie die Teilnahme insgesamt waren freiwillig. Da die Bearbeitung als sehr zeitaufwändig und auch kognitiv anspruchsvoll eingeschätzt wurde, wurde den Teilnehmenden die Möglichkeit des Zwischenspeicherns eingeräumt. Im letzten Drittel des Befragungszeitraums wurde eine Erinnerungsemail an alle Personen in der Zielgruppe verschickt. Deren Effektivität lässt sich jedoch nicht bemessen, da aufgrund der technischen Einstellungen zur Datenschutzkonformität bei der Möglichkeit des Zwischenspeicherns keine Aussagen über Beginn und Ende der Teilnahme gemacht werden können.

Da ein erklärtes Ziel der Expertenbefragung die Quantifizierung der Ergebnisse war, wurden die Antwortskalen so formuliert, dass der von Lawshe (1975) entwickelte Index zur

Relevanzbemessung einzelner Items, der „Content Validity Ratio (CVR)“ (131), später in der Literatur häufiger als „content validity index (CVI)“ bezeichnet (132–135), ermittelt werden konnte. Dieser Index wurde entwickelt, um in Situationen, in denen unter Experten kein Konsens bezüglich der Eignung eines Items besteht, anhand des Durchschnitts aller Urteile den Grad der Inhaltsvalidität zu quantifizieren (131). Dafür müssen alle befragten Experten auf einer standardisierten Ratingskala die Relevanz des Items bewerten. Hierbei werden häufig vierstufige Skalen eingesetzt, die für die spätere Berechnung des CVR bzw. CVI dichotomisiert werden (133, 134). Lawshe (1975) stellte folgende Formel für die Berechnung des CVR auf:

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

Formel 1: Content Validity Ration nach Lawshe (131).

Dabei ist n_e die Anzahl der Experten, die das Item als relevant einstufen und N beschreibt die Gesamtzahl aller befragten Experten (131). Der Wertebereich des CVR bzw. CVI reicht somit von – 1 bis 1. Negative Werte implizieren, dass die Mehrheit der Befragten das Item für nicht Relevant hält, positive Werte ergeben sich, wenn die Mehrheit das Item als relevant bewertet hat und der CVR beträgt 0, wenn jeweils genau die Hälfte ein positives oder negatives Urteil abgegeben hat (131). Dieses Prinzip kann nicht nur für den Aspekt der Relevanz, sondern auch auf andere Qualitätskriterien von Items wie bspw. Widerspruchsfreiheit und Verständlichkeit angewandt werden (135). Für die Befragung der Beschäftigten im Geschäftsbereich BMVg wurde die Methodik daher nicht nur für die Relevanz der einzelnen Items, sondern auch für die Frage nach vollständiger Repräsentation der Subskalen durch die Items angewandt. Anstelle einer vierstufigen, wurde eine sechsstufige Antwortskala gewählt, da nachgewiesen werden konnte, dass Befragungsteilnehmende mit steigender Zahl der Antwortkategorien das Gefühl haben, ihre Meinung besser ausdrücken zu können (136). Gleichzeitig stellt eine sechsstufige Skala keine Überforderung für die Probanden dar. Auf die Mittelkategorie wurde weiterhin verzichtet, da es zum einen der Methodik des CVR bzw. CVI entspricht und im vorliegenden Fall auch leichte Meinungstendenzen der Befragten erfasst werden sollten, wofür eine Mittelkategorie hinderlich wäre (124). Da Wertungen erhoben werden sollten, war es zudem besonders wichtig, die Antwortoptionen horizontal darzustellen, um Primacy-Effekte, also eine Tendenz zur Beantwortung der oberen Antwortkategorien, zu vermeiden (137).

Der CVR bzw. CVI kann sowohl auf Item- als auch auf Skalenebene angewandt werden. Gilbert & Prion (2016) schlagen vor, Lawshe’s Content Validity Ratio CVR als Bezeichnung für die Itemebene zu verwenden und von Content Validity Index CVI zu sprechen, wenn eine ganze Skala bewertet werden soll (132). Diese begriffliche Unterscheidung soll für die weiteren

Erläuterungen übernommen werden, um eine eindeutige sprachliche Abgrenzung zu ermöglichen⁷. Ein wesentlicher Kritikpunkt an beiden Maßzahlen ist, dass bei der Berechnung die Wahrscheinlichkeit für zufällig positive Urteile nicht berücksichtigt wird. So liegt die bloße Wahrscheinlichkeit, dass die Experten ein Item als relevant bewerten bei einer zufälligen Verteilung der Antworten bereits bei 50% (133). Ayre & Scally (2014) schlagen in ihrer Veröffentlichung daher zur Bestimmung geeigneter Grenzwerte für den CVR eine Korrektur um die jeweiligen Wahrscheinlichkeitswerte der Antwortverteilungen vor (138). Der Binomialverteilung folgend ergeben sich dabei je nach Stichprobengröße unterschiedliche Grenzwerte. Der Wert, den der korrigierte CVR mindestens erreichen muss, sinkt dabei mit zunehmender Anzahl der Experten.

Ein praktikableres Vorgehen, das einen stichprobengrößenunabhängigen Grenzwert verwendet, greift auf das Reliabilitätsmaß Kappa zurück. Dieser Koeffizient wird häufig zur Bestimmung der Inter-coder-Übereinstimmung bei qualitativen Inhaltsanalysen verwendet (139). Auch hier erfolgt eine Korrektur um zufällige Übereinstimmungen: „Rather than comparing the total proportion of agreements (P_o) to a maximum value of 100%, the total P_o is compared to a maximum possible value that accounts for agreements occurring by chance alone ($1 - P_e$), given the marginal distribution of item ratings assigned by each expert panelist“ (133). Die Berechnung des CVI stellt dabei ein mehrstufiges Vorgehen dar, da zunächst stichprobengrößenadäquate Parameter für die Korrekturformel ermittelt werden müssen. In einem ersten Schritt wird die Wahrscheinlichkeit zufälliger Zustimmung unter der Annahme von binomialverteilten Ergebnissen ermittelt:

$$p_c = \left[\frac{N!}{A! (N - A)!} \right] \cdot 5^N$$

Formel 2: Wahrscheinlichkeit zufällig positiver Antworten (140).

N entspricht dabei der Anzahl an Experten und A der Anzahl der Antworten, die die Relevanz als hoch bewerten (140). In einem zweiten Schritt kann Kappa als Maß der Expertenübereinstimmung berechnet werden:

$$k = \frac{CVR - p_c}{1 - p_c}$$

Formel 3: Übereinstimmungskoeffizient Kappa (140).

⁷ Es sei jedoch darauf verwiesen, dass diese Begriffsunterscheidung keinen Konsens in der Fachliteratur darstellt und in anderen Veröffentlichungen andere Begriffe gewählt werden bzw. CVR und CVI gelegentlich synonym verwendet werden.

CVR entspricht dabei dem Anteil der Antworten, die das Item als relevant bewerteten. Für den hier eingesetzten Fragebogen entspricht das den Antworten 4 bis 6 („4 – Ziemlich relevant“, „5 – sehr relevant“, „6 – Höchst relevant“). Werte über 0,78 gelten als „excellent“ (140). Hier liegt die Wahrscheinlichkeit einer nicht zufälligen Übereinstimmung in einem 95%-Konfidenzintervall.

Für die Bewertung der vier Subskalen wurden die CVR-Werte der jeweils zugehörigen Items addiert und durch die Anzahl aller Items der Skala dividiert. Hier gilt ein strengerer Grenzwert von 0,90 für die Gesamtskala (140). Der Vorteil bei diesem Vorgehen ist, dass anhand der CVR auf Itemebene erkannt werden kann, welche Items zu einer Verschlechterung des Mittelwerts führen. Polit et al. (2007) empfehlen bei Items mit einem CVR niedriger als 0,78 eine Überarbeitung zu erwägen und Items mit einem sehr niedrigen CVR zu entfernen (140).

Neben den geschlossenen Antworten konnten im Rahmen der Befragung viele Freitextantworten gegeben werden. Das entstandene Textmaterial wurde mittels quantitativer Inhaltsanalyse ausgewertet, welche eine „klassische Auswertungsmethode für offene Fragen“ (32) darstellt. Diese Methode ermöglicht eine hohe Abstraktion und Quantifizierung von Aussagen. Der Unterschied zwischen qualitativer und quantitativer Inhaltsanalyse liegt letztlich ausschließlich in der Verwendung klassifizierender oder metrischer Begriffe. Die zugrundeliegende Methode, also das *qualitative* inhaltsanalytische Vorgehen, ist bei beiden Verfahren gleich; jeder quantitativen geht eine qualitative Inhaltsanalyse voraus (141). Die durch die Textcodierung generierten Kategorien konnten für die quantitative Analyse verwendet werden, sodass Häufigkeitsauszählungen möglich waren, um eine Aussage darüber zu treffen, welche Aspekte besonders häufig von den Befragten genannt wurden.

Um die Objektivität und Reliabilität der Ergebnisse zu garantieren, wurden zu Beginn der Analyse ein Ablaufmodell und ein Kategorienschema festgelegt. Ersteres definierte die einzelnen Analyseschritte und die Fragestellung (141), letzteres gibt vor, nach welcher Systematik das Material zu codieren ist (142). Um die Codierung möglichst reliabel zu gestalten, wurde ein Drittel des Textmaterials von zwei unabhängigen Personen nach dem gleichen Kategorienschema codiert und anschließend die Inter-coder-Übereinstimmung ermittelt (143). Die Inter-coder-Übereinstimmung wurde als relativer Anteil der Codes, die übereinstimmend für eine Textpassage gegeben wurden, im Verhältnis zur Gesamtzahl der vergebenen Codes bemessen. Diese Maßzahl kann zu einer Überschätzung der Inter-coder-Reliabilität (ICR) führen (35), da hier keine Korrekturformel für zufällige Übereinstimmungen angewandt wird, wie es z. B. beim alternativen Maß Kappa der Fall ist. Letzteres hat jedoch den Nachteil, dass es nur angewandt werden kann, wenn für Textpassagen genau ein Code verwendet werden kann (26). Da dies beim vorliegenden

Datenmaterial nicht der Fall war⁸, konnte Kappa in diesem Fall nicht als Maß der InterCoderübereinstimmung verwendet werden. Die InterCoderübereinstimmung kann gleichzeitig als Maß der Objektivität und der Reliabilität gesehen werden (144). Um beiden Kriterien gerecht zu werden und eine möglichst hohe InterCoderübereinstimmung zu erreichen, sollte das Kodierschema gut definiert und in einem iterativen Prozess von den Codierenden entwickelt werden. Wenn ein Kodierschema auf eine Stichprobe angewandt wurde und eine hohe ICR erreicht hat, ist es zu rechtfertigen, dass das Schema auf das restliche Material nur von einem Codierer angewandt wird, um Ressourcen zu sparen (143).

Kategorienschemata zur Codierung von Freitexten sollten, wenn sie eine hohe ICR ermöglichen sollen, bestimmte Kriterien erfüllen (32). Zunächst sollten die Kategorien mit den Zielen der Erhebung korrespondieren. Um dies zu gewährleisten, wurde für jede Frage im Fragebogen ein eigenes Set an Kategorien gemeinsam von den Codierenden festgelegt. Weiterhin sollte das Kategoriensystem vollständig sein, d. h. es sollte den durch die Forschungsfrage vorgegebenen Bedeutungsaspekt sowohl auf der Begriffsebene als auch auf der Ebene des Datenmaterials vollständig erfassen. Ersteres wurde ebenfalls durch das eben beschriebene Vorgehen erreicht, zweiteres wurde als gegeben betrachtet, als die Codierenden keine neuen Kategorien mehr zum Verschlagworten der Texte benötigten. Da der Inhalt der Antwortmöglichkeiten potenziell alles betreffen konnte, wurden hierfür keine zeitlichen und mengenmäßigen Obergrenzen definiert, da bis zur letzten Textpassage mit einer neuen Kategorie gerechnet werden musste. Allerdings zeigte sich retrospektiv, dass der Möglichkeitsraum potenziell notwendiger Kategorien zum Zeitpunkt der Überprüfung der InterCoderübereinstimmung bereits nahezu ausgeschöpft war und danach nur noch sehr vereinzelt neue Kategorien hinzukamen. Ein drittes Kriterium verlangt sogenannte Ankerbeispiele, also Beispiele aus den gewonnenen Daten, an denen für alle Codierenden deutlich wird, welche Aussagen der entsprechenden Kategorie zuzuordnen sind (32). Die Ankerbeispiele wurden im Rahmen der ersten Phase der gemeinsamen Codierung gesammelt und für zukünftige Texte in einer Tabelle unter dem entsprechenden Schlagwort gespeichert. Damit konnte auch einem weiteren Kriterium, dem der Eindeutigkeit von Kategorien, begegnet werden. Wenn die Zuordnung einer Textpassage zu einer bestimmten Kategorie nicht ganz eindeutig war, konnte anhand der Ankerbeispiele überprüft werden, ob eventuell bereits inhaltsähnliche Aussagen unter dieser Kategorie subsumiert wurden.

⁸ Es kam häufiger vor, dass einzelne Freitextantworten Aussagen zu mehreren Aspekten enthielten, die entsprechend mehreren Kategorien zugeordnet werden mussten. Zum Beispiel wurde der Hinweis gegeben: „Die Rückmeldung zu meinen Arbeitsergebnissen sollte aufgeführt sein“. Dies betrifft sowohl die Feedbackkultur bzw. Rückmeldung im Allgemeinen als auch den Führungsstil bzw. die Beziehung zum Vorgesetzten.

In Anlehnung an Mayring & Fenzel (2014) wurde das in Abbildung 6 dargestellte Ablaufmodell entwickelt (144). Zunächst galt es – wie bei jeder wissenschaftlichen Erhebung – zunächst den Forschungsgegenstand klar einzugrenzen, indem das Ziel der Analyse und die verfolgten Fragestellungen formuliert wurden. Die Erläuterungen dazu finden sich in Abschnitt 3.3.3.2. Eine Differenzierung der Fragestellung ergab sich im Rahmen der Entwicklung des eingesetzten Fragebogens (vgl. Abschnitt 3.3.5). Die Struktur und inhaltliche Gliederung des Fragebogens gab gleichsam die Orientierung für die laut Mayring & Fenzl (2014) zu bestimmenden Analyseeinheiten (144). Die Kodiereinheit bildet jedes Wort oder eine Wortgruppe, welches einer einzigen konkreten Bedeutung zugeordnet werden kann bzw. auf eine Aussage, reduziert werden kann, die als Antwort auf die jeweilige Fragestellung zu verstehen ist. Die Kontexteinheit ist dabei jedes Wort bzw. jede Wortgruppe, die sich zu genau einer Paraphrase zusammenfassen lässt. Als Auswertungseinheit wurde eine fallweise Auswertung bei möglicher Mehrfachkodierung festgelegt. Das Selektionskriterium wurde allgemein an der Zielsetzung der Erhebung ausgerichtet und durch keine weiteren Aspekte ergänzt. Dabei war in die Analyse einzubeziehen, was der unmittelbaren Beantwortung der jeweiligen Fragestellung dient. Die gemeinsame Entwicklung des Kategorienschemas durch die Codierenden sowie die nachfolgenden Schritte zur Qualitätssicherung wurde oben bereits beschrieben. Nach Bearbeitung von ca. 60% des Textmaterials sollte nochmals überprüft werden, ob eine Überarbeitung des Kodierleitfadens notwendig ist. Ist das nicht der Fall, sollte die Interkoderreliabilität (ICR) ermittelt werden, um zu entscheiden, ob das übrige Material von einem einzelnen Codierenden anhand des bestehenden Kodierleitfadens kodiert werden kann. Für das vollständig bearbeitete Datenmaterial musste zur abschließenden Bewertung der Reliabilität erneut die ICR berechnet werden, um gewährleisten zu können, dass sich das Datenmaterial zur Ergebnisinterpretation eignet.

Abbildung 6: Ablaufmodell der quantitativen Inhaltsanalyse



(Eigene Darstellung)

3.3.6 Zusammenfassung zum methodischen Vorgehen bei den expertenbeteiligten Verfahren

Die Beteiligung von Experten an der Fragebogenentwicklung und -evaluation hat nachweislich viele Vorteile, die andere, standardisierte oder statistische Methoden, nicht bedienen können. So identifizieren Experten nicht nur häufiger problematische Items hinsichtlich der Verständlichkeit, sondern auch Items, die später in der statistischen Analyse eine schlechte Datenqualität aufweisen (116). Viel wichtiger ist jedoch, dass die inhaltliche Güte eines Instruments nur durch Experten im jeweiligen Fachgebiet beurteilt werden kann (1). Aus diesem Grund empfiehlt sich für die Überprüfung der psychometrischen Güte eines

Instruments eine Kombination von qualitativen und quantitativen Verfahren, um die Nachteile bzw. „blinden Flecken“ des jeweils anderen auszugleichen (121). Ein solcher Mixed-Method-Ansatz wurde daher auch im Rahmen dieser Arbeit verfolgt, sodass neben der statistischen Überprüfung der Reliabilität und Validität expertenbeteiligende Verfahren zur Überprüfung der Inhaltsvalidität zum Einsatz kamen.

Die Befragung von Experten verfolgte mehrere Fragestellungen. Einerseits soll die Eignung des Fragebogens für den Einsatzzweck aus Sicht der Zielgruppe bewertet werden. Hierfür gilt es die Relevanz einzelner Aspekte und die inhaltliche Vollständigkeit des AIGScreen zu bewerten. Um dem aktuellen Stand der Forschung und allgemeinen Qualitätsstandards bei der Fragebogenentwicklung zu entsprechen, sollte zudem das Erhebungsdesign nochmals kritisch geprüft werden. Aufgrund der Komplexität des Erhebungsziels war die Beteiligung mehrerer Expertengruppen und interdisziplinärer Teams angezeigt.

Da zwar eine Vielzahl möglicher Auswahlkriterien für die Bestimmung von Experten existiert, aber keines für sich allein, eindeutig verlässliche Ergebnisse liefert, empfahl sich eine Kombination mehrerer Kriterien zur Auswahl der Experten und eine Kombination von Einzelmeinungen zu einem Konsens, wodurch verlässlichere Ergebnisse erreicht werden (111). Die Kriterien, an denen sich die Rekrutierung von Experten für das Fragebogenreview und für die Expertenbefragung waren persönliche Involviertheit, Zuschreibung von Expertise, Spezial- und Detailwissen (discrimination ability) und externale Eigenschaften. Die Zuschreibung erfolgte dabei durch die Autorin dieser Arbeit im Austausch mit dem interdisziplinären Team, welches im Forschungsprojekt mitwirkte. Externale Eigenschaften wie berufliche Position und formale Qualifikation dienten der Objektivierung der Auswahl. Primär sollten Personen angesprochen werden, denen man ein persönliches Interesse an der Thematik unterstellte. Für die erste Gruppe der Fragebogenreviewer konnte dies mit der arbeitsbedingten persönlichen Nähe zum Thema und zur Fragestellung begründet werden. Für die zweite Gruppe, die zu einer teilstandardisierten Online-Befragung eingeladen wurde, war eine unmittelbare Betroffenheit durch den Forschungsgegenstand selbst gegeben. Da es im Interesse der Beschäftigten im Geschäftsbereich des BMVg liegt, ein möglichst gesundheitsförderliches Arbeitsumfeld zu haben, wurde angenommen, dass eine hohe Mitwirkungsbereitschaft bei der Bedarfsanalyse und Fragen zu deren Gestaltung bestand. Ob die ausgewählten Personen die notwendige Expertise aufwiesen, lies sich anhand ihrer Fähigkeit profunde Fragen angemessen zu beantworten, feststellen. Für die Gesamtheit aller Antworten der jeweiligen Befragung wurde zudem ein Konsens angestrebt (Fragebogenreview) bzw. anhand der Befragungsergebnisse ermittelt (Expertenbefragung), um mögliche Verzerrungseffekte von einzelnen Antworten und Meinungen auszugleichen. Bezüglich des Erhebungsdesigns wird Konsens in der Expertengruppe als eindeutiger

Handlungsbedarf bewertet. Die Relevanz von Themen wurde ebenfalls am Mittel aller abgegebenen Einschätzungen bemessen.

Die Gruppe der Experten sollte sich aus Experten hinsichtlich des Inhalts, also Arbeitsmediziner und Sicherheitsingenieure, Experten auf dem Gebiet der Fragebogenkonstruktion sowie Experten für das spezifische Arbeitssetting des Geschäftsbereichs des BMVg zusammensetzen, sodass die Prüfkriterien der inhaltlichen Vollständigkeit, der adäquaten Fragebogengestaltung und Operationalisierung sowie der Verständlichkeit und Relevanz für die Zielgruppe (Augenscheininvalidität) überprüft werden können.

Aufgrund der Vielfältigkeit der verfolgten Fragestellungen und der dafür benötigten Expertengruppen wurde die Erhebung in zwei unabhängige Verfahren geteilt. In einem Expertenreview sollte eine Gruppe von Experten im Bereich Fragebogendesign die Qualität des Instruments evaluieren. Die Relevanz und inhaltliche Vollständigkeit sollte hingegen von Experten im Bereich der Arbeitsmedizin und betrieblichen Gesundheitsförderung, die gleichzeitig Vertreter der Zielgruppe der Befragung darstellen, bewertet werden. Hierfür wurde eine teilstandardisierte Online-Befragung durchgeführt, deren Ergebnisse teils quantitativ und teils mittels qualitativer Inhaltsanalyse ausgewertet wurden.

3.4 Reliabilitäts- und Validitätsanalyse

Während Aspekte der inhaltlichen Güte durch die expertenbeteiligten Verfahren evaluiert werden sollen, bedarf es zur Beurteilung der psychometrischen Güte quantitative Methoden. Zur Beurteilung der Güte einzelner Items wurden zunächst Itemanalysen durchgeführt. Dabei werden verschiedene Kennwerte aus der Antwortverteilung ermittelt sowie die statistische Beziehung zwischen einem einzelnen Item und der jeweiligen Antwortskala analysiert. Nur Items, die bestimmte Kriterien erfüllen, eignen sich für die anschließenden multivariaten Analysen wie die Faktorenanalyse. Diese wurde durchgeführt, um die zuvor theoretisch postulierte Dimensionalität des Instruments zu überprüfen. Wie im Kapitel 2.3.3 erläutert, dient die Faktorenanalyse zudem der Überprüfung der Konstruktvalidität, da mit ihr auch die interne Konsistenz der Skalen überprüft wird. Eine Kriteriumsvalidierung im Sinne einer prognostischen oder konkurrenten Validität ist für reflektive Messmodelle erst dann sinnvoll, wenn Reliabilität und Konstruktvalidität als gegeben angenommen werden können.

Um eine effiziente und präzisere Parameterschätzung (145) zu ermöglichen, wurden fehlende Werte mittels multipler Imputation ergänzt. Für die Analysen wurden die Statistikprogramme R 3.2, SPSS 18 und Stata 14.0 verwendet.

3.4.1 Itemanalyse

Bevor die Dimensionalität der Itemskalen bewertet wird, soll zunächst ein Überblick über die Qualität der einzelnen Items gewonnen werden. Dieser Schritt setzt voraus, dass eine Datenerhebung mit einer umfassenden Stichprobe unter realen Rahmenbedingungen bereits stattgefunden hat (70). Da die Ergebnisse der Itemanalyse stichprobenabhängig sind, ist insbesondere darauf zu achten, dass die Stichprobe auch der Personengruppe zukünftiger Zielgruppen entspricht (71). Die Itemanalyse berücksichtigt die Antwortverteilung und Streuung sowie fehlende Antworten. Dadurch sollen Rückschlüsse auf die Itemschwierigkeit, die Trennschärfe, Itemvarianz und Homogenität ermöglicht werden (146). Ziel der Itemanalyse ist es, problematische Items zu identifizieren. Damit sind Items gemeint, deren Verteilungswerte darauf hindeuten, dass die Reliabilität oder die Validität des gesamten Instruments verbessert werden kann, indem man das jeweilige Item entfernt (61).

In einem ersten Schritt werden sowohl grafische Auswertungen als auch deskriptive Statistiken genutzt, um die Normalverteilung der Antworten zu überprüfen, Decken- und Bodeneffekte zu identifizieren und die Schiefe und Kurtosis zu inspizieren (28). Zudem sollten insbesondere Items mit einem hohen Anteil fehlender Antworten genauer überprüft werden (70). Zur Überprüfung der Normalverteilung kann zusätzlich der auf Schiefe und Kurtosis basierende Mardia-Test angewandt werden (147, 148). Da dieser in einigen Fällen von der Stichprobengröße und der Zahl der Freiheitsgrade beeinflusst ist (149), wurde zusätzlich der Doornik-Hansen-Test auf Normalverteilung durchgeführt (150). Die Überprüfung auf Normalverteilung dient einerseits der Beurteilung der einzelnen Items und ihrer Eignung für den Einsatzzweck und erfasst andererseits eine notwendige Vorbedingung für die anschließend geplante Faktorenanalyse. Hierfür ist eine notwendige Voraussetzung die multivariate Normalverteilung (151).

Ob Normalverteilung vorliegt, hängt von Schiefe und Kurtosis ab. Die Schiefe beschreibt, inwiefern die Verteilung nicht symmetrisch ist (152). Bei einer grafischen Inspektion können so auch Decken- und Bodeneffekte erkannt werden (123). Eine Verteilung gilt als linksschief oder rechtssteil, wenn sich eine hohe Zahl von Antworten am rechten Ende der Verteilungskurve kumulieren. Entspricht das hohen Werten auf der Antwortskala spricht man auch von Deckeneffekten. Umgekehrt liegen Bodeneffekte bzw. eine rechtsschiefe oder auch linkssteile Verteilung vor, wenn vermehrt Antworten am unteren Ende Skala gegeben wurden bzw. wenn das Histogramm am linken Rand Verteilungskurve hohe Antwortanteile erkennen lässt (152). Hingegen spräche eine möglichst symmetrische Antwortverteilung für Vorliegen einer Normalverteilung. Eine zweite notwendige Bedingung für Normalverteilung ist die Form der Verteilungskurve, die Kurtosis. Diese muss unimodal („eingipflig“) und mesokurtisch, also nicht zu flach- und nicht zu steilgipflig verlaufend sein (152, 61).

Nicht nur für Leistungstests interessant ist die Itemschwierigkeit. Der Begriff ist im Zusammenhang arbeitsbedingter psychischer Belastung etwas irreführend, da es hierbei keine richtigen und falschen Antworten gibt. Die Itemschwierigkeit ist jedoch definiert als „der Anteil an Personen, die das Item im Sinne des Merkmals beantwortet haben“ (71). Wichtig ist dabei, dass bei der wahrheitsgemäßen Beantwortung Varianz entsteht. Würden alle Personen die Frage gleich beantworten, würde sie keinen Erkenntnisgewinn liefern und auf die Verwendung des Items sollte verzichtet werden. Da es sich bei der vorliegenden Erhebung um subjektives Empfinden handelt, kann schwer objektiviert werden, inwiefern das jeweilige Item im Sinne des Merkmals beantwortet wurde. Da es sich um eine mehrstufige Antwortskala handelt und den Abständen zwischen den Skalenpunkten Gleichabständigkeit unterstellt wird, kann das arithmetische Mittel als Itemschwierigkeit verwendet werden (71).

Ein zweiter Indikator, inwiefern ein Item gut zwischen Personen mit hoher und Personen mit niedriger Merkmalsausprägung unterscheidet, ist die Trennschärfe. Sie ist definiert als die Korrelation eines Items mit der Skala bzw. Subskala (71). Allerdings empfiehlt es sich bei der Berechnung der Korrelation des Itemwertes mit dem Gesamtscore der Skala, zu überprüfen, inwiefern sich die Korrelation unterscheidet, wenn der Wert des betreffenden Items nicht mit einbezogen wird. Denn passt ein Item schlecht zur Skala oder misst es generell schlecht, dann verzerrt dies den Gesamtwert der Skala und die Trennschärfe des problematischen Items wird künstlich verbessert, weil die Verzerrung der Messergebnisse in die Berechnung des Gesamtwertes einfließt. Aus diesem Grund sollte neben der Korrelation des jeweiligen Items mit seiner Skala auch die Korrelation des Items mit allen anderen Items, die sogenannte Item-Rest-Korrelation berücksichtigt werden (153). Für Items, die eine geringe Trennschärfe aufweisen, sollte erwogen werden, diese zu eliminieren (123).

Im Rahmen der Itemanalyse können bereits erste Erkenntnisse zur Reliabilität der Subskalen gewonnen werden. Wie im Abschnitt 2.3.2 „Reliabilität“ bereits erläutert wurde, sollte eine Konsistenzanalyse anstelle der Testhalbierungsmethode zur Beurteilung der Äquivalenz vorgenommen werden, da diese eine stabilere Schätzung der Reliabilität ermöglicht (60). Der hierfür am häufigsten verwendete Kennwert ist der Koeffizient Cronbachs Alpha (76). Er ergibt sich aus dem Durchschnitt von Paralleltests aller möglichen Itemkombinationen einer Skala oder Subskala (61). Eine Einschränkung, die es bei der Ergebnisinterpretation zu berücksichtigen galt, ist jedoch, dass er bei niedrigen Itemzahlen oder heterogenen Konstrukten zu einer Unterschätzung der Reliabilität führt (60, 76). Umgekehrt bedeutet dies, dass hohe Konsistenz-Koeffizienten der Subskalen für eine gute interne Konsistenz sprechen. Als gut gelten Werte des Reliabilitätskoeffizienten ab 0,80 (63). Da es sich beim AIGScreen um einen mehrdimensionalen und heterogenen Test handelt, empfiehlt sich die Berechnung von Cronbachs Alpha für die einzelnen Subskalen und auch niedrige Koeffizienten ab 0,70 sollten aufgrund der Heterogenität als akzeptabel betrachtet werden (63, 60). Umgekehrt galt,

dass eine zu hohe interne Konsistenz der Subskalen auch nicht erwünscht war, da die vier Merkmalsbereiche Arbeitsumgebung, Arbeitsinhalt und -aufgabe, Soziale Aspekte und Arbeitsorganisation bewusst heterogen in Form formativer Messmodelle operationalisiert wurden (71).

Zusätzlich bieten Statistikprogramme die Möglichkeit, die interne Konsistenz zu ermitteln, wenn das jeweilige Item nicht in der Skala enthalten wäre. Verbessert sich Cronbachs Alpha der Subskala bei Weglassen des jeweiligen Items, wäre dies ein eindeutiges Indiz für eine Elimination oder zumindest Modifikation des betreffenden Items (154, 76). Items, die die interne Konsistenz einer Skala negativ beeinflussen, sollten vor Durchführung einer Faktorenanalyse identifiziert werden, da sie die Schätzungen verzerren können bzw. zu unbefriedigenden Ergebnissen führen (155).

Ein weiteres Kriterium, dass die Eignung einzelner Items für die Faktorenanalyse anzeigt, ist das Kaiser-Meyer-Olkin-Kriterium (KMO), gelegentlich auch MSA (measure of sampling adequacy) - Kriterium. Dieser Kennwert beurteilt auf Basis der Anti-Image-Korrelationsmatrix⁹ die Zusammengehörigkeit von Items zu der zugewiesenen Skala (156). Es kann davon ausgegangen werden, dass Variablen das gleiche Konstrukt messen bzw. einen gemeinsamen Faktor abbilden, wenn das KMO wenigstens 0,5 bis 1,0 in einem möglichen Wertebereich von 0 bis 1 beträgt (157).

3.4.2 Multiple Imputation

Fehlende Werte können zu erheblichen Problemen bei der statistischen Analyse führen. So beeinträchtigen sie häufig die Effizienz der Schätzung der Parameter, können diese Ergebnisse mitunter sogar verzerren und erschweren die Durchführung vieler statistischer Standardverfahren, da diese für vollständige Datensätze konzipiert sind (145). Da im Rahmen der Erhebung mit dem AIGScreen sämtliche Angaben freiwillig waren, d. h. die Beantwortung bestimmter Fragen auch übersprungen werden konnte und zudem die Möglichkeit der Auswahl einer „unzutreffend“-Antwort bestand, musste insbesondere bei Items, die aufgrund des betreffenden Merkmals nicht alle Befragten beantworten konnten, mit teilweise hohen Missingquoten umgegangen werden. Dabei ist es inzwischen Konsens in der Fachliteratur, dass traditionellere Verfahren zum Umgang mit Missings wie z. B. listenweiser Fallausschluss ebenfalls die Gefahr verzerrter Schätzungen bergen und mittlerweile wesentlich verlässlichere und effizientere Verfahren zum Umgang mit fehlenden Werten entwickelt und erprobt sind

⁹ Die Anti-Image-Kovarianzmatrix beschreibt laut Guttman den Anteil der Varianz, die von den übrigen Variablen unabhängig ist, d. h. der sich im Rahmen einer multiplen Regressionsanalyse durch diese nicht erklären ließe. Da der Zusammenhang zwischen Variablen des gleichen Faktors möglichst hoch sein sollte, ist es wünschenswert, dass das Anti-Image möglichst gering ausfällt (156).

(145, 158–162). Hierzu zählen einerseits imputationsbasierte Verfahren und andererseits modellbasierte Verfahren (145). Bezüglich letzterer wird besonders häufig FIML (Full Information Maximum Likelihood Estimation) vorgeschlagen, welches interessierende Parameter auf Grundlage des unvollständigen Datensatzes schätzt (161). Der Nachteil dabei ist, dass für fehlende Werte keine Imputation stattfindet, also dass kein vollständiger Datensatz erzeugt wird (163). Dies kann zu Einschränkungen der Schätzung führen, wenn die Zahl der Items die Zahl der verbleibenden vollständigen Fälle übersteigt (158). Im Gegensatz zu den Maximum Likelihood (ML)-Verfahren, erlauben Imputationen die fehlenden Werte vor der statistischen Analyse, für die Daten aufbereitet werden sollen, zu behandeln, sodass der Umgang mit fehlenden Werten von der eigentlichen statistischen Analyse abgekoppelt ist (161). Für imputationsbasierte Verfahren wird häufig die multiple Imputation (MI) vorgeschlagen, welche zwar ein vergleichsweise aufwendiges Verfahren darstellt, aber sich als robust gegenüber Verletzungen von Modellannahmen erwiesen hat und es ermöglicht, aus einem unvollständigen Datensatz einen vollständigen Datensatz zu generieren (163, 145). Speziell für kategoriale bzw. ordinale Daten wird MICE (multiple imputation with chained equations) empfohlen (164, 145). Dieses Verfahren liefert auch bei fehlender multivariater Normalverteilung und einem hohen Anteil fehlender Werte noch verlässliche Ergebnisse (145, 165). MICE zeichnet sich durch die schrittweise und wiederholte Schätzung mittels Regressionen auf Basis aller anderen Variablen aus (164). Dabei werden üblicherweise mehrere Schätz-Zyklen durchlaufen, sodass am Ende auch mehrere geschätzte Werte für einen fehlenden Wert vorhanden sind. Die Empfehlungen, wie häufig die Schätzung wiederholt werden sollte, divergieren dabei stark. So empfehlen von Ginkel & Kroonenberg (2014), dass bereits bei einer Missingquote von 15% mindestens 100 Imputationen durchgeführt werden sollten, wohingegen Royston & White (2011) in Simulationen zu dem Ergebnis kommen, dass bereits 10 Imputationszyklen vollkommen ausreichend sind, wenn die Variablen mit fehlenden Werten nicht hoch korreliert sind.

Bevor multiple Imputation angewandt werden kann, sollte die Erfüllung der Modellprämissen überprüft werden. Hier ist insbesondere die Verteilungsstruktur der Missings zu berücksichtigen (159). Während früher im Umgang mit fehlenden Werten lediglich zwischen item nonresponse (fehlende Angaben bei einzelnen Fragen) und unit nonresponse (fehlende Angaben bei ganzen Frageblöcken oder für das gesamte Erhebungsinstrument) unterschieden wurde, betrachten neuere Verfahren zur Analyse fehlender Werte ihr Fehlen als ein probabilistisches Phänomen (161). Das bedeutet, dass die Annahme zugrunde liegt, dass die fehlenden Werten einer bestimmten Verteilung folgen. Hierbei werden in der Literatur grundsätzlich drei Verteilungsformen unterschieden: MCAR – missing completely at random, MAR – missing at random und MNAR – missing not at random (ebd.). Jede Struktur hat ihre Vor- und Nachteile für statistische Analysen, die mit dem Datensatz durchgeführt werden.

MCAR und MAR liefern unverzerrte Parameter, wobei die statistische Power bei MCAR eingeschränkt ist. MNAR wird hingegen zu verzerrten Schätzern führen (160).

Die Voraussetzung für MICE ist MAR (159). Das bedeutet, die Wahrscheinlichkeiten der fehlenden Werte können von den beobachteten Werten abhängig sein, aber dürfen nicht von den fehlenden Werten abhängen (161). Allerdings führt Graham (2009) zwei wichtige Argumente für die Verwendung von bzw. statistische Analyse mit Datensätzen, deren Missingstruktur nicht zufällig verteilt ist, an. Erstens konnte in Simulationsstudien gezeigt werden, dass allein die nicht zufällige Verteilung fehlender Werte in der Regel nicht die Verlässlichkeit der Schätzer beeinträchtigt. Zweitens führen traditionelle Verfahren zum Umgang mit fehlenden Werten wie z. B. der listenweise Fallausschluss zu verzerrteren Ergebnissen als wenn für die gleiche Datenstruktur multiple Imputation oder Maximum-Likelihood-Schätzverfahren eingesetzt werden: "In short, MI and ML methods are always at least as good as the old procedures [...] and MI/ML methods are typically better than old methods, and often very much better" (160). Er verweist außerdem darauf, dass eine trennscharfe Unterscheidung von Missingstrukturen in die genannten drei Gruppen nicht möglich ist und man vielmehr von einem Kontinuum möglicher Strukturen von MCAR bis MNAR ausgehen sollte. Es kommt demnach nicht so sehr auf die Frage an, ob bestimmte Modellvoraussetzungen wie MCAR oder MAR verletzt werden, sondern ob die Verletzung dieser Annahme einen wesentlichen Effekt auf die Ergebnisse haben kann.

Auf Basis welcher Variablen imputiert wird, kann nicht statistisch hergeleitet werden, sondern muss im Vorfeld theoretisch begründet werden (166). Dabei sind alle Variablen einzubeziehen, die im späteren Analysemodell eingesetzt werden, sowie alle weiteren Variablen, die zusätzliche Informationen über die fehlenden Werte enthalten können (ebd.). Hierin besteht ein weiterer wesentlicher Vorteil von MI gegenüber Likelihood-Methoden. Da es MI ermöglicht, unterschiedliche Methoden für die Phase der Imputation und für die anschließende Analyse anzuwenden, können auch die verwendeten Daten zwischen beiden Verfahren variieren. Es können also, anders als bei ML-Verfahren, mehr Daten zur Vorhersage fehlender Werte verwendet werden, da diese nicht auf die für die statistische Analyse notwendigen Variablen beschränkt ist (161). Dies macht das Verfahren auch robuster gegenüber Verletzungen der MAR-Annahme bzw. macht die Annahme vom Vorliegen von MAR plausibler, insbesondere wenn Hilfsvariablen in die Schätzung einbezogen werden können, welche einen Teil der fehlenden Werte erklären oder mit dem betreffenden Merkmal korreliert sind (159). Dies ist insofern wichtig, als dass die MAR-Annahme in Diagnoseverfahren höchstens widerlegt, nicht jedoch bestätigt werden kann und es zudem in realen Datensätzen unwahrscheinlich ist, eine vollständig zufällige Verteilung von fehlenden Werten vorzufinden (167, 166, 161).

Für die Imputation fehlender Werte im vorliegenden Datensatz wurde das ice-Programm für die Analysesoftware Stata eingesetzt, welches die Schätzung fehlender Werte mittels MICE ermöglicht (168). Zur Schätzung wurden neben den zu analysierenden Items des AIGScreen soziodemographische Merkmale (Alter, Geschlecht, Bildung), Variablen zur Erfassung der beruflichen Tätigkeit (Berufsgruppe, Berufserfahrung in Jahren, Beamtenstatus, Personalverantwortung, Vollzeitbeschäftigung) und als Outcome-Variable die allgemeine Arbeitszufriedenheit eingesetzt.

3.4.3 Explorative Faktorenanalyse

Wie bereits in Kapitel 2.3 beschrieben, dienen Faktorenanalysen der Strukturprüfung latenter Variablen und sind somit geeignete Verfahren, um die Reliabilität und Validität eines Instruments zu überprüfen. Für die Reliabilitätsanalyse eignet sich die explorative Faktorenanalyse (EFA), da sie der Überprüfung der Replizierbarkeit der Faktorenstruktur dient (64). Im Gegensatz zur Konfirmatorischen Faktorenanalyse (CFA – Confirmatory Factor Analysis) ermöglicht sie jedoch keine Hypothesenprüfung (169). Da die EFA auf der Annahme linearer Zusammenhänge zwischen den Items und den zu extrahierenden Faktoren basiert, müssen die Variablen Intervallskalenniveau oder wenigstens ein quasi-metrisches, d. h. gleichabständiges Skalenniveau aufweisen (151, 169). Ziel des Verfahrens ist es, Sparsamkeit in der Analyse und der Interpretation der Daten zu erzeugen (170). Das bedeutet, dass die Zahl der betrachteten Variablen auf eine überschaubarere Anzahl zu interpretierender Faktoren reduziert wird, welche die Struktur der Daten und die Beziehung der Variablen untereinander am besten zusammenfassen (171). Ziel ist zudem, dass die extrahierten Faktoren möglichst eine Einfachstruktur aufweisen. Das bedeutet, die Items sollten möglichst nur auf einen Faktor laden und die Faktoren sollten untereinander nicht oder kaum korreliert sein (172). Die Faktorenextraktion kann mit verschiedenen Methoden durchgeführt werden. Häufig wird die Hauptkomponentenanalyse (PCA – Principal Components Analysis) als eine Extraktionsmethode der EFA aufgeführt und sogar aufgrund der Voreinstellungen in vielen Statistikprogrammen standardmäßig angewandt (151, 173). Im Rahmen dieser Arbeit wurde jedoch die Faktorenextraktion mittels Maximum-Likelihood-Verfahren vorgenommen. Methodisch birgt dieses Verfahren den Vorteil, dass eine Vielzahl an Goodness of Fit-Indizes zum Modellvergleich kalkuliert werden können (173). Zudem stimmen die Modellprämissen deutlich besser mit den Vorüberlegungen zur Analyse überein. So stellt die PCA keinerlei Grundannahmen und zielt ausschließlich auf die Komplexitätsreduktion auf Basis der Korrelationsmatrix ab, während die EFA ein inferenzstatistisches Verfahren darstellt, welche die zugrunde liegende latente Struktur möglichst erklärend abzubilden strebt (151). Des Weiteren erscheint das Postulat der vollständigen Varianzaufklärung bei der PCA, wonach

keine Restvarianz durch Messfehler besteht (172), für die vorliegenden Daten unplausibel. Schließlich spricht für die ML-Schätzung, dass diese auch bei der Überprüfung der faktoriellen Validität mittels Konfirmatorischer Faktorenanalyse (CFA) angewandt wird und andernfalls im Ergebnisvergleich nicht abschließend beurteilt werden könnte, ob Ergebnisunterschiede auf eine unterschiedliche Methodenwahl zurückzuführen sind. In diesem Sinne wurden beide Verfahren auch mit imputierten Daten durchgeführt. Hierfür wurde der Gesamtdatensatz zufällig in zwei Stichproben für das jeweilige Analyseverfahren geteilt. Damit ergab sich für die EFA eine Stichprobe von $n_{EFA} = 8.134$. Die Analyse wurde mit dem R-Paket „psych“ (174) durchgeführt.

Zur Bestimmung der Anzahl zu extrahierender Faktoren stehen eine Vielzahl von Tests zur Verfügung. Häufig wird zunächst eine grafische Inspektion mittels Scree-Tests empfohlen, Die dafür verwendeten Diagramme, Scree-Plots, bilden die Eigenwerte der extrahierten Faktoren in absteigender Reihenfolge ab. Im Verlauf der Eigenwertkurve sollte sich ein eindeutiger Knick abzeichnen. Alle Faktoren, die sich vor diesem Knick im Diagramm befinden, können als inhaltlich relevant erachtet werden (172). Liefert diese graphische Inspektion kein eindeutiges Ergebnis, sollte die Parallelanalyse hinzugezogen werden (173). Sie trägt dem Umstand Rechnung, dass Zufallskorrelationen die Eigenwerte der Faktoren fälschlicherweise erhöhen könnten (172). Hierfür werden Zufallsdatensätze erzeugt, die in Variablenanzahl und Stichprobengröße dem empirischen Datensatz entsprechen. Diese Zufallsdaten werden zufällig miteinander korreliert und jeweils einer Faktorenanalyse unterzogen. Aus den gemittelten Eigenwerten dieser Faktorenanalysen werden die Daten gewonnen, welche im Paralleltest in das Eigenwertdiagramm einfließen (ebd.).

Zusätzlich können Eigenwert-Kriterien und Modell-Fit-Indizes zur Bestimmung der idealen Faktorenzahl herangezogen werden. Eigenwerte drücken den Anteil erklärter Varianz durch den jeweiligen Faktor aus (172). Nach dem Kaiser-Kriterium sind Faktoren mit einem Eigenwert größer als 1,0 zu extrahieren (156), nach dem (weniger strengen) Jolliffe-Kriterium können bereits Faktoren ab einem Eigenwert von 0,7 extrahiert werden (170). Beide Kriterien können jedoch zu einer Überschätzung der Zahl zu extrahierender Faktoren führen und sollten daher nur in Kombination mit weiteren Kriterien Anwendung finden (173). Eine Möglichkeit die ideale Faktorenzahl auf Basis der Kovarianzmatrix zu bestimmen, bietet Valicer's MAP (Minimum Average Partial Correlation) (175). Hierfür wird die mittlere Partial-Korrelation nach jeder Komponenten-Extraktion berechnet und Extraktion gilt als beendet, wenn der kleinste Wert erreicht ist (ebd.). Somit eignet sich der MAP auch für den Modellvergleich. Ein weiteres Kriterium, welches nur im Modellvergleich interpretierbar ist, stellt das BIC (Bayesian Information Criterion) dar. Auch hier gilt, je kleiner das BIC ist, desto besser ist der Modellfit, da der Wert mit zusätzlichen, wenig Varianz erklärenden Variablen steigt (155). Allerdings ist auch das niedrigste BIC unter allen Modellen kein Garant dafür, dass das Modell per se gut

ist und sollte entsprechend ebenfalls nur in Kombination mit weiteren Indizes betrachtet werden (64). Ein weiteres Maß für den Modellfit stellt der RMSEA (Root Mean Square Error of Approximation) dar, welcher ebenfalls möglichst niedrig ausfallen sollte (30). Der „Error of Approximation“ bezieht sich dabei auf die fehlende Anpassung des Modells an die Populations-Kovarianzmatrix, wobei das Minimum bei null liegt, Werte bis 0,05 als gut und bis 0,08 als akzeptabel gelten (176).

Nachdem die Kriterien für die Faktorenextraktion bestimmt wurden, ist eine weitere Entscheidung, die im Rahmen der Faktorenanalyse zu treffen ist, die Wahl der Rotationsmethode (171). Faktorrotationen dienen dem Ziel der Einfachstruktur und können in rechtwinklige (orthogonale) und schiefwinklige (oblique) Verfahren unterschieden werden (172). Rechtwinklige Verfahren werden vor allem dann eingesetzt, wenn keinerlei theoretische Vorannahmen über die Beziehung der Daten bestehen und die Faktorenextraktion ausschließlich der Datenreduktion dient (ebd.). Wird hingegen eine Korrelation der Faktoren vermutet, empfiehlt sich die schiefwinklige Rotation (156). Letzteres ist plausibel, insbesondere bei Merkmalen menschlichen Verhaltens und menschlicher Einstellungen (171), also auch bei den mit dem AIGScreen erhobenen Merkmalen. Aus diesem Grund wurde die oblique Rotation gewählt. Eine weitere Herausforderung bestand darin, eine Methode zu finden, die die Ergebnisse aller imputierten Datensätze übersichtlich zusammenfasst. Hierfür eignet sich die Procrustes-Rotation, welche die Rotationen der einzelnen Datensätze kombiniert (164). Hierbei werden die Faktorladungen so rotiert, dass die Unterschiede zwischen den Datensätzen minimiert werden, wodurch die Genauigkeit erhöht wird (177). Die Faktorladungen werden dann als Durchschnitt der einzelnen Ladungen ausgegeben (178). Dieser Ansatz wird als Alternative zur Verwendung der durchschnittlichen Kovarianzmatrix empfohlen, die den Nachteil hat, dass die Ergebnisse wahrscheinlich fehlerhaft sind, wenn die Reihenfolge der erhaltenen Komponenten der einzelnen unterstellten Datensätze nicht die gleiche ist (158).

3.4.4 Konfirmatorische Faktorenanalyse

Während die explorative Faktorenanalyse (EFA) ein strukturentdeckendes Verfahren darstellt, ist die Konfirmatorische Faktorenanalyse (CFA) ein hypothesenprüfendes Verfahren, d. h. Ziel der CFA ist es, zu prüfen, „ob eine hinreichende Übereinstimmung (Modellfit) zwischen den empirischen Daten und dem theoretischen Modell besteht“ (172). Sie ist somit eine geeignete Methode zur Überprüfung der faktoriellen Validität (178). Sie wurde im Rahmen dieser Arbeit angewandt, um die postulierte vierdimensionale Struktur mit alternativen Modellen, die im Rahmen der EFA plausibel erschienen, zu vergleichen. Die Modellspezifikation basierte entsprechend aufbauend auf den Ergebnissen der EFA. Für die Parameterschätzung wurde

die ML-Methode angewandt, um eine Vergleichbarkeit mit den Ergebnissen der EFA zu gewährleisten.

Die CFA basiert auf dem Postulat reflektiver Messmodelle, d. h. den in der Analyse verwendeten Indikatorvariablen wird unterstellt, dass sie das zu messende latente Konstrukt repräsentieren (156). Um eine ausreichende Modellsättigung erzielen zu können, bedarf es hierfür mindestens drei Variablen pro latentem Konstrukt (64). Welche Items welchem Faktor zuzuordnen sind, wurde auf Basis der Ergebnisse der EFA entschieden.

Da es keinen einzigen statistischen Signifikanztest gibt, der ein korrektes Modell für die Stichprobe ausweist, ist es notwendig, mehrere Kriterien der Modellanpassung zu berücksichtigen und die Modelle auf der Grundlage verschiedener Maße gleichzeitig zu bewerten (176). Neben den in Abschnitt 3.4.3 erläuterten Modelfit-Indikatoren RMSEA und BIC können zum Modellvergleich bei der CFA der Comparative Fit Index (CFI), der Tucker-Lewis-Index (TLI), das Akaike Information Criterion (AIC) sowie das Standardized Root Mean Square Residual (SRMR) herangezogen werden. Letzteres wird auch als „badness-of-fit measure“ (179) bezeichnet, da es die verbleibende Diskrepanz zwischen empirischer und geschätzter Kovarianzmatrix nach Schätzung der Modelparameter misst. Entsprechend entspricht ein Wert von Null einem perfekten Modelfit. Darüber hinaus existieren keine eindeutigen Cut-off-Werte, allerdings kann als Daumenregel ein Wert bis 0,05 als gut und bis 0,10 als akzeptabel angesehen werden (ebd.). Für den CFI und den TLI hingegen gilt ein Cut-off-Wert von circa 0,95 (64). Bei beiden Indizes handelt es sich um komparative Indizes, die auf Basis der Chi-Quadrat-Verteilung das zu testende Modell mit dem Nullmodell vergleichen (64, 180). Voraussetzung ist, dass die Modellschätzung mittels ML-Methode vorgenommen wurde (180). Bentler & Bonett (1980) nennen als einen wesentlichen Vorteil dieser Indizes, dass sie es erlauben, neben der statistischen Signifikanz auch die praktische Signifikanz eines Schätzmodells im Vergleich zum Nullmodell zu beurteilen. Der Unterschied zwischen beiden Indizes ist, dass der TLI im Gegensatz zum CFI nicht normiert ist, d. h. der Wertebereich des CFI liegt zwischen 0 und 1 während der TLI auch außerhalb dieses Wertebereichs liegen kann (181). Obwohl der TLI dadurch etwas schwerer zu interpretieren ist, empfiehlt sich die Anwendung beider Indizes, da der TLI sensitiv gegenüber komplexen Modellspezifikationen fehlspezifizierten Faktorladungen ist (ebd.). Ein drittes Maß für den Vergleich mehrerer Modelle stellt das AIC dar. Es kann nur in Relation zu den Werten der anderen Modelle interpretiert werden, wobei gilt, dass das am besten saturierte Modell den niedrigsten Wert erzielt (64). Genauso wie das BIC bevorzugt das AIC das am wenigstens komplexe Modell (ebd.). Allerdings kann auch der niedrigste Wert erheblich von einem idealen Modell abweichen, weshalb zusätzliche Tests, insbesondere der Signifikanztest zur Beurteilung der Modellgüte notwendig sind (181, 64).

Das einzige Modelfit-Maß, das es erlaubt einen Signifikanztest zur Beurteilung, welches Modell sich am besten eignet, durchzuführen, ist die Chi-Quadrat(χ^2)-Test-Statistik, welche die empirische Kovarianzmatrix mit der vom Modell implizierten Kovarianzmatrix vergleicht (176). Da die Normalverteilungsannahme verletzt war, wurden zur Beurteilung der Modelgüte der nach Satorra-Bentler skalierte χ^2 -Tests (SB-Statistik) angewendet (182). Die Entwickler dieses Verfahrens konnten nachweisen, dass die damit berechneten Kennwerte robust gegenüber der Verletzung der Normalverteilungsannahme sind (183). Durch die Anpassung der Modellschätzung an die SB-Statistik werden auch die Parameter des Modells robust geschätzt (184). Allerdings werden die Ergebnisse des χ^2 -Tests von der Stichprobengröße beeinflusst. Mit steigender Fallzahl steigt auch die Zahl der Freiheitsgrade. Das wiederum kann einerseits dazu führen, dass plausible Modelle abgelehnt werden, weil das Signifikanzniveau nicht ausreichend hoch ist, obwohl der Unterschied zwischen empirischer und modellierter Kovarianzmatrix nicht wesentlich ist (176). Umgekehrt kann aber auch der Fall eintreten, dass schlecht passende Modelle ein ausreichendes Signifikanzniveau erreichen, da der χ^2 -Wert ebenfalls mit steigender Fallzahl zunimmt (ebd.). Aus diesem Grund sollte auch der χ^2 -Test-Statistik allein nicht zu viel Gewicht beigemessen werden. Zudem wird empfohlen als Daumenregel nur Modelle als ausreichend passend zu betrachten, die die Bedingung $\chi^2 \leq 2df$ (df = degrees of freedom, Freiheitsgrade) erfüllen, also bei denen die Zahl der Freiheitsgrade mindestens doppelt so hoch ist wie der χ^2 -Wert (172, 181).

Die CFA wurde mit Stata 14.0 für die zweite Teilstichprobe $n_{CFA} = 8.058$ des Gesamtdatensatzes durchgeführt.

4 Ergebnisse

4.1 Ergebnisse des internen Expertenreviews

Am Review des AIGScreenBw beteiligten sich sieben Personen, darunter zwei Arbeitsmediziner, zwei Psychologen und jeweils ein wissenschaftlicher Mitarbeiter mit Abschlüssen in Epidemiologie, Soziologie und Public Health.

Bezüglich der Objektivität des Fragebogens bestanden keine Einwände oder Verbesserungsvorschläge, da diese aufgrund des hohen Standardisierungsgrades sowohl bei der Datenerhebung als auch bei der Auswertung als gegeben betrachtet werden kann. Im Rahmen der Datenverarbeitung konnte lediglich bei der manuellen Eingabe von Freitextangaben in den Papierfragebögen eine Objektivitätsgefährdung durch Interpretation der Datenerfassenden identifiziert werden. Da hier jedoch mindestens zwei Personen bei uneindeutigen Angaben oder nicht bzw. schwer lesbaren Texten an der Datenerfassung beteiligt waren und im Vorfeld eine Schulung aller Personen, die an der Datenerfassung beteiligt waren, durchgeführt wurde und es genaue Anweisung zur Datenerfassung gab, wurde dieses Risiko als äußerst gering eingestuft.

Wie in Kapitel 3.3 erläutert, stellt Konsens zwischen den Experten eine notwendige Bedingung dafür dar, dass die Ergebnisse überhaupt als verlässlich gelten können. Daher wurde nach Bearbeitung der Item-Checkliste durch die einzelnen Experten in einem zweiten Schritt der Konsens zwischen allen Beteiligten gesucht und in einer Tabelle festgehalten. Bei zunächst widersprüchlichen Einschätzungen konnte in der Diskussion eine gemeinsame Position gefunden werden. Die Ergebnisse dieser zweiten Expertenrunde sind in Tabelle 2 **Fehler! Verweisquelle konnte nicht gefunden werden.** zusammengefasst. Dabei wurden nur die 17 Items aufgeführt, für die die Experten Hinweise gaben. Für die anderen 17 Items wurden keine qualitativen Mängel oder Verbesserungsbedarf festgestellt. Dies betrifft unter anderem die gesamte Subskala zur Arbeitsumgebung.

In Tabelle 2 nicht dargestellt sind die Punkte aus der Fragebogen-Checkliste „Probleme im Zusammenhang mit der Aufgabenbeschreibung“ und „Komplexität der mentalen Leistung“, da es zu beiden Punkten keine Anmerkungen oder Verbesserungsvorschläge seitens der Experten gab. Hingegen sind die Anmerkungen zu den Antwortskalen und der „Unzutreffend“-Kategorie nicht in der Tabelle dargestellt, weil sie nicht für einzelne Items sondern für den gesamten Fragebogen gemacht wurden. Demnach war die Antwortoption „Unzutreffend“ zwar notwendig und angemessen, stand jedoch in deutlichem Widerspruch zum Anleitungstext zu Beginn des Frageblocks. Dort hieß es: „Bitte geben Sie an, inwiefern die Aussage auf Sie zutreffend ist.“ Diese Instruktion ist insofern problematisch, als dass sie zu einer missverständlichen Nutzung der Antwortoption „Unzutreffend“ führen kann. Aus Sicht der

Probanden wäre in Konsequenz auf diese Anleitung die Antwort „Unzutreffend“ beispielsweise bei der Beantwortung des Items „Meine Arbeit ist in angenehmem Maße abwechslungsreich“ zu wählen, wenn die Befragten der Ansicht sind, dass ihre Arbeit nicht abwechslungsreich ist. Tatsächlich sollte jedoch eine Antwortoption auf der vierstufigen Skala von „nein“, „eher nein“, „eher ja“ und „ja“ gewählt werden. Die „Unzutreffend“-Kategorie ist für Befragte vorgesehen, die die Frage nicht beantworten können, weil sie nicht ihrer Arbeitssituation entspricht. Das wäre beispielsweise der Fall, wenn Personen ohne Personalverantwortung nach der Beziehung zu den ihnen unterstellten Mitarbeitenden gefragt werden. Daher wurde auch nicht die Antwortkategorie selbst von den Experten kritisiert, jedoch eine dringende Überarbeitung des Einleitungstextes empfohlen, da die Gefahr einer missverständlichen Beantwortung bei allen Items im Kernfragebogen bestand. Während bezüglich dieses Punktes Einigkeit unter den Experten bestand, wurden verschiedene Aspekte der Antwortskala kontrovers diskutiert. Dies betraf erstens die Verwendung einer zusätzlichen „Weiß nicht“-Kategorie, zweitens die Zahl der Skalenpunkte und drittens die Frage, ob die Antwortskala einheitlich verbalisiert sein muss oder genauer auf die Frageinhalte abgestimmt werden sollte.

Für variierende Wortlaute der Antwortskalen sprach einerseits die Vermeidung von Monotonie bzw. Langeweile der Befragten. Einige Experten argumentierten, dass bei zu großer Monotonie der Antwortskalen, Probanden dazu tendieren verschiedene Antwortoptionen auszuwählen, um monotone Muster aufzubrechen und das unabhängig davon, was ihrer wahren Einstellung oder Meinung entspricht. Ein weiteres Argument für eine genauere Anpassung der Verbalisierungen an den jeweiligen Frageinhalt stellte die Zuverlässigkeit des Antwortverhaltens dar. Je eindeutiger Antwortkategorien interpretiert werden, so geringer sei der Messfehler. Dagegen sprach, dass man in vielen Fällen auch die Items entsprechend umformulieren müsste, also eine vollständige Fragebogenrevision durchführen müsste. Des Weiteren wurde mit satisficing-Effekten argumentiert, denn zu häufig wechselnde Antwortkategorien und Frageformate könnten die kognitiven Anforderungen an die Befragten deutlich erhöhen und zu Ermüdung führen. Dies könnte wiederum dazu führen, dass bestimmte Antwortmuster (z. B. immer die zweite Antwortoption) verfolgt werden und auch häufiger zu versehentlich falsch gegebenen Antworten führen, was die Reliabilität der Daten erheblich beeinträchtigen würde. Schließlich wurde entschieden, die bestehende Antwortskala mindestens bis zur statistischen Analyse der psychometrischen Güte beizubehalten. Da der Kernfragebogen auch nur einen Teil einer längeren Befragung darstellte, wurde angenommen, dass den Befragten ein übersichtliches (da einheitliches) Antwortschema mit Blick auf die zu erbringende mentale Gesamtleistung eher entgegenkommt als sie zu langweilen.

Auch die Zahl der Antwortoptionen wurde kritisiert, jedoch auch hier im Konsens beschlossen, die statistischen Analysen abzuwarten. Hier bestand Einigkeit in der Expertengruppe, dass die aktuelle Forschung uneingeschränkt mehr Antwortoptionen im Sinne der wissenschaftlichen

Güte einer Erhebung empfiehlt. Gegen eine Erweiterung der Antwortskala sprach jedoch die Vergleichbarkeit mit bereits erhobenen Daten. Zudem fürchtete man eine höhere kognitive Belastung der Befragten und schnellere Ermüdung. Einigkeit bestand ferner darin, dass eine Mittelkategorie vermieden werden sollte, um satisficing-Effekten vorzubeugen. Bei sämtlichen erhobenen Merkmalen konnte man davon ausgehen, dass die Befragten eine, wenn auch nur leichte, Meinungstendenz haben. Vollkommen neutrale Positionen waren nur dann wahrscheinlich, wenn der erfragte Aspekt die Arbeitssituation des Befragten nicht betraf. Dafür wird im Fragebogen jedoch die „Unzutreffend“-Kategorie angeboten, weshalb eine zweite zusätzliche neutrale Antwortoption, zumindest für diesen Grund nicht notwendig schien. Der gleichen Argumentation wurde bei der Frage, ob eine zusätzliche „Weiß nicht“-Antwort ergänzt werden sollte, gefolgt. Befürworter dieser zusätzlichen Kategorie argumentierten, dass es einen Unterschied macht, ob ein bestimmtes Item nicht beantwortet werden kann, weil es grundsätzlich nicht auf den Befragten zutrifft oder ob dieser es nicht beantworten kann, weil er die Antwort darauf nicht kennt. Dieser feine, jedoch wichtige Unterschied, würde mit dem derzeitigen Antwortformat nicht abgebildet. Allerdings kam man in der Diskussion zu dem Ergebnis, dass dies bei den mit dem AIGScreen erhobenen Merkmalen kaum der Fall sein kann. Schließlich ginge es bei der Befragung auch nicht um konkretes Wissen, wo zwischen richtig und falsch unterschieden werden kann, sondern um das subjektive Empfinden.

Weitere Anmerkungen bezogen sich auf einzelne Items. So stand zur Diskussion, ob zur Aussage „Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen“ ein einheitliches Antwortverhalten gegeben ist, da nicht klar war, ob dieses Merkmal überhaupt auf alle Probanden zutrifft. In Rücksprache mit Vertretern der Zielgruppe wurde jedoch erläutert, dass die Formulierung über alle Organisationseinheiten hinweg gut und einheitlich verständlich sei. Die Items 20, 21, 29, 30, 37, 38 und 39 hingegen treffen nicht auf alle Beschäftigten zu. Eine mögliche Lösung wäre, eine entsprechende Filterfrage zu ergänzen, die dazu führt, dass nur Personen die Frage beantworten, auf die diese Merkmale (Personalverantwortung, interner Kundenkontakt, Verantwortung für Objekte) zutreffen. Da dies jedoch mindestens drei zusätzlicher Fragen bedürfte, was die Bearbeitung für die Personen, auf die das Merkmal zutrifft, weiter erhöht und die Personen, auf die das Merkmal nicht zutrifft, statt zwei Items immer noch eines bearbeiten müssen, wurde dies als keine gute Lösung im Sinne der Fragebogenökonomie bewertet. Es wurde sich darauf verständigt, dass die „Unzutreffend“-Kategorie ausreichend ist, um die Beantwortung der Frage durch Personen, die dazu gar keine Einschätzung abgegeben können, zu vermeiden. Allerdings wurde seitens der Experten auch davor gewarnt, dass diese Items eine höhere Missing-Quote aufweisen könnten, da die Beantwortung aller Fragen freiwillig ist und Probanden dazu tendieren könnten, die Frage einfach zu überspringen, wenn sie feststellen, dass der Fragegehalt sie nicht betrifft.

Für zwei Items, Nummer 18 („Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit“) und Nummer 23 („Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit“), wurde angemerkt, dass Kontexteffekte durch vorausgegangene Fragen möglich seien. Bei Item Nr. 18 wird zuvor die Aussage „Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet“ bewertet. Trifft das nicht zu, ist eine tendenziell negativere Bewertung des darauffolgenden Items wahrscheinlicher und umgekehrt kann davon ausgegangen werden, dass Personen, die sich gut qualifiziert fühlen auch die Weiterbildungsmöglichkeiten positiver bewerten. Die gleiche Problematik trifft auf Item Nr. 23 zu, dem das Item „Ich kann erkennen, ob ich meine Arbeit gut gemacht habe“ vorgeschaltet ist. Personen, die selbst eine Ergebniskontrolle ihrer Arbeit vornehmen können, sind wahrscheinlich weniger auf Rückmeldung angewiesen und erachten daher auch einen geringeren Grad an Rückmeldung bereits als angemessen. Allerdings sei die Frage zu klären, inwiefern es sich bei beiden Items um einen Kontexteffekt oder grundsätzliche Interkorrelation handelt. Mögliche Kontexteffekte ließen sich durch randomisierte Fragereihenfolgen überprüfen und ggf. auch damit vermeiden. Eine Vergleichserhebung sollte daher für zukünftige Befragungen erwogen werden.

Ein weiteres Problem, welches zwei Items betraf, ist die doppelte Verneinung. Hierdurch können durch Fehlinterpretation versehentlich falsche Antworten gegeben werden, was eine systematische Verzerrung der Ergebnisse bedingen würde. Betroffen sind Item Nr. 33 „Mit meinen Mitarbeitenden gibt es keine Konflikte“, Item Nr. 36 „Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte“ und das Item Nr. 39 „Mit den Untergebenen/den mir unterstellten Beschäftigten gibt es keine Konflikte“. Bei allen drei Items stellte sich aus Sicht der Experten die Frage, welche Antwortoption die Befragten im Falle keiner oder weniger Konflikte wählen würden: „Ja“ im Sinne von es gibt keine Konflikte oder „nein“ im Sinne von Konflikte treffen auf meine Tätigkeit nicht zu? Doppelte Verneinungen sind bei Frageformulierungen grundsätzlich zu vermeiden (124). Allerdings sollten alle Items im AIGScreen positiv formuliert sein (vgl. Abschnitt 3.1), um einen Wechsel der Polung bei der Beantwortung der Fragen zu vermeiden. Aufgrund dieser Kontinuität wurde in der Gruppendiskussion auch befunden, dass die Beantwortung der betroffenen Items für die Befragten unproblematisch sein sollte. Denn alle drei Items befinden sich im hinteren Teil des Frageblocks und es wurde davon ausgegangen, dass bis dahin ein gewisser Lerneffekt bei den Probanden eingesetzt hat, wonach die rechten Antwortoptionen („nein“, „eher nein“) den negativen Pol und die rechten Antwortoptionen („eher ja“, „ja“) den positiven Pol darstellen. Damit sollte auch ohne größere Überlegungen für die Probanden eine adäquate Beantwortung der Frage möglich sein.

Konsens bestand darin, dass das Item Nr. 30 „Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt“ im Rahmen der

nächsten Fragebogenrevision überarbeitet bzw. zweigeteilt werden muss. Denn Respekt und Unterstützung sind zwei unterschiedliche Aspekte, die nicht mit einer Antwort bewertet werden können. Auch das Item Nr. 25 „Ich werde bei meiner Arbeit selten unterbrochen“ sollte aus Sicht der Experten überarbeitet werden. Durch die Verwendung des Wortes „selten“ wird die Möglichkeit, dass es niemals zu Arbeitsunterbrechungen kommt, ausgeschlossen. Dies mag bezüglich der beruflichen Belastungssituation möglicherweise keinen wesentlichen Unterschied machen, es könnte jedoch bei den Probanden Unzufriedenheit auslösen, wenn diese ihre Arbeitsrealität nicht im Fragebogen abgebildet sehen. Zudem sollte immer der gesamte Möglichkeitsraum potenzieller Antworten im Rahmen einer Frage bzw. der Antwortoptionen abgedeckt sein.

Für einzelne Experten irritierend war die Verwendung des Begriffs „Mitarbeitende“ für Kolleginnen und Kollegen bei den Items Nr. 31 und 32. Auch hier ergab allerdings die Rücksprache mit Vertretern der Zielgruppe, dass dies dem internen Sprachgebrauch im Geschäftsbereich BMVg entspreche und für alle verständlich sei. Unklar war zudem, inwiefern die Verwendung des Begriffs „angenehm“ intersubjektiv verständlich in Bezug auf die Arbeitssituation bei Item Nr. 13 und Item Nr. 19 sei. Es wurde beschlossen, die Ergebnisse der quantitativen Analysen abzuwarten. Sollte eine Überarbeitung dieser Items notwendig sein, so sollte auch über eine Veränderung dieses Wortlautes nachgedacht werden. Zusätzlich wurde bei Item Nr. 13 diskutiert, welche Kriterien die Probanden zur Bewertung Ihrer Arbeitszeitverteilung anwenden, da hier eine Vielzahl von Aspekten subsumiert werden kann. Allerdings kam man in der Gruppendiskussion zu dem Ergebnis, dass es sich beim AIGScreen um ein Grobscreening handelt, welche Schwerpunkte für die vertiefte Arbeitsplatzbeurteilung aufzeigen soll. Für diesen Anwendungsfall ist es zumindest für den Erhebungszweck im Rahmen des BGM und der Beurteilung psychischer Belastungen nachrangig, welche Aspekte im Detail zur Bewertung der Arbeitszeitverteilung von den einzelnen Befragten einbezogen wurden. Allerdings wiesen die Experten darauf hin, dass es für die Reliabilität der erhobenen Daten durchaus einen negativen Effekt haben könnte.

Tabelle 2: Ergebnisse des Fragebogenreviews

Item	Frageverständnis/Wortlaut	Unzutreffende Annahmen über Befragte	Fragekontext und Nebeneffekte
Arbeitsmenge, zeitliche Aspekte, Arbeitspausen			
11 Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen.	ok	Trifft dieses Merkmal auf alle zu? Kommt es hier evtl. zur Unterschätzung, da nur wenige hier überhaupt Gefährdung angeben können und alle, auf die es nicht zutrifft, "angemessen" angeben bzw. positive Antwortoptionen wählen?	ok
13 Meine Arbeitszeitverteilung ist mir angenehm.	Hier fließen mehrere Aspekte ein: Schichtdienst, Pausenzeiten, Überstunden, Arbeitszeitregelung (Gleitzeit, Option Homeoffice etc.). Für die Befragten könnte es schwer sein, eine Antwort zu finden. Zudem könnte Intersubjektivität nicht gegeben sein.	ok	ok
Arbeitsanforderungen, Kontrolle, organisatorische Arbeitsbedingungen			
18 Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit.	ok	ok	Beeinflussung durch vorangehende Frage möglich
19 Meine Arbeit ist in angenehmem Maße abwechslungsreich.	Angenehmn ist sehr unspezifisch, Intersubjektivität fraglich	ok	ok

20	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z. B. Personalverantwortung).	ok	nicht auf alle zutreffend	ok
21	Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material).	ok	nicht auf alle zutreffend	ok
23	Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit.	ok	ok	Beeinflussung durch vorangehende Frage möglich
25	Ich werde bei meiner Arbeit selten unterbrochen.	ok	"Selten" schließt "nie" aus, volles Antwort- bzw. Möglichkeitsspektrum wird hier nicht abgedeckt	ok
Soziale Aspekte der Arbeit				
29	Mein Kundenkreis für interne Dienstleistungen (z. B. auch Patienten, Leistungsempfänger) verhält sich angemessen.	ok	nicht auf alle zutreffend	ok
30	Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.	Respekt und Unterstützung sind zwei unterschiedliche Aspekte und sollten daher nicht mit einer Frage erfasst werden	nicht auf alle zutreffend	ok
31	Die Anerkennung meiner Arbeit durch meinen Mitarbeiterkreis ist angemessen.	Gemeint sind KollegInnen - Wird das von allen Probanden so verstanden?	ok	ok

32	Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis.	Gemeint sind KollegInnen - Wird das von allen Probanden so verstanden?	ok	ok
33	Mit meinen Mitarbeitenden gibt es keine Konflikte.	doppelte Verneinung Gemeint sind KollegInnen - Wird das von allen Probanden so verstanden?	ok	ok
36	Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte.	doppelte Verneinung	ok	ok
37	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Beschäftigten ist angemessen.	ok	nicht auf alle zutreffend	ok
38	Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Beschäftigten.	ok	nicht auf alle zutreffend	ok
39	Mit den Untergebenen/den mir unterstellten Beschäftigten gibt es keine Konflikte.	doppelte Verneinung	nicht auf alle zutreffend	ok

4.2 Ergebnisse der teilstandardisierten Online-Befragung

Insgesamt nahmen an der Expertenbefragung 173 Personen teil. Da keine genauen Informationen über die Anzahl der potenziell zu Befragenden vorlagen, sondern nur ein Richtwert von 1.000 bekannt war, kann die Rücklaufquote auf rund 17% geschätzt werden. Von den Befragten haben 92 Personen alle Fragen vollständig beantwortet. Fehlende Angaben liegen vor allem für die Fragen vor, die am Ende des Fragebogens platziert waren. Wie Abbildung 7 zu entnehmen ist, war die Teilnahme derjenigen, die unmittelbar im BGM involviert sind, besonders hoch. BGM-Koordinierende und BGM-Beauftragte stellen gemeinsam die Hälfte der Befragungsteilnehmenden dar, die eine Angabe zu ihrer Funktion machten. Unter Sonstige subsumierten sich unter anderem Personalratsmitglieder, Vorgesetzte oder Psychologen, aber auch Personen, die Angaben in einem Angestelltenverhältnis ohne weitere fachspezifische Qualifikation zu sein.

Wie Abbildung 8 zeigt, war ein Drittel der Befragten in Kommandobehörden und ein Fünftel in Sanitätseinrichtungen tätig. Den dritt- und vierthäufigsten Einsatzbereich der befragten Experten stellen die Verwaltung und sonstige Bereiche dar. Letzteres wurde insbesondere dann angegeben, wenn sich die Befragten mehreren Bereichen gleichzeitig zugehörig sahen.

Abbildung 7: Berufsgruppenzugehörigkeit der Teilnehmer an der Expertenbefragung (n = 92)

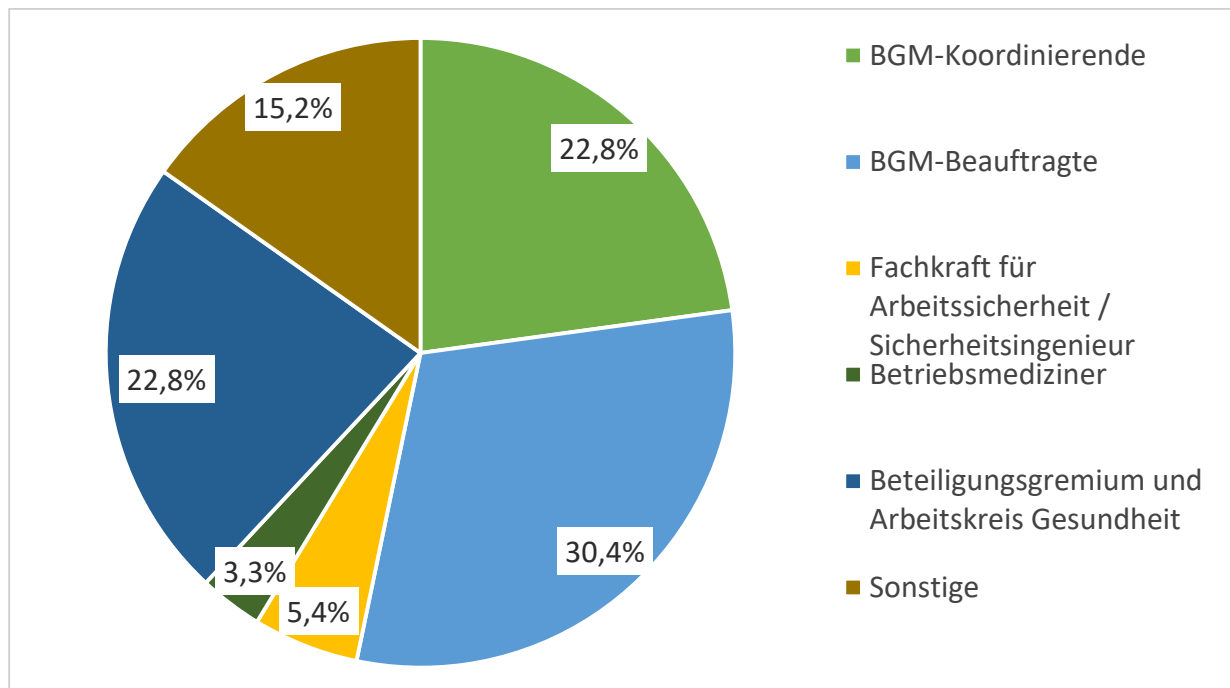
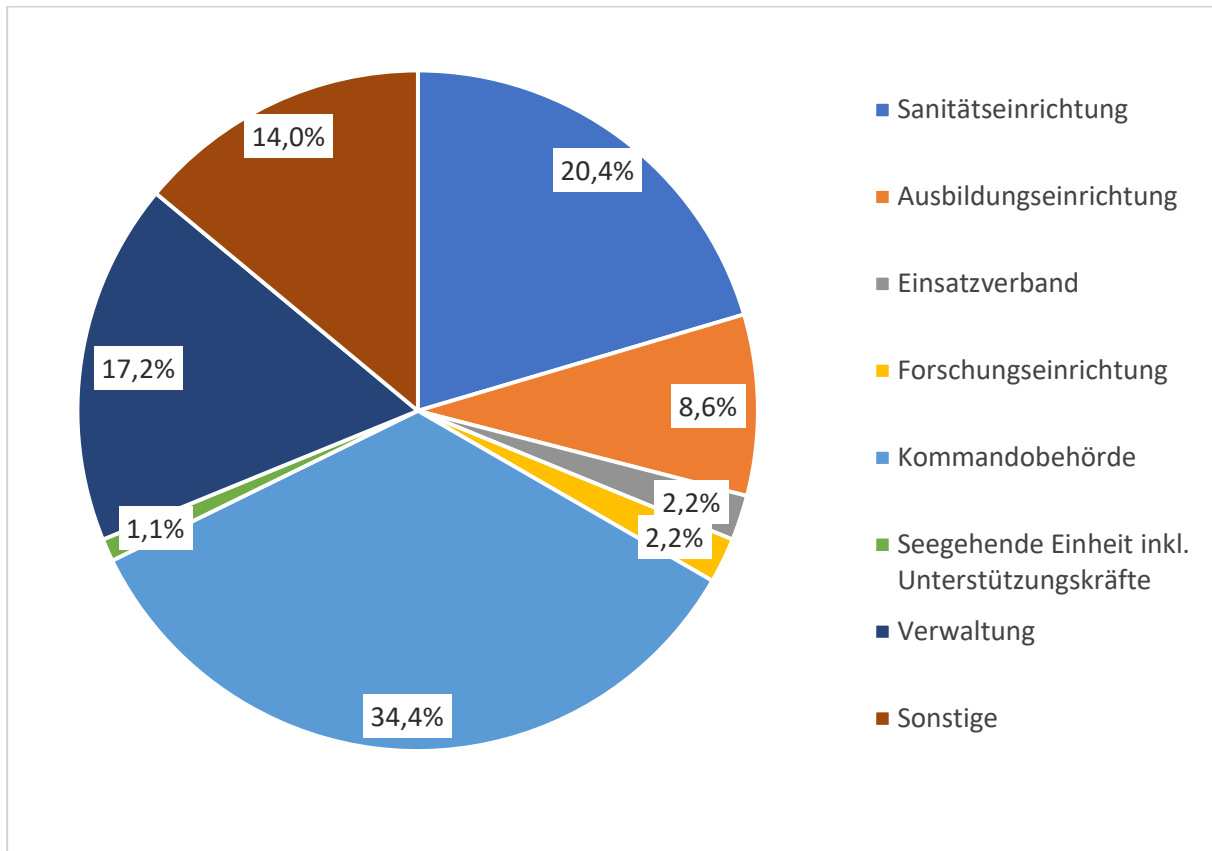


Abbildung 8: Einsatzbereich der Teilnehmenden an der Expertenbefragung (n = 93)



4.2.1 Quantitative Auswertungen zur Inhaltsvalidität

Die Bewertungen der Relevanz der einzelnen Items durch die befragten Experten ist in Tabelle 3 zusammengefasst. Zur Beurteilung wird wie in Kapitel 3.3.5 erläutert der Übereinstimmungskoeffizient Kappa berechnet, welcher den Anteil positiver Wertungen um die Wahrscheinlichkeit zufälliger positiver Bewertungen korrigiert. Im Durchschnitt bewerteten 86 Personen ein Item (min. 69 bis max. 103 Bewertungen). Werte größer als 0,78 gelten dabei als „Excellent“ und Werte über 0,59 als „Good“. Insgesamt wurden 27 Items mit „Excellent“ bewertet und 8 Items mit „Good“. Kein Item wurde schlechter bzw. als unzureichend relevant bewertet. Der niedrigste Wert, der erreicht wurde, betrug 0,63 für das Item Nr. 21 „Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material)“. Dieses Item ist im Durchschnitt aller Expertenurteile das am wenigsten relevante. Der höchste Wert wurde für Item Nr. 34 „Die Anerkennung meiner Arbeit durch meine unmittelbaren Vorgesetzten ist angemessen“ ermittelt und betrug 0,96. Dies scheint aus Sicht der Experten das relevanteste Item zu sein. Der Gesamtdurchschnitt über alle Items beträgt $k = 0,85$ und kann somit ebenfalls als „Exzellente“ bewertet werden.

Tabelle 3: Bewertung der Relevanz der Items im Rahmen der Expertenbefragung

Item-Nr.	Wortlaut	Anzahl der Bewertungen	Anzahl der positiven Bewertungen	Kappa	Evaluation
aa1	Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit ausgewogen.	103	95	0,92	Excellent
aa2	Die Anzahl der von mir verlangten Vertretungen / Bereitschaftsdienste ist angemessen.	103	71	0,69	Good
aa3	Meine Arbeitszeitverteilung ist mir angenehm.	103	86	0,83	Excellent
aa4	Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen.	103	75	0,73	Good
aa5	Meine Arbeitspausen sind ausreichend.	103	77	0,75	Good
aa6	Meine Arbeitspausen sind störungsfrei.	103	84	0,82	Excellent
am1	Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet.	92	81	0,88	Excellent
am2	Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit.	92	81	0,88	Excellent
am3	Meine Arbeit ist in angenehmem Maße abwechslungsreich.	92	69	0,75	Good
am4	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z. B. Personalverantwortung).	92	72	0,78	Excellent
am5	Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material).	92	58	0,63	Good
am6	Ich kann erkennen, ob ich meine Arbeit gut gemacht habe.	92	82	0,89	Excellent
am7	Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit.	69	64	0,93	Excellent
am8	Ich habe einen angemessenen Handlungsspielraum (z. B. Entscheidungsfreiheit) bei meiner Arbeit.	92	81	0,88	Excellent
am9	Ich werde bei meiner Arbeit selten unterbrochen.	103	88	0,85	Excellent
am10	Die nötigen Arbeitsmittel sind vorhanden.	91	85	0,93	Excellent

Item-Nr.	Wortlaut	Anzahl der Bewertungen	Anzahl der positiven Bewertungen	Kappa	Evaluation
am11	Arbeitsräume sind in ausreichender Anzahl vorhanden.	91	83	0,91	Excellent
am12	Die Einrichtung an meinem Hauptarbeitsplatz ist angemessen.	91	82	0,90	Excellent
sa1	Mein Kundenkreis für interne Dienstleistungen (z. B. auch Patienten, Leistungsempfänger) verhält sich angemessen.	69	50	0,72	Good
sa2	Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.	69	52	0,75	Good
sa3	Die Anerkennung meiner Arbeit durch meinen Mitarbeiterkreis ist angemessen.	69	63	0,91	Excellent
sa4	Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis.	69	64	0,93	Excellent
sa5	Mit meinen Mitarbeitenden gibt es keine Konflikte.	69	58	0,84	Excellent
sa6	Die Anerkennung meiner Arbeit durch meine unmittelbaren Vorgesetzten ist angemessen.	69	66	0,96	Excellent
sa7	Ich erfahre genügend Unterstützung von meinen unmittelbaren Vorgesetzten.	69	64	0,93	Excellent
sa8	Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte.	69	59	0,86	Excellent
sa9	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Beschäftigten ist angemessen.	69	59	0,86	Excellent
sa10	Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Beschäftigten.	69	59	0,86	Excellent
sa11	Mit den Untergebenen/den mir unterstellten Beschäftigten gibt es keine Konflikte.	69	57	0,83	Excellent
sa12	Konflikte werden offen besprochen.	69	65	0,94	Excellent
um1	Den Geräuschpegel an meinen Hauptarbeitsplatz empfinde ich als angenehm.	91	79	0,87	Excellent

Item-Nr.	Wortlaut	Anzahl der Bewertungen	Anzahl der positiven Bewertungen	Kappa	Evaluation
um2	Die äußeren Bedingungen (Klima, Beleuchtung, Belüftung etc.) an meinem Hauptarbeitsplatz sind angenehm.	91	81	0,89	Excellent
um3	Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen.	91	75	0,82	Excellent
um4	Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann (z. B. Arbeitshöhe, Sitzhöhe, Beinfreiheit, Durchgangsbreiten/Sicherheitsabstände).	91	84	0,92	Excellent

*Anmerkung: Kappa = Übereinstimmungskoeffizient Kappa (vgl. Kapitel 3.3.5);
Evaluation = Bewertung von Kappa nach den Empfehlungen von Polit et al. (2007)*

Auf Basis der Bewertungen der einzelnen Items konnte der CVI für die jeweilige Subskala ermittelt werden. Der Schwellenwert für ein akzeptables Ergebnis liegt mit 0,90 deutlich höher als auf Itemebene (140). Keine der vier Skalen der Arbeitsmerkmale erreichte diesen Wert. Folgende CVI wurden für die Skalen ermittelt:

- Arbeitsorganisation CVI = 0,784
- Arbeitsinhalt und Arbeitsaufgabe CVI = 0,802
- Soziale Aspekte CVI = 0,877
- Arbeitsumgebung CVI = 0,876

Neben den einzelnen Items waren die Experten aufgefordert, die gesamte Skala hinsichtlich ihrer Vollständigkeit zu bewerten. Die Antwortverteilung für die vier Dimensionen Arbeitsorganisation, Arbeitsinhalt und -aufgabe, Soziale Aspekte und Arbeitsumgebung sind in Abbildung 9 bis Abbildung 12 dargestellt. Für alle vier Skalen wurde die jeweilige Skala von einer absoluten Mehrheit als vollständig bewertet. Während der Anteil derjenigen, die die Skala als nicht vollständig beurteilte bei Arbeitsumgebung, Soziale Aspekte und Arbeitsinhalt und -aufgabe zwischen jeweils rund 7 und 8% lag, war der Stimmenanteil negativer Bewertungen für die Subskala Arbeitsorganisation deutlich höher und lag bei rund 23%. Es verwundert auch daher nicht, dass dies die Skala war, bezüglich der die meisten Befragten auf die Frage „Würden Sie weitere Aspekte (Items) zum Merkmalsbereich erfassen?“ mit „Ja“ beantworteten. Dieser Anteil betrug 28,7%, dicht gefolgt jedoch von der Arbeitsumgebung mit 28%. Für den

Arbeitsinhalt schlugen 19,3% eine Erweiterung vor, bezüglich der Sozialen Aspekte sahen die wenigsten Befragten, nämlich 16,5% einen Ergänzungsbedarf.

Abbildung 9: Bewertung der Vollständigkeit der Subskala Arbeitsorganisation

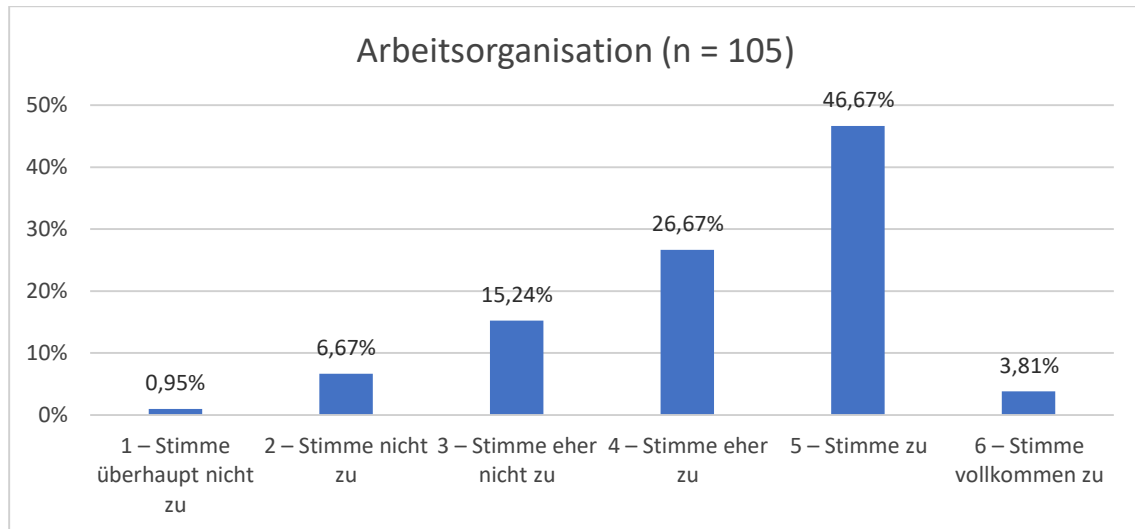


Abbildung 10: Bewertung der Vollständigkeit der Subskala Arbeitsinhalt und -aufgabe

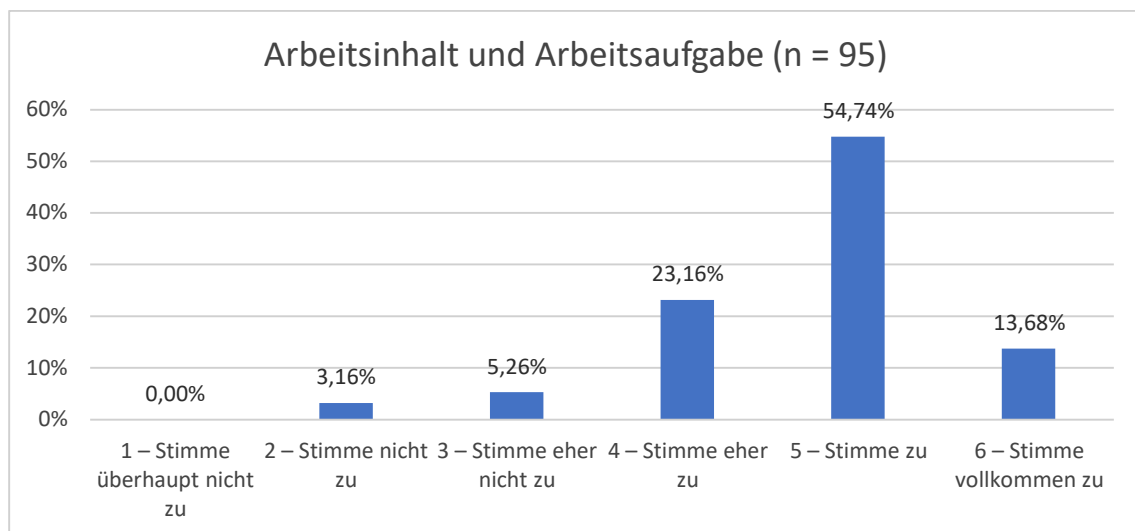


Abbildung 11: Bewertung der Vollständigkeit der Subskala Soziale Aspekte

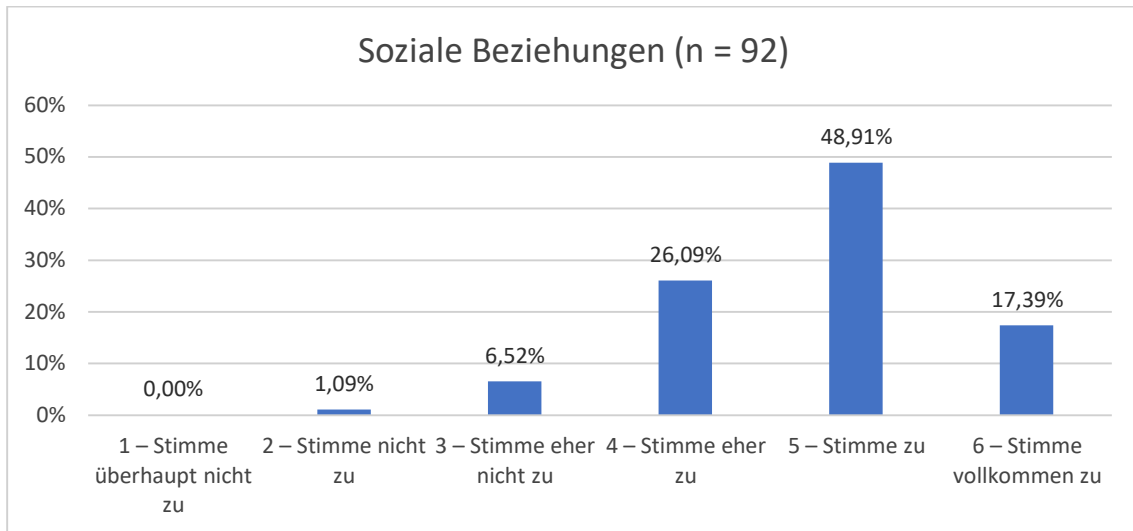
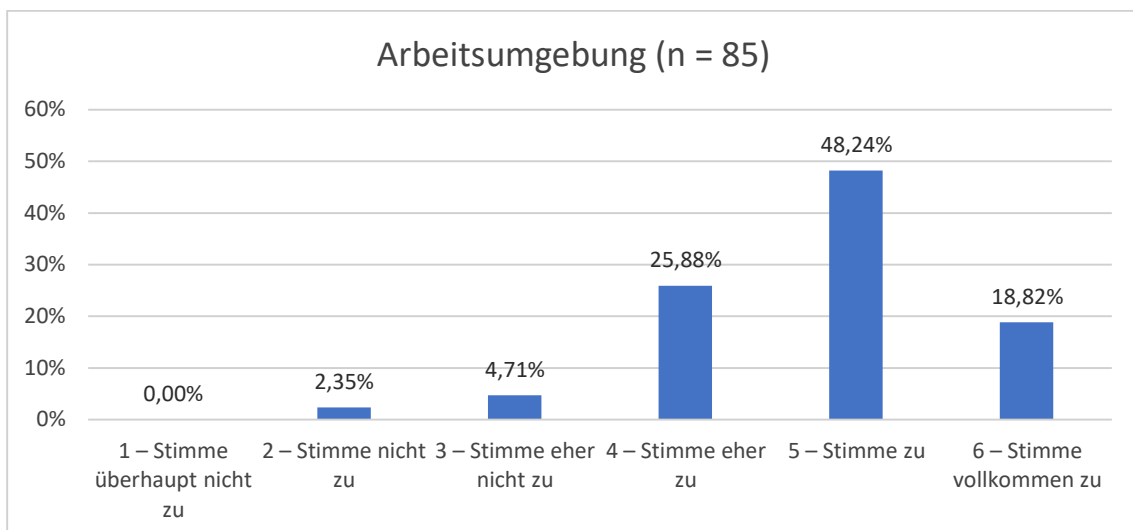


Abbildung 12: Bewertung der Vollständigkeit der Subskala Arbeitsumgebung



Im Rahmen der Datenauswertung wurde auch geprüft, ob diese Bewertungen berufsgruppenspezifisch variierten. Hier zeigte sich kein klares Muster. Lediglich die Gruppe der Personen, die für ihren Einsatzbereich die Antwortoption „Sonstiges“ auswählten, wiesen ein tendenziell kritischeres Antwortverhalten auf. Bei drei der vier Subskalen wies diese Gruppe den geringsten Anteil von Zustimmung bei der Frage nach der inhaltlichen Vollständigkeit („Der Merkmalsbereich [XY] ist vollständig repräsentiert“) der jeweiligen Skala auf. So betrug der Anteil der gegebenen Antworten im Bereich 1 „Stimme überhaupt nicht zu“ bis 3 „Stimme eher nicht zu“ für die Arbeitsorganisation bei der Gruppe „Sonstige“ 38,5% während alle anderen Berufsgruppen im Mittel nur mit 17% nicht zustimmten. Bezüglich der Sozialen Aspekte standen 21,4% einem Durchschnitt von 2,6% bei allen anderen Berufsgruppen gegenüber und bezüglich der Arbeitsumgebung kam die Gruppe der „Sonstigen“ auf 15,4%, die restlichen Berufsgruppen im Mittel auf 3,8%. Lediglich für die

Subskala Arbeitsinhalt und Arbeitsaufgabe lag der Anteil der Betriebsmediziner, die die Skala nicht vollständig repräsentiert sahen unter allen Berufsgruppen am höchsten (33,3%). Umgekehrt konnte kein Muster bezüglich überproportional positiver Antworten einer Berufsgruppe festgestellt werden.

Der Anteil derjenigen, die weitere Aspekte im Merkmalsbereich erheben würden, ist deutlich höher als der Anteil derjenigen, die den Merkmalsbereich nicht vollständig repräsentiert sehen wie Tabelle 4 zu entnehmen ist. Die Definition der jeweiligen Merkmalsbereiche wurde hingegen von nahezu allen Befragten als inhaltlich vollständig bewertet. Jeweils 95% bewerteten die Definition der Dimensionen Soziale Beziehung und Arbeitsinhalt und Arbeitsaufgabe als vollständig, bei der Arbeitsumgebung waren es 96%. Der Bereich „Arbeitsorganisation“ erzielt mit 90% Zustimmung den niedrigsten Wert. Noch etwas höher liegen die Zustimmungswerte bei der gegenteiligen Frage „Ist die Definition des Merkmalsbereichs Ihrer Meinung nach zu weit gefasst?“. Hier verneinten 98% (n = 82) bezüglich der Sozialen Aspekte, 97% (n = 91) bezüglich der Arbeitsorganisation, 96% (n = 78) bezüglich der Arbeitsumgebung und 95% (n = 82) bezüglich Arbeitsinhalt und Arbeitsaufgabe.

Tabelle 4: Antwortverteilung auf die Frage „Würden Sie weitere Aspekte (Items) zum Merkmalsbereich erfassen?“ (n = 116)

<i>Subskala</i>	<i>Antwort</i>	<i>n</i>	<i>%</i>
Arbeitsorganisation (n = 94)	Ja	27	28,7%
	Nein	67	71,3%
Arbeitsinhalt und Arbeitsaufgabe (n = 88)	Ja	17	19,3%
	Nein	71	80,7%
Soziale Aspekte (n = 85)	Ja	14	16,5%
	Nein	71	83,5%
Arbeitsumgebung (n = 82)	Ja	23	28,1%
	Nein	59	72,0%

Tabelle 5: Antwortverteilung auf die Frage „Ist die Definition des Merkmalsbereichs Ihrer Meinung nach vollständig?“ (n = 96)

<i>Subskala</i>	<i>Antwort</i>	<i>n</i>	<i>%</i>
Arbeitsorganisation (n = 90)	Ja	81	90,0%
	Nein	9	10,0%
Arbeitsinhalt und Arbeitsaufgabe (n = 84)	Ja	80	95,2%
	Nein	4	4,8%
Soziale Aspekte (n = 81)	Ja	77	95,1%
	Nein	4	4,9%
Arbeitsumgebung (n = 75)	Ja	72	90,4%
	Nein	3	9,6%

4.2.2 Ergebnisse der quantitativen Inhaltsanalyse

Die Diskussion der Codierenden nach Bearbeitung von ca. 60% des Textmaterials ergab, dass eine Überarbeitung des Kodierleitfadens nicht notwendig war. In Tabelle 6 sind die Arbeitsschritte aus dem gemeinsam festgelegten Kodierleitfaden zusammengefasst. Zudem wurde sich im Vorfeld darauf verständigt, dass die gebildeten Kategorien sich am Kategorienschema der GDA für psychische Belastung orientieren sollen. In einem ersten Arbeitsschritt wurde das Textmaterial auf inhaltstragende Elemente reduziert. Nicht inhaltstragende Elemente sind z. B. Füllwörter, rhetorische Fragen oder auch unverständliche Formulierungen. Nicht relevante Elemente sind Äußerungen, die nicht der Beantwortung der Fragestellungen dienen und auch keinen Hinweis darauf geben, wie der Fragebogen verbessert werden kann. Anschließend erfolgte eine Paraphrasierung der Aussagen auf eine einheitliche, neutrale Sprachebene, welche ausschließlich sachbezogene Aussagen enthielt. Diese Aussagen wurden anschließend in einem dritten Schritt nochmals verkürzt unter einem bestimmten Schlagwort subsumiert. Diese Schlagworte entsprechen dem Kategorienschema, auf das sich die Codierer gemeinsam verständigt hatten. Dieses Kategorienschema war jedoch für die spätere Auswertung zu differenziert, sodass eine weitere Generalisierung der einzelnen Schlagworte vorgenommen werden musste, um zu viele Einzelnennungen zu vermeiden. Beispielsweise wurden die Schlagworte „Tiefergehende und umfassendere Informationserhebung wünschenswert“ und „Frage erfasst nicht vollständigen Bereich“ weiter zu „Antwortmöglichkeiten müssen erweitert und detaillierter werden“ generalisiert. Häufig wurden von den Befragten aber auch kurze Stichworte als Antwort gegeben, die sich nicht

weiter in ihrem Inhalt verkürzen oder abstrahieren ließen. In solchen Fällen wurde von einer weiteren Abstraktion abgesehen. Auf eine Generalisierung der Kategorien über alle Teilfragen wurde abgesehen, da die Fragen bereits so spezifisch waren und unterschiedliche Inhaltsbereiche thematisierten, dass wenige vergleichbare Kategorien hätten entwickelt werden können.

Tabelle 6: Im Kodierleitfaden definierte Arbeitsschritte

Arbeitsschritt	Beispiel
1. Paraphrasieren	
a. Streichung aller nicht inhaltstragenden oder relevanten Elemente*	<i>„Da hier aus meiner Bewertung nicht wirklich ein zielführender Weg beschritten wird, würde ich mich sehr darüber freuen, wenn man über Ursachen sprechen würde um eher die Möglichkeit zu finden in einem Fragebogen solche zu manifestieren um dementsprechend als Folgerung besser in die richtigen Richtungen zu fahren. Warum sollte man feststellen wollen wo alles toll ist? Für mich wird vom Fragebogen schon sehr viel erreicht, aber leider zu oberflächlich. Natürlich könnte man hier und da dann auch weiterführende Gespräche oder sonstiges durchführen, ist das denn immer und überall so gewollt. Aus meiner Sicht würde eine Beurteilung von unten nach oben klarer die Handlungsfelder erkennen lassen. Beide natürlich, von oben und unten.“</i>
b. Übersetzung der inhaltstragenden Elemente auf eine einheitliche Sprachebene	Kommentar 1: „... zu oberflächlich. Natürlich könnte man hier und da dann auch weiterführende Gespräche oder sonstiges durchführen“ → Abstraktion 1: Tiefergehende und umfassendere Informationserhebung wünschenswert Kommentar 2: „Die Formulierung "Belastungen und Beanspruchungen am Arbeitsplatz führen bei mir zu ..." beschränkt sich nicht nur auf psychische Belastungen“ → Abstraktion 2: Frage erfasst nicht vollständigen Bereich Kommentar 3: „Die letzten beiden Fragen setzen voraus, dass die Beschwerden auf die Arbeit zurückzuführen sind. Weder ist das notwendig der Fall ("Freizeitstress"), noch ist das durch den Befragten immer zweifelsfrei zu erkennen. In den Fragen ist das in Ordnung - bei der Auswertung sollte hier aber entsprechende Vorsicht walten.“ → Abstraktion 3: Beanspruchungen außerhalb des Arbeitssettings werden nicht kontrolliert
2. Generalisierung innerhalb der Dimension (induktive Kategorienbildung)	Sind im Kategorienschema der jeweiligen Subdimension festgelegt
a. Alle Paraphrasen sollen auf die gleiche, zuvor definierte Abstraktionsebene	Abstraktion 1: Tiefergehende und umfassendere Informationserhebung wünschenswert Abstraktion 2: Frage erfasst nicht vollständigen Bereich → Generalisierung 1 & 2: Antwortmöglichkeiten müssen erweitert und detaillierter werden

gebracht bzw. übertragen werde	Abstraktion 3: Beanspruchungen außerhalb des Arbeitssettings werden nicht kontrolliert → Generalisierung 3: Keine Kontrolle von Confoundern
b. Paraphrasen, die bereits über dem angestrebten Abstraktionsniveau liegen, werden vorerst so belassen	Kommentar 4: „Selbstwertgefühl und Akzeptanz“ → Abstraktion 4: „Selbstwertgefühl“ → Abstraktion 5: „Akzeptanz“ → Generalisierung 4: Akzeptanz → Generalisierung 5: Selbstwertgefühl
c. Entfällt: Generalisierung über alle Teilfragen	Wenige Kategorien werden dimensionsübergreifend verwendet, inhaltlich nicht sinnvoll

Da der Kodierleitfaden keiner Überarbeitung aus Sicht der Codierenden bedurfte, konnte unmittelbar die Überprüfung der InterCoderreliabilität (ICR) vorgenommen werden. Da hier eine zufriedenstellende mittlere ICR von 84% erreicht wurde, konnte das Datenmaterial vollständig codiert werden. Insgesamt wurden 552 Freitextantworten abgegeben, die im Rahmen der quantitativen Inhaltsanalyse ausgewertet werden konnten. Die mittlere ICR für die vollständige Codierung des Datenmaterials lag bei 89%.

Unter den Freitextantworten entfielen 176 auf Filterfragen, die den geschlossenen Fragen zu den Arbeitsmerkmalsbereichen folgten. Tabelle 7 gibt eine Übersicht, wie häufig welche Aspekte in welchen Merkmalsbereichen genannt wurden. Der am häufigsten kommentierte Merkmalsbereich ist die Arbeitsumgebung mit insgesamt 52 Nennungen, wobei insbesondere die Kategorien Sozialräume (n = 12) und Arbeitssicherheit (n = 14) häufig thematisiert wurden. Ein gutes Viertel aller Kommentare entfällt auf den Bereich Arbeitsorganisation, wo Kommunikation und Kooperation (n = 14) besonders häufig angesprochen wurden. Für den Bereich Arbeitsinhalt und –aufgabe wurden insgesamt 40 Freitexte codiert, ein Drittel davon in der Kategorie Qualifikation. Der Merkmalsbereich Soziale Aspekte weist mit 23 Freitexten etwas weniger Nennungen auf, unter anderem auch, weil der Bereich Führung mit insgesamt 10 Kommentaren separat ausgewertet wurde. Zudem ließen sich 5 weitere Antworten keinem Merkmalsbereich zuordnen und wurden unter Sonstiges zusammengefasst.

Weitere Anmerkung zur Erfassung der Arbeitsmerkmale wurden bereits im Rahmen der ersten Frage „Bitte notieren Sie, welche Aspekte Ihrer Meinung nach bezüglich der beschriebenen Anforderungen im Fragebogen enthalten sein müssen“ gemacht. Für diese Frage wurden insgesamt 376 Textsegmente zu folgenden Merkmalsbereichen ausgewertet:

- Bedarfs-/Wirksamkeitsanalyse, Evaluation und Umsetzung des BGM (n = 70)
- Erhebungs- und Auswertungsdesign (n = 69)
- Arbeitssituation (n = 83)

- Psychische Belastungen und Gefährdungen (n = 40)
- Individuelles Gesundheitsverhalten (n = 15)
- Gesundheitssituation (n = 41)
- Individuelle Beanspruchungsreaktionen (n = 26)
- Soziale Aspekte (n = 19)
- Führung (n = 13)

Die genaue Antwortverteilung und die jeweiligen Unterkategorien sind in Anlage 4 dargestellt. Wesentliche Ergebnisse dieser Frage sind in Bezug auf den ersten Punkt, dass die Experten sowohl der Analyse der Ist-Situation und der Identifikation von Handlungsschwerpunkten eine hohe Bedeutung zuschrieben, aber auch die Erwartungshaltung an das Instrument hatten, dass es die BGM-Maßnahmen evaluiert. Inhaltlich wurden bereits zentrale Aspekte der Beurteilung der Arbeitssituation benannt. Dies betrifft alle vier Hauptdimensionen, Arbeitsorganisation, Soziale Aspekte, Arbeitsinhalt und -aufgabe sowie die Arbeitsumgebung. In 40 Textsegmenten wurde explizit die Beurteilung der psychischen Belastungssituation gefordert. Hier wurde mehrheitlich (n = 31) die Beurteilung psychischer Gefährdung als Notwendigkeit einer vollständigen Arbeitsplatzanalyse genannt.

Die Experten sahen zudem Bedarf, auch Faktoren auf Individualebene zu berücksichtigen. Dies betraf einerseits das individuelle Gesundheitsverhalten (n = 15), andererseits auch die Gesundheitssituation der Beschäftigten (n = 41). Zudem wurde in 26 Aussagen gefordert, nicht nur die Belastungssituation, sondern auch individuelle Beanspruchungsreaktionen zu erfassen.

Ferner wurden verschiedenen Vorschläge zur Verbesserung des Erhebungsdesigns oder zur inhaltlichen Ausgestaltung des Fragebogens gemacht. So forderte ein Drittel der Antwortenden (n = 23), zusätzliche Fragebogeninhalte wie beispielsweise die Zufriedenheit mit den bisherigen angebotenen Maßnahmen, Fragen zur Telearbeit oder detaillierte Informationen zum ausgeübten Dienstposten und Nebenfunktionen zu erheben. Letzterer Aspekt wurde so häufig genannt (n = 22), dass er als eigene Kategorie codiert wurde. Unter der Kategorie Fragebogengestaltung wurden auch die fünf Kommentare erfasst, die den Fragebogen in seiner aktuellen Fassung als inhaltlich vollständig empfinden und keine Veränderungen vornehmen würden. Andere Experten wiesen darauf hin, dass sich bestimmte Einheiten (z. B. Schiffsbesatzung) nicht mit den Fragen identifizieren konnten und regten eine inhaltliche Abstimmung auf bestimmte Organisationsbereiche an.

Tabelle 7: Kommentare zu den Arbeitsmerkmalen im Rahmen der Expertenbefragung (n = 174)

Merkmalsbereich	Kategorie	Nennungen absolut	Anteil im Merkmalsbereich	Anteil an allen Nennungen (n = 174)	Beispiel (Originalformulierungen)
Arbeitsorganisation (n = 46)	Transparenz	3	6,5%	1,7%	„z.B. Zufriedenheit mit Transparenz/Teilhabe von/an Entscheidungen“
	Personalplanung	3	6,5%	1,7%	„Durch hohen Personalwechsel entstehende mehr Belastungen“
	Kommunikation und Kooperation	14	30,4%	8,0%	„Möglichkeiten (regelmäßig) zu Absprache/ Abstimmung/ Austausch/ Koordination/ Teambesprechungen mit Kollegen und Dienststellenleiter“
	Ergebniskontrolle / Rückmeldung durch die Tätigkeit	4	8,7%	2,3%	„Rückmeldungen zum Arbeitsergebnis und ggf. Anpassung des Arbeitsprozesses“
	Arbeitsablauf	9	19,6%	5,1%	„störungsfreies Arbeiten“
	Arbeits-/Pausenzeit	8	17,4%	4,5%	„Einflussnahme/Planbarkeit von Arbeitszeiten und Pausen“
	Verhältnis Arbeitsmenge Arbeitszeit	5	10,9%	2,8%	„Belastungsspitzen und Zeiten in denen viele parallele Aufträge eingehen“
Arbeitsinhalt und -aufgabe (n = 40)	Information, Ergebniskontrolle, Transparenz und Vollständigkeit	9	22,5%	5,1%	„Mir wird der Gesamtzusammenhang meiner Tätigkeit im Großen Ganzen erläutert bzw. sehe/verstehe mich als Teil des Ganzen.“
	Qualifikation (inkl. Über-/ Unterforderung)	13	32,5%	7,4%	„Bin ich überhaupt richtig eingesetzt, oder führe ich Aufgaben durch, welche eigentlich nicht auf meinem Dienstposten/Arbeitsplatz vorhanden sind.“

Soziale Aspekte (n = 23)	Handlungsspielraum / Autonomie und Mitsprache	10	25,0%	5,7%	„Handlungsspielraum ist nicht nur "Entscheidungsfreiheit" (Sachentscheidungen zum Arbeitsgebiet), sondern vor Allem Spielraum in der Arbeitsverteilung“
	Emotionale Inanspruchnahme	4	10,0%	2,3%	„Emotionale Inanspruchnahme, weil dies großen Einfluss auf den Arbeitsauftrag hat. Speziell bei Situationen des täglichen Arbeitsalltags, in Bezug auf die sich immer verschlechternde Gesamtsituation in der Bundeswehr.“
	Sinnhaftigkeit	4	10,0%	2,3%	„Ich sehe einen Sinn in der Arbeit, die ich verrichte“
	Arbeitsklima	8	34,8%	4,5%	„Soziale Eingebundenheit mit den Mitarbeitern in der Freizeit.“
	Kommunikation / Kooperation	4	17,4%	2,3%	„Anzahl der Interaktionen mit Kollegen und/oder Vorgesetzten pro Tag um die täglichen Aufgaben erfüllen zu können. (Hintergrund: Einfluss der Häufigkeit von zwischenmenschlicher Interaktion extravertierter / introvertierter Menschen hat Einfluss auf Arbeitszufriedenheit)“
	Mobbing	3	13,0%	1,7%	„werden einzelne Personen ausgeschlossen/ gemobbt“
	Soziale Beziehungen zu Dritten	2	8,7%	1,1%	„Die gleichen Fragen, aber im Bezug zu Kunden, Partner“
	Organisatorische Einbindung in das Arbeitssystem	2	8,7%	1,1%	„die organisatorische Einbindung der eigenen Arbeit in das Arbeitssystem“
	Umgang mit Konflikten und Fehlerkultur	4	17,4%	2,3%	„Wie wird im Kollegenkreis mit starken Emotionsschwankungen einzelner Mitarbeiter umgegangen? Wie regelt die Gemeinschaft solche Herausforderungen?“

Führung (n = 10)	Führungsstil	3	30,0%	1,7%	„Ansprechbarkeit, Vertrauenswürdigkeit und Führungsfähigkeit von Vorgesetzten“
	Faire Arbeitsverteilung	2	20,0%	1,1%	„muss man Aufgaben für andere erledigen, fühlt man sich bei der Arbeits- und Aufgabenverteilung benachteiligt gegenüber anderen etc.“
	Rückmeldung	5	50,0%	2,8%	„Feedback der Ergebnisse/Arbeitsweise/Zwischenergebnisse“
Arbeitsumgebung (n = 52)	Arbeitssicherheit	14	26,9%	8,0%	„Wird die persönliche Schutzausrüstung als ausreichend und gut empfunden (Bsp.: Gehörschutz, Mundschutz,..)“
	Infrastrukturelle Anbindung	3	5,8%	1,7%	„Liegt die Dienststelle für Kasernenschläfer in einem Bereich mit guter Infrastruktur?“ „Haben sie die Möglichkeit ihren Arbeitsplatz selbst zu gestalten?“
	Gestaltungsspielraum bzgl. Arbeitsumgebung	3	5,8%	1,7%	„Haben sie die Möglichkeit andere Einrichtungen anzufordern?“ „Dürfen sie eigene Einrichtungen zu ihrem Arbeitsplatz mitbringen?“
	Sozialräume	12	23,1%	6,8%	„Sozialräume sind ausreichend und ansprechend Sanitäranlagen sind ausreichend dimensioniert und geschlechterabhängig vorhanden (m-w-d)“
	Ergonomie	7	13,5%	4,0%	„individuelle leidens- oder behindertengerechte Arbeitsplatzgestaltung“
	biologische, chemische und physikalische Faktoren	7	13,5%	4,0%	„die persönliche Belastung mit Gefahrstoffen und Biostoffen ist so gering wie möglich“
	Arbeitsmittel	6	11,5%	3,4%	„Arbeitsmittel vorhanden oder gibt es Mängel?“
Sonstiges (n = 5)	Kontakt zum Betriebsarzt	1	20,0%	0,6%	„Ich kenne den zuständigen Betriebsarzt und war binnen der letzten 12 Monate dort vorstellig.“
	Vereinbarkeit mit Familie und Privatleben	3	60,0%	1,7%	„Work-Life Balance? Erreichbarkeit außerhalb des Dienstes? Diensthandy? Unterbrechungen von Freistellungen?“

Organisationsaufbau	1	20,0%	0,6%	<i>Eine Frage nach der grundsätzlichen aufbauorganisatorischen Ausrichtung (Einlinien-, Stablinien-, Matrixorganisation) der Dienststelle wäre wünschenswert.</i>
---------------------	---	-------	------	---

4.2.3 Vergleich von Expertenmeinungen und GDA-Empfehlung

Neben der Überprüfung der inhaltlichen Vollständigkeit und der Relevanz einzelner Items, sollte der Frage nachgegangen werden, inwiefern sich die Expertenmeinungen im Geschäftsbereich des BMVg mit offiziellen Empfehlungen der GDA zur Beurteilung psychischer Belastung decken. Zur Beantwortung dieser Frage sind zusätzlich zu den 174 Kommentaren, die zur Erhebung der Arbeitsmerkmale gemacht wurden, weitere 115 Kommentare zu berücksichtigen, welche bereits im Rahmen der Eingangsfrage („Bitte notieren Sie, welche Aspekte Ihrer Meinung nach bezüglich der beschriebenen Anforderungen im Fragebogen enthalten sein müssen“) gegeben wurden und den Bereichen Arbeitsorganisation, Soziale Aspekte und Führung zugeordnet wurden. Für die Aspekte, die bereits in Form eines Items im AIGScreen repräsentiert waren, wurde zusätzlich die Bewertung der Relevanz durch die Experten in Form des Koeffizienten Kappa (k) berücksichtigt.

Die hier dargestellten Ergebnisse mussten im Sinne der Übersichtlichkeit in ihrem Informationsgehalt stark reduziert werden. Es sei jedoch darauf hingewiesen, dass es häufig auch widersprüchliche Einschätzungen zu den einzelnen Aspekten gab. So forderten beispielsweise zwei Befragte, die Anerkennung durch die Mitarbeitenden zu streichen, wohingegen die Mehrheit dieses Items als sehr relevant ($k = 0,91$) bewertete. Gab es solche diametralen Positionen wurde in den Ergebnissen die Mehrheitsmeinung berichtet und ggf. spezifische Anregungen zu Teilaspekten gesondert berücksichtigt und in Tabelle 8 ausgewiesen.

Für die Mehrheit der Aspekte wurden vor allem Vorschläge gemacht, welche zusätzlichen Merkmale berücksichtigt werden sollten, allerdings gab es auch Aspekte, die aus Expertensicht ersatzlos gestrichen werden können. Sehr eindeutig fielen die Meinungen bezüglich der Aspekte Variabilität und Verantwortung aus, deren Streichung von mehreren Personen gefordert wurde und die auch keinen zufriedenstellenden Wert für den Übereinstimmungskoeffizienten Kappa erreichten. In einigen Bereichen wurden Teilaspekte kritisch bewertet. Dies betrifft die physischen Faktoren, die Arbeitszeit und den Aspekt Verantwortung. Letztere sollte aus Sicht einiger Experten um die Thematik Objektverantwortung gekürzt werden. Positive Meinungen gab es zu diesem Teilaspekt nicht und auch Kappa ist der niedrigste im Vergleich zu allen anderen Items. Besonders häufig wurde die Streichung des Items "Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen" gefordert. Bezüglich der physischen Arbeitsbedingungen spielt die körperliche Anforderung in den Augen der Experten keine relevante Rolle.

Besonders häufig wurde im Merkmalsbereich Arbeitsumgebung vorgeschlagen, das Vorhandensein, die Einrichtung, den Zustand und die hygienischen Bedingungen von Sozial-

und Gemeinschaftsräumen zu berücksichtigen. Aber auch die Erhebung von Gefahrstoffexpositionen wurde in diesem Bereich häufiger gefordert.

Im Bereich Soziale Aspekte wurden besonders viele Ergänzungsvorschläge vor allem zu den Aspekten Führung und Arbeitsklima gemacht. Letzteres wurde gelegentlich auch im Zusammenhang mit den Vor- und Nachteilen von Teamarbeit thematisiert und floss in die Auswertung der Vielzahl von Kommentaren zum Aspekt Kommunikation und Kooperation im Merkmalsbereich Arbeitsorganisation ein, da es vordergründig um arbeitsorganisatorische Themen ging. So wurde in diesem Zusammenhang beispielsweise häufiger angemerkt, dass auch eine zu hohe Interaktionsdichte bzw. Kommunikationserfordernisse eine Belastung darstellen können. Außerdem war im Bereich Arbeitsorganisation die Möglichkeit der flexiblen Arbeitszeitgestaltung häufiger ein Thema.

Bezüglich des Arbeitsinhalts und der Arbeitsaufgabe spielte die Verfügbarkeit von tätigkeitsrelevanten Informationen in den Kommentaren eine große Rolle und sollte aus Sicht der Experten umfassender erfragt werden. Die meisten Erweiterungsvorschläge erfolgten in diesem Merkmalsbereich allerdings zum Aspekt Qualifikation. Hier wurde in 10 Textsegmenten die Erfassung von Über- oder Unterforderung bzw. allgemein die Übereinstimmung von individuellen Kompetenzen und Anforderungen der Tätigkeit thematisiert. Darüber hinaus spielt in diesem Zusammenhang aus Expertensicht die Sinnhaftigkeit der Aufgabe eine wesentliche Rolle. Des Weiteren wurde von Einzelpersonen die Erhebung von Aufgabenkomplexität und Qualität der Einarbeitung gefordert.

Zusätzliche Themen, die genannt wurden, aber weder Gegenstand des AIGScreen noch der GDA-Empfehlungen sind, waren Arbeitssicherheit, verschiedene Aspekte bzgl. Work-Life-Balance, infrastrukturelle Anbindung des Arbeitsplatzes sowie mehrere zusätzliche Themen im Bereich Soziale Aspekte (Arbeitsklima, Mobbing, Führungsstil).

Tabelle 8: Zusammenfassung der Kommentare (n = 289) in der Expertenbefragung zu einzelnen Aspekten

Kategorie	Begriffsbestimmung	Item im AIGScreen	Einschätzung der Experten
Arbeitsorganisation			
Kommunikation und Kooperation^a	Kommunikation und Kooperation als Teilaspekt der Arbeitsorganisation umfassen den Informationsaustausch und die Abstimmung mit anderen Personen innerhalb der Organisation. Gefährdungen ergeben sich insbesondere durch eingeschränkte oder fehlende Kommunikation, z. B. bei Einzelarbeit oder überwiegend indirekte Kommunikation sowie durch unklare Verantwortlichkeiten oder fehlende Unterstützung (98, 31).		(+) Für diesen Merkmalsbereich das am häufigsten thematisierte und am relevantesten bewertete Thema. Allerdings sollte aus Sicht der Befragten auch eine mögliche Gefährdung durch zu viele Kommunikations- und Abstimmungserfordernisse erfasst werden.
Arbeitsablauf^a	Störungen und Unterbrechungen im Arbeitslauf sind Hindernisse, die bedingen, dass die Arbeitskraft ihrer Tätigkeit aufgrund von vorhersehbaren oder unvorhersehbaren Barrieren in der Arbeitsorganisation, d. h. der planmäßigen Ausübung von Tätigkeiten, abbrechen muss. Das Störungsausmaß lässt sich in Abhängigkeit davon bestimmen, ob und wie eine Handlung nach der Unterbrechung wieder aufgenommen wird (98).	Ich werde bei meiner Arbeit selten unterbrochen	(+) Item („Ich werde bei meiner Arbeit selten unterbrochen“) wurde im Durchschnitt von allen Experten als relevant bewertet (k = 0,85) und Arbeitsablauf wird in vielen Kommentaren vertieft thematisiert und als wichtig hervorgehoben

Arbeitszeit^a	Die Arbeitszeit ist gemäß dem Arbeitszeitgesetz definiert als die Zeit von Beginn bis zum Ende der Arbeit ohne Pausen. Neben der reinen Expositionsdauer bemisst sich die Belastungsintensität durch die Lage und Verteilung der Arbeitszeit und Bereitschaftsdiensten sowie der Pausen- und Ruhezeiten (185).	Meine Arbeitspausen sind ausreichend. Meine Arbeitspausen sind störungsfrei. Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen. Meine Arbeitszeitverteilung ist mir angenehm.	(+) flexible Arbeitszeit sollte aufgenommen werden (-) Verhältnis von Arbeits- und Freizeit, ausreichende Pausen und Arbeitszeitverteilung sind gesetzlich geregelt, können daher objektiv beurteilt werden und bedürfen keiner subjektiven Einschätzung (-) Die durchschnittliche Bewertung der Relevanz für ausreichende Arbeitspausen fiel niedrig aus (k = 0,75) und auch das Verhältnis von Arbeitszeit zu Freizeit ist in den Augen der Experten nicht sehr relevant (k = 0,73) (+) Die übrigen Items erreichten gute bis sehr gute Bewertungen der Relevanz (störungsfreie Arbeitspausen: k = 0,82; Arbeitszeitverteilung: k = 0,83)
Überstunden^d	Jegliche Mehrarbeit über die vertraglich festgelegte Arbeitszeit hinaus geht	Die Anzahl der von mir verlangten Überstunden ist angemessen. ¹⁰	(-/+) Überstunden werden zwar nicht als ausreichend relevant bewertet (k = 0,75), wurden aber gleichzeitig an mehreren Stellen in den Freitextkommentaren als wichtiger Aspekt genannt

¹⁰ Dieses Item wurde erst im Rahmen einer Überarbeitung im Fragebogen im November 2019 ergänzt und ist daher nicht in allen Auswertungen zu finden. Da es jedoch im Rahmen der Expertenbefragung Gegenstand der Analyse war, ist es hier der Vollständigkeit halber mit aufgeführt.

Bereitschaftsdienste^d	Zeitspannen, in denen sich Arbeitnehmende für einen sofortigen oder zeitnahen Arbeitseinsatz bereit halten müssen	Die Anzahl der von mir verlangten Vertretungen / Bereitschaftsdienste ist angemessen.	(-) Item wird als wenig relevant bewertet (k = 0,69), ein Experte forderte explizit die Streichung des Items wegen Irrelevanz
Arbeitsintensität^b	Die Arbeitsintensität beschreibt die quantitativen Anforderungen einer Tätigkeit im Hinblick auf das Verhältnis von geforderter Arbeitsmenge zu verfügbarer Zeit und das Verhältnis von Aufgabenkomplexität zur verfügbaren Zeit (98).	Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit ausgewogen.	(+) Hat unter allen Items die höchste Relevanzbewertung (k = 0,92) und wurde mehrfach als wichtig in den Freitextkommentaren genannt

Arbeitsinhalt und Arbeitsaufgabe

Handlungsspielraum^a	Der Handlungsspielraum ist definiert durch den Umfang, in dem zeitliche Abfolge, Zeitpunkt und Dauer sowie Arbeitsweise, Auswahl der Arbeitsmittel und Festlegung der Arbeitsziele selbst bestimmt sind. Es kann weiter nach Autonomie in den Subdimensionen a) Planung b) Methode und c) Entscheidungen differenziert werden (96).	Ich habe einen angemessenen Handlungsspielraum (z. B. Entscheidungsfreiheit) bei meiner Arbeit.	(+) Wurde mehrheitlich als relevant bewertet ($k = 0,88$). 1 Person fordert Streichung, ein Viertel aller Ergänzungsvorschläge bzgl. dieser Skala entfallen jedoch auf diesen Aspekt.
Vollständigkeit^a	Vollständigkeit umfasst sowohl die hierarchische Vollständigkeit, also unterschiedlich komplexe und kognitiv anspruchsvolle Aufgaben, sowie sequenzielle Vollständigkeit, was voraussetzt, dass die Aufgabenerledigung Planung, Vorbereitung, Ausführung und Kontrolle der Aufgabe umfasst. Beschäftigte sollten demnach in sich geschlossene, sinnvolle Arbeitseinheiten erledigen können, anstatt in Fragmenten zu arbeiten oder nur Teilstücke der Aufgabe zu erkennen (96).	Ich kann erkennen, ob ich meine Arbeit gut gemacht habe.	(+) Eine Person forderte die Streichung des Items, aber insgesamt wurde es als sehr relevant bewertet ($k = 0,89$). (+) Alle damit verbundenen Aspekte werden positiv und relevant bewertet. Ergänzend wird immer wieder das Thema Sinnhaftigkeit genannt.

Variabilität^a	Abwechslungsreichtum einer Tätigkeit ist gegeben, wenn die Aufgaben („Breite“) und Anforderungen („Tiefe“) variieren und verschiedene Fähigkeiten, Handlungen und Arbeitsmittel zur Aufgabenerledigung notwendig sind (31, 186).	Meine Arbeit ist in angenehmem Maße abwechslungsreich.	(-) als nicht ausreichend relevant bewertet (k = 0,75), in Freitextkommentaren nicht weiter thematisiert
Information /-sangebot^a	Die Qualität des Informationsangebots definiert sich durch die Vollständigkeit, Relevanz, Aktualität, Verständlichkeit, Zugänglichkeit und Eindeutigkeit der Informationen bezüglich Arbeitsaufgaben, den Arbeitsgegenstand, das gewünschte Ergebnis oder bereits zuvor erledigte und nachfolgende Prozessschritte welche die Arbeitskraft zur Aufgabenerledigung benötigt (98).		(+) 6 explizite Forderungen, diesen Aspekt aufzunehmen, z. B. "Ich bekomme genaue Anweisungen was meine Tätigkeiten und meine Aufgaben betrifft." oder "Im Fragebogen wird die Verfügbarkeit der für die Ausübung der Tätigkeit relevanten Informationen nicht explizit abgefragt."
Verantwortung^a	Verantwortung beschreibt die Pflichten, die ein Arbeitsplatzinhaber hat, wenn er eine bestimmte Tätigkeit ausführt. Verantwortung kann sich auf die Arbeitsergebnisse und die Leistung oder auf die Sicherheit und Gesundheit anderer Personen, zum Beispiel Mitarbeiter, Schüler, Patienten oder zu Pflegende beziehen. Arbeitsplatzinhaber können aber auch im Rahmen ihrer Tätigkeit für Sachwerte, zum Beispiel Maschinen und Anlagen oder die Umwelt verantwortlich sein (186).	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z. B. Personalverantwortung). Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material).	(- / +) Objektverantwortung wird nicht als relevant bewertet (k = 0,63) und sollte nach Experteneinschätzung gestrichen werden. Die Verantwortung für Personen wird hingegen als ausreichend relevant eingestuft (k = 0,78)

Emotionale Inanspruchnahme^a	Zur emotionalen Inanspruchnahme von Arbeitsplatzinhabern kommt es vor allem bei Tätigkeiten im Dienstleistungssektor. Einerseits können die Bedürfnisse und Interessen sowie die verbalen oder nonverbalen Äußerungen der Kunden, Klienten oder Patienten emotional belasten. Andererseits kann die Arbeitsaufgabe das Zeigen gewünschter Emotionen erfordern, die teilweise im Widerspruch zu eigenen Gefühlen stehen (98).		(+) Wurde von vier befragten Experten explizit als zu erhebender Aspekt gefordert
Qualifikation^a	Qualifikation umfasst die Fachkompetenzen, wie zum Beispiel Fachkenntnisse, Arbeitstechniken, Fähigkeiten und Fertigkeiten, aber auch die sozialen und kommunikativen Kompetenzen, die für die fachgerechte Ausführung einer bestimmten Tätigkeit erforderlich sind sowie die Angemessenheit der Einarbeitung in die Tätigkeit und Schulungs- und Fortbildungsmöglichkeiten (186).	Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet. Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit. Die körperliche Leistungsfähigkeit ist wichtig, damit ich die Anforderungen meines Dienstpostens erfüllen kann.	(+/-) Berufliche Anforderungen und Weiterbildungsmöglichkeiten wurden mehrheitlich als relevant bewertet ($k = 0,88$), aber körperliche Leistungsfähigkeit ist aus Sicht der Experten nicht relevant ($k = 0,76$). Für letzteres wurde zweimal explizit die Streichung des Items gefordert. (+) Weitere Aspekte, die mehrfach von Befragten für die Erfassung vorgeschlagen wurden: Über-/Unterforderung, Übereinstimmung von Kompetenzen und Anforderungen, Komplexität der Aufgabe und Einarbeitung

Soziale Beziehungen

Kolleginnen und Kollegen^a	"Soziale Beziehungen zu den Kollegen ergeben sich durch die arbeitsbezogene Interaktion innerhalb oder außerhalb der eigenen Gruppe und den Kontakt mit den Kollegen. Soziale Unterstützung ist vorhanden, wenn die Kollegen, zum Beispiel bei Problemen oder Schwierigkeiten in der Arbeit, um Rat gefragt und/oder um Hilfe gebeten werden können" (186).	Die Anerkennung meiner Arbeit durch meinen Mitarbeiterkreis ist angemessen. Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis. Mit meinen Mitarbeitenden gibt es keine Konflikte.	(+) Sämtliche Items wurden als relevant bewertet (Anerkennung MA: k = 0,91, Unterstützung MA: k = 0,93, Konflikte MA: k = 0,84, Rückmeldung V: k = 0,93, Anerkennung V: k = 0,96, Unterstützung V: k = 0,93, Konflikte V: k = 0,86)
Vorgesetzte^a	"Die sozialen Beziehungen zu dem oder den Vorgesetzten ergeben sich aus der organisatorischen Einbindung der eigenen Arbeit in das Arbeitssystem, der erforderlichen arbeitsbedingten Interaktion und dem Kontakt mit dem oder den Vorgesetzten. Soziale Unterstützung ist vorhanden, wenn der oder die Vorgesetzten zum Beispiel bei Problemen oder Schwierigkeiten in der Arbeit um Rat gefragt und/oder um Hilfe gebeten werden können" (186).	Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit. Die Anerkennung meiner Arbeit durch meine unmittelbaren Vorgesetzten ist angemessen. Ich erfahre genügend Unterstützung von meinen unmittelbaren Vorgesetzten. Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte.	(-/+) Konflikte sind nicht zwangsläufig etwas Negatives, entscheidend ist der Umgang damit. (+) Mobbing sollte erfasst werden
Kunden^e	Gemeint sind im Arbeitskontext BMVg interne Kunden, also Personen mit denen arbeitsbedingt eine Dienstleister-Kunden-Beziehung entsteht, die gleichsam die Merkmale arbeitsbedingter sozialer Beziehungen (s. o.) aufweist.	Mein Kundenkreis für interne Dienstleistungen (z. B. auch Patienten, Leistungsempfänger) verhält sich angemessen. Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.	(-) Zwar wurde der Aspekt von 2 Experten als zu erfassen benannte, jedoch verfehlen beide Items eine ausreichende Relevanzbewertung (Respekt: k = 0,75, Verhalten: k = 0,72), die Unterstützung durch Kunden wurde zudem von 2 Experten als nicht relevant/notwendig bewertet.

			(-) "angemessenes Verhalten" ist sehr interpretativ und nicht disjunkt mit Frage nach Respekt
Mitarbeiter/Untergebene^e	„Untergebene“ entspricht dem bundeswehrinternen Sprachgebrauch für unterstellte Mitarbeiter. Relevant sind bei diesem Aspekt die arbeitsbedingten sozialen Beziehungen (s. o.), die sich in diesem hierarchischen Verhältnis ausbilden.	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Beschäftigten ist angemessen. Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Beschäftigten. Mit den Untergebenen/den mir unterstellten Beschäftigten gibt es keine Konflikte.	(+/-)Vorhandensein von Konflikten wird von 4 Personen als nicht zwangsläufig negativ empfunden, da diese bei Zusammenarbeit entstehen können/dazu gehören, entscheidend ist Umgang/Lösung von Konflikten Unterstützung: k = 0,86 Anerkennung k = 0,86 Konflikte k = 0,83 mehrfacher Hinweis die Formulierung von "Untergebene" zu ändern
Kommunikation und Kooperation^d	Enthalten sowohl bei sozialen Beziehungen zu Kollegen als auch zu Vorgesetzten sowie Untergebenen und Kunden. Vorschlag zur Erfassung als eigenständige Kategorie durch die Experten.		(+) Erfassung wurde viermal gefordert; Hinweis, dass Kooperation nicht nur positiv ist, sondern das hohe Interaktionsdichte und Kommunikationserfordernisse auch Belastung darstellen können
Konfliktkultur^d	im Verständnis der Experten: Umgang mit Konflikten bzw. (konstruktive) Konfliktlösung/ Konfliktlösungskompetenz, aber auch Fehlerkultur sowohl bezogen auf den Vorgesetzten als auch durch Kolleg*innen	Konflikte werden offen besprochen.	(+) hohe Relevanz (k = 0,94), wurde viermal als wichtig und umfassend erhebenswert genannt
Arbeitsklima^d	Ist bisher nicht Gegenstand der Empfehlungen. In der Literatur und seitens der Befragten wurden hierunter die Aspekte Teamgeist, Teambildung, Zusammenhalt, Sozialklima oder empathischer Umgang genannt. In Studien zum Sozialklima am		(+) Dieser Aspekt wurde am häufigsten im Merkmalsbereich Soziale Beziehung adressiert und sollte laut Expertenvotum mit erhoben werden. (+) Mehrfach genannt wurden auch die Fehler- und Konfliktkultur (s. o.) sowie negative soziale Interaktionen wie

	Arbeitsplatz wird dieses häufig über Skalen zu Vertrauen, Kooperation, Reziprozität, Teamklima und Konflikte erfasst (187).		Mobbing oder sexuelle Belästigung als wichtige soziale Aspekte. 3-mal wurde explizit die Erfassung von Mobbing gefordert.
Führung^d	Führung stellt ebenfalls keinen eigenen Aspekt, sondern nur einen Teil der Sozialen Beziehungen in den Empfehlungen von GDA und BAuA dar. Unter gesunder Führung werden sehr umfangreiche Faktoren in den Bereichen Aufgabenorientierung, Interaktion und Mitarbeitendenorientierung subsumiert. Relevante Ressourcen sind dabei unter anderem Partizipation, Gesundheitsorientierung bzw. –förderung, emotionale Intelligenz, Wertschätzung und Rückmeldung, Rücksichtnahme, Erreichbarkeit Transparenz, Gerechtigkeit oder strukturierte Arbeitsgestaltung (188).		(+) Das Thema Führung wurde in den Kommentaren als höchst relevant bewertet und die Erhebung dieses Aspekts hinsichtlich Führungsstil, Fairness und Feedback wurde mehrfach explizit seitens der Befragten gefordert.

Arbeitsumgebung			
Physikalische und chemische Faktoren^a	Mit physikalischen Faktoren sind Faktoren wie Beleuchtung, Klima oder Lärm gemeint. Chemische Faktoren umfassen alle festen, flüssigen oder gasförmigen (Gefahr-)Stoffe, die zu Explosionen oder Bränden führen können oder die giftig, gesundheitsschädlich, ätzend oder umweltgefährlich sind (189).	Die äußeren Bedingungen (Klima, Beleuchtung, Belüftung etc.) an meinem Hauptarbeitsplatz sind angenehm. Den Geräuschpegel an meinem Hauptarbeitsplatz empfinde ich als angenehm.	(+) Gute Bewertung der Relevanz (Geräuschpegel: k = 0,87, äußere Bedingungen: k = 0,89) Ein Hinweis empfahl zusätzlich Gefahrstoffe aufzunehmen
Physische Faktoren^a	Zu den physischen Faktoren werden die ergonomische Gestaltung der Arbeit und die physische Belastung bei der Arbeit gezählt (186, 31).	Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen. Die Einrichtung an meinem Hauptarbeitsplatz ist angemessen.	(+/-) körperliche Anforderungen sollten einer Meinung nach gestrichen werden, wurde jedoch als relevant bewertet (k = 0,82) (+) Einrichtung am Hauptarbeitsplatz aus Expertensicht sehr relevant (k = 0,90) (+) Ergonomie sollte unter Berücksichtigung der besonderen Bedarfe von Personen mit Behinderungen erfasst werden
Arbeitsplatz- und Informationsgestaltung^a	Zur Arbeitsplatz- und Informationsgestaltung gehören die räumlichen Arbeitsverhältnisse und die Informationsgestaltung (186).	Arbeitsräume sind in ausreichender Anzahl vorhanden. Die Einrichtung an meinem Hauptarbeitsplatz ist angemessen.	(+) Beide Items wurden als sehr relevant bewertet (Arbeitsräume: k = 0,91, Einrichtung: k = 0,90) (+) weitere zu erfassende Aspekte nach Einzelmeinungen: individuelle Gestaltungsmöglichkeiten des Arbeitsplatzes, Hygiene, Einrichtung und Zustand der Sozialräume
Arbeitsmittel^a	Der Aspekt Arbeitsmittel umfasst einerseits die Verfügbarkeit von Maschinen, Geräten und Anlagen, an denen bzw. mit denen gearbeitet wird	Die nötigen Arbeitsmittel sind vorhanden.	(+) Eine Forderung nach Streichung, jedoch 6-mal explizit als relevant benannt und insgesamt sehr relevant bewertet (k = 0,93)

	und andererseits die Einschätzung des Zustandes der Arbeitsmittel durch die Beschäftigten. Bewertungskriterien sind u.a. Zustand und Vorhandensein aller notwendigen Arbeitsmittel, Positionierung der Arbeitsmittel oder Gestaltung und Vielfalt der Arbeitsmittel, so dass eine optimale Aufgabenerledigung ermöglicht wird oder ob notwendige Schutzausrüstung vorhanden ist und bei Bedarf verwendet wird (189).		
Weitere Aspekte			
Arbeitssicherheit^d	Gemeint ist hier der technische Bereich des Arbeitsschutzes, welcher mechanische, biologische, elektrische oder physikalische Gefährdungen sowie Gefahrstoffe umfasst. Er stellt in den Empfehlungen von BAuA und GDA keinen zu beurteilenden Aspekt im Rahmen der Gefährdungsbeurteilung psychischer Belastung dar.		(+) Wurde als zusätzlich zu erfassender Aspekt von 14 Experten vorgeschlagen
Infrastrukturelle Anbindung^d	Hier wurden seitens der Befragten Anmerkungen zur Verkehrsanbindung und Wegstrecken zu Bedarfen des alltäglichen Lebens bzw. Pendelzeiten subsumiert. Dieser Aspekt ist nicht Gegenstand der Empfehlungen von BAuA und GDA.		(+) sollte nach Meinung von 3 Experten erfasst werden

Work-Life-Balance	Meistens wird mit Work-Life-Balance der Rollenkonflikt zwischen verschiedenen Lebensbereichen gemeint (97), der dadurch entsteht, dass Ressourcen nur begrenzt zur Verfügung stehen und z. B. Ermüdung durch Stress bei der Arbeit dazu führt, seinen privaten Verpflichtungen nicht mehr vollständig nachkommen zu können. Es gibt inzwischen auch neuere Ansätze, die neben dem Konflikt der verschiedenen Lebensbereiche und die Möglichkeit zur gegenseitigen Bereicherung berücksichtigen (95).		(+) sollte nach Meinung von 3 Experten erfasst werden
--------------------------	--	--	---

Anmerkung: a: Empfehlung von GDA und BAuA (31, 98), b: Empfehlung der BAuA (98), c: Empfehlung der GDA (31), d: Ergänzungsvorschlag von den Experten, e: Zusätzlicher Aspekt im AIGScreen

4.3 Itemanalyse, Reliabilitätsanalyse und Konstruktvalidierung

Nachfolgend werden die Ergebnisse der statistischen Auswertungen zur Qualität und Güte der Skalen und Items berichtet. Die Schritte der Itemanalyse, explorativen Faktorenanalyse und konfirmatorischen Faktorenanalyse bauen sukzessive aufeinander auf. Die Itemanalyse ergab, dass sich die verwendeten Daten für die Durchführung von Faktorenanalysen eignen. Sie zeigte allerdings auch einige Schwachstellen auf Itemebene auf. Insbesondere überproportional viele positive Antworten bei nahezu allen Items führten zur Abweichung der Normalverteilung, was die anschließenden Analysen beeinflusste. Zudem musste mit einer hohen Zahl fehlender Werte umgegangen werden. Diese waren zwar überwiegend plausibel in ihrer Entstehung und stellten nicht die betreffenden Items an sich in Frage, für sie musste jedoch vor Durchführung der Faktorenanalyse eine Lösung gefunden werden. 284 Fälle wurden wegen zu hoher Missingquoten, d.h. fehlender Angaben in mehr als 50% der Fragen, gänzlich von der Analyse ausgeschlossen. Für alle Fälle mit einer Missingquote unter 50% wurden fehlende Werte mittels multipler Imputation für die Analysen ersetzt.

Itemanalysen und explorative Faktorenanalysen (EFA) wurden zur Bestimmung der Itemcharakteristika und der Reliabilität durchgeführt. Zudem sollte die latente Struktur der Items zur Messung der Arbeitsmerkmale eruiert werden. Um die postulierten Konstrukte sowie die im Rahmen der EFA entdeckten latenten Strukturen zu überprüfen, wurden konfirmatorische Faktorenanalysen (CFA) durchgeführt. Um eine Verzerrung der Ergebnisse („Overfitting“) der CFA durch die Überschneidung der Stichproben zu verhindern, wurde die Stichprobe vor der Analyse zufällig in zwei Teilstichproben aufgeteilt ($n_{CFA} = 8.058$, $n_{EFA} = 8.134$).

4.3.1 Beschreibung der Stichprobe

Die Stichprobe umfasst insgesamt 16.476 Mitarbeitende im Geschäftsbereich des BMVg, die zwischen Januar 2018 und Oktober 2019 freiwillig und anonym an der Befragung teilgenommen haben. Mit der Studienteilnahme waren keine Vor- oder Nachteile verbunden. Die Befragung wurde online in 127 Dienststellen des BMVg durchgeführt. Die Befragung wurde unter der Verantwortung des jeweiligen Gesundheitsmanagers der einzelnen Standorte durchgeführt, der im Vorfeld genaue Instruktionen zur Durchführung der Befragung erhalten hatte. Die Studie wurde von der Ethikkommission der Landesärztekammer Rheinland-Pfalz genehmigt.

Die durchschnittliche Teilnahmequote lag bei 37 %. Da die Kennwerte stichprobenabhängig sind, sollte die gewählte Stichprobe möglichst der Zielgruppe entsprechen (71). Nur so kann eine Aussage über die Eignung für den geplanten Einsatzzweck getroffen werden. Die

spezifische Repräsentativität der Stichprobe wurde anhand der zur Verfügung stehenden soziodemographischen Informationen, d. h. bezüglich des Beschäftigungsstatus (zivile Angestellte vs. Soldaten), des Alters (kategorial) und des Geschlechts mittels Chi²-Anpassungstest überprüft. Für alle drei betrachteten Merkmale musste die Nullhypothese, dass es keine Abweichungen zwischen der Grundgesamtheit und der Stichproben gibt, eindeutig verworfen werden. Die demografische Zusammensetzung der Stichprobe und der Grundgesamtheit ist in Tabelle 9 zu finden. Wegen eines zu hohen Anteils (>50%) fehlender Werte mussten 284 Fälle aus der Stichprobe ausgeschlossen werden, sodass für die Analyse am Ende ein Datensatz mit n = 16.192) zur Verfügung stand.

Tabelle 9: Soziodemographische Verteilung in der Stichprobe (n = 16.192) im Vergleich zur Grundgesamtheit (n = 44.642)

		Grundgesamtheit		Stichprobe		Chi ² - Anpassungs- test
Beschäfti- gungsstatus	Soldat	27.207	60,94%	10.884	67,22%	$\chi^2 = 2815,00$ df = 2 p = 0,00000
	Zivil	17.435	39,06%	4.471	27,61%	
	Fehlende Werte	0	0,00%	837	5,17%	
Alters- gruppe	Bis 29 Jahre	11.502	25,76%	4.301	26,56%	$\chi^2 = 2027,05$ df = 4 p = 0,00000
	30 - 39 Jahre	12.300	27,55%	4.356	26,90%	
	40 - 49 Jahre	8.182	18,33%	3.298	20,37%	
	50 Jahre und älter	12.646	28,33%	3.558	21,97%	
	Fehlende Werte	12	0,03%	679	4,19%	
Geschlecht¹¹	Männlich	35.217	78,89%	11.136	68,77%	$\chi^2 = 7078,10$ df = 2 p = 0,00000
	Weiblich	9.425	21,11%	2.602	16,07%	
	Fehlende Werte	0	0,00%	2.454	15,16%	
Total		44.642		16.192		

4.3.2 Itemanalyse und multiple Imputation

Bei der grafischen Verteilungsanalyse fiel auf, dass 13 der Items eine unimodale, aber eindeutig linksschiefe Antwortverteilung aufweisen. Alle Histogramme sind in Anlage 5 zu finden. Für alle vier Subskalen konnten Deckeneffekte beobachtet werden. Lediglich die Frage nach Arbeitsunterbrechungen wies mehr negative als positive Antworten auf. Entsprechend hoch fiel auch der Mittelwert über alle Items mit $\bar{X}=3,0$ bei einem Wertebereich von 1,0 bis 4,0 aus. Den niedrigsten Mittelwert verzeichnete das bereits erwähnte Item am9 zur Arbeitsunterbrechung mit $\bar{X} = 2,3$. Der höchste Wert wurde für Item um3 „Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen“ mit $\bar{X} = 3,4$ ermittelt.

¹¹ Die Kategorie „Divers“ wurde erst zu einem späteren Zeitpunkt im Fragebogen ergänzt.

Überdurchschnittlich positiv wurden zudem die Items aa5 (ausreichende Arbeitspausen) mit $\bar{x} = 3,23$, am6 (Ergebniskontrolle) mit $\bar{x} = 3,32$, sa4 (Unterstützung durch Mitarbeitende) mit $\bar{x} = 3,20$, sa9 (Anerkennung durch unterstellte Beschäftigte) mit $\bar{x} = 3,22$ und sa10 (Unterstützung durch unterstellte Beschäftigte) mit $\bar{x} = 3,26$ von der Mitarbeitendenstichprobe beantwortet. Die Verteilungswerte sind für die einzelnen Items in Tabelle 10 aufgeführt.

Wie bereits in Abschnitt 3.4.1 erläutert, kann die Itemschwierigkeit nicht hinsichtlich richtig oder falsch gegebener Antworten bewertet werden, da es im Falle des AIGScreen kein Richtig oder Falsch bezüglich der Antworten gibt. Allerdings kann der Mittelwert als Indikator der Itemschwierigkeit bzw. im vorliegenden Fall vor allem als Indikator einer angemessenen Antwortverteilung herangezogen werden. Da die Mittelwerte der einzelnen Items sich tendenziell am oberen Ende der Antwortskala bewegten, würde man allgemein von „leichten“ Items sprechen (71). In den Kontext arbeitsbedingter psychischer Belastungen übertragen bedeutet das, dass die Mehrheit dem jeweiligen Item zustimmte bzw. dass das jeweilige Merkmal mehrheitlich als positiv eingeschätzt wurde. Gleichzeitig konnten hohe Standardabweichungen festgestellt werden. Diese fiel am niedrigsten für das gleichsam „leichteste“ Item um3 mit $SD = 0,73$ aus. Item um4 („Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann“) wies die höchste Standardabweichung von $SD = 1,09$ auf. Hier variierten die Antworten besonders häufig. Ein Zusammenhang zwischen Itemschwierigkeit und Standardabweichung ließ sich nicht feststellen.

Aufgrund der überwiegend rechtssteilen Verteilung bei den meisten Items, ist es auch nicht verwunderlich, dass der Mardia-Skewness-Test ($\chi^2=21643,04$), der Mardia-Kurtosis-Test ($\chi^2=16500,48$) und der Doornik-Hansen-Test ($\chi^2=5829,35$) allesamt multivariate Nicht-Normalität ($p < 0,001$) indizierten.

Die Trennschärfe fiel bei Anwendung der Part-Whole-Korrektur, also bei Ausschluss des betreffenden Items aus der Summe des Gesamtscores, erwartungsgemäß niedriger aus als bei der Item-Skalen-Korrelation. Die durchschnittliche Item-Skalen-Korrelation betrug $r_S = 0,65$, die durchschnittliche Item-Rest-Korrelation betrug $r_R = 0,52$. Die Items am6 („Ich kann erkennen, ob ich meine Arbeit gut gemacht habe.“) und am9 („Ich werde bei meiner Arbeit selten unterbrochen.“) wiesen die niedrigsten Trennschärfekoeffizienten von $r_S = 0,50$ und $r_R = 0,37$ bzw. $r_S = 0,47$ und $r_R = 0,33$ auf. Eine besonders hohe Item-Skalen-Korrelation ließ sich für die Items der Subskala Arbeitsumgebung, mit Ausnahme des Items um3 (körperliche Anforderungen der Tätigkeit) feststellen. Die hat jedoch nur für das Item um2 (äußere Bedingungen) Bestand, wenn man die Part-Whole-Korrektur anwendet. Für Item um1 (Geräuschpegel) und um3 (Arbeitsplatzgestaltung) fiel sie von $r_S = 0,79$ auf $r_R = 0,59$ bzw. von $r_S = 0,80$ auf $r_R = 0,58$ ab.

Mit Blick auf KMO und Cronbachs Alpha konnte festgestellt werden, dass sich alle Items sehr gut für eine Faktorenanalyse eignen. Das KMO-Kriterium erzielte Werte von mindestens 0,85, mehrheitlich sogar über 0,90, was in der Literatur als „marvelous“ eingestuft wird (156). Demnach eigneten sich sämtliche Items für die Durchführung einer Faktorenanalyse.

Die interne Konsistenz aller vier Subskalen erzielte akzeptable bis gute Werte. Die geringste interne Konsistenz, welche aufgrund der geringen Itemzahl und der Heterogenität des Konstrukts aber immer noch als gut erachtet werden kann, betrug $C-\alpha = 0,759$ für die Arbeitsumgebung. Die anderen Merkmalsbereiche wiesen mit $C-\alpha = 0,871$ für Soziale Aspekte, $C-\alpha = 0,799$ für Arbeitsinhalt und -aufgabe und $C-\alpha = 0,823$ für Arbeitsorganisation hohe interne Konsistenzen auf. Ausschließlich für Item um3 konnte eine sehr geringfügige potenzielle Verbesserung der internen Konsistenz der Subskala Arbeitsumgebung bei Eliminierung des Items nachgewiesen werden. Alle anderen Items trugen zur Erhöhung der internen Konsistenz ihrer jeweiligen Subskala bei.

Die Analyse fehlender Antworten zeigte keine Auffälligkeiten. Der Anteil fehlender Antworten lag zwischen 0,4% (für die Items am1 „Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet“ und um1 „Den Geräuschpegel an meinen Hauptarbeitsplatz empfinde ich als angenehm“) und 5,1% (Item sa1 „Meine Kunden für interne Dienstleistungen (z.B. auch Patienten, Leistungsempfänger) verhalten sich angemessen“). Allerdings müssen zusätzlich die „unzutreffend“-Antworten berücksichtigt werden, da diese ebenfalls in den weiteren Analysen nicht als Teil der Likert-Skalen berücksichtigt werden können. Hierdurch erhöht sich der Anteil fehlender bzw. im Rahmen der Analysen nicht auswertbarer Daten für bestimmte Items erheblich. Dies betrifft vor allem die Fragen, die sich auf den Kundenkontakt und unterstellte Beschäftigte beziehen. Diese konnten von rund einem Drittel der Befragten jeweils nicht beantwortet werden (vgl. Tabelle 11). Auch Personal- oder Objektverantwortung sind Merkmale, die auf mehr als ein Fünftel der Befragten nicht zutreffen. Etwas weniger hoch, aber mit 17,9% immer noch sehr hoch fällt die Missingquote für Item aa2 („Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen“) aus.

Wie bereits im Kapitel 3.4.2 erläutert, kann MAR statistisch nicht abschließend nachgewiesen werden. Es kann lediglich mittels Zusammenhangsanalysen überprüft werden, ob sich anhand der vorhandenen Daten eine Verletzung der MAR-Annahme nachweisen lässt. Dies war nicht der Fall. Darüber hinaus obliegt es den Forschenden durch gründliche Vorüberlegungen und eine begründete Auswahl von Hilfsvariablen das Vorliegen von MCAR auszuschließen und die Schätzung gegenüber Verletzungen der MAR-Annahme möglichst robust zu machen (166, 161, 159). Zusätzliche Hilfsvariablen für die Imputation standen im verwendeten Datensatz zur Verfügung und konnten, wie in Kapitel 3.4.2 beschrieben, allesamt in die MICE einbezogen werden. Da keine hohe Inter-Item-Korrelation festgestellt werden konnte, musste die Zahl der

Imputationszyklen nicht besonders hoch ausfallen, sondern konnte sich an der Daumenregel orientieren, wonach die Zahl der Imputationen dem Anteil unvollständiger Datensätze entsprechen sollte (164). Die Korrelationsmatrix ist in Anlage 6 zu finden. Die meisten Autoren argumentieren zudem, dass mehr als 5 bis maximal 10 Imputationen zu keiner Verbesserung der Schätzung führen, zumal die Rechenleistung von Hard- und Software auch begrenzt ist (159, 167, 161, 164, 166). Letztere konnte maximal 30 Imputationen durchführen, was in Anbetracht der Empfehlungen der Fachliteratur als eine mehr als ausreichend hohe Zahl für eine robuste und verlässliche Schätzung gesehen werden kann.

Tabelle 10: Itemstatistiken

Skala / Item		MW	SD	r_S	r_R	C- α ohne Item	KMO
Arbeitsorganisation ($\alpha=.823$)		2.94	.69				
aa1	Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit ausgewogen.	2.81	.97	.74	.59	.80	.93
aa2	Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen.	3.16	.87	.69	.55	.81	.93
aa3	Meine Arbeitszeitverteilung ist mir angenehm.	3.11	.90	.76	.64	.79	.93
aa4	Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen.	2.81	.99	.76	.61	.79	.93
aa5	Meine Arbeitspausen sind ausreichend.	3.23	.88	.74	.61	.79	.92
aa6	Meine Arbeitspausen sind störungsfrei.	2.56	1.04	.72	.55	.81	.92
Arbeitsinhalt und Arbeitsaufgabe ($\alpha=.799$)		2.96	.52				
am1	Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet.	3.15	.84	.58	.47	.78	.93
am2	Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit.	2.80	.96	.58	.45	.78	.93

<i>am3</i>	Meine Arbeit ist in angenehmem Maße abwechslungsreich.	3.01	.89	.55	.43	.79	.93
<i>am4</i>	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z. B. Personalverantwortung).	3.21	.84	.57	.47	.78	.87
<i>am5</i>	Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material).	3.16	.87	.57	.46	.78	.88
<i>am6</i>	Ich kann erkennen, ob ich meine Arbeit gut gemacht habe.	3.32	.82	.50	.37	.79	.94
<i>am7</i>	Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit.	2.87	1.01	.56	.42	.79	.91
<i>am8</i>	Ich habe einen angemessenen Handlungsspielraum (z. B. Entscheidungsfreiheit) bei meiner Arbeit.	3.07	.90	.57	.46	.78	.96
<i>am9</i>	Ich werde bei meiner Arbeit selten unterbrochen.	2.31	.93	.47	.33	.79	.93
<i>am10</i>	Die nötigen Arbeitsmittel sind vorhanden.	2.82	.99	.59	.47	.78	.94
<i>am11</i>	Arbeitsräume sind in ausreichender Anzahl vorhanden.	2.91	1.07	.59	.46	.78	.92
<i>am12</i>	Die Einrichtung an meinem Hauptarbeitsplatz ist angemessen.	2.92	1.04	.61	.49	.78	.89
	Soziale Aspekte ($\alpha=.871$)	3.08	.56				
<i>sa1</i>	Mein Kundenkreis für interne Dienstleistungen (z. B. auch Patienten, Leistungsempfänger) verhält sich angemessen.	3.09	.74	.54	.44	.85	.92

sa2	Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.	3.16	.76	.63	.55	.84	.92
sa3	Die Anerkennung meiner Arbeit durch meinen Mitarbeiterkreis ist angemessen.	3.07	.81	.70	.62	.84	.93
sa4	Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis.	3.20	.77	.68	.60	.84	.94
sa5	Mit meinen Mitarbeitenden gibt es keine Konflikte.	2.91	.91	.62	.50	.85	.89
sa6	Die Anerkennung meiner Arbeit durch meine unmittelbaren Vorgesetzten ist angemessen.	2.94	.92	.72	.62	.84	.89
sa7	Ich erfahre genügend Unterstützung von meinen unmittelbaren Vorgesetzten.	3.03	.90	.70	.60	.84	.91
sa8	Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte.	3.11	.95	.64	.52	.85	.9
sa9	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Beschäftigten ist angemessen.	3.22	.75	.67	.59	.84	.89
s10	Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Beschäftigten.	3.26	.75	.67	.59	.84	.87
s11	Mit den Untergebenen/den mir unterstellten Beschäftigten gibt es keine Konflikte.	3.10	.89	.60	.49	.85	.85
s12	Konflikte werden offen besprochen.	3.05	.94	.63	.51	.85	.97

Arbeitsumgebung ($\alpha=.759$)		2.95	.74				
<i>um1</i>	Den Geräuschpegel an meinen Hauptarbeitsplatz empfinde ich als angenehm.	2.86	1.00	.79	.59	.68	.92
<i>um2</i>	Die äußeren Bedingungen (Klima, Beleuchtung, Belüftung etc.) an meinem Hauptarbeitsplatz sind angenehm.	2.69	1.04	.83	.64	.65	.92
<i>um3</i>	Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen.	3.40	.73	.62	.43	.76	.96
<i>um4</i>	Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann (z. B. Arbeitshöhe, Sitzhöhe, Beinfreiheit, Durchgangsbreiten/Sicherheitsabstände).	2.84	1.09	.80	.58	.69	.92

Anmerkung: MW = arithmetisches Mittel, SD = Standardabweichung, r_s = Item-Skalen-Korrelation (Korrelation der Itemwerte mit Summenwert der Gesamtskala), r_R = Item-Rest-Korrelation (Korrelation der Itemwerte mit Summenwert der Gesamtskala ohne das betreffende Item), C- α ohne Item = Skalenwert für Cronbachs Alpha ohne betreffendes Item, KMO = Kaiser-Meyer-Olkin- Kriterium

Tabelle 11: Items mit hohen Missingquoten, wenn „unzutreffend“-Antwort als fehlender Wert behandelt wird

Item		Fehlend	Anteil Unzu- treffend	Gesamt Missing
am4	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z.B. Personalverantwortung).	4,79%	18,0%	22,79% (n = 3.690)
am5	Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z.B. Infrastruktur, Material).	3,43%	17,9%	21,33% (n = 3.453)
sa1	Meine Kunden für interne Dienstleistungen (z.B. auch Patienten, Leistungsempfänger) verhalten sich angemessen.	0,06%	35,05%	35,65% (n = 5.773)
sa2	Ich werde von den internen Kunden (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.	1,08%	30,01%	31,09% (n = 5.034)
sa9	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Mitarbeitern ist angemessen.	6,25%	25,61%	31,86% (n = 5.158)
sa10	Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Mitarbeitern.	6,25%	25,61%	31,86% (n = 5.158)
sa11	Mit den Untergebenen/mir unterstellten Mitarbeitern gibt es keine Konflikte.	5,42%	26,03%	31,45% (n = 5.093)
aa2	Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen.	2,64%	15,22%	17,86% (n = 2.892)

4.3.3 Explorative Faktorenanalyse

Die Zahl zu extrahierender Faktoren im Rahmen der EFA wurde anhand von Eigenwert-Kriterien, Scree-Test, Parallel-Analyse und Modelfit-Vergleichen bestimmt. Hieraus konnte jedoch keine eindeutige Lösung abgeleitet werden, da die Tests jeweils unterschiedliche Faktorenzahlen nahelegen. Die Anzahl der Eigenwerte größer als 1,0 (Kaiser-Kriterium) (156) deutet auf eine Gesamtzahl von 3 Faktoren hin. Die Anzahl der Eigenwerte größer als 0,7 (Jolliffe-Kriterium) (170) deutet auf eine Gesamtzahl von 5 Faktoren hin, während der Scree-Plot (Abbildung 13) zwei Faktoren als beste Lösung anzeigt. In diesem Fall wird empfohlen, die Parallelanalyse (

Abbildung 14) heranzuziehen (173, 172), die in der vorliegenden Stichprobe für die Beibehaltung von 11 Faktoren spricht. Weniger Faktoren sollten nach Valicer's MAP (Minimum Average Partial Correlation) beibehalten werden, der mit 5 Faktoren ein Minimum von 0,01 erreicht. Das BIC (Bayesian Information Criterion) sowie das stichprobenangepasste BIC erreichen beide ihr Minimum bei 8 Faktoren. Der RMSEA (Root Mean Square Error of Approximation) zeigt ebenfalls, dass die beste Anpassung bei 8 Faktoren erreicht wird, wobei auch 4 bis 7 Faktoren noch akzeptabel wären (Tabelle 12). Die zuvor postulierten 4 Skalen scheinen auch mit Blick auf die Ladungsmatrizen (Anlage 7) die Datenstruktur nicht gut abzubilden.

Abbildung 13: Ergebnisse der explorativen Faktorenanalyse – Scree-Plot (n = 8.134)

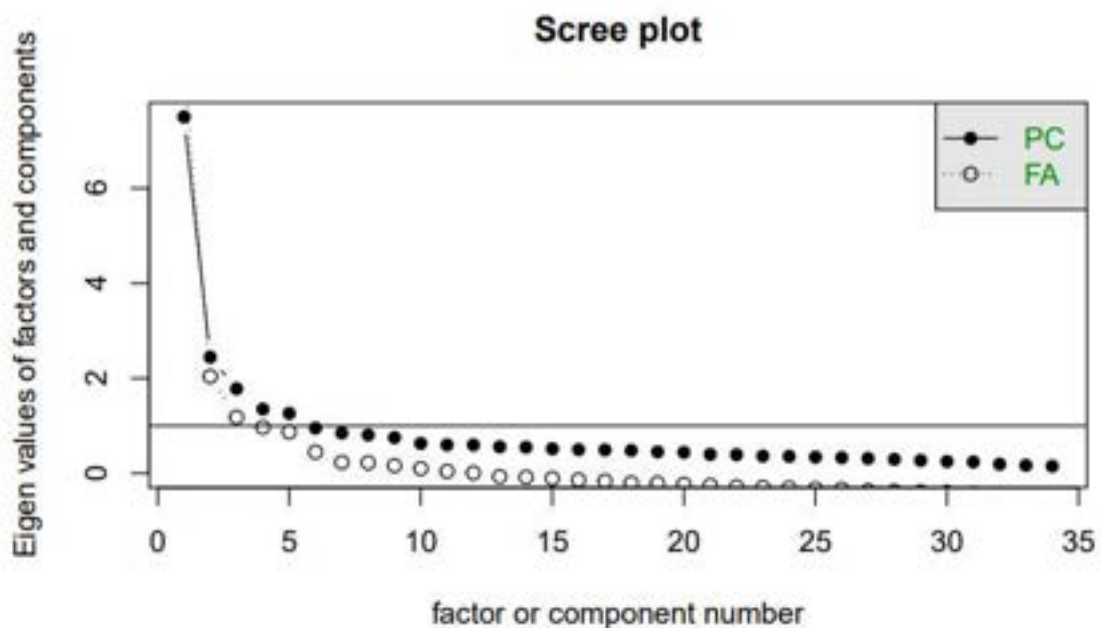


Abbildung 14: Ergebnisse der explorativen Faktorenanalyse – Parallel-Analyse (n = 8.134)

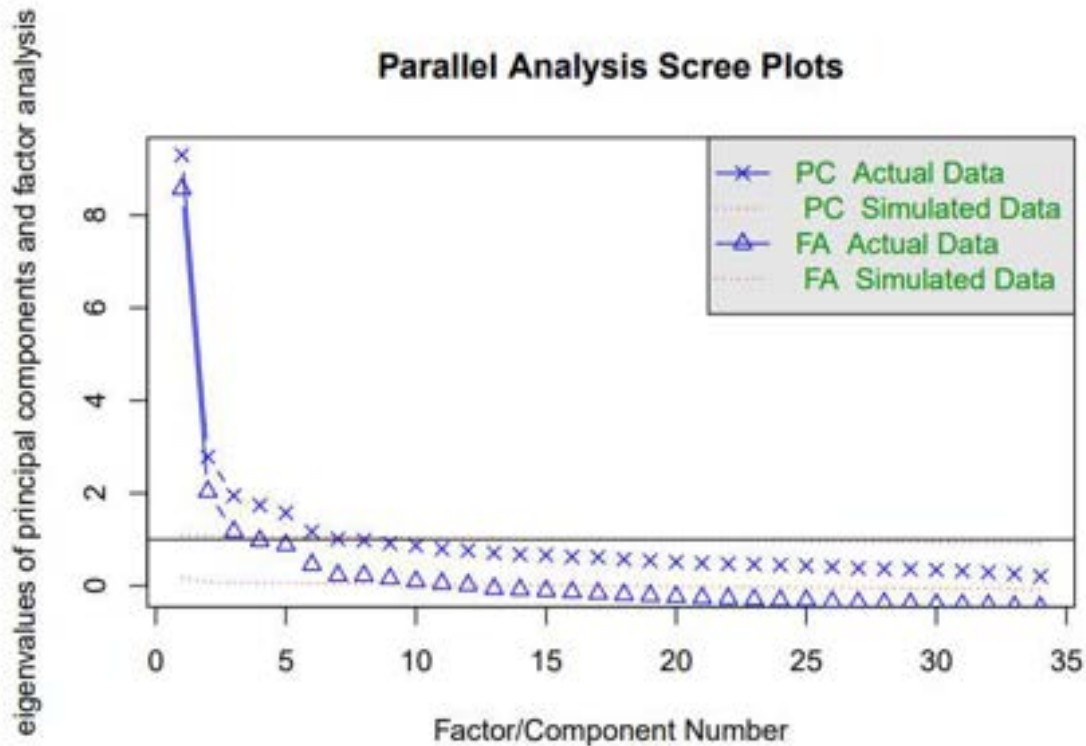


Tabelle 12: Ergebnisse der explorativen Faktorenanalyse – Teststatistiken nach Faktorenzahl für die explorative Faktorenanalyse (n = 8.134)

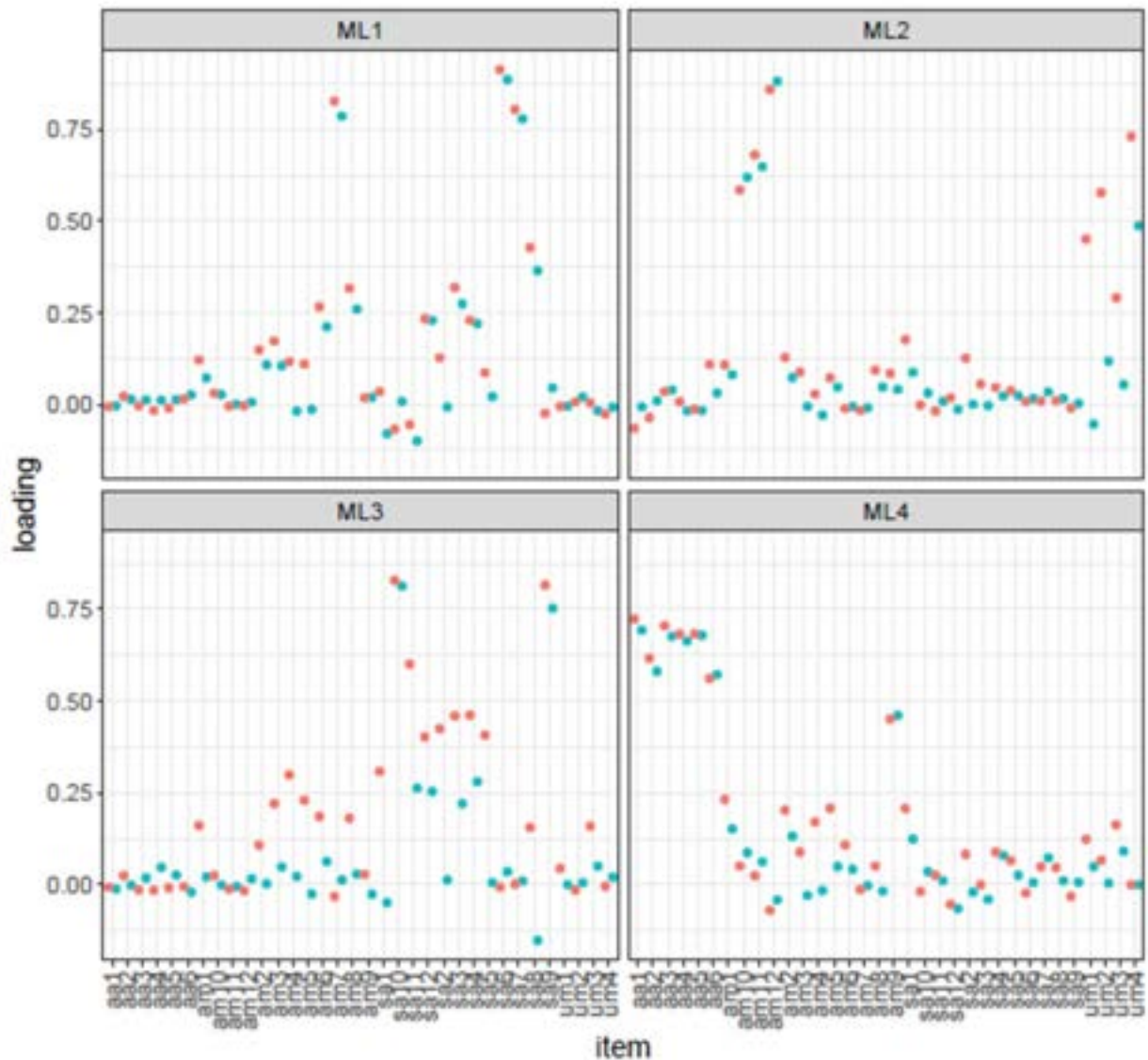
Faktorenzahl	MAP	RMSEA	BIC	SABIC
1	0,018	0,109	96480	98155
2	0,014	0,097	70172	71741
3	0,012	0,086	51374	52842
4	0,012	0,074	34668	36037
5	0,010	0,066	24768	26042
6	0,011	0,060	18410	19592
7	0,012	0,054	13186	14279
8	0,013	0,050	9989	10996

Anmerkung: MAP = Velicer's Minimum Average Partial Correlation, RMSEA = Root Mean Square Error of Approximation, BIC = Bayesian Information Criterion, SABIC = sample adjusted Bayesian Information Criterion

Für das weitere Vorgehen wurde eine Zahl von 8 Faktoren gewählt, da sich diese inhaltlich am besten begründen lassen und das beste Verhältnis zwischen Sparsamkeit und Modellpassung ermöglichen. Die Lösung mit 4 Skalen, wie sie zuvor postuliert wurde, erscheint weder hinsichtlich der Kriterien zur Bestimmung der Faktorenanzahl angemessen, noch ließ sich in den entsprechenden Ladungen eine Einfachstruktur erkennen, und zwar auch dann nicht, wenn eine oblique Rotation vorgenommen wurde (vgl. Abbildung 15: Faktorladungen bei 4 Faktoren (EFA, Oblimin, n = 8.134). Hingegen verteilen

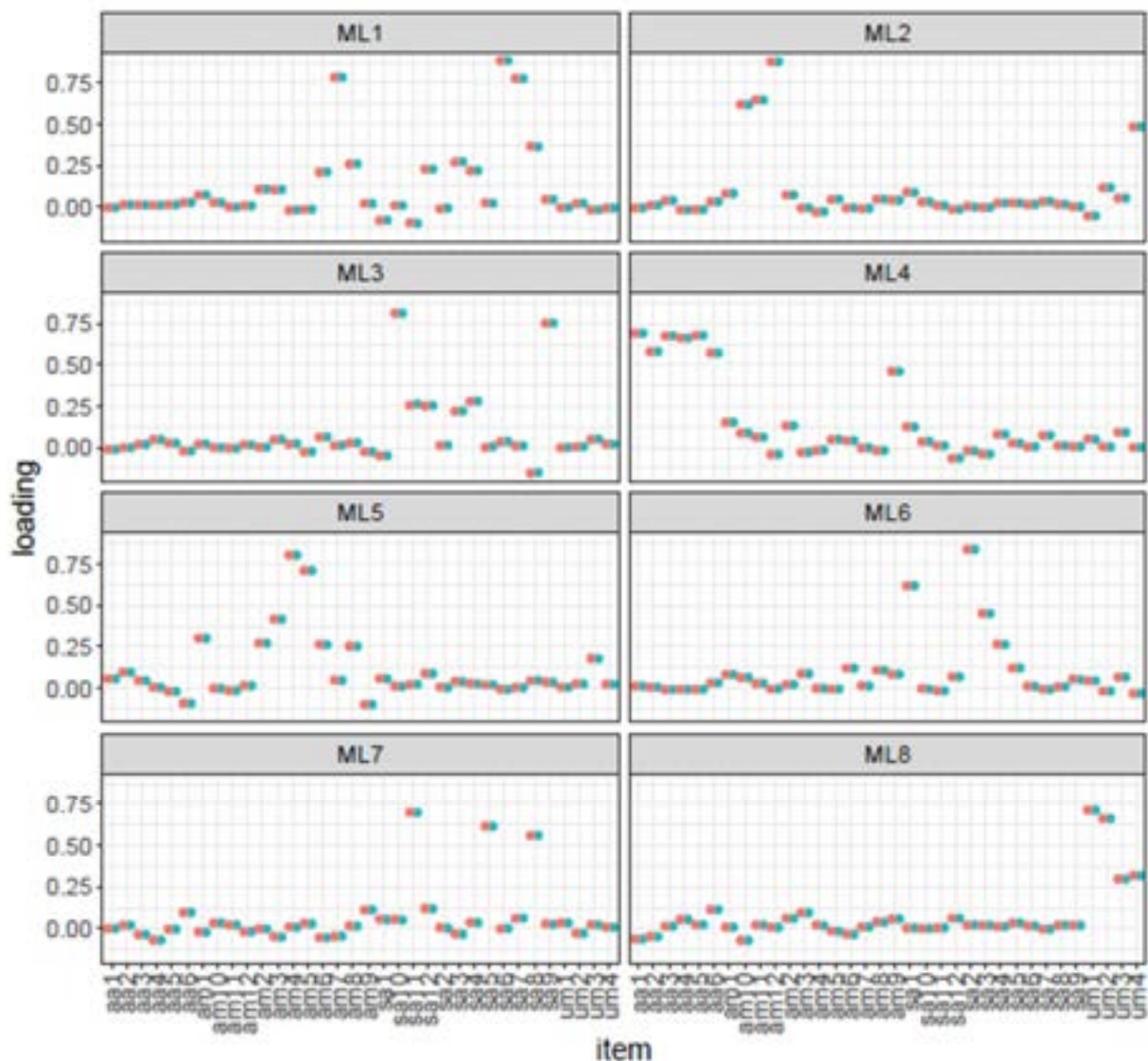
sich die Ladungen der einzelnen Items für die 8-Faktoren-Lösung deutlich eindeutiger bzw. weisen weniger Querladungen auf (vgl. Abbildung 16).

Abbildung 15: Faktorladungen bei 4 Faktoren (EFA, Oblimin, $n = 8.134$)



Anmerkung: Blaue Punkte zeigen die mittleren Faktorladungen aller 30 imputierten Datensätze; rote Punkte zeigen die Faktorladungen nach Procrustes-Rotation

Abbildung 16: Faktorladungen bei 8 Faktoren (EFA, Oblimin, n = 8.134)



Anmerkung: Blaue Punkte zeigen die mittleren Faktorladungen aller 30 imputierten Datensätze; rote Punkte zeigen die Faktorladungen nach Procrustes-Rotation

Mit Blick auf die 8-Faktoren-Lösung (Tabelle 13) zeigte sich aber auch, dass die Skalen Arbeitsorganisation und Arbeitsumgebung durchaus wie postuliert jeweils einen Faktor bildeten, mit Ausnahme von Item um4 („Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann“), welches auf Faktor 2 höher lädt. Da jedoch für die Arbeitsumgebung keine Einfachstruktur festgestellt werden konnte, wären auch andere Skalenzusammensetzungen plausibel. Die Subskala Arbeitsorganisation ist sehr homogen und wies keine relevanten Querladungen auf. Die beiden Skalen Soziale Aspekte und Arbeitsinhalt und –aufgabe konnten mit Blick auf die Ladungsstruktur nicht wie theoretisch postuliert als jeweils eine Skala verifiziert werden und bedürfen einer Neuordnung der Items. Die Ladungen einzelner Items, z. B. am8 („Ich habe einen angemessenen

Handlungsspielraum (z. B. Entscheidungsfreiheit) bei meiner Arbeit“) oder sa4 („Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis“) verteilen sich gleichmäßig auf mehrere Faktoren. Die höchsten Ladungen und, im Falle der Items sa4, sa12 und wtc8, die inhaltlich plausibelsten Zuordnungen sind in Tabelle 13 fett gedruckt.

Tabelle 13: Faktorladungen nach Extraktionen von 8 Faktoren (EFA, ML-Methode, Oblimin, n = 8.134)

Item	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8	K
<i>Arbeitsorganisation</i>									
aa1				0,69					0,48
aa 2				0,58					0,38
aa 3				0,68					0,51
aa 4				0,66					0,46
aa 5				0,68					0,46
aa 6				0,57			0,10	0,11	0,43
<i>Arbeitsmenge und Arbeitsaufgabe</i>									
am1				0,15	0,30				0,26
am 2	0,11			0,13	0,27				0,22
am 3	0,11				0,42				0,30
am 4					0,80				0,64
am 5					0,71				0,53
am 6	0,21				0,26	0,12			0,24
am 7	0,79								0,64
am 8	0,26				0,25	0,11			0,28
am 9				0,46			0,11		0,30
am 10		0,62							0,43
am 11		0,65							0,49
am 12		0,88							0,76
<i>Soziale Aspekte</i>									
sa1				0,12		0,62			0,49
sa2						0,84			0,71
sa3	0,27		0,22			0,45			0,56
sa4	0,22		0,28			0,26			0,45
sa5						0,12	0,62		0,48
sa6	0,89								0,82
sa7	0,78								0,70
sa8	0,36						0,56		0,53
sa9			0,75						0,68
sa10			0,81						0,73
sa11			0,26				0,70		0,67
sa12	0,23		0,25				0,12		0,28
<i>Arbeitsumgebung</i>									
um1								0,71	0,52
um2		0,12						0,66	0,55
um3					0,18			0,30	0,27
um4		0,49						0,32	0,52
<i>s_F</i>	17%	13%	11%	18%	12%	11%	9%	9%	

Anmerkung: K = Kommunalität; s_F = durch den Faktor erklärte Varianz; fett gedruckte Werte drücken die Zuordnung des Items zum jeweiligen Faktor aus; Faktorladungen < 0,1 wurden zur besseren Übersicht ausgeblendet

4.3.4 Konfirmatorische Faktorenanalyse

Da die Ergebnisse der EFA das 8-Faktoren-Modell am plausibelsten erscheinen ließen, wurde im Rahmen der CFA überprüft, ob die 8-Faktoren-Lösung gegenüber den Modellen mit den vier postulierten Skalen, Arbeitsumgebung, Arbeitsinhalt und -aufgabe, Arbeitsorganisation und Soziale Aspekte, oder ein Faktor den besseren Modelfit aufweist. Der Vergleich zum 1-Faktor-Modell wurde gewählt, um auszuschließen, dass eine Einfachstruktur für mehrere Faktoren nur deshalb nicht gefunden werden kann, weil es sich bei den Items zu den Arbeitsmerkmalen letztlich um ein homogenes Gesamtkonstrukt handelt. In Tabelle 14 sind die Itemzuordnungen zum jeweiligen Konstrukt, wie es sich aus den Faktorladungen in Tabelle 13 ableiten ließe, beschrieben. Die Faktortitel entsprechen der inhaltlichen Interpretation der Itemskalen.

Problematisch ist, dass Faktor 3 nur aus zwei Items besteht, was die Modellsättigung negativ beeinflusst (64).

Tabelle 14 - Skalenbildung bei einer 8-Faktoren-Lösung

Itemnr.	Faktor 1 „Feedback / Führungsstil“
am7	Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit.
sa6	Ich erfahre genügend Unterstützung von meinen unmittelbaren Vorgesetzten.
sa7	Die Anerkennung meiner Arbeit durch meine unmittelbaren Vorgesetzten ist angemessen.
sa12	Konflikte werden offen besprochen.
Faktor 2 „Arbeitsplatzgestaltung“	
am10	Die nötigen Arbeitsmittel sind vorhanden.
am11	Arbeitsräume sind in ausreichender Anzahl vorhanden.
am12	Die Einrichtung an meinem Hauptarbeitsplatz ist angemessen.
um4	Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann (z.B. Arbeitshöhe, Sitzhöhe, Beinfreiheit, Durchgangsbreiten/Sicherheitsabstände).
Faktor 3 „Soziale Beziehung zu Untergebenen“	
sa9	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Beschäftigten ist angemessen.
sa10	Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Beschäftigten.
Faktor 4 „Zeitliche Aspekte / Quantitative Anforderungen“	
aa1	Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit als ausgewogen.
aa2	Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen.
aa3	Meine Arbeitszeitverteilung ist mir angenehm.
aa4	Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen.

aa5	Meine Arbeitspausen sind ausreichend.
aa6	Meine Arbeitspausen sind störungsfrei.
am9	Ich werde bei meiner Arbeit selten unterbrochen.
Faktor 5 „Qualitative Anforderungen“	
am1	Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet.
am2	Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit.
am3	Meine Arbeit ist in angenehmem Maße abwechslungsreich.
am4	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z.B. Personalverantwortung).
am5	Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z.B. Infrastruktur, Material).
am6	Ich kann erkennen, ob ich meine Arbeit gut gemacht habe.
am8	Ich habe einen angemessenen Handlungsspielraum (z.B. Entscheidungsfreiheit) bei meiner Arbeit.
Faktor 6 „Anerkennung / respektvoller Umgang“	
sa1	Mein Kundenkreis für interne Dienstleistungen (z.B. auch Patienten, Leistungsempfänger) verhält sich angemessen.
sa2	Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.
sa3	Die Anerkennung meiner Arbeit durch den Mitarbeiterkreis oder Beschäftigte ist angemessen.
sa4	Ich erfahre genügend Unterstützung durch den Mitarbeiterkreis.
Faktor 7 „Konflikte“	
sa5	Mit meinen Mitarbeitenden gibt es keine Konflikte.
sa8	Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte.
sa11	Mit den Untergebenen/mir unterstellten Beschäftigten gibt es keine Konflikte.
Faktor 8 „Umgebungsbedingungen“	
um1	Den Geräuschpegel an meinem Hauptarbeitsplatz empfinde ich als angenehm.
um2	Die äußeren Bedingungen (Klima, Beleuchtung, Belüftung etc.) an meinem Hauptarbeitsplatz sind angenehm.
um3	Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen.

Die Modellfitstatistiken der CFA für einen, vier und acht Faktoren sind in Tabelle 15 und Tabelle 16 dargestellt. Alle Modelle erfüllten das Kriterium von $\chi^2 \leq 2df$ (176), sodass auch dem Signifikanzniveau der χ^2 -Tests vertraut werden kann. Keines der getesteten Modelle erreichte hier annähernd einen guten Anpassungswert. Für alle drei Modelle musste die Nullhypothese, dass die gewählte Kovarianzmatrix des jeweiligen Modells gleich der wahren Kovarianzmatrix ist, verworfen werden.

Da der χ^2 -Wert jedoch mit zunehmendem Stichprobenumfang ansteigt, sollte den deskriptiven Modellvergleichswerten und Parsimoniemaßen mehr Bedeutung beigemessen werden (176). Hinsichtlich des RMSEA (Root Mean Square Error of Approximation) erreicht

das 8-Faktoren-Modell den besten Wert, welcher auch als einziger noch knapp als akzeptabel betrachtet werden könnte. Auch diesem Kriterium nach muss somit davon ausgegangen werden, dass keine gute Übereinstimmung zwischen empirischer und implizierter Kovarianzmatrix der jeweiligen Modelle erreicht wurde. Umgekehrt verhielt es sich mit der SRMR (Standardized Root Mean Square Residual), die die geringste Diskrepanz zwischen der empirischen Kovarianzmatrix und der durch das 1-Faktor-Modell implizierten Kovarianzmatrix anzeigt. Allerdings lag die SRMR für das Ein-Faktor-Modell immer noch knapp über dem empfohlenen Grenzwert von 0,08 und kann nicht mehr als akzeptabel gelten. Sowohl der CFI (Comparative fit index) als auch der TLI (Tucker-Lewis Index) erreichten beide deutlich nicht den Schwellenwert von 0,95. Entsprechend muss auch anhand dieser Maße davon ausgegangen werden, dass die getesteten Modelle keinen ausreichend guten Erklärungsbeitrag gegenüber dem Nullmodell aufweisen.

Die kleinsten Informationskriterien (AIC und BIC) wurden für das 8-Faktoren-Modell beobachtet (vgl. Tabelle 16). Dies ist zwar kein Garant für ein ausreichend spezifiziertes Modell, allerdings deutet es darauf hin, dass mit zunehmender Faktorenzahl eine bessere Varianzaufklärung erreicht wird.

Tabelle 15: Modellvergleich der mit CFA geschätzten Modelle für 1, 4 und 8 Faktoren Teil I (n = 8.058)

Modell	SB Chi ²	df	RMSEA	SRMR	TLI	CFI
1 Faktor	43.962,49 ***	527	0,101	0,092	0,513	0,543
4 Faktoren	39.534,65 ***	527	0,096	0,196	0,563	0,589
8 Faktoren	23.575,666 ***	527	0,081	0,210	0,698	0,718

SB Chi² = chi²-skalierte Teststatistik nach Satorra-Bentler; ***p<.001.; df = Freiheitsgrade; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index; CFI = Comparative fit index

Tabelle 16: Modellvergleich der mit CFA geschätzten Modelle für 1, 4 und 8 Faktoren Teil II (n = 8.058)

Modell	AIC	BIC
1 Faktor	655.329,785	656.043,216
4 Faktoren	649.740,991	650.454,422
8 Faktoren	587.136,465	587.768,947

AIC = Akaike's information criterion; BIC = Bayesian information criterion

5 Diskussion der Ergebnisse und Implikationen für das weitere Vorgehen

Zur Beurteilung der wissenschaftlichen Güte des AIGScreen liegen verschiedene Daten aus mehreren Erhebungen vor. Hinsichtlich der Inhaltsvalidität sowie der Qualität des Fragebogens wurden qualitative Verfahren gewählt, welche Experten in den jeweiligen Fachgebieten wie z. B. Fragebogendesign oder Arbeitsmedizin beteiligten. Die Kriterien der Reliabilität und Konstruktvalidität sowie die Itemqualität ließen sich hingegen mittels quantitativer Methoden beurteilen. Entsprechend unterschiedlich wurde auch bei der Auswertung und Interpretation der Daten vorgegangen. Während für die expertenbeteiligenden Verfahren überwiegend unstandardisiertes Textmaterial vorlag, welches zunächst abstrahiert, objektiviert und vergleichbar gemacht werden musste, konnten die mithilfe statistischer Analysen gewonnen Kennwerte unmittelbar mit Vergleichsmodellen oder ggf. auch eindeutigen Grenzwerten verglichen und interpretiert werden. Letztlich waren beide methodischen Ansätze einerseits notwendig, da sich zum einen bestimmte Aspekte wissenschaftlicher Güte nicht standardisiert messen bzw. statistisch beurteilen lassen und umgekehrt andere Kriterien nicht ausschließlich statistisch überprüft werden können. Zum anderen zeigen die Ergebnisse aus den verschiedenen Erhebungen, dass der hier verfolgte mixed methods-Ansatz einen erheblichen Mehrwert für die gewonnenen Erkenntnisse darstellt. Erstens können die Ergebnisse aus den verschiedenen Datenquellen gegenübergestellt werden und stellen somit auch eine Evaluation der gewählten Methoden und damit gewonnenen Ergebnisse dar. Zum Beispiel können die Expertenurteile zur Qualität und Relevanz einzelner Items sowie der Skalen ihrer jeweiligen statistischen Performanz gegenübergestellt werden. Zweitens blieben bestimmte Informationen unentdeckt, hätte man sich auf eine einzige Methode verlassen. Insbesondere die Einblicke in die Bedürfnisse der Zielgruppe, welche im Rahmen der Expertenbefragung gewonnen werden konnten, lieferten zusätzlichen Informationsgewinn wie der Fragebogen für die Zielgruppe attraktiver gestaltet werden könnte und welche zusätzlichen Aspekte jenseits des aktuellen Forschungsstands und der Empfehlungen von GDA und BAuA berücksichtigt werden sollten. Dies kann zur Steigerung des Rücklaufs und somit zur Verlässlichkeit der Datenbasis beitragen.

Trotz einiger Einschränkungen zeichnen die Ergebnisse insgesamt ein positives Bild vom AIGScreen und er kann bereits jetzt als geeignetes Instrument für ein Grobscreening arbeitsbedingter psychischer Belastung und Steuerungsinstrument im Rahmen des Betrieblichen Gesundheitsmanagements bewertet werden, solange die Grenzen des Einsatzbereiches und der Interpretationsmöglichkeiten der Daten berücksichtigt werden. Da jedoch auch einige Verbesserungsmöglichkeiten identifiziert wurden und die Ergebnisse

möglichst umfassend genutzt und umgesetzt werden sollen, um eine noch bessere Fragebogenversion zu erarbeiten, liegt der Schwerpunkt der nun folgenden Diskussion auf den Defiziten des Instruments.

5.1 Fragebogenreview

Das interne Expertenreview des AIGScreen diente dazu, Schwächen im Fragebogendesign und Qualitätsmängel einzelner Items zu identifizieren und Verbesserungsmöglichkeiten zu erarbeiten sowie die Objektivität des Erhebungsinstrumentes zu bewerten. Diese wurde aufgrund des hohen Standardisierungsgrades sämtlicher Datenverarbeitungsschritte übereinstimmend von allen Experten und ohne Einschränkungen als gegeben betrachtet.

Bezüglich der sieben Kategorien, in denen die Experten den Fragebogen überprüfen sollten, wurden für einzelne Items in vier Kategorien potenzielle Probleme identifiziert. Dies betraf mögliche Fragereihenfolge- und Kontexteffekte, die impliziten Voraussetzungen, die die Erhebung an die Zielgruppe stellt, diverse Anmerkungen zu potenziell missverständlichen Formulierungen sowie der Antwortoption „Unzutreffend“, welche in einem missverständlichen Verhältnis zum Instruktionstext gesehen wurde. Für alle identifizierten Probleme konnten im Konsens Lösungen zur besseren Fragebogengestaltung gefunden werden. Deren konkrete Umsetzung wird im Kapitel „5.5.1 Anpassungen im Fragebogendesign“ im Detail beschrieben. Viele Anmerkungen bezogen sich auf Aspekte, für die nur eine potenzielle Gefährdung der Fragebogenqualität gesehen wurde. Z. B. konnten für bestimmte Items Beeinflussungen im Antwortverhalten durch vorausgehende Fragen nicht ausgeschlossen, aber auch nicht mit hoher Wahrscheinlichkeit angenommen werden. Hier bestand der beschlossene Lösungsansatz darin, vermutete Effekte der Fragereihenfolge zunächst statistisch zu überprüfen, bevor Veränderungen am Instrument vorgenommen werden sollen. Auch die Zahl der Antwortoptionen wurde kritisiert, jedoch auch hier im Konsens beschlossen, die statistischen Analysen abzuwarten. Hier bestand Einigkeit in der Expertengruppe, dass die aktuelle Forschung uneingeschränkt mehr Antwortoptionen im Sinne der wissenschaftlichen Güte einer Erhebung nahelegt. Gegen eine Erweiterung der Antwortskala sprach jedoch die Vergleichbarkeit mit bereits erhobenen Daten. Zudem fürchtete man eine höhere kognitive Belastung der Befragten und schnellere Ermüdung. Da die deskriptiven Statistiken, wie von den Experten vermutet, ungünstige Antwortverteilungen (nicht normalverteilt, hohe Deckeneffekte, z. T. geringe Streuung) nachwiesen, wurde zu einem späteren Zeitpunkt die Erarbeitung einer alternativen Antwortskala vorgenommen. Dabei wurde auch die Verbalisierung überarbeitet, welche im Rahmen des Reviews einen weiteren Diskussionspunkt darstellte. Hier wurde, wie von den

Experten vorgeschlagen, eine bessere Abstimmung der ausformulierten Antwortoptionen auf den jeweiligen Frageinhalt vorgenommen. Damit soll erreicht werden, dass der Interpretationsspielraum bei der Beantwortung des jeweiligen Items sinkt bzw. die Intersubjektivität im Frageverständnis steigt und somit auch die Reliabilität der Messung verbessert wird. Gegen die in der Diskussion geäußerte Befürchtung von satisficing-Effekten durch häufig wechselnde Antwortskalen sprach, dass Ermüdung auch durch eine unzureichende Passung von Antwortoptionen und Frageinhalt, welche eine höhere Interpretationsleistung verlangt und höhere kognitive Anforderungen an die Probanden stellt, auftreten kann.

Einige Items erschienen problematisch, weil sie nicht auf alle Probanden zutreffen, wenn diese z. B. nicht in Führungspositionen sind oder keinem sog. internen Kundenkreis zuarbeiten. Diesem Problem könnte nur sinnvoll durch den Einsatz von Filterfragen begegnet werden, wenn man die Lösung durch eine „Unzutreffend“-Antwortoption für nicht ausreichend bewertet. Die Vermutung der Experten, dass Items, die auf bestimmte Personengruppen nicht zutreffen, eine höhere Missingquote aufweisen, konnte jedoch nur nicht in der Itemanalyse bestätigt werden. Das z. B. als problematisch thematisierte Item aa2 („Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen“) wurde nur von 2,6% der Befragten nicht beantwortet. Die Antwortoption „Unzutreffend“ wurde hingegen von 15,2% gewählt, sodass davon ausgegangen werden kann, dass die Frageinstruktion von den Befragten verstanden und die für sie jeweils adäquate Antwortmöglichkeit gewählt wurde. Insgesamt zeigten die Antworthäufigkeiten der „Unzutreffend“-Kategorie im Verhältnis zur Missingquote, dass scheinbar auch in solchen Fällen die überwiegende Mehrheit der Personen, auf die ein Merkmal nicht zutraf, entsprechen die Antwortoption „unzutreffend“ wählten. Der Einsatz von Filterfragen sollte in Zukunft ggf. dort erwogen werden, wo mehrere Fragen aufeinanderfolgen, die sich auf dasselbe Merkmal beziehen (z. B. Mitarbeiterverantwortung), von welchem nicht angenommen werden kann, dass es auf alle Befragten zutrifft, da dann der Bearbeitungsaufwand für diese Personengruppe tatsächlich reduziert werden kann. Solange sich durch den Einsatz von Filterfragen jedoch die Gesamtmenge von zu bearbeiteten Items nicht reduzierte, erschien dieser Lösungsansatz wenig zielführend.

Durch die Modifikation der Antwortskalen konnte auch dem Problem doppelter Verneinungen bei den Items, die Konflikte mit Kollegen, Mitarbeitern und Vorgesetzten erfassen, gelöst werden. Die Richtung der Polung konnte beibehalten werden, sodass kein Umdenken der Probanden notwendig ist, aber indem anstelle von Zustimmung die Häufigkeit erhoben werden soll, sind die Antwortoptionen eindeutiger. Da die Antwortverteilung aller drei Items jedoch wie die übrigen Items eine rechtssteile Verteilung aufwies, ist es unwahrscheinlich, dass viele Befragte Probleme bei Beantwortung im

bisherigen Format hatten. Hier würde man eine stärkere Verschiebung zur Mitte bzw. eine häufigere Nutzung beider Randkategorien erwarten, wenn es durch die doppelte Verneinung zu einer Verwechslung von positivem und negativem Antwortpol gekommen wäre.

5.2 Expertenbefragung in der Zielgruppe

Zur Beurteilung der Inhaltsvalidität des AIGScreens wurde eine Befragung von Fachexperten, die sich zugleich aus der Zielgruppe rekrutierten, durchgeführt. Als Experte gilt in diesem Zusammenhang jede Person, die detailliertes oder spezielles Wissen in einem spezifischen und abgrenzbaren Bereich hat (126). Dieses Kriterium kann für alle adressierten Berufsgruppen (Arbeitsmedizin, Gesundheitsmanager, Fachkräfte für Arbeitssicherheit etc.) im Geschäftsbereich des BMVg als gegeben betrachtet werden. Der Vorteil einer externen Beratung bei der Durchführung einer Bedarfs- und Gefährdungsanalyse ist, dass diese Perspektive neutral und von betrieblichen Belangen unbeeinflusst ist. Der Nachteil besteht jedoch darin, dass Externen häufig wichtige Informationen zu den spezifischen Rahmenbedingungen eines Unternehmens fehlen. Hier sind die Mitarbeitenden selbst gewissermaßen die besseren Experten, da sie ihr Arbeitssetting am besten kennen und potentielle Gefährdungen und zu berücksichtigende Aspekte aufzeigen, die unentdeckt bleiben könnten, wenn ausschließlich eher abstrakten, generellen Empfehlungen bei der Umsetzung gefolgt wird. Ein Vorteil von großen Unternehmen ist, dass sich dort häufig Personen finden, die beides abdecken: Sie sind als Fachkraft für Arbeitssicherheit, Betriebsarzt, Gesundheitsmanager oder in einer ähnlichen Profession ausreichend qualifiziert, um allgemeine Gefährdungsfaktoren der Arbeit erkennen und benennen zu können und sie verfügen als Angehörige des Betriebes über umfassendes Wissen über das Arbeitssetting. Dies ersetzt eine neutrale Außenperspektive nicht, ergänzt jedoch die allgemeinen Empfehlungen um wertvolle Informationen, die für Außenstehende andernfalls unerkannt blieben. Die Datenerhebung fand im Rahmen eines Forschungsprojekts zum BGM im Geschäftsbereich des BMVg statt. Hierbei handelt es sich um einen Arbeitgeber mit sehr unterschiedlichen Standorten im In- und Ausland und sehr unterschiedlichen Tätigkeitsfeldern, welche neben dem militärischen Bereich z. B. auch Verwaltung, Krankenhäuser, IT, Logistik oder Forschungseinrichtungen umfassen. Daher kann davon ausgegangen werden, dass die Ergebnisse auf andere Arbeitssettings übertragbar sind und nicht exklusiv für den militärischen Kontext gelten.

Der gegenüber expertenbasierten Erhebungen häufig geäußerten Kritik, dass Expertenmeinungen immer subjektiv sind und Ergebnisse stark verzerrt sein können, wurde mit einer für dieses Verfahren sehr großen Stichprobe ($n = 173$) begegnet. Mögliche

Verzerrungseffekte durch Einzelmeinungen verlieren so im Mittel aller Bewertungen an Gewicht und gleichen sich aus. Zudem erfolgte die Auswahl der Experten anhand mehrerer Kriterien, sodass auch Verzerrungseffekte hinsichtlich der Stichprobenzusammensetzung weitgehend ausgeschlossen werden konnten. Grundsätzlich müssen abweichende Einzelmeinungen jedoch nicht falsch oder irrelevant sein. Im Rahmen der Auswertung wurde sich jedoch im Sinne einer Konsensfindung bezüglich der zukünftigen Fragebogengestaltung an der Mehrheitsmeinung orientiert. Allerdings sollten bei zukünftigen Anpassungen immer wieder auch die Einzelmeinungen aufgegriffen und berücksichtigt werden.

Während die standardisiert erhobenen Variablen zu Relevanz und Vollständigkeit der Items und Skalen direkt ausgezählt werden konnten, musste das entstandene Textmaterial aus den Freitextantworten zu den sich anschließenden vertiefenden Fragen zunächst im Rahmen einer quantitativen Inhaltsanalyse auf eine zählbare Abstraktionsebene codiert werden. Die mittlere Interdecoderübereinstimmung für die vollständige Codierung des Datenmaterials lag bei 89%, sodass davon ausgegangen werden konnte, dass das angewandte Kategorienschema sich gut für den Einsatz am Textmaterial eignete und die durch Codierung entstandenen Kategorien eine verlässliche Datenbasis darstellen (141, 143).

Für die Aspekte, die in geschlossenen Antwortformaten erhoben wurden, wurden auch Vergleiche zwischen den Berufsgruppen vorgenommen. Hier ließen sich keine wesentlichen Unterschiede bei der Bewertung von Relevanz und Vollständigkeit zwischen den einzelnen Gruppen feststellen. Vermutet wurde, dass bestimmte Berufsgruppen bestimmte Merkmalsbereiche und Aspekte überproportional positiv bewerten könnten (z. B. Sicherheitsingenieure und Fachkräfte für Arbeitssicherheit den Merkmalsbereich Arbeitsumgebung). Es konnte jedoch für keine Gruppe eine Häufung positiver Antworten in einem bestimmten Merkmalsbereich erkennen lassen. Lediglich die Berufsgruppe „Sonstige“ wies ein tendenziell negativeres Bewertungsmuster bezüglich der Vollständigkeit der Skalen auf. Denkbar ist, dass es sich bei dieser Gruppe um eine besonders kritische Stichprobe von Beschäftigten im Geschäftsbereich BMVg handelt, welche nicht unbedingt Zielgruppe der Befragung waren, die sich jedoch Zugang zur Umfrage verschafften. Für diese Vermutung spricht, dass die von dieser – sehr kleinen Personengruppe (n = 11) – angegebene Berufsgruppenzugehörigkeit häufig sehr vage ist oder auch nicht der nach den formulierten Auswahlkriterien definierten Zielgruppe entspricht (z. B. „Offizier im Truppendienst mit M.Sc.“, „BGM-Teilnehmerin“). Obwohl eine breite Beteiligung aller Personengruppen aus der Zielgruppe des AIGScreen zur Beurteilung der Augenscheinvalidität hilfreich ist, ist für die Teilnahme an der Befragung spezifische Fachexpertise im Bereich Betriebliches Gesundheitsmanagement erforderlich.

Dass sich dennoch Personen, die dieses Kriterium nicht erfüllen, ermutigt sahen, an der Befragung teilzunehmen kann entweder damit erklärt werden, dass die Zugangslinks und Teilnahmebedingungen zur Befragung nicht ausreichend restriktiv vergeben bzw. klar kommuniziert wurden oder einige wenige Personen wollten bestimmte Kritikpunkte kommunizieren, für die sie keinen anderen geeigneten Adressaten gefunden haben. Letzteres sollte auch in Zukunft im Rahmen des Erhebungsdesigns berücksichtigt werden. Für die Augenscheinvalidität und damit verbundene Teilnahmebereitschaft und Rücklaufquote kann es förderlich sein, wenn auch allen Personen aus der Zielgruppe Raum für kritische Rückmeldungen gegeben wird. Eine Fragestellung, die für die Expertenbefragung implizit mit verfolgt wurde, war, ob aus den Einschätzungen der einzelnen Gruppen auch Schlüsse bezüglich der Zusammensetzung von Expertengruppen bzw. Gesundheitsgremien in der betrieblichen Praxis gezogen werden können. Die Teilnehmendenstruktur und die Rückmeldungen aus der Gruppe „Sonstige“ sprechen dafür, dass in Gesundheitsgremien auch Beschäftigte vertreten sein sollten, die nicht zur Expertengruppe gehören bzw. dass für alle Beschäftigten offene Beteiligungsstrukturen angeboten werden sollten, um offen für Rückmeldung und Kritik aus allen Bereichen zu bleiben und die Akzeptanz von BGM insgesamt zu erhöhen.

Insgesamt wurde der AIGScreen jedoch sehr positiv bewertet und erhielt mit Blick auf die mittleren Relevanz- und Vollständigkeitsbewertungen mehrheitlich positive Beurteilungen. Dies spricht auch für eine hohe Augenscheinvalidität und Akzeptanz des Instruments in der Zielgruppe, da die Bewertungen ausschließlich von Vertretern der Zielgruppe vorgenommen wurden. Kein Item wurde bezüglich seiner Relevanz auf der Bewertungsskala nach Polit et al. (2007) schlechter als „good“ (140) bewertet. Allerdings lassen sich diese positiven Einschätzungen auf Itemebene nicht vollständig auf die vier Itemskalen übertragen, für die strengere Grenzwerte bezüglich des Relevanzindikators gelten. Die Relevanz der Dimensionen Arbeitsumgebung (CVI = 0,876) und Soziale Aspekte (CVI = 0,877) liegen relativ nah am kritischen Schwellwert von 0,90. Problematischer, das heißt hinsichtlich ihrer Inhaltsvalidität überarbeitungsbedürftig, sind die Subskalen Arbeitsorganisation (CVI = 0,784) und Arbeitsinhalt und -aufgabe (CVI = 0,802). Bei Arbeitsorganisation fällt auf, dass bis auf das Item „Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit ausgewogen“ mit $k = 0,92$ alle Items deutlich unter dem notwendigen Schwellwert für eine als relevant einzustufende Gesamtskala liegen. Das bedeutet, dass auch der Austausch oder die Optimierung einzelner Items nicht zu einer Verbesserung des CVI für die gesamte Skala führen würde. Mit Blick auf die Skala Arbeitsinhalt und Arbeitsaufgabe fällt insbesondere das Item „Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material)“ negativ mit $k = 0,63$ auf. Allein durch Ausschluss dieses Items würde sich die Gesamtskala

auf einen $CVI = 0,84$ verbessern. Fragen zur Ergebniskontrolle und der eigenen Qualifikation und Weiterbildungsmöglichkeiten werden mit $k = 0,88$ bzw. $k = 0,89$ deutlich besser bewertet und scheinen aus Sicht der Experten eine wesentlich höhere Relevanz zu haben. Hinweise darauf, wie Items in Zukunft formuliert werden sollten, um eine höhere Relevanzbewertung zu erhalten oder welche zusätzlichen Aspekte erfasst werden müssten bzw. welche nicht mehr erfasst werden sollten, um die Inhaltsvalidität einer Skala zu verbessern, konnten aus den vielzähligen Freitextkommentaren gezogen werden. Eine ausführliche Diskussion dazu findet sich im nachfolgenden Abschnitt „5.2.1 Vergleich der Expertenmeinungen mit Empfehlungen der GDA und wissenschaftlichen Erkenntnissen“.

Auffallend ist, dass die Relevanz der Items mit Fortschreiten der Befragung zunehmend positiv ausfällt. Dies kann schlicht darauf zurückzuführen sein, dass die Experten die hinteren Items als relevanter bewerten als die erstgenannten Items. Es kann auch damit zusammenhängen, dass man den beiden letztgenannten Dimensionen Soziale Aspekte und Arbeitsumgebung grundsätzlich mehr Relevanz zuschreibt und sich das entsprechend positiv auf die Bewertung der einzelnen Items auswirkt. Eine dritte Erklärung wäre aber auch, dass die Befragten ihre Kritik und Ergänzungsvorschläge zum Zeitpunkt der Bewertung der dritten und vierten Skala bereits getan haben und diese beiden Dimensionen von ihrer Position im Fragebogen „profitieren“. Für diesen Ansatz spricht auch, dass ein Teil der Befragten die Befragung hier bereits abgebrochen hatte und die Bewertung möglicherweise von einer motivierteren und wohlwollenderen, da stark am BGM interessierten Gruppe vorgenommen wurde. Diese Annahmen sind jedoch nur spekulativ und ließen sich ausschließlich mit einer randomisierten Fragenreihenfolge überprüfen. Gegen mögliche Ermüdungseffekte spricht wiederum, dass der zuletzt präsentierte Merkmalsbereich Arbeitsumgebung die meisten Freitextkommentare aufweist. Hierbei enthalten die meisten Antworten Ergänzungsvorschläge für weitere Aspekte.

Obwohl die absolute Mehrheit der Befragten diesen sowie auch die anderen drei Merkmalsbereiche als vollständig repräsentiert bewertet, wurden von den Befragten weitere Aspekte vorgeschlagen, die erhoben werden sollten. Dabei liegt bei allen vier Skalen der Anteil derjenigen, die Ergänzungsvorschläge für zusätzliche Aspekte formulieren, deutlich über dem Anteil derjenigen, die die Skala als nicht vollständig repräsentiert bewerten.

Einige wenige Experten wiesen darauf hin, dass sich bestimmte Berufsgruppen bzw. Einheiten nicht mit den Fragen identifizieren konnten und regten eine inhaltliche Abstimmung auf bestimmte Organisationsbereiche an. Dies widerspricht allerdings dem Ziel der Vergleichbarkeit der Befragungsergebnisse. Zu prüfen wäre vielmehr, ob es sich hierbei um Einzelmeinungen handelt oder tatsächlich bestimmte Berufsgruppen oder Einheiten ihre berufliche Situation im Fragebogen nicht abgedeckt sehen. In diesem Fall

müssten die Fragen weiter abstrahiert werden, sodass sie auf sämtliche Arbeitssettings übertragbar sind. Da dieses Problem jedoch nur von einem sehr kleinen Teil der Experten ($n = 2$) benannt wurde und die Mehrheit der Experten den Fragebogen in seiner inhaltlichen Ausgestaltung überwiegend positiv bewertete, wurde vermutet, dass es sich hierbei um Einzelmeinungen handelt. Dafür spricht auch, dass sich bei keiner Einheit, Berufsgruppe oder Organisationseinheit ein systematisch niedrigerer Rücklauf oder ein anderweitig problematisches Antwortverhalten feststellen ließ. Würden sich bestimmte Personengruppen mit dem Fragebogen nicht identifizieren können, wäre dies hingegen zu erwarten.

Mehrere Experten sahen zudem Bedarf, das individuelle Gesundheitsverhalten und die Gesundheitssituation der Mitarbeitenden zu erfassen. Der Ansatz, neben Belastungen und Beanspruchungen auch Ressourcen zu berücksichtigen, entspricht dem Belastungs-Beanspruchungs-Modell (vgl. Kapitel 2.1.2). Er wird auch bereits im Rahmen der Mitarbeitendenbefragung berücksichtigt, indem umfassend das physische und psychische Wohlbefinden gemessen am Vorliegen bestimmter Erkrankungen und Beschwerden sowie das individuelle Gesundheitsverhalten gemessen an verschiedenen Indikatoren zu Ernährungs-, Bewegungs- und Arbeitsverhalten erfasst werden. Hierbei handelt es sich jedoch um andere Merkmalsbereiche, die auch anders als die subjektiv empfundene Arbeitssituation operationalisiert wurden. Beispielsweise handelt es sich bei den Fragen zum Vorliegen bestimmter Erkrankungen nicht um latente Merkmale, sondern um objektiv beurteilbare Aspekte, welche sich in einem formativen Messmodell zusammenfassen ließen. Für die Validierung dieser Skalen müsste ein anderer methodischer Ansatz verfolgt werden und es ist nicht zielführend, die verschiedenen Merkmalsbereiche zu mischen. Der AIGScreen dient ausschließlich der Erfassung der Arbeitssituation und soll als BGM-Instrument insbesondere für die Maßnahmenplanung und -entwicklung im Rahmen der Verhältnisprävention eingesetzt werden. Insofern können Ergänzungsvorschläge, die sich auf Aspekte beziehen, welche primär im Rahmen der Verhaltensprävention relevant sind, an dieser Stelle nicht berücksichtigt werden.

5.2.1 Vergleich der Expertenmeinungen mit Empfehlungen der GDA und wissenschaftlichen Erkenntnissen

Insgesamt lässt sich eine hohe Übereinstimmung zwischen den von den Experten als relevant identifizierten Aspekten und den Empfehlungen von GDA (130, 190) und BAuA (186, 97) feststellen. Dennoch finden sich mit Blick auf einige Aspekte auch deutliche Unterschiede. Dabei muss für bestimmte Aspekte jedoch hinterfragt werden, inwiefern es sich hierbei um settingspezifische Einschätzungen handelt. Beispielsweise fällt auf, dass

der Aspekt **Sinnhaftigkeit**, welcher nicht Gegenstand der BAuA- und GDA-Empfehlungen ist, häufig im Zusammenhang mit Problemen der Qualifikation, insbesondere Über- bzw. Unterforderung, sowie der Personal- und Aufgabenplanung erwähnt wurde. Häufige Personalrotationen im Geschäftsbereich BMVG lassen zumindest vermuten, dass es fluktuationsbedingt häufiger dazu kommt, dass Personen nicht oder zumindest nicht vollständig ihrer Qualifikation entsprechend eingesetzt werden und diese Erfahrung in die Expertenurteile einfließt. Eine hohe Diskrepanz zwischen Fähigkeiten und Arbeitsanforderungen lässt es wiederum wahrscheinlicher werden, die Sinnhaftigkeit des eigenen Tuns infrage zu stellen. Daher sollte bei der Durchführung einer Gefährdungsbeurteilung erwogen werden, ob oder wie wahrscheinlich Mitarbeitende nicht ihrer Qualifikation entsprechend eingesetzt werden. Diese Gefährdung könnte auch in anderen Betrieben auftreten, die weniger Personalfluktuations- und -rotation verzeichnen, wo jedoch die Aufgaben in Quantität und Qualität häufiger variieren.

Häufig stimmten die Einschätzungen der Experten mit den inhaltlichen Empfehlungen der GDA jedoch überein. So wurden beispielsweise die drei Items am schlechtesten, das heißt als irrelevant und aus dem Fragebogen zu streichen, bewertet, welche auch nicht als zu berücksichtigende Gefährdungen im Rahmen der Leitlinie zur Überwachung psychischer Belastung am Arbeitsplatz (50) aufgeführt sind. Dies betrifft die Objektverantwortung, die körperliche Leistungsfähigkeit sowie das Verhältnis von Arbeitszeit zu Freizeit. Diese Rückmeldungen gaben den Anlass, die jeweiligen Items im internen Expertenreview zu evaluieren und die Kritik bei der Modifikation des AIGScreen zu berücksichtigen.

Bezüglich des Gefährdungsfaktors **Verantwortung** wurde unter Berücksichtigung der aktuellen Empfehlungen und des Forschungsstandes deutlich, dass es für die Bewertung der psychischen Belastung nicht entscheidend ist, wofür eine Person Verantwortung trägt, sondern vor allem, ob Verantwortlichkeiten klar definiert sind oder vielmehr Rollenambiguität besteht (50, 191). Inwiefern die Menge an Verantwortung zu viel oder zu wenig ist, kann nicht anhand von allgemeinen Grenzwerten bestimmt werden, sondern hängt stets von individuellen Ressourcen und dem subjektiven Empfinden ab (192). Da die Frage nach der Angemessenheit des Umfangs an Personalverantwortung jedoch als relevant von den Experten bewertet wurde und neben unklaren Verantwortlichkeiten auch eine Überforderung durch zu viel bzw. Monotonie oder eine verminderte Arbeitsmotivation durch zu wenig Verantwortung (96, 193) als kritische Ausprägungen dieses Aspekts ausgeschlossen werden sollten, sollte auch eine Frage zum Umfang von Verantwortung im künftigen Fragebogen enthalten sein. Weitere Kommentare zu den Gefährdungsfaktoren Über- und Unterforderung bezogen sich auf die Übereinstimmung von individuellen Kompetenzen und Anforderungen der Aufgabe, aber auch auf die Qualität der Einarbeitung sowie die Komplexität der Aufgabe. Die beiden erstgenannten Aspekte werden von der

GDA ebenfalls als kritische Ausprägungen des Aspekts Qualifikation benannt (50) und sollten entsprechend auch genauso im Fragebogen abgebildet sein. Aufgabenkomplexität wird vor allem durch Abwechslungsreichtum und die Vollständigkeit einer Tätigkeit bestimmt.

Dass das Item zur Erfassung der Aufgabenvielfalt bzw. **Variabilität** nicht als relevant eingestuft wird, überraschte daher insofern, als dass diese nicht nur bei den Kommentaren in Bezug auf die potenziellen Gefährdungen im Bereich Qualifikation eine wichtige Rolle spielte, sondern auch, weil die Studienlage hier sehr eindeutig einen positiven Effekt auf die Arbeitszufriedenheit und einen negativen Zusammenhang mit depressiven Symptomatiken aufzeigt (96). Ausführliche Kommentare zu diesem Item lassen jedoch vermuten, dass die mit dem Item („Meine Arbeit ist in angenehmem Maße abwechslungsreich“) gewonnen Informationen nicht als zielführend für die weitere Maßnahmenplanung erachtet wurden und die negativen Bewertungen eher auf die mangelnde Spezifität der Formulierung zurückzuführen sind. Allerdings ließen auch manche Kommentare erkennen, dass einzelne Experten die Monotonie bestimmter Tätigkeiten als zwangsläufig gegeben und unveränderlich betrachteten, sodass die Erhebung von Variabilität in diesen Tätigkeitsbereichen keinen Informationsgewinn bringen bzw. keinen Mehrwert bzgl. der Verbesserung der Arbeitssituation leisten kann. Diese Einschätzung, dass manche Tätigkeiten genuin monoton seien, übersieht jedoch, da Aufgabenvielfalt nicht nur vertikal, sondern auch horizontal variabel gestaltet werden kann und da auch der Arbeitsablauf an sich Gestaltungsspielraum ermöglicht (97). Dass es sich lohnt, diesen Aspekt im Rahmen der Gefährdungsbeurteilung zu berücksichtigen und ggf. Interventionsmaßnahmen abzuleiten, zeigt das Scoping Review von Rosen (194). Danach können selbst die Tätigkeiten, die durch ein hohes Maß an Monotonie gekennzeichnet sind, durch Veränderungen in der Ablauforganisation mittelbar die Variabilität beeinflussen.

Auch der Aspekt **Vollständigkeit** wurde von den Befragten als relevant bewertet. Die Vollständigkeit einer Tätigkeit umfasst sowohl die hierarchische Vollständigkeit, also unterschiedlich komplexe und kognitiv anspruchsvolle Aufgaben, sowie die sequenzielle Vollständigkeit, welche voraussetzt, dass die Aufgabenerledigung Planung, Vorbereitung, Ausführung und Kontrolle der Aufgabe umfasst. Beschäftigte sollten demnach in sich geschlossene, sinnvolle Arbeitseinheiten erledigen können, anstatt in Fragmenten zu arbeiten oder nur Teilstücke der Aufgabe zu erkennen (96). Dies würde wiederum die Aufgabenkomplexität bzw. die damit verbundenen Qualifikationsanforderungen beeinflussen und auch die Sinnhaftigkeit einer Aufgabe in der Wahrnehmung der Ausführenden erhöhen. Der Aspekt Sinnhaftigkeit wurde auch seitens der Experten immer wieder in diesem Zusammenhang genannt. Eine vollständige, das heißt, in ihrem Gesamtprozess erkennbare Aufgabe, stellt eine notwendige Voraussetzung für das

Sinnerleben einer Aufgabe dar, weshalb die Nennung dieses Aspekts im Zusammenhang mit Vollständigkeit nachvollziehbar ist und weshalb die Erhebung dieses Aspekts erwogen werden sollte, auch wenn die Notwendigkeit nicht explizit aus der GDA-Leitlinie hervorgeht.

Insgesamt wurden die **arbeitszeitbezogenen Aspekte** zwar von den befragten Experten im Geschäftsbereich des BMVg als relevant bewertet und häufig und wurden sogar zusätzliche zu erfassende Aspekte von den Befragten vorgeschlagen, gleichzeitig wurde aber die Streichung einzelner Aspekte verlangt. Das betrifft einmal die Frage nach ausreichenden Arbeitspausen sowie die Aussage „Meine Arbeitszeitverteilung ist mir angenehm“ und das oben bereits genannte Verhältnis von Arbeitszeit und Freizeit. Ein Einwand der bezüglich aller Aspekte mehrfach aufgeführt wurde, ist, dass hier gesetzliche Regelungen bestehen und daher subjektives Empfinden nicht relevant sei. Dies widerspricht jedoch dem Ansatz der psychischen Gefährdungsbeurteilung, wo das subjektive Belastungsempfinden durchaus im Fokus steht (45). Die Einschätzung von 10 Befragten, dass das Verhältnis von Arbeits- und Freizeit, nicht erhoben werden sollte, ist eventuell eher ein Problem der missverständlichen Formulierung ("Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen."), da hier das rein quantitative Verhältnis von Arbeits- und Freizeit verstanden werden kann. Das häufiger angeführte Argument, wonach die Arbeitszeiten gesetzlich geregelt sind und somit eine Höchstgrenze der Arbeitsdauer festgelegt ist, legt das zumindest nahe. Allerdings spricht das eher dafür, das Item anders bzw. eindeutiger zu formulieren. Denn der Einfluss des Arbeitszeitregimes auf die Möglichkeiten der Freizeitgestaltung hat einem BAuA-Gutachten zufolge einen Effekt auf die Gesundheit und sollte damit auch Teil der psychischen Gefährdungsanalyse sein (185). Den Aspekt flexible Arbeitszeitgestaltung zu erfassen, wurde zudem von fünf Experten gefordert. Dass dieser Aspekt mit in die Gefährdungsbeurteilung aufgenommen werden sollte, deckt sich mit aktuellen Befunden der BAuA (195). In den Empfehlungen zur Gefährdungsbeurteilung taucht der Aspekt Flexibilität jedoch höchstens als Randnotiz auf, wenn es um Gefährdungen durch eine einseitige Flexibilisierung seitens des Arbeitgebers z. B. durch Arbeiten auf Abruf oder durch die Digitalisierung bedingte räumliche und zeitliche Entgrenzung von Arbeit geht (185, 50, 196). Eine umfassende Analyse mit Blick auf eine verbesserte Work-Life-Balance ist bisher nicht empfohlen. Dennoch erscheint es sinnvoll, eine flexible Arbeitszeitgestaltung zu erfassen, da sie auch einen Aspekt des persönlichen Handlungsspielraums darstellt und zudem ermöglicht, die individuell wahrgenommene Verteilung von Arbeitszeit zu erfassen (197). Das direkte Erfragen der Einhaltung gesetzlicher Bestimmungen zur Arbeitszeit erscheint hingegen weder hinsichtlich der Fragebogenökonomie noch hinsichtlich der Verlässlichkeit der Antworten sinnvoll, da deren Überprüfung und Einhaltung zu einer Aufgabe des Arbeitgebers ist und zum anderen nicht davon ausgegangen werden kann, dass Arbeitnehmende alle

gesetzlichen Regelungen im Detail kennen. Da das Vorhandensein gesetzlicher Regelungen und die damit verbundenen Pflichten des Arbeitgebers jedoch kein Garant für deren Einhaltung sind, sollten zumindest kritische Ausprägungen des Aspekts Arbeitszeit wie ein unzureichendes Pausenregime oder häufige Mehrarbeit ebenfalls erfasst werden. Sie stellen gewissermaßen auch die Kehrseite flexibler Arbeitszeitgestaltung dar. Ein zukünftiges Item zum Aspekt Überstunden sollte jedoch auch den Einfluss individuellen Verhaltens berücksichtigen, um im Rahmen vom BGM eine Entwicklung von Maßnahmen zu ermöglichen, die unter anderem auch auf die Verhaltensebene abzielen und Ressourcen stärken können (z. B. Zeitmanagement).

Ergonomie und physische Belastung stellen sowohl in den offiziellen Empfehlungen als auch in der Einschätzung der befragten Experten einen wichtigen Aspekt dar. Was bisher in den Veröffentlichungen von GDA und BAuA jedoch nicht thematisiert wird, wurde gleich mehrfach seitens der Experten gefordert: Die ergonomische Arbeitsgestaltung sollte auch darauf überprüft werden, ob sie adäquat für Menschen mit Behinderungen ist. Nicht passend scheint in diesem Kontext hingegen das Item "Die körperlichen Anforderungen meiner Tätigkeit sind angemessen". Dieses empfanden nicht alle Experten relevant und einige forderten explizit dessen Streichung. Hier wäre zu klären, ob die Frage nach der Angemessenheit als irrelevant von den Experten bewertet wird, da es für die physische Belastung konkrete Richt- und Schwellenwerte gibt oder ob grundsätzlich der Aspekt körperliche Anforderung als unwichtig empfunden wurde. Letzteres würde den Empfehlungen von BAuA und GDA diametral gegenüberstehen. Zwar könnte man hier argumentieren, dass schwere körperliche Arbeit primär eine physiologische Belastung darstellt, allerdings ist diese auch mit psychischer Ermüdung assoziiert (98). Aus diesem Grund sollte die physische Belastung auch im künftigen Screening enthalten sein. Hier bietet sich sogar eine dezidiertere Erhebung des Merkmals an, da es sich um eine der wenigen Gefährdungen handelt, für die konkrete Beurteilungsmaßstäbe definiert sind (198) und für die objektive, personenunabhängige kritische Ausprägungen erhoben werden können (199).

An diesem Aspekt zeigt sich ein grundsätzliches Dilemma bei der Beurteilung psychischer Gefährdungen am Arbeitsplatz: Eine Eindeutige Trennung zwischen psychischen und allen anderen Faktoren ist nicht möglich. Insbesondere im Merkmalsbereich der Arbeitsumgebung gibt es viele Gefährdungen, deren Wirkung man zunächst als physiologisch (z. B. Lärm, Gefahrstoffe, Unfallgefahren) betrachten würde und deren Analyse vermutlich bereits in anderer Form (z. B. Begehungen durch Fachkräfte für Arbeitssicherheit) abgedeckt ist, die aber eben auch Effekte auf die psychische Gesundheit haben können. Es erscheint wenig sinnvoll diese Informationen mehrfach zu erheben und außerdem würde es die Beschäftigten sehr wahrscheinlich in ihren zeitlichen und kognitiven

Ressourcen überfordern sämtliche Aspekte, die im Rahmen einer vollständigen Gefährdungsbeurteilung erfasst werden müssen, im Rahmen eines Fragebogens zu erheben. Andererseits kann eine gewisse Redundanz in den Daten auch den Vorteil mit sich bringen, dass die Belastbarkeit einer Informationsquelle durch Verwendung eines zweiten Erhebungsformats überprüft wird und man somit eine verlässlichere Datenbasis generiert. Um hier ein angemessenes Verhältnis zwischen Ressourcenaufwand und Nutzen zu wahren, ist die Beteiligung von Experten vor Ort sehr gewinnbringend, denn diese kennen sowohl die spezifischen Rahmenbedingungen, wissen also, welche Ergänzungen sinnvoll sein können und welche Informationen bereits an anderer Stelle vorliegen und deren erneute Erhebung keinen Mehrwert bringen würde.

Vor diesem Hintergrund war es zunächst etwas überraschend, dass seitens der Experten mehrfach vorgeschlagen wurde, Fragen zum Merkmalsbereich **Arbeitssicherheit** wie z. B. Unfallverhütung, Schutzausrüstung oder Gefahrstoffe zu ergänzen. Da, wie oben bereits erläutert, ein Einfluss physikalischer, chemischer oder biologischer Faktoren aus der Arbeitsumgebung auch die individuelle Psyche beeinflussen kann, ist ein Zusammenhang zwischen psychischen und anderen Gefährdungen jedoch zu berücksichtigen (50, 186). Dies gilt nicht nur hinsichtlich potenzieller psychischer Belastungen durch das Auftreten anderer Faktoren, sondern auch umgekehrt haben psychische Belastungen einen Einfluss auf den Arbeitsschutz, da sie das Unfallgeschehen nachweislich beeinflussen (200–203). Insofern ist es durchaus sinnvoll technischen Arbeitsschutz und psychische Gefährdungen nicht vollkommen getrennt voneinander zu betrachten, da in beiden Bereichen Maßnahmen auch auf die Gefährdungen im jeweils anderen Bereich einen positiven Effekt haben können. Überraschend war, dass sich keine berufsgruppenspezifischen Schwerpunkte oder Häufungen bei den Kommentaren zeigten. Beispielsweise wäre zu erwarten gewesen, dass die häufigen Nennungen im Bereich Arbeitsumgebung auf die Berufsgruppe der Sicherheitsingenieure und Fachkräfte für Arbeitssicherheit zurückzuführen sind, da dies deren Arbeitsschwerpunkt darstellt. Faktisch kamen aus dieser Gruppe die wenigsten Vorschläge zu diesem Merkmalsbereich, was sehr wahrscheinlich nicht darauf zurückzuführen ist, dass sie diesen Bereich mittels AIGScreen als vollständig erfasst betrachten, sondern vielmehr, dass diesen Personen bekannt ist, dass es für diese spezifische Gefahrengruppe weitere Erhebungen gibt. Es ist wichtig, auch solches Hintergrundwissen in das Erhebungsdesign einfließen zu lassen, um überflüssigen Ressourcenaufwand und Ermüdung seitens der Teilnehmenden zu vermeiden.

Auffallend war die überdurchschnittlich hohe Bewertung der Relevanz des Items „Arbeitsräume sind in ausreichender Anzahl vorhanden“ in Verbindung mit vielen weiteren Kommentaren zu **Sozial- und Gemeinschaftsräumen**. Hier stellte sich im Rahmen der Ergebnisinterpretation die Frage, ob die Bewertung der Relevanz einzelner Aspekte

insofern zeitlichen Schwankungen unterliegt, als dass sie von aktuellen Problemen am Arbeitsplatz bzw. im individuellen Umfeld beeinflusst wird. Dies würde auch erklären, warum die infrastrukturelle Anbindung des Arbeitsortes dreimal explizit als wichtiger Aspekt genannt wurde, ohne dass dieser Aspekt durch den Fragebogen selbst aufgeworfen wurde. Es ist zu vermuten, dass die Nennung dieses Aspekts auf das spezifische Arbeitssetting zurückzuführen ist, wo insbesondere die Soldaten über längere Zeiträume auf dem Kasernengelände leben.

Besonders häufig (n = 12) wurde vorgeschlagen im Rahmen der **Arbeitsplatzgestaltung** auch die Sozialräume einzubeziehen. In diesem Zusammenhang wurde auch häufiger der Aspekt Hygiene angeführt sowie die Möglichkeit, individuelle Einflussmöglichkeiten auf die Gestaltung des Arbeitsplatzes zu haben. Letzteres scheint zwar plausibel, da die Anpassung des Arbeitsplatzes an die individuellen Bedürfnisse das Wohlbefinden am Arbeitsplatz erhöhen könnte, jedoch sind hierzu keine wissenschaftlichen Erkenntnisse bekannt. Der Aspekt Hygiene taucht zwar nicht unter den psychischen Gefährdungen auf, ist jedoch im Kontext biologischer Gefährdungen zu berücksichtigen und wird somit grundsätzlich auch in den offiziellen Empfehlungen zur Gefährdungsbeurteilung als relevanter Aspekt benannt (186). Sozialräume haben im Arbeitssetting in multipler Weise eine hohe Bedeutung: Sie sind einerseits im Rahmen der Arbeitsstättenverordnung (204) ab einer Beschäftigtenzahl von 10 Personen vorgeschrieben und wichtiger Rückzugs- und Ruheort und sie können zudem zur sozialen Kohärenz und somit einem besseren Arbeitsklima unter den Beschäftigten beitragen. Insofern scheint es angemessen, über die offiziellen Empfehlungen hinaus die Erhebung dieses Aspekts zu erwägen.

Dass Lärm nicht als relevanter Aspekt psychischer Belastung gesehen wird, überrascht, da dieser nachweislich die kognitiven Anforderungen einer Tätigkeit erhöht, die Konzentrationsfähigkeit negativ beeinflusst bzw. Ursache für Arbeitsunterbrechungen sein kann sowie einen Stressor darstellt, der mit erhöhter Erschöpfung einhergeht (56). Er sollte daher neben anderen Umgebungsfaktoren wie z. B. der eben genannten Hygiene berücksichtigt werden.

Die seitens der Experten als zu erheben geforderte **Vereinbarkeit von Beruf und Privatleben** ist bisher nicht Gegenstand konkreter Empfehlungen in Bezug auf die Gefährdungsbeurteilung psychischer Belastung, allerdings beschreibt die BAuA, dass diese Thematik verstärkt an Aufmerksamkeit und Evidenz in der Forschung gewinnt und insbesondere für junge Generationen einen wichtigen Faktor in Bezug auf die Arbeitszufriedenheit darstellt (195). Die bisherigen Befunde belegen demnach, dass Konflikte zwischen Arbeit und Privatleben, die psychische Gesundheit negativ beeinflussen und umgekehrt eine bessere Work-Life-Balance auch einen positiven Effekt auf die

psychische Gesundheit hat. Die GDA (2018) hat diesen Aspekt bereits unter einem fünften Merkmalsbereich "Neue Arbeitsformen" subsumiert. Hier werden negative Auswirkungen wie beispielsweise Entgrenzung zwischen Arbeit und Privatleben oder auch zeitliche und räumliche Flexibilisierung, welche ebenfalls das Privatleben negativ beeinflussen können, aufgeführt. Da mit einem Zuwachs neuer Arbeitsformen, die durch genau solche Merkmale gekennzeichnet sind, die die Vereinbarkeit von Arbeits- und Privatleben beeinflussen wie z. B. Arbeitsintensivierung, gestiegene Verfügbarkeitsanforderungen, Flexibilisierung und Entgrenzung, zu rechnen ist (205), scheint es durchaus sinnvoll, diesen Merkmalsbereich in die Analyse psychischer Gefährdungen aufzunehmen. Hierfür müssen jedoch zunächst eine theoretische Herleitung und Begründung der Zusammenhänge sowie eine klare Definition der zu erfassenden Merkmale erfolgen. Es wäre wenig zielführend und der wissenschaftlichen Güte abträglich, „irgendwelche“ Merkmale zu erheben, von denen man denkt, dass sie neue Arbeitsformen charakterisieren, ohne gut begründen zu können, wie sie auf die individuelle Gesundheit wirken oder in welchem Zusammenhang sie mit anderen Aspekten des Arbeitens und arbeitsbedingter Gefährdungen stehen.

Im Kontext Soziale Aspekte wurde auch immer wieder gefordert, dass **Arbeitsklima** zu erfassen. Darunter subsumiert wurden Schlagworte wie Teambuilding, Teamgeist, Arbeitsatmosphäre, Motivation oder auch empathischer Umgang. Diese Schlagworte finden sich bisher nicht in den Empfehlungen von GDA und BAuA. In ihrem Bericht zum Einfluss sozialer Beziehungen auf die psychische Gesundheit kommen Drössler et al. (2016) zu dem Schluss, dass das Sozialklima und ein Zusammenhalt im Team nicht nur eine Ressource sein können, sondern dass deren Nicht-Vorhandensein durchaus auch eine Gefährdung darstellen kann (187). Insofern ist der Vorschlag der Experten, die Analyse der Sozialen Aspekte um dieses Merkmal zu erweitern, sicherlich erwägenswert. Wichtig wäre jedoch im Vorfeld eine eindeutige und trennscharfe Definition des Konstrukts vorzunehmen.

In Bezug auf die **sozialen Beziehungen zu Kollegen und Vorgesetzten** stimmen die Befragten überein, dass es sich um wichtige Aspekte in Bezug auf die Arbeitssituation handelt. Die in der Leitlinie „Beratung und Überwachung bei psychischer Belastung am Arbeitsplatz“ genannte kritische Ausprägung "häufige Streitigkeiten und Konflikte" (50) bewerten die Experten hingegen nicht als grundsätzlich problematisch, sondern verwiesen darauf, dass vielmehr der Umgang mit Konflikten und die Konfliktlösung(skultur) die kritische Ausprägung darstellen können. Das Argument ist, dass in der beruflichen Interaktion mehrerer Personen zwangsläufig unterschiedliche Meinungen auftreten, eine Gefährdung jedoch erst dort entsteht, wo Konflikte nicht gelöst und keine für alle Betroffenen zufriedenstellenden Kompromisse gefunden werden. In diesem Zusammenhang wurde auch mehrfach gefordert, neben der Konfliktkultur Mobbing in das Grobscreening zu integrieren. Mobbing stellt einen arbeitsbedingten Stressor da, jedoch

„fehlt eine einheitliche theoretische Grundlage und valide Erfassung“ (187). Darüber hinaus müsste dann konsequenterweise jede Form negativer sozialer Interaktion wie z. B. verbale Gewalt oder sexuelle Belästigung ebenfalls erhoben werden, was dem Rahmen eines Grobscreenings nicht angemessen wäre. Ökonomischer und praxistauglicher scheint es in diesem Zusammenhang zu sein, abstraktere Aspekte, welche die meisten Formen negativer sozialer Interaktion gemeinsam haben, zu erheben. Dies betrifft neben einer unzureichenden Konfliktbewältigung beispielsweise fehlende Unterstützung oder fehlende Wertschätzung. Um den Vorgaben der Leitlinie gerecht zu werden und gleichzeitig die Augenscheinvalidität zu verbessern, sollten sowohl Konflikthäufigkeit als auch Konfliktbewältigung erfasst werden. Dies ist vertretbar, da es sich um einen Aspekt handelt, welcher mehrere Teildimensionen betrifft und nachweislich einen Effekt auf die psychische Gesundheit hat. Zudem können die beiden Merkmale als Indikator für weitere negative soziale Interaktionen stehen.

Zwar werden fehlende Führungskompetenzen und fehlende Wertschätzung als mögliche kritische Ausprägung der sozialen Beziehung zu den Vorgesetzten von GDA (50, 53) und BAuA (98) aufgeführt, allerdings stellt **Führung** bzw. der Führungsstil derzeit keine eigene Kategorie bei den Empfehlungen zur Beurteilung von Gefährdungen im Bereich sozialer Beziehung dar. Dass diesem Thema aus Sicht der Befragten jedoch eine besonders hohe Relevanz zugesprochen wird, drückt sich nicht nur im überdurchschnittlichen Relevanzindikator ($k = 0,93$) des einzigen Items zum Thema Führung („Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit“) aus, sondern auch darin, dass allein 10 Kommentare diesen Aspekt unter den Gesichtspunkten Rückmeldung, Fairness oder Führungsstil unmittelbar thematisierten. Eine differenzierte Erfassung des Führungsstils kann jedoch nicht im Rahmen eines Grobscreenings umgesetzt werden, zumal unterschiedliche Konzepte von gesundheitsförderlichen Führungsstilen in der Fachliteratur existieren (97, 206–209) und eindeutige Kausalzusammenhänge zwischen bestimmten Ausprägungen eines Führungsstils und der Mitarbeitendengesundheit zunächst theoretisch begründet und empirisch nachgewiesen werden müssten sowie sich gegenüber konkurrierenden Ansätzen als überlegen erweisen müssten.

Während BAuA und GDA in Bezug auf die arbeitsorganisatorische Gestaltung von **Kommunikation und Kooperation** ausschließlich die Gefährdung durch unzureichende Kommunikationsmöglichkeiten und Einzelarbeit fokussieren, adressieren die befragten Experten hingegen die potenzielle Belastung durch zu viele bzw. zu häufige Kommunikationserfordernisse. In diesem Kontext wird insbesondere der Nutzen von Besprechungen in Frage gestellt, aber auch der Mehraufwand bei zusätzlichen Abstimmungsbedarfen in Teamarbeit angeführt. Diese Problematik betrifft wiederum auch

den Merkmalsbereich Sozialer Aspekte, wo neben zu wenigen auch durchaus zu viele soziale Kontakte eine Gefährdung darstellen können (50). Allerdings ist die Kontaktintensivität bisher nicht Gegenstand des Grobscreenings. Denkbar ist, dass diese Punkte insbesondere dann eine höhere Belastung darstellen, wenn die dafür benötigte Zeit nicht in das Arbeitspensum eingeplant und somit die Arbeitsintensität erhöht wird, aber auch, wenn es durch viele Abstimmungsbedarfe zu häufigen Arbeitsunterbrechungen kommt. Insofern erscheint es angebracht, diesen Aspekt in Zukunft bei der Beurteilung der Arbeitsorganisation zu berücksichtigen.

5.3 Itemqualität, Reliabilität und Konstruktvalidität des AIGScreen

Die statistischen Analysen dienten der Beurteilung der Qualität der einzelnen Items und der Reliabilität und Validität des Instruments insgesamt. Im Rahmen der Itemanalyse standen die Verteilung der erhobenen Daten, fehlende Antworten sowie die Beziehung der Items untereinander im Fokus. Die Itemanalyse diente zudem dazu, die Eignung der Items für die anschließenden Faktorenanalysen zu überprüfen. Faktorenanalysen stellen wie in Abschnitt 2.3.2 erläutert eine gute Methode für Äquivalenzmessungen dar und können so ein Urteil über die Reliabilität der Antwortskalen ermöglichen. Sie dienen aber auch der Bewertung der internen Konsistenz und Eindimensionalität, weshalb sie zudem eine gute Methode darstellen, um die Konstruktvalidität zu überprüfen. Die explorative Faktorenanalyse wurde zuerst mit der Hälfte der Stichprobe durchgeführt, um die postulierte latente Struktur zu überprüfen und mit alternativen Lösungen zu vergleichen. Die zweite Hälfte der Stichprobe diente im Rahmen der Konfirmatorischen Faktorenanalyse dazu, den Modelfit der unterschiedlichen Lösungsansätze zu vergleichen und somit das valideste Modell zu identifizieren.

Für die Aussagekraft der hier präsentierten Ergebnisse war es wichtig, dass die Erhebung unter realen Feldbedingungen erfolgte. Da dies der Fall war, können die Schlüsse ohne Einschränkungen für den Einsatz des AIGScreen, die daraus resultierende Datenbasis und die darauf aufbauenden Ergebnisse als gültig betrachtet werden.

5.3.1 Qualität der Items

Es wurde bereits mehrfach darauf hingewiesen, dass es bezüglich des vorliegenden Instruments keine richtigen oder falschen Antworten gibt, weshalb auch der Begriff der Itemschwierigkeit etwas irreführend ist. Wenn im Folgenden von „schweren“ oder „leichten“ Items die Rede ist, sind damit solche Items gemeint, die überproportional häufig negativ oder positiv bezüglich des erfassten Merkmals beantwortet wurden. Die Mehrheit der im

AlGSscreen erhobenen Items ist als eher leicht einzustufen, da ihre Antwortmittelwerte über dem Mittelpunkt der Antwortskala liegen. Besonders häufig positiv bewertet wurden die Items um3 (körperlichen Anforderungen der Tätigkeit), aa5 (ausreichende Arbeitspausen), am6 (Ergebniskontrolle), sa4 (Unterstützung durch Mitarbeitende), sa9 (Anerkennung durch unterstellte Beschäftigte) und sa10 (Unterstützung durch unterstellte Beschäftigte). Da hier die Antwortverteilungen im Vergleich zu den anderen Items auch weniger Streuung aufweist, also Deckeneffekte festgestellt werden können, ist zu überlegen, ob die Qualität der Skala ohne das jeweilige Item bzw. durch dessen Modifikation verbessert werden könnte. Es sollte jedoch auch überprüft werden, ob eine Erhöhung der Antwortoptionen zu besseren Verteilungswerten führt. So entspricht die aktuelle Anzahl der Antwortoption von vier Antwortkategorien auf einer Skala von „Ja“ bis „Nein“ plus einer „Unzutreffend“-Antwortoption nicht den Empfehlungen zur Gestaltung von Antwortskalen, die sich aus dem aktuellen Stand der Forschung ableiten. So berichten Preston & Colman (2000) von einer geringen Diskriminanz bei weniger als 5 Antwortoptionen und auch Menold & Bogner (2015) empfehlen im Sinne einer guten Reliabilität und Validität mindestens 5 Skalenpunkte (136, 125). Eine Erhöhung der Antwortkategorien könnte auch dazu beitragen, multivariate Normalverteilung zu erreichen, welche für die vorliegenden Items nicht nachgewiesen werden konnte. Allerdings sollte auch die inhaltliche Eignung der Items überprüft werden. Denn neben Modifikationsbedarf im Fragebogendesign ist es möglich, dass die erfassten Merkmale tatsächlich überwiegend positiv bzw. in der Wahrnehmung der Befragten zufriedenstellend sind. Eine dritte Möglichkeit, die Boden- und Deckeneffekte begünstigen kann, betrifft die Formulierung der Items. Die Verwendung von Begriffen wie „angemessen“ oder „ausreichend“ bieten viel Interpretationsspielraum seitens der Probanden und schränken die Intersubjektivität des Frageverständnisses und somit die Reliabilität der Ergebnisse ein. Hier muss überprüft werden, inwiefern durch Frageformulierungen konkretere und trennschärfere Merkmale und Merkmalsausprägungen vorgegeben werden können.

Die Trennschärfe, gemessen an der Item-Skalen-Korrelation variiert deutlich zwischen Werten von 0,47 bis 0,76. Dies ist bei heterogenen Konstrukten durchaus zu erwarten und nicht zwangsläufig problematisch (71). Allerdings sind Items mit höherer Trennschärfe zu bevorzugen, da sie besser zwischen Probanden mit extremen Ausprägungen diskriminieren (30). Dies setzt jedoch auch eine ausreichend hohe Varianz voraus. Diese wiederum kann die Itemtrennschärfe positiv beeinflussen (ebd.), was sich jedoch in den vorliegenden Daten nicht nachweisen ließ. Zudem sollten auch die Item-Rest-Korrelationen bzw. die korrigierte Itemtrennschärfe berücksichtigt werden. Hier sind niedrige Werte kritischer zu bewerten, da niedrige Werte zwar nicht unbedingt nachteilig für das jeweilige Item, aber zumindest für die Zugehörigkeit des Items zur jeweiligen Skala zu interpretieren

sind. Beispielsweise zeigt das Item um3 eine, im Vergleich zu den übrigen Items der Skala Arbeitsumgebung, niedrige Item-Rest-Korrelation. Wenn man zusätzlich berücksichtigt, dass sich die interne Konsistenz der Skala bei Entfernen dieses Items verbessern würde, liegt der Schluss nahe, dass das Item besser einem anderen Konstrukt zugeordnet werden sollte bzw. dass es zumindest bezüglich des Konstrukts „Arbeitsumgebung“ nicht ausreichend gut diskriminiert. Auch die Items am6 (Ergebniskontrolle) und am9 (Arbeitsunterbrechungen) weisen vergleichsweise niedrige Item-Rest-Korrelationen auf. Da es sich bei der Skala Arbeitsinhalt und Arbeitsaufgabe um ein Konstrukt handelt, welches insgesamt 12 zum Teil sehr heterogene Merkmale umfasst, ist es nicht ganz verwunderlich, dass bestimmte Items aufgrund geringer Korrelationen mit einzelnen Items der Skala eine geringere Item-Rest-Korrelation aufweisen, auch wenn deren Item-Skalen-Korrelation insgesamt als gut zu bewerten ist. Dennoch sollte auch in solchen Fällen überprüft werden, ob eine differenziertere Aufteilung in Subskalen oder eine Neuordnung dieser Items zu anderen Konstrukten der Datenstruktur nicht besser gerecht werden würde. Insgesamt sind jedoch aufgrund der Konstruktion des Fragebogens keine sehr hohen Trennschärfekoeffizienten für sämtliche Items zu erwarten. Denn bezüglich der Trennschärfe von Items existiert eine Konkurrenz zwischen Testgütekriterien und Nebengütekriterien, wenn heterogene, vielschichtige Konstrukte mit möglichst wenig Items abgebildet werden sollen. In diesem Fall können im Sinne von Fragebogenökonomie und Praxistauglichkeit nicht mehr alle Aspekte des Konstrukts erfasst werden, was sich jedoch negativ auf die Itemkorrelationen und die interne Konsistenz von Skalen auswirkt.

Die interne Konsistenz, bemessen am Koeffizienten Cronbachs Alpha, kann für wenigstens zwei Skalen, Arbeitsumgebung und Arbeitsinhalt und Arbeitsaufgabe als verbesserungswürdig eingestuft werden. Ein hoher Alpha-Koeffizient ist erstrebenswert, da mit ihm auch die Reliabilität der Skala steigt (63). Ein Wert von 0,8 bedeutet, dass 80% der Varianz durch die Skala repräsentiert werden und nur 20% der Varianz auf zufällige Messfehler zurückzuführen sind. Voraussetzung für diese Annahme ist jedoch, dass die Skala auch valide ist (154). Daher soll dieser Aspekt ausführlich im Zusammenhang mit den Ergebnissen der Faktorenanalyse diskutiert werden (siehe Abschnitt 5.3.2).

Der geringe Anteil fehlender Antworten spricht für eine hohe Verständlichkeit und gute Praktikabilität der Items. Für kein Item konnte systematische non-response festgestellt werden. Allerdings erhöht sich der Anteil fehlender bzw. nicht auswertbarer Datensätze erheblich, wenn zusätzlich die „unzutreffend“-Antworten als fehlende Werte behandelt werden. Dies war für die anschließenden multivariaten statistischen Analysen notwendig, da die angewandten Verfahren ausschließlich für mindestens ordinalskalierte Daten anwendbar sind. Der Antwortkategorie „unzutreffend“ lässt sich jedoch keine Position in der Rangfolge der übrigen Antwortoptionen zuordnen. Sie ist nicht Teil derselben Dimension,

sondern drückt explizit die Abwesenheit der jeweiligen Merkmalsausprägung aus. Für die Items, für die ein hoher Anteil fehlender Werte – gemeint sind damit im Folgenden nicht gegebene Antworten und „unzutreffend“-Antworten – festgestellt werden konnte, ist dieser hohe Anteil durchaus plausibel. Es handelt sich hierbei mit Ausnahme von Item aa2 um Items, die Merkmale erfassen bzw. voraussetzen, welche nicht auf alle Mitarbeitenden zutreffen können wie z. B. Personalverantwortung. Weniger plausibel ist der hohe Anteil fehlender Werte bei Item aa2 („Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen“). Zwar ist es nachvollziehbar, dass nicht alle Mitarbeitenden im Bereitschaftsdienstsysteem arbeiten, aber die Angemessenheit von Vertretungstätigkeiten sollten alle Befragten beurteilen können. Denkbar ist, dass die gleichzeitige Nennung von Bereitschaftsdiensten hier zu Unsicherheiten oder Verständnisproblemen bei den Befragten führt und diese meinen, dass es sich dabei um einen Aspekt handelt, der auf sie nicht zutrifft. Hier scheint es angebracht, eine Neuformulierung des Items zu erwägen, um Antwortausfälle aufgrund der schwer verständlichen Frage zu vermeiden. Dies könnte auch zu einer Verbesserung der Relevanz aus Sicht der Experten führen, welche dieses Item im Vergleich zu den übrigen Items als deutlich weniger relevant bewerteten.

Ein Item fiel während der Itemanalyse besonders problematisch auf. Für Item um3 („Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen.“) konnte ein überdurchschnittlich hoher Mittelwert bei gleichzeitig vergleichsweise geringer Standardabweichung festgestellt werden, was dafür spricht, dass das Item „zu leicht“ ist bzw. nur wenig Erkenntnis liefert, da es von der Mehrheit der Befragten gleich beantwortet wird. Zudem scheint es nicht gut in die Subskala Arbeitsumgebung zu passen. Zwar erreichen sowohl die Item-Skalen-Korrelation als auch die Item-Rest-Korrelation ausreichend gute Werte hinsichtlich der Empfehlungen wie hoch die Korrelationen mindestens sein sollten, doch ist der Zusammenhang mit der Skala im Vergleich zu den anderen Items deutlich geringer und die Item-Rest-Korrelation weist den drittniedrigsten Wert unter allen Items auf. Gleichzeitig ist es das einzige Item für das eine, wenn auch nur geringfügige, Verbesserung der Homogenität der Skala ermittelt wurde, wenn man das Item aus der Skala eliminierte. Für die Beibehaltung des Items, ggf. als Aspekt einer anderen Skala spricht, dass die Experten im Rahmen der teilstandardisierten Befragung sich für die Erhebung dieses Merkmals aussprachen.

Insgesamt kann keine hohe Übereinstimmung zwischen der Qualität eines Items bemessen an den hier diskutierten Kennwerten und der Einschätzung seitens der befragten Experten für die Beurteilung der Inhaltsvalidität festgestellt werden. So wurden zwar die Items aa2 („Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen“) und aa5 („Meine Arbeitspausen sind ausreichend“) als irrelevant von den befragten

Experten bewertet und weisen in der Skala Arbeitsorganisation auch die höchsten Deckeneffekte auf, aber das ebenfalls von Experten kritisch bewertete Item aa4 („Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen“) weist eine gute Verteilung auf und verhält sich auch bezüglich der übrigen Kennwerte zufriedenstellend. In der Skala Arbeitsinhalt und -aufgabe wurde hingegen das Item mit den höchsten Deckeneffekten und der geringsten Trennschärfe, am6 („Ich kann erkennen, ob ich meine Arbeit gut gemacht habe“) sehr positiv bewertet. Andere, weniger problematische Items wie z. B. am3 („Meine Arbeit ist in angenehmem Maße abwechslungsreich“) müssten nach Expertenmeinung gestrichen werden. Allerdings waren die befragten Experten im Geschäftsbereich BMVg auch nicht dazu aufgefordert, statistische Verteilungen oder Zusammenhänge der Items zu bewerten. Dass die Ergebnisse aus der Itemanalyse und der Expertenbefragung übereinstimmen, war weder auszuschließen noch zu erwarten, da beide Verfahren völlig unterschiedliche Aspekte beurteilen sollten. Dennoch muss für die weitere Fragebogenentwicklung berücksichtigt werden, dass Itemqualität und Reliabilität auf der einen Seite und Inhaltsvalidität auf der anderen Seite in keinem Zusammenhang stehen und der Fragebogen dennoch beiden Kriterien gerecht werden muss.

Ein deutlicherer Zusammenhang zwischen Ergebnissen der Itemanalyse und den Resultaten des Expertenreviews wäre hingegen zu erwarten gewesen, da hier Aspekte des Fragebogendesigns und der Eignung der einzelnen Items im Fokus standen. Allerdings konnte auch hier kaum ein Zusammenhang festgestellt werden. Zwar kritisierten die Experten für Fragebogendesign einige Items, die eine stark rechtssteile Verteilung und / oder geringe Trennschärfekoeffizienten aufwiesen, jedoch wurden auch viele Items mit problematischen Itemkennwerten nicht im Fragebogenreview diskutiert und umgekehrt weisen Items, die bezüglich bestimmter Kriterien von den reviewenden Experten kritisch bewertet wurden, nach statistischen Kriterien eine gute Itemqualität auf. Damit soll nicht gesagt werden, dass das Expertenreview oder die Expertenbefragung nicht zielführend oder hilfreich bei der Verbesserung der wissenschaftlichen Güte des AIGScreen waren. Vielmehr zeigt es den Vorteil der Kombination von quantitativen und qualitativen Methoden. Weder das eine noch das andere Vorgehen hätte alle Probleme identifizieren können. Mit einem Methodenmix wurde hingegen ein umfassendes Bild von Verbesserungsmöglichkeiten der einzelnen Items gewonnen.

Schließlich diene die Itemanalyse auch der Vorbereitung der Faktorenanalysen. Die Itemanalyse ermöglicht es, zu überprüfen, inwiefern sich die Daten für die Faktorenanalyse eignen. Einen Hinweis auf das Vorhandensein latenter Strukturen liefern die Korrelationen zwischen den Items. Hier sollten sich möglichst hohe Korrelationen zwischen den Items erkennen lassen, die demselben Konstrukt zugeordnet wurden und die Korrelationen mit konstruktfernden Items sollten möglichst gering ausfallen (156). Dies war nicht für alle

Items der Fall und ließ erwarten, dass die explorative Faktorenanalyse die zunächst postulierte 4-Faktoren-Struktur möglicherweise nicht vollständig reproduzieren können. Als Maß für die Itemkorrelationen der gesamten Skala kann zudem Cronbachs Alpha berücksichtigt werden, welches auch ein Maß der Reliabilität der Skala darstellt. Die Zuverlässigkeit der 4 postulierten Skalen kann als akzeptabel (Arbeitsumgebung, Arbeitsinhalt und -aufgabe) bzw. als gut (Soziale Aspekte, Arbeitsorganisation) bewertet werden (63), sodass die faktoranalytische Überprüfung der postulierten Struktur gerechtfertigt ist.

Ein drittes Maß, welches die Eignung der Korrelationsmatrix für die Faktorenanalyse bemisst, ist das KMO- oder auch MSA-Kriterium. Die Untergrenze von 0,8 (156) wird von allen Items weit übertroffen, sodass davon ausgegangen werden kann, dass sich die Korrelationsmatrix sehr gut für die Durchführung der Faktorenanalyse eignet.

5.3.2 Resümee aus den Ergebnissen der Faktorenanalyse

Nachdem die Eignung der AIGScreen-Items für die Faktorenanalyse festgestellt wurde, fehlende Antworten mittels multipler Imputation ersetzt wurden und der Datensatz in zwei zufällige Stichproben geteilt wurde, konnten die Reliabilitäts- und Validitätsüberprüfungen mittels Faktorenanalysen durchgeführt werden. Zunächst sollte mittels explorativer Faktorenanalyse die Verlässlichkeit der vier postulierten Skalen Arbeitsorganisation, Arbeitsinhalt und Arbeitsaufgabe, Soziale Aspekte und Arbeitsumgebung, überprüft und ggf. alternative Strukturen identifiziert werden. Bereits im Rahmen der Itemanalyse konnte mit Cronbachs Alpha ein erster Indikator für die Verlässlichkeit der vier Skalen berechnet werden. Zwei Skalen, Arbeitsorganisation und Soziale Aspekte, wiesen eine sehr gute interne Konsistenz auf während die Ergebnisse für Arbeitsinhalt und Arbeitsaufgabe sowie Arbeitsumgebung nur als akzeptabel betrachtet werden können. Allerdings führt eine niedrige Itemzahl zu niedrigeren Werten für Cronbachs Alpha (63). So ist insbesondere das vergleichsweise schlechte Ergebnis für die Skala Arbeitsumgebung immer noch als gut zu bewerten, berücksichtigt man den Umstand, dass diese Skala gerade einmal vier Items umfasst. Laut Rammstedt (2004) sind Reliabilitätskoeffizienten von mindestens $\alpha = 0,70$ ausreichend reliabel für Gruppenvergleiche. Dennoch sprechen die Ergebnisse der Faktorenanalyse insgesamt gegen die ursprünglich postulierten vier Skalen, sodass eine Neuuzuordnung der Items angemessen erscheint.

Die explorative Faktorenanalyse (EFA) konnte das ursprünglich präferierte 4-Faktoren-Modell nicht bestätigen. Da die zuvor postulierten 4 Skalen Arbeitsumgebung, Soziale Aspekte, Arbeitsinhalt und –aufgaben sowie Arbeitsorganisation sehr heterogene Konstrukte darstellen, ist das Ergebnis der EFA auch nicht besonders überraschend.

Allerdings ließ sich aus den Ergebnissen der EFA auch nicht eindeutig eine bestimmte zu extrahierende Faktorenzahl ableiten. Je nach Kennwert waren zwischen 1 und 11 Faktoren abzuleiten. Nach Kaiser-Kriterium wären nur 3 Faktoren zu extrahieren, also weniger als die vier postulierten Hauptskalen. Das ist im Hinblick auf die theoretische Heterogenität des Konstrukts nicht sinnvoll. Andererseits sind 11 Faktoren, wie es einige Kriterien nahelegen, ebenfalls nicht zielführend. Einerseits würden so Skalen mit nur zwei oder drei Items entstehen. Ein weiteres Gegenargument ist das Ziel, ein Kennzahlensystem mit wenigen Indikatoren zu entwickeln. In der Gesamtschau der Teststatistiken, unter Berücksichtigung eines möglichst optimalen Verhältnisses von Modellsparsamkeit und Modellanpassung und auch aufgrund der inhaltlichen Plausibilität der Ladungsstruktur wurde schließlich ein Modell mit 8 Faktoren für die weitere Analyse ausgewählt, welches mit dem ursprünglich postulierten 4-Faktoren-Modell sowie einer 1-Faktor-Lösung im Rahmen der Konfirmatorischen Faktorenanalyse (CFA) verglichen werden sollte. Die gestiegene Anzahl an Skalen ergibt sich vor allem durch die Aufteilung der Dimensionen Arbeitsinhalt und -aufgabe sowie Soziale Aspekte. Letztere zeigte bereits im 4-Faktoren-Modell erhebliche Querladungen, während für erstere überhaupt keine Einfachstruktur festzustellen war. Die Dimension Arbeitsumgebung konnte weitgehend in ihrer faktoriellen Struktur repliziert werden, allerdings wies das Item um4 im 8-Faktoren-Modell eine höhere Ladung auf einen anderen Faktor auf, welcher die neu zu bildende Skala Arbeitsplatzgestaltung darstellt. Eine deutlichere Einfachstruktur für die ursprünglich postulierte Skala Arbeitsumgebung konnte im orthogonal rotierten 4-Faktoren-Modell erreicht werden (siehe Anlage 7), allerdings wies hier das bereits häufiger als für diese Skala problematisch diskutierte Item um3 Querladungen auf die anderen Faktoren auf. Hingegen verbesserte sich im 8-Faktoren-Modell das ohnehin schon sehr gute Ergebnis für die Dimension Arbeitsorganisation noch einmal leicht. Hier konnte bereits im 4-Faktoren-Modell eine Einfachstruktur nachgewiesen werden, welche sich auch im 8-Faktoren-Modell bestätigte (vgl. Anlage 9). Allerdings zeigte sich für beide Modelle auch, dass neben den ursprünglich der Skala zugeordneten Items das Item am9 („Ich werde bei meiner Arbeit selten unterbrochen“) dem Faktor zuzuordnen ist.

Um zu überprüfen, ob die 8-Faktoren-Lösung die Datenstruktur tatsächlich besser abbildet als es das 4-Faktoren-Modell oder auch eine Gesamtskala aller AIGScreen-Items können, wurde die Konfirmatorische Faktorenanalyse (CFA) als hypothesenprüfendes Verfahren zur Beurteilung der faktoriellen Validität durchgeführt. Mit Blick auf die Modellfitstatistiken konnte jedoch kein Modell vollständig überzeugen. Wenig überraschend ist, dass das 1-Faktoren-Modell die schlechtesten Werte bezüglich der Modellanpassung erzielt, da CFA reflektive Messmodelle unterstellen (156). Reflektive Messmodelle postulieren ein bestimmtes Konstrukt zu repräsentieren und sollten daher eindimensional und konsistent

sein. Dies ließ sich weder in der Ladungsstruktur erkennen, noch ist es für die verwendeten Items inhaltlich begründbar. Die Heterogenität der Items des AIGScreen berücksichtigend ist es insofern auch plausibel, dass das 8-Faktoren-Modell den besten Modellfit erzielt. Nur dieses Modell erreicht mittelmäßige Werte in Bezug auf Modell-Sparsamkeit und den relativen Modellfit. Letztlich kann aber auch diese Lösung nicht als vollständig akzeptabel angesehen werden. Denkbar ist jedoch, dass aufgrund der fehlenden (multivariaten) Normalverteilung, eine inflationierte Schätzung der χ^2 -Werte erfolgte, was wiederum zu einer Verzerrung der Modellfit-Maße führen würde. Aichholzer (2017) zufolge, kann Nicht-Normalverteilung dazu führen, dass richtig spezifizierte Modelle fälschlicherweise abgelehnt werden (155). Allerdings wurde aus diesem Grund die robuste ML-Schätzung nach Satorra & Bentler vorgenommen, sodass die damit geschätzten Fit-Maße weniger verzerrt sein sollten. Nicht beeinflusst bzw. korrigiert werden können hingegen die Ladungsstrukturen. Beauducel & Wittmann (2005) konnten zeigen, dass bei niedrigen und mittleren Ladungen ($<.60$) und bereits bei geringen Querladungen ($<.20$) die Modellfit-Indizes teilweise widersprüchliche Ergebnisse liefern und zu einer Ablehnung des Modells tendieren, auch wenn dieses richtig spezifiziert ist (210). Die Autoren stellen grundsätzlich infrage, ob die Anwendung komparativer Fit-Maße bei Fragebögen zur Erhebung latenter Merkmale geeignet ist oder eine Ablehnung des Modells im Rahmen der CFA nahezu erwartbar wäre: „If one accepts the perspective that, due to low reliabilities, the main loadings of trait items are often not much larger than .40 [...] and that there will always be some small distortion of simple structure, this study suggests that trait models would only have a chance to be accepted when incremental fit indexes and the GFI are not used for model evaluation. Small distortions of simple structure will be penalized in incremental fit indexes and in the GFI when main loadings are low. This may explain the observation that trait models assuming simple structure tend to be rejected with CFA“ (210). Tatsächlich ist die Annahme einer Einfachstruktur für die vorliegenden Items auch problematisch. Das zeigt sich auch in der Neuordnung der 8-Faktoren-Lösung. Einige Items lassen sich nicht eindeutig einem Faktor zuordnen und die entsprechenden Querladungen sind durchaus plausibel und theoretisch begründbar. Beispielsweise lädt das Item sa12 („Konflikte werden offen besprochen“) in etwa gleich auf die Faktoren Soziale Beziehung zu Vorgesetzten und Soziale Beziehung zu Kollegen.

Die vorgeschlagene Neuordnung der Items, wenn man acht anstelle der bisher vier Skalen verwenden wollte, wurde in Tabelle 14 präsentiert. Aber auch hier ist die inhaltliche Interpretation nicht für alle Faktoren eindeutig. Während die Faktoren 2, 4 und 5 theoretisch plausible Konstrukte darstellen, ist die inhaltliche Interpretation der Faktoren 1, 3, 6 und 7 nicht selbstverständlich. Obwohl sich Faktor 8 „Arbeitsumgebung“ nicht wie erwartet verhält, ist das Ergebnis plausibel. Ursprünglich würde man erwarten, dass alle

Arbeitsumgebungs-Items einen Faktor bilden. Da Faktor 2 jedoch als Skala für die Arbeitsplatzgestaltung interpretiert werden kann, ist eine höhere Korrelation des Items „Mein Hauptarbeitsplatz [...] ist so gestaltet, dass ich gut arbeiten kann“ (um4) auf dieser Skala ebenso plausibel wie die Querladung auf die Arbeitsumgebungs-Skala, welche in Faktor 8 repräsentiert ist.

Faktor 4 „Zeitliche Aspekte / Quantitative Anforderungen“ repräsentiert die ursprüngliche Arbeitsorganisations-Skala. Die Ergebnisse zeigen jedoch sehr deutlich, dass das Item „Ich werde bei der Arbeit selten unterbrochen“ (am9) ebenfalls dieser Skala zuzuordnen ist. Eine Umbenennung des Konstrukts erfolgte zum einen zur besseren Abgrenzung gegenüber der ursprünglichen Skala, zum anderen, weil aufbauend auf den Ergebnissen der Expertenbefragung zukünftig weitere Aspekte im Merkmalsbereich Arbeitsorganisation im AIGScreen ergänzt werden sollten, welche sich inhaltlich nicht dieser Skala zuordnen ließen. Faktor 5 enthält die Hälfte der ursprünglichen Skala Arbeitsinhalt- und aufgabe-Skala ab und wurde als „Qualitative Anforderungen“ betitelt, da er Handlungsspielraum, Qualifikation, Information, Verantwortung und Variabilität umfasst. Obwohl die Ladung von Item am8 („Ich habe einen angemessenen Handlungsspielraum (zum Beispiel Entscheidungsfreiheit) in Bezug auf meine Arbeit“) auf Faktor 1 „Feedback / Führungsstil“ marginal höher ist, ist eine Zuordnung zu Faktor 5 inhaltlich plausibler. Faktor 6 ist inhaltlich nicht ganz eindeutig und könnte am besten mit „respektvollem Umgang und Anerkennung“ überschrieben werden. Faktor 1 „Feedback / Führungsstil“ weist eine deutliche Einfachstruktur auf und entspricht damit dem mehrfachen Wunsch, der im Rahmen der Expertenbefragung formuliert wurde, diesem Aspekt mehr Gewicht beizumessen und es als eigenständige Dimension zu erheben. Faktor 3 „Soziale Beziehung zu unterstellten Mitarbeitenden“ enthält nur zwei Items. Zum einen ist es plausibel, dass sich diese beiden Items, die sich auf unterstellte Mitarbeitende beziehen, einen eigenen Faktor bilden. Andererseits ist es fraglich, ob eine 2-Item-Skala zur Reduzierung der Komplexität hilfreich ist. Costello & Osborne (2005) empfehlen Skalen mit mindestens 5 Items, um stabile Skalen zu erhalten (173). In einem 6-Faktor-Modell, das hier nicht weiter diskutiert wurde, bildeten diese beiden Items einen gemeinsamen Faktor mit den Items, die hier Faktor 6 bilden. Eventuell wäre es sinnvoller, diese Skalen zusammenzufassen, da alle Elemente beider Faktoren zur Messung des sozialen Klimas des Personals interpretiert werden können. Faktor 7 repräsentiert einen weiteren Aspekt der sozialen Beziehungen, nämlich Konflikte bei der Arbeit. Sowohl die Ergebnisse des 6-Faktor- als auch des 8-Faktor-Modells weisen eindeutig darauf hin, dass diese Items eine eigene Skala bilden und als unabhängig vom Faktor der sozialen Beziehung betrachtet werden sollten. Auch das wird der Forderung aus der Expertenbefragung gerecht, dem Aspekt Konfliktkultur mehr Raum im Fragebogen zu

geben und Konflikte nicht grundsätzlich als negative Ausprägung Sozialer Beziehungen zu erfassen.

Allerdings stellen acht Faktoren nur eine geringe Komplexitätsreduktion in Form eines Kennzahlensystems dar und einige Skalen sind auch dann nicht ausreichend konsistent, um als verlässlicher Indikator für eine bestimmte inhaltliche Dimension dienen zu können. Insbesondere einige Items zu Sozialen Aspekten weisen immer noch Querladungen im oblique rotierten 8-Faktoren-Modell auf. Mit Blick auf die Konstruktvalidität muss zudem festgestellt werden, dass für eine ausreichende Modellsättigung mindestens drei Items pro Skala empfohlen sind (64). Dies trifft auf Faktor 2 nicht zu und auch die Faktoren 7 und 8 erreichen gerade einmal dieses Minimum.

Von den 34 Items haben 9 eine Kommunalität $< .40$. Dies kann auf zwei Arten interpretiert werden. Entweder steht das Item in keinem Zusammenhang mit den anderen und passt nicht gut in das Konstrukt, oder es werden einfach mehr Faktoren benötigt, weil die Items in keine Skala passen. Im ersten Fall sollte man erwägen, das betreffende Item zu streichen. Im zweiten Fall sollte man erwägen, weitere Items hinzuzufügen, die ähnliche Aspekte messen (173). Festzuhalten bleibt, dass weder das bevorzugte Acht-Faktoren-Modell noch die angenommenen vier Faktoren oder ein Ein-Faktoren-Modell sich in der CFA als insgesamt gut geeignet erwiesen. Lediglich das Acht-Faktoren-Modell erreicht in Bezug auf die Sparsamkeit und den inkrementellen Modelfit mittelmäßige Werte. Folglich kann auch dieses Modell nicht als vollständig akzeptabel angesehen werden, auch wenn sich Argumente anführen lassen, warum die verwendeten Modelfit-Maße grundsätzlich zu einer Ablehnung des Modells führen. Eine Verbesserung von Reliabilität und Validität kann durch Anpassungen des Fragebogendesigns und -inhalts erzielt werden. Bezüglich der inhaltlichen Anpassungen können die Ergebnisse der Expertenbefragung eine gute Orientierung geben.

Auf Skalenebene lassen sich wenig Zusammenhänge zwischen den Ergebnissen der Expertenbefragung und der Faktorenanalyse feststellen. So steht eine gute Bewertung der Inhaltsvalidität der Skala Arbeitsumgebung, gemessen an Relevanz und Vollständigkeit, einer aufgrund der geringen Itemzahl noch tolerablen internen Konsistenz ($\alpha = 0,759$) und fehlender Einfachstruktur gegenüber. Beides könnte durch eine Erhöhung der Itemzahl und somit auch einer Verbreiterung des erfassten Merkmalsbereichs verbessert werden. Hingegen kann die Dimension Arbeitsorganisation gemessen an der internen Konsistenz ($\alpha = 0,823$) und aufgrund ihrer Einfachstruktur in allen Modellen als reliabel und valide betrachtet werden, erreichte jedoch keine gute Bewertung bezüglich ihrer Relevanz durch die Experten und wurde von der Hälfte auch nicht als inhaltlich vollständig beurteilt. Näher beieinander liegen die Ergebnisse für die Inhaltsvalidität aus der Expertenbefragung und

die Ergebnisse aus der statistischen Überprüfung der Konstruktvalidität für die beiden Skalen Soziale Aspekte und Arbeitsinhalt und Arbeitsaufgabe. Die Dimension Soziale Aspekte wurde von den Experten positiv hinsichtlich Relevanz und Vollständigkeit bewertet und weist auch eine gute interne Konsistenz ($\alpha = 0,871$) auf. Dennoch legen die Ergebnisse der Faktorenanalyse eine Aufteilung der Skala in mehrere Subskalen nahe. Umgekehrt kann für die Dimension Arbeitsinhalt und Arbeitsaufgabe keine ausreichende Relevanz festgestellt werden und die interne Konsistenz kann mit $\alpha = 0,799$ lediglich als akzeptabel bewertet werden.

5.4 Schlussfolgerungen für die wissenschaftliche Güte des AIGScreen

Sowohl die Datenerhebung als auch die Analyse, der mit dem AIGScreen erhobenen Daten kann als objektiv bewertet werden, da keine Verzerrungseffekte durch Testleiter möglich sind, keine Suggestivformulierungen oder andere Gefahren systematischer Verzerrungen aufgrund des Fragebogendesigns oder einzelner Frageformulierungen seitens der befragten Experten für Fragebogendesign identifiziert werden konnten und die Datenauswertung standardisiert erfolgt, also keine Interpretationsspielräume oder Variation bei der Operationalisierung seitens der Auswertenden zulässt. Eine Beeinflussung des Befragungsergebnisses durch Ort und Zeitpunkt der Erhebung sowie sonstige die Befragungsergebnisse beeinflussende Faktoren (z. B. die Anwesenheit Dritter während der Beantwortung des Fragebogens) kann aufgrund des Erhebungsdesigns zwar nicht ausgeschlossen werden, jedoch wurden im Rahmen der dreijährigen Erhebungsphase auch keine systematischen Störgrößen bekannt, deren Kontrolle im Sinne der Objektivität notwendig wäre. Zudem ermöglichen die Rahmenbedingungen der Erhebung eine flexible Wahl von Durchführungszeitpunkt und -ort sowie Endgerät seitens der Probanden, sodass diese selbst mögliche Störfaktoren ausschließen können und von einer Durchführungsobjektivität ausgegangen werden kann.

Die Reliabilität der vier Skalen kann als gut bis akzeptabel bewertet werden, wenn man ausschließlich die Ergebnisse der Itemanalyse berücksichtigt. Mit Blick auf die Faktorladungen, die sich aus der explorativen Faktorenanalyse ergaben, kann jedoch nicht an der 4-faktoriellen Struktur festgehalten werden. Zwar ließen sich zwei von vier Skalen, Arbeitsumgebung sowie Arbeitsorganisation, sehr gut im Rahmen der EFA replizieren, jedoch konnte die Eindimensionalität für Soziale Aspekte und Arbeitsinhalt und -aufgabe nicht bestätigt werden. Es muss jedoch auch hinterfragt werden, inwiefern die Messung des heterogenen Merkmalsbereichs „wahrgenommene Arbeitssituation“ mit möglichst wenigen Items überhaupt in wenige Skalen zusammengefasst werden kann, welche eine

Einfachstruktur aufweisen sollen. Bei der Messung von latenten Merkmalen wie Meinungen, Persönlichkeitsmerkmalen oder Einstellungen ist in der Regel immer mit einer gewissen Unschärfe in der Datenstruktur aufgrund von komplexen Zusammenhangsstrukturen zu rechnen (210, 76). Querladungen bzw. die Verletzung einer Einfachstruktur sind erwartbar und inhaltlich plausibel, widersprechen jedoch den Modellannahmen und führen somit schneller zur Ablehnung des jeweiligen Modells. Hinzukommt, dass die meisten Verfahren Normalverteilung voraussetzen und für metrisches Skalenniveau konzipiert sind. Beides ist im vorliegenden Fall, wie auch in den meisten anderen Fällen, in denen personenbezogene (latente) Merkmale erhoben werden, nicht vollständig gegeben (211, 62). Zwar konnte in vielzähligen Studien nachgewiesen werden, dass die meisten parametrischen Verfahren robust gegenüber der Verletzung der Normalverteilungsannahme bzw. bei der Verwendung pseudo-metrisch skalierten Variablen sind (212–218) und auch vergleichende, hier nicht präsentierte Analysen, konnten keine substantiellen Unterschiede in den Ergebnissen feststellen. So wiesen beispielsweise weder die nicht nach Santorra & Bentler korrigierten Modelfit-Maße größere Unterschiede zum korrigierten Pendant auf, noch konnten durch Anwendung polychorischer Faktorenanalysen abweichende Ladungsstrukturen ermittelt werden. Auch die in Anlage 6 dargestellte Korrelationsmatrix zeigt keine bedeutsamen Abweichungen zwischen Pearson- und Spearman-Korrelationen. Allerdings zeigen empirische Analysen auch, dass bei Verletzung der Normalverteilungsannahme oder bei Verwendung von ordinalskalierten Daten Modelle häufiger abgelehnt werden (210, 219, 220). Das bedeutet für die hier verwendeten Daten, dass Fehler erster Art, also die fälschliche Annahme des Alternativmodells trotz unzureichender Spezifikation, deutlich unwahrscheinlicher sind als Fehler zweiter Art, nämlich die Ablehnung des Alternativmodells, obwohl es die empirische Datenstruktur angemessen beschreibt. Insofern kann nicht ausgeschlossen werden, dass die wissenschaftliche Güte aufgrund der angewandten Methodik unterschätzt wurde.

Die Validität des AIGScreen wurde anhand der Konstrukt- und der Inhaltsvalidität beurteilt. Ersteres konnte mittels statistischer Verfahren, letzteres durch teilstandardisierte Experteninterviews mit einem hohen Anteil qualitativer Analysen durchgeführt werden. Die zuvor postulierte theoretische Struktur des AIGScreen in vier Hauptdimensionen Arbeitsorganisation, Arbeitsinhalt und -aufgabe, Soziale Aspekte und Arbeitsumgebung konnte, wie oben bereits erläutert, nur teilweise durch die Item- und Faktorenanalyse verifiziert werden. So wiesen die beiden Skalen Arbeitsorganisation und Soziale Aspekte zwar eine ausreichend gute interne Konsistenz gemessen am Koeffizienten Cronbachs Alpha und zufriedenstellende Iteminterkorrelationen auf, jedoch verfehlte die Skala Arbeitsinhalt und Arbeitsaufgabe den Schwellenwert von 0,8 für Cronbachs Alpha mit 0,799 knapp und die Skala Arbeitsumgebung deutlich mit 0,759. Letzteres könnte auch auf die

niedrigere Itemzahl von vier zurückgeführt werden, weshalb Rammstedt (2004) auch dafür plädiert, Skalen mit einem Cronbachs Alpha ab 0,70 als ausreichend zu bewerten (63). Allerdings wiesen die Itemkorrelationen, die Item-Rest-Korrelation und die Ladungsstruktur dieses Faktors eher darauf hin, dass mindestens ein Item (um3 „Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen“) kein homogenes Konstrukt mit den übrigen Items bildet und eine Neuaufteilung dieser Skala angebracht ist.

Aber auch für die beiden als zufriedenstellend zu bewerteten Skalen kann diskutiert werden, inwiefern eine Einteilung in die vier Hauptdimensionen, die sich an der Klassifizierung der GDA-Empfehlungen orientieren, der inhaltlichen Heterogenität der Konstrukte gerecht wird und ob nicht eine differenziertere Aufteilung der Skalen der Datenstruktur besser gerecht würde. So bewies lediglich der Faktor Arbeitsorganisation im 8-Faktoren-Modell seine Einfachstruktur, aber für die Sozialen Aspekte kann eine Einfachstruktur nicht als gegeben betrachtet werden. Dies überrascht wenig, da es sich hierbei um ein sehr heterogenes Konstrukt handelt, welches unterschiedliche Aspekte (z. B. Unterstützung, Konflikte) und unterschiedliche soziale Beziehungen (z. B. zu Vorgesetzten, zu Kunden) subsumiert. Eine Aufteilung in mehrere, inhaltlich homogenere Skalen sollte die interne Konsistenz und Einfachstruktur deutlich verbessern. Dafür sprechen auch die Ergebnisse des 8-Faktoren-Modells, wo eine Aufteilung nach sozialen Beziehungen mit bestimmten Personengruppen und nach bestimmten Aspekten sozialer Beziehungen zu konsistenteren und eindeutigeren Ergebnissen führt. Dies steht mitunter im Widerspruch zum Ziel möglichst weniger Indikatoren und effizienten Möglichkeiten der Ergebnisinterpretation im Rahmen der strategischen Steuerung des BGM. Allerdings muss dieses Kriterium hier zurückgestellt werden, da eine verlässliche Maßnahmenplanungen nur auf Basis verlässlicher Daten möglich ist. Und dies setzt eine valide und reliable Erhebung und Datenauswertung voraus.

Die Inhaltsvalidität kann mit Blick auf die Ergebnisse der Expertenbefragung als gut bewertet werden. Zwar wurden viele Verbesserungs- und Ergänzungsvorschläge gemacht, aber die Relevanz und Vollständigkeit der Skalen wurde von der Mehrheit positiv bewertet. Viele Vorschläge bezogen sich auf Aspekte, welche nicht in den Merkmalsbereich arbeitsbedingte psychische Belastung fallen (z. B. Infrastrukturelle Anbindung des Arbeitsplatzes, individuelles Gesundheitsverhalten) und konnten für Anpassungen am AIGScreen nicht berücksichtigt werden. Für die Aufnahme einiger Aspekte, deren zusätzliche Erhebung vorgeschlagen wurde, spricht weniger, dass sie zur inhaltlichen Vollständigkeit des Fragebogens beitragen, sondern vielmehr, dass sie die Akzeptanz bei der Zielgruppe erhöhen würden, was sich auch positiv auf die Teilnahmebereitschaft auswirken kann. Es wurde auch eine hohe Übereinstimmung von Empfehlungen der GDA und Aspekten, die die Experten als wichtig bewerten festgestellt. Somit beeinflussen sich theoretische Fundierung und Augenscheinvalidität positiv.

Schließlich bestehen mit den zu erfüllenden Haupt- und Nebengütekriterien auch teilweise konkurrierende Zielsetzungen. Beim AIGScreen handelt es sich um einen Fragebogen, der mit möglichst wenigen Items ein breites und heterogenes Spektrum an Arbeitsmerkmalen erfassen soll. Die Reduktion auf so wenige Items wie möglich dient der Ökonomie und Praxistauglichkeit des Instruments, führt aber auch zu einer geringen inhaltlichen Nähe der damit erfassten Merkmale und wirkt sich negativ auf die Reliabilität und Validität der Skalen aus. Dadurch, dass viele latente Merkmale durch nur ein Item repräsentiert werden, ergibt sich mit dem AIGScreen ein sehr heterogenes Gesamtkonstrukt. Daher war zu erwarten, dass die interne Konsistenz für die einzelnen Skalen nicht sehr hoch ausfallen würde. Dieser Mangel an Homogenität wurde bewusst in Kauf genommen und kann bis zu einem gewissen Grad für heterogene latente Konstrukte akzeptiert werden (30). Allerdings stellen zu niedrige Korrelationen zwischen den Items einer Skala die Konstruktvalidität in Frage. Eine Aufteilung in homogene Subskalen erfordert für den vorliegenden Fragebogen, Skalen mit nur wenigen Items zu bilden. Niedrige Itemzahlen wirken sich jedoch negativ auf die Gütemaße einer Skala aus und beeinflussen so auch das Gesamtergebnis negativ, weil dann zum Beispiel keine ausreichende Modellsättigung erreicht werden kann (64). Es ist möglich, dass auch in Zukunft Items im Fragebogen enthalten sein werden, welche der Reliabilität und Konstruktvalidität einer Skala eher schaden, obwohl sie inhaltlich notwendig und gut begründbar sind. Für solche Items sollte erwogen werden, sie als Einzelindikatoren anstatt als Teil einer Skala zu verwenden. Dies erleichtert zwar nicht die Ergebnisinterpretation, da die Daten nicht auf einige wenige Kennwerte reduziert werden können, ist jedoch für den geplanten Einsatzzweck (Längsschnitt- und Subgruppenvergleiche) zwingend notwendig. Würden hier Summenindikatoren interpretiert, die nicht die wissenschaftlichen Gütekriterien erfüllen, besteht die Gefahr falscher Schlussfolgerungen, da die Ergebnisse, auf denen sie beruhen nicht als verlässlich betrachtet werden können.

Zusammenfassend kann festgehalten werden, dass der AIGScreen die Hauptgütekriterien Objektivität vollständig, Reliabilität und Validität jedoch nur eingeschränkt erfüllt. So konnte die postulierte vierdimensionale Struktur des Fragebogens nicht vollständig bestätigt werden. Eine Aufteilung in acht Subskalen wird der Datenstruktur wesentlich besser gerecht und erhöht die Reliabilität der Skalen erheblich. Allerdings kann auch damit keine vollständige Einfachstruktur generiert werden und die Modelfit-Maße der Konfirmatorischen Faktorenanalyse sprechen gegen die Konstruktvalidität. Auch wenn man die weiter oben diskutierten methodischen Einschränkungen berücksichtigt, die eine grundsätzliche Ablehnung des Modells trotz korrekter Spezifikation wahrscheinlicher machen, sollten Anpassungen im Fragebogen erwogen werden.

5.5 Notwendige weitere Anpassungen

Aus den Ergebnissen aller Erhebungen gingen Anpassungsbedarfe und Möglichkeiten der Verbesserung des Fragebogens hinsichtlich bestimmter Gütekriterien hervor. Diese betrafen sowohl inhaltliche Aspekte als auch die Gestaltung des Fragebogens, insbesondere die Antwortskalen. Hier sehen nicht nur die am Fragebogenreview beteiligten Experten Probleme bei den derzeitigen Antwortskalen, sondern auch der Stand der Forschung spricht für verschiedene notwendige Anpassungen. Hingegen kann über viele inhaltliche Änderungen, welche aufgrund der Rückmeldung aus der Expertenbefragung in der Zielgruppe, in Abschnitt 5.5.2 präsentiert werden diskutiert werden. Diese sind zumindest den Empfehlungen der GDA und BAuA folgend nicht alle notwendig. Da einige Aspekte jedoch häufig von verschiedenen Personen aus der Zielgruppe thematisiert wurden, schien es im Sinne der Verbesserung der Akzeptanz des AIGScreen bei der Zielgruppe nützlich, diese Aspekte aufzunehmen, sofern sie sich dem Merkmalsbereich arbeitsbedingter psychischer Belastung zuordnen ließen.

5.5.1 Anpassungen im Fragebogendesign

Im Fragebogenreview fiel auf, dass aufgrund des Instruktionstextes: „Bitte geben Sie an, inwiefern die Aussage auf Sie zutreffend ist“ eine missverständliche Auswahl der Antwortoption „unzutreffend“ erfolgen kann. Daher wurde die Formulierung korrigiert in „Die folgenden Aussagen beziehen sich auf Ihre Arbeitsmerkmale. Bitte schätzen Sie ein, inwiefern die folgenden Arbeitsmerkmale an Ihrem Hauptarbeitsplatz vorhanden sind“.

Um die Reliabilität des Items aa3 („Meine Arbeitszeitverteilung ist mir angenehm“) zu erhöhen, wurde eine konkretere Formulierung zur Erfassung der Selbstbestimmung („Ich bin frei in der zeitlichen Einteilung meiner Arbeit“) vorgeschlagen, da andere Aspekte, die in die Bewertung einer „angenehmen Arbeitszeitverteilung“ wie z. B. Überstunden durch andere Items abgedeckt werden und hier nicht erneut in die Bewertung des Merkmals einfließen sollten.

Das Item am9 („Ich werde bei meiner Arbeit selten unterbrochen“) war aus Sicht der Experten dringend zu überarbeiten, da durch die Verwendung des Begriffs „selten“ der Möglichkeitsraum beschränkt wurde. Die Angabe „trifft zu“ würde dann Arbeitsunterbrechungen immer noch nicht vollständig ausschließen, sondern nur den Schluss zulassen, dass diese zumindest nicht häufig auftreten. Aus diesem Grund wurde für zukünftige Befragungen die Formulierung „Störungen, die eine Unterbrechung oder einen Tätigkeitswechsel erzwingen, kommen bei meiner Arbeit nicht vor“ vorgeschlagen.

Schließlich führte die Verwendung des Begriffs „Untergebene“ in der Gruppe der Experten zu Irritationen, was jedoch für den speziellen Kontext Bundeswehr der angemessene Wortlaut ist und bei der Zielgruppe zu keinerlei Irritationen führen würde. Da der AIGScreen jedoch auch für andere Berufsgruppen eingesetzt werden können soll und eine alternative Formulierung gefunden werden konnte („unterstellte Beschäftigte“), wurde diese auch für zukünftige Erhebungen vorgeschlagen, um keine berufsgruppenspezifischen Formulierungen anwenden zu müssen und damit die Vergleichbarkeit der Daten zu gefährden. Genauso wurde auch der Begriff „Mitarbeitende“ durch „Kolleg*innen“ im neuen Fragebogenentwurf ersetzt. Auch hier war die Überlegung, dass die Verwendung eines universell einheitlich verstandenen Begriffs besser geeignet ist, wenn berufsgruppen- und arbeitssettingübergreifende Vergleiche auf Basis der mit dem AIGScreen erhobenen Daten möglich sein sollen.

Einen Hauptkritikpunkt stellte die Gestaltung der Antwortskalen dar. Die vierstufige Likert-Skala, auf der sämtliche Items zu bewerten sind, wurde im Fragebogenreview mit Blick auf den aktuellen Stand der Forschung zur optimalen Zahl von Skalenpunkten als problematisch bewertet. Da jedoch auch einige Argumente für die Beibehaltung der Antwortskala sprachen, sollten die Ergebnisse der statistischen Analysen abgewartet werden. Diese bestätigten jedoch, was bereits von den Experten vermutet wurde: Starke Abweichungen von der Normalverteilung, Decken- und Bodeneffekte und entsprechende Einschränkungen der Skalenreliabilität. Bei der derzeitigen vierstufigen Antwortskala zeigt sich eine Häufung positiver Antworten. In einigen Fällen stellt die höchste bzw. positivste Antwortoption die am häufigsten gewählte Antwort dar. Bei vielen Items wird jedoch die dritte bzw. die zweitpositivste Antwortoption gewählt. Es kann daher angenommen werden, dass es sich hierbei nicht um Deckeneffekte aufgrund einer zu niedrigen Itemschwierigkeit handelt, sondern dass sich die Antworten mehr der Normalverteilung annähern würden, wenn eine differenziertere Antwortskala eingesetzt würde. Dies kann insgesamt die Qualität, Reliabilität und Validität der Items und Antwortskalen positiv beeinflussen. Dafür sprechen auch zahlreiche Studienergebnisse. So müssen nachweislich geringere Trennschärfen, häufigere Abweichungen von der Normalverteilung und eine schlechtere Reliabilität und Validität auf Item- wie auf Skalenebene in Kauf genommen werden, wenn weniger als fünf Skalenpunkte eingesetzt werden (136, 221, 222, 125).

Bezüglich der Frage, wie viele Antwortoptionen optimal sind, gibt es mehrere Kriterien zu berücksichtigen. Zum einen muss geklärt werden, ob eine Mittelkategorie erwünscht ist. Und zum anderen spielt die Verbalisierung der Antwortskalen eine erhebliche Rolle. Grundsätzlich gilt, dass die Daten- und Messqualität bei vollverbalisierten Skalen besser ist als bei teilverbalisierten, wenn beispielsweise nur die Skalenendpunkte benannt werden (223). Durch die Benennung jeder Antwortoption wird die Beantwortung für die Probanden

erleichtert, Nebeneffekte durch andere visuelle Einflüsse reduziert und damit auch die Reliabilität und Validität erhöht (125). Insofern ist an der Vollverbalisierung der Antwortskalen auch in Zukunft festzuhalten. Hier gilt es jedoch zu berücksichtigen, dass vollverbalisierte Skalen äquidistant sein sollten, also gleiche Abstände zwischen den einzelnen Antwortoptionen aufweisen müssen (224). Das ist vor allem dann von Bedeutung, wenn die damit erhobenen (ordinalskalierten) Daten wie metrische Daten behandelt werden sollen bzw. für Analysen verwendet werden sollen, die metrisches Skalenniveau voraussetzen. Bei den vorliegenden vier Antwortoptionen – „ja“, „eher ja“, „eher nein“, „nein“ – kann diese Gleichabständigkeit jedoch nicht angenommen werden, da der Abstand zwischen der zweiten und dritten Antwortoption von den Befragten wahrscheinlich größer wahrgenommen wird als der zwischen der ersten und zweiten oder der dritten und vierten. Eine bessere Äquidistanz ließe sich hier durch die Verwendung einer Mittelkategorie herstellen. In der Fachliteratur lassen sich sowohl Argumente für als auch gegen die Verwendung einer Mittelkategorie finden und entsprechend kontrovers wurde deren zukünftiger Einsatz auch im Rahmen des Expertenreviews diskutiert. Gegen den Einsatz einer Mittelkategorie spricht, dass sie insbesondere bei schwach ausgeprägten Einstellungen oder Desinteresse am Thema die Tendenz der Befragten fördern, die Mittelkategorie als einen „easy way out“ (223) zu nutzen und damit den kognitiven Bearbeitungsaufwand möglichst gering zu halten. Diese Hypothese ließ sich im Rahmen empirischer Studien bisher jedoch nicht belegen, wohingegen umgekehrt nachgewiesen werden konnte, dass die Datenqualität bei Antwortskalen ohne Mittelpunkt geringer ausfällt (ebd.). Schließlich sollte auch eine neutrale Meinung als potenziell mögliche Ausprägung im Gesamtspektrum möglicher Antworten vertreten sein. Wenn man Personen, die tatsächlich keine Meinung haben, zwingt, sich für eine bestimmte Position zu entscheiden, werden die Ergebnisse dadurch häufig verzerrt sein (125). Zudem kann es die Akzeptanz des Fragebogens bei der Zielgruppe negativ beeinflussen, wenn diese den Eindruck gewinnen, dass sie ihre Meinung nicht adäquat abbilden können. Das Argument, dass die Mittelkategorie ein „easy way out“ sein kann, trifft im Falle des AIGScreen nicht zu, da zusätzlich die „unzutreffend“-Antwortoption angeboten wird. Zudem ist die Befragungsteilnahme freiwillig und es kann davon ausgegangen werden, dass Personen mit einem geringen Interesse am Thema oder ohne Meinung zur Fragestellung gar nicht erst an der Befragung teilnehmen werden. Die „unzutreffend“-Option sollte auch weiterhin angeboten werden, um zu vermeiden, dass ahnungslose Probanden sich gedrängt fühlen, ein Votum abzugeben, auch wenn das Merkmal auf sie eventuell gar nicht zutrifft und somit zu einer Verfälschung der Ergebnisse führen würde (103). Einen weiteren potenziellen Nachteil von Mittelkategorien können Schwierigkeiten bei der Interpretation im Rahmen der Auswertung darstellen. So könnte die Wahl der Mittelkategorie sowohl bedeuten, dass der

Befragte eine neutrale Meinung hat oder dass das betreffende Merkmal auf ihn nicht zutrifft oder Ambivalenz bezüglich des Sachverhalts ausdrücken. Die zweite Interpretationsmöglichkeit kann auch hier durch das Anbieten der „unzutreffend“-Antwortmöglichkeit weitestgehend ausgeschlossen werden. Die dritte Interpretationsmöglichkeit ist eher ein Problem bipolarer Skalen und kann daher für den AIGScreen als nachrangig betrachtet werden (125). Insgesamt sprechen mehr Argumente für die Verwendung einer Mittelkategorie, weshalb diese auch mit Blick auf den aktuellen Stand der Forschung in Zukunft als Antwortoption im AIGScreen enthalten sein sollte.

Bezüglich der absoluten Anzahl an Skalenpunkten lässt sich mit Blick auf die Studienlage sehr einheitlich feststellen, dass fünf bis sieben Skalenpunkte ideal bezüglich Datenqualität, Reliabilität, Validität und Präferenz der Befragten sind (223, 125, 136). Allerdings gilt das nicht uneingeschränkt für Zustimmungsskalen, wo eine Erhöhung der Antwortoptionen nachweislich zu einer geringeren Itemqualität und -validität führte und das Optimum möglicher Skalenpunkte bereits bei fünf erreicht war (221). Allerdings konnte auch nachgewiesen werden, dass die Datenqualität und wissenschaftliche Güte bei Verwendung von Zustimmungsskalen grundsätzlich niedriger ist als bei Verwendung itemspezifischer Antwortskalen und zwar unabhängig von der Anzahl der Antwortoptionen (222). Oftmals wird argumentiert, dass häufig wechselnde Antwortskalen von den Befragten eine zu hohe Anpassungsleistung erfordern und somit schneller zu Ermüdungseffekten führen. Saris et al. (2010) konnten jedoch nachweisen, dass das Gegenteil der Fall ist und die kognitiven Anforderungen für Probanden bei allgemeinen Zustimmungsskalen höher sind als bei passenden, itemspezifischen Antwortskalen. Zudem weisen Probanden ein wesentlich unsichereres Antwortverhalten bei Zustimmungsskalen auf, was sich negativ auf die Reliabilität auswirkt (ebd.). Um die Qualität der Items und Skalen sowie die wissenschaftliche Güte des AIGScreen insgesamt zu verbessern, sollten demnach in Zukunft itemspezifische Antwortskalen verwendet werden. Diese sollten nach aktuellem Forschungsstand wenigstens fünf Skalenpunkte enthalten. Preston & Colman (2000) konnten zeigen, dass die Diskriminanz bei weniger Antwortoptionen eingeschränkt ist und die Retest-Reliabilität sowie die interne Konsistenz mit zunehmender Zahl von Skalenpunkten steigt, wobei sich ab 7 Skalenpunkten ein gewisser Sättigungseffekt einstellt (136). Darüber hinaus könnten mehr Antwortoptionen die Probanden überfordern und wiederum negative Effekte auf Datenqualität und wissenschaftliche Güte haben. Allerdings konnte auch nachgewiesen werden, dass Probanden sich mit mehr Antwortoptionen wohler fühlen bzw. Skalen mit mindestens neun Antwortoptionen am besten bewerteten, wenn es um die Möglichkeit ging, die eigenen Gefühle adäquat auszudrücken (125, 136). Hingegen wurden Einfachheit der Anwendung und Schnelligkeit bei der Beantwortung insbesondere fünfstufige Antwortskalen von den Anwendern positiv bewertet (136).

Um ein gutes Verhältnis zwischen wissenschaftlicher Güte, Itemqualität, Ökonomie und Befragtenpräferenz bei den Antwortskalen zu erreichen, scheinen fünfstufige Antwortskalen mit itemspezifischen Formulierungen am besten geeignet. Die im Rahmen des Expertenreviews geäußerte Befürchtung, dass eine Vergleichbarkeit mit bereits erhobenen Daten bei Veränderung der Antwortskalen verloren ginge, kann entgegengehalten werden, dass Transformationen zwischen verschiedenen Skalen möglich sind und zu vergleichbaren Ergebnissen führen (225). Dies gilt auch, wenn zwischen geraden und ungeraden Skalen umgerechnet werden soll. Schließlich muss auch eingestanden werden, dass es nicht im Sinne der Erhebung sein kann, an einem Instrument festzuhalten, welches nicht in vollem Umfang den daran gestellten Anforderungen gerecht wird, obwohl Verbesserungsmöglichkeiten bekannt und umsetzbar sind. Dies betrifft sowohl Aspekte des Fragebogendesigns als auch die inhaltliche Ausgestaltung, welche inhaltlich valide, d. h. vollständig in Bezug auf die Konstrukte, die es misst, sein soll als auch relevant aus Sicht der Zielgruppe bewertet werden soll und bei dieser akzeptiert sein muss. Aus diesem Grund werden für die Fragebogenrevision auch einige Vorschläge für inhaltliche Anpassungen des AIGScreen im nachfolgenden Kapitel vorgestellt.

5.5.2 Inhaltliche Anpassungen im Fragebogen

Aufgrund des hohen Modifikationsbedarfs, der sich sowohl aus den Expertenbefragungen wie auch aus der statistischen Analyse ergab, wurde eine umfassende Revision des Fragebogens vorgenommen. Hierbei wurden einige Items gänzlich gestrichen und durch neue ersetzt, andere in ihrer Formulierung angepasst. Für sämtliche Items wurden zudem neue, itemspezifische Antwortskalen formuliert. Eine Übersicht der Neuformulierungen findet sich in Tabelle 17. Zur besseren Übersichtlichkeit wurden die Items wieder den vier Merkmalsbereichen Arbeitsorganisation, Arbeitsinhalt und -aufgabe, Soziale Aspekte und Arbeitsumgebung zugeordnet. Dies entspricht jedoch nicht der postulierten latenten Struktur.

Der Neuentwurf des Fragebogens, „AIGScreen 3.0“, umfasst alle von der GDA empfohlenen Aspekte zur Beurteilung der psychischen Belastung in den Subdimensionen Arbeitsumgebung, Arbeitsorganisation, Soziale Aspekte sowie Arbeitsinhalt und Arbeitsaufgabe. Aspekte, die neue Formen der Arbeit betreffen bzw. solche, die von den Experten als zusätzlich zu erfassende Themen vorgeschlagen wurden (Work-Life-Balance, räumliche und infrastrukturelle Anbindung des Arbeitsplatzes), können nicht universell für alle möglichen Arbeitsbereiche erfasst werden und stellen zudem an sich so komplexe Konstrukte dar, dass sich für deren Erhebung die Entwicklung separater Instrumente empfiehlt. Diese können, in Bereichen, in denen es sinnvoll erscheint, den AIGScreen

modular ergänzen. Einzelne Aspekte neuer Arbeitsformen, wie z. B. Flexibilisierung von Arbeitszeiten, sind jedoch durch die inhaltlichen Modifikationen der Items zur Erfassung der vier Hauptdimensionen im neuen Fragebogen enthalten.

Einige Fragen stellen zudem filtergeführte Fragen dar, d. h. sie werden nur Personen gestellt, die in einer vorausgegangenen Frage bestimmte Antwortoptionen gewählt haben. Gibt beispielsweise eine Person an, keine unterstellten Beschäftigten zu haben, sind weiterführende Fragen dazu nicht sinnvoll. Gleiches gilt für die Ausprägung von Arbeitsmerkmalen. Treten keine Arbeitsunterbrechungen auf oder werden keine physischen Arbeiten verrichtet, können die Befragten auch keine Angaben zu konkreten Belastungen, die sich daraus ergeben würden, machen. Im Sinne der Fragebogenökonomie und der Anwenderfreundlichkeit bzw. Praxistauglichkeit und um die Augenscheinvalidität nicht zu gefährden, werden Items, die an bestimmte Merkmalsausprägungen anknüpfen daher nur den betreffenden Personen präsentiert.

Wie Tabelle 17 zu entnehmen ist, wurden einheitliche Antwortskalen für alle AIG-Items aufgegeben und fragespezifische Antwortoptionen formuliert. Eine Neuformulierung der Antwortoptionen wäre in jedem Fall notwendig gewesen, da, wie in 5.5.1 erläutert, eine Erhöhung der Anzahl der Antwortkategorien als notwendig erachtet wurde.

Einige Items wurden aus dem Fragebogen entfernt und der betreffende Aspekt auf einem höheren Abstraktionsniveau erfasst wie z. B. das Item „Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfängern) respektiert und in meiner Arbeit unterstützt“. Dieses Item war vor allem problematisch, weil es zwei verschiedene Aspekte (Respekt, Unterstützung) erfragte, was die Reliabilität der Antworten erheblich beeinträchtigte. Des Weiteren handelt es sich um ein sehr spezifisches Merkmal, welches nicht auf alle Beschäftigten zutrifft und daher auch hohe Antwortausfälle verzeichnete. Daher wurde entschieden, die zugrundeliegenden Aspekte, nämlich Wertschätzung, Unterstützung und emotionale Belastungen in allgemeiner formulierten Items zu erfassen, welche von möglichst allen Befragten bewertet werden können. Da sich im Arbeitskontext bzw. im Rahmen von BGM die Wertschätzung und Unterstützung ausschließlich von Seiten der Vorgesetzten und Kollegen beeinflussen und verändern lässt, nicht aber von Kunden und Leistungsempfängern, wurden diese beiden Aspekte auch ausschließlich für die ersten beiden Personengruppen erhoben. Psychische Belastungen durch den Kontakt mit Kunden und Leistungsempfängern wurden durch die emotionale Inanspruchnahme in den Items 32 bis 34 erfasst.

Auch das Item „Die körperliche Leistungsfähigkeit ist wichtig, damit ich die Anforderungen meines Dienstpostens erfüllen kann“ ist im neuen Fragebogen ebenfalls nicht mehr enthalten, da stattdessen die physische Belastung mit Item Nr. 48 und ggf. vier

anschließenden Filterfragen erfasst werden sollte. Die Anpassung wurde als sinnvoll erachtet, da mit dem ursprünglichen Item keine Gefährdung identifiziert werden konnte und durch die Umformulierung objektivierbare Belastungsindikatoren wie z. B. Lastwichtung, Zeitwichtung oder Ausführungsbedingungen erhoben werden können.

Das Item „Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet“ wurde seitens der Experten sehr kritisch bewertet. Einerseits sei der Fragegegenstand nicht klar erkennbar. So können hier sowohl individuelle Qualifikationen wie auch die Qualität der Einarbeitung und Einweisung bewertet werden. Andererseits wurde der Aspekt Qualifikation in der Zielgruppe insgesamt sehr relevant bewertet und sollte nach Meinung der befragten Experten mit mehreren differenzierten Items anstelle eines globalen Items erhoben werden. Aus diesem Grund wurde auch das Item „Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit“ beibehalten, obwohl es keinen explizit zu erfassenden Aspekt psychischer Gefährdung gemäß GDA-Empfehlung darstellt. Zusätzlich wurden drei Neuformulierungen zur Beurteilung des Aspekts Qualifikation im AIGScreen 3.0 aufgenommen. Dies betrifft das Item Nr. 22 zur individuellen fachlichen Qualifikation, das Item Nr. 23, welches die sozialen und kommunikativen Fähigkeiten der Befragten erfasst und das Item Nr. 24, welches nun explizit die Qualität der Einarbeitung berücksichtigt. Da seitens der Experten mehrfach auf den Zusammenhang zwischen geeigneter Qualifikation für die jeweiligen Arbeitsaufgaben und der erlebten Sinnhaftigkeit der eigenen Tätigkeit hingewiesen wurde und dabei auch empfohlen wurde, neben einer potenziellen Überforderung auch die Möglichkeit der Unterforderung im Fragebogen abzudecken (Vgl. Diskussion in Abschnitt 5.2.1), wurde Item Nr. 22 entsprechend so formuliert, dass es bipolar beantwortet werden kann.

Ergebniskontrolle als Indikator für die Vollständigkeit einer Tätigkeit („Ich kann erkennen, ob ich meine Arbeit gut gemacht habe“) wurde aus dem Fragebogen entfernt. Vielmehr sollten als zentrale Aspekte einer vollständigen die sequenzielle Vollständigkeit sowie die hierarchische Vollständigkeit erhoben werden, welche sich in Aufgabengeschlossenheit und Abwechslungsreichtum ausdrücken (96). Um neben der bereits im Fragebogen enthaltenen Variabilität auch die Aufgabengeschlossenheit zu erfassen, wurde das neu entwickelte Item Nr. 27 im Fragebogen ergänzt. Zusätzlich, wenn auch nicht als potenzieller Gefährdungsfaktor, sondern eher als Ressource zu betrachten, wurde mit Item Nr. 31 die Sinnhaftigkeit erfasst. Obwohl dieser Aspekt bisher nicht im Fragebogen oder in den den Experten präsentierten Definitionen enthalten war, wurde er mehrfach explizit von diesen nachgefragt und daher aufgrund seiner augenscheinlich hohen Relevanz für die Zielgruppe im Fragebogen ergänzt. Die Sinnhaftigkeit einer Tätigkeit kann zudem andere gefährdungsrelevante Faktoren, insbesondere soziale Aspekte wie Unterstützung,

beeinflussen und ist somit nicht irrelevant für die Auswirkung der Arbeit auf die individuelle Gesundheit (226).

Ebenfalls aus dem Fragebogen entfernt wurde das Item „Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen“. Die Augenscheinvalidität dieses Items war mit Blick auf die Ergebnisse der Expertenbefragung nicht gegeben. Eine Notwendigkeit zur Erfassung dieses Aspekts ergibt sich auch nicht aus den Empfehlungen der GDA. Die damit zusammenhängende Gefährdung, nämlich die Massierung von Arbeitszeit ohne Ausgleich, wird über die Frage zu Überstunden erfasst.

Das Item „Meine Arbeitspausen sind störungsfrei“ wurde zwar in beiden Expertenerhebungen nicht kritisiert und auch nicht als irrelevant bewertet, allerdings war die Rückmeldung der Zielgruppe, dass ein Item zur Pausengestaltung ausreichend sei, zumal es hierfür ja klare gesetzliche Vorgaben gäbe. Diese Anregungen wurden aufgegriffen, indem die beiden Items zum Aspekt Arbeitspausen in Item Nr. 13 kombiniert und die gesetzlich festgelegten Mindestzeiten in die Formulierung des Items aufgenommen wurden.

Das Item „Meine Arbeitszeitverteilung ist mir angenehm“ wurde im Rahmen des Fragebogenreviews von den Experten hinsichtlich seiner Intersubjektivität kritisch bewertet, im Rahmen der Expertenbefragung aber als relevant eingestuft. Allerdings wurde auch kritisch hinterfragt, bezüglich welches konkreten Aspektes diese Aussage zu interpretieren sei. Es wurde daher entschieden, die zugrundeliegenden Aspekte in Bezug auf die Arbeitszeit in anderer Form weiter zu erheben und konkretere Formulierungen zu wählen. Hierfür wurde einerseits gemäß der mehrfachen Forderung der Experten aus der Zielgruppe das Item Nr. 12 zur Erfassung der individuellen zeitlichen Flexibilität neu im Fragebogen aufgenommen und andere Aspekte, welche die Arbeitszeit betreffen wie z. B. das Pausenregime, Mehrarbeit oder Arbeitsintensität dezidiert erfasst. Das Arbeitszeit-Item „Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen“ wurde ebenfalls gestrichen. Auch hier stand zur Diskussion, inwiefern „angenehm“ eine eindeutige Formulierung darstellt, die von allen Befragten gleich interpretiert wird. Zudem wurde die Relevanz dieses spezifischen Aspekts im Rahmen der Expertenbefragung in Frage gestellt und stellt auch im Rahmen der GDA-Empfehlungen keinen explizit zu erfassenden Aspekt dar. Die damit verbundenen psychischen Belastungen wie Arbeitsmenge und -intensität sowie Mehrarbeit werden vom Fragebogen erhoben, sodass dadurch bedingte mögliche kritische Ausprägungen auch im neuen Fragebogen identifiziert werden können.

Ebenfalls hinsichtlich der Bedingung einer intersubjektiv einheitlichen Lesart kritisiert wurde das Item „Meine Arbeit ist in angenehmem Maße abwechslungsreich“. Variabilität der

Arbeitstätigkeit stellt einen zentralen Aspekt in den GDA-Vorgaben dar und sollte daher trotz niedrig bewerteter Relevanz aus Sicht der Zielgruppe weiter erhoben werden. Um den methodischen Mängeln zu begegnen und die Messgenauigkeit zu verbessern wurde in Form von Item Nr. 25 eine neue Formulierung gewählt, sodass eindeutig wird, dass der Fragegegenstand Abwechslungsreichtum bezüglich Aufgaben und Anforderungen ist.

Auch die Formulierung „Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z. B. Personalverantwortung)“ wurde im Fragebogenreview als änderungsbedürftig bewertet. Auch hier stand eine allgemein einheitliche Frageinterpretation aufgrund des Wortes „angemessen“ zur Diskussion. Hierfür wurde eine Neuformulierung mit Item Nr. 28 aufgenommen. Da der Aspekt Verantwortung jedoch nicht nur kritische Ausprägungen hinsichtlich des Grades bzw. der Menge zu tragender Verantwortung aufweisen kann, sondern auch Rollenklarheit und Transparenz entscheidend für eine gesunde Arbeitsgestaltung sind (50), wurde ein zusätzliches Item Nr. 29 aufgenommen, um diesen Aspekt vollständig zu erfassen. Vor diesem Grund konnte auch das von der Zielgruppe häufig als irrelevant kritisierte Item „Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material)“ auch im Sinne der Fragebogenökonomie, aber vor allem zum Zweck der Augenscheinvalidität aus dem Fragebogen entfernt werden.

Das Item „Ich werde bei meiner Arbeit selten unterbrochen“ wurde in der Expertenbefragung zwar als sehr relevant bewertet, im Expertenreview jedoch aufgrund seiner Formulierung kritisch bewertet. Daher sollte der Aspekt beibehalten, aber in einem anderen Wortlaut erfasst werden. Aufgrund der hohen Bedeutung, die diesem Aspekt aus Sicht der Zielgruppe beigemessen wurde, wurde neben einem Core-Item eine weitere Filterfrage ergänzt. Der Vorteil von Item Nr. 18 gegenüber der Formulierung in der Vorgängerversion ist, dass sie das vollständige Möglichkeitsspektrum bezüglich der Häufigkeit erfasst. Das Item Nr. 19 wurde ergänzt, weil neben der Quantität auch die qualitative Beeinträchtigung durch Arbeitsunterbrechungen als relevant für etwaige psychische Beanspruchungen erachtet wurde.

Mehrarbeit und Überstunden sollten trotz durchwachsener Bewertungen seitens der Experten aus dem Geschäftsbereich des BMVg beibehalten werden, da sie gemäß GDA-Empfehlungen und aktuellem Forschungsstand einen wichtigen Aspekt arbeitsbedingter psychischer Belastung darstellen. Allerdings erschien eine Neuformulierung des Items aus mehreren Gründen notwendig. Zum einen impliziert die Formulierung „Die Anzahl der von mir verlangten Überstunden ist angemessen“, dass diese ausschließlich von außen angeordnet oder eingefordert werden. Da der Arbeitgeber jedoch in der Pflicht ist, Rahmenbedingungen zu überprüfen und ggf. anzupassen, wenn es zu häufigen

Überstunden kommt, sollten individuelle Verhaltensmuster der betreffenden Beschäftigten ebenso berücksichtigt werden wie Mängel in der Arbeitsplanung (227). Wie schon bei vorausgegangenen Items wurde auch die Frage seitens der Experten gestellt, inwiefern „angemessen“ von allen Befragten gleichermaßen interpretiert und bewertet werden kann und ob dadurch nicht die Reliabilität negativ beeinflusst wäre. Da zudem die kritische Ausprägung nicht im grundsätzlichen Auftreten von Überstunden, sondern in der Häufigkeit der Notwendigkeit dazu gesehen wurde, musste die Formulierung für Item Nr. 11 so gewählt werden, dass sie auf einer Häufigkeitsskala bewertet werden kann.

Das Problem der intersubjektiven Verständlichkeit wurde auch in Bezug auf das Item „Ich habe einen angemessenen Handlungsspielraum (z. B. Entscheidungsfreiheit) bei meiner Arbeit“ im Fragebogenreview aufgeworfen. Allerdings ist hier neben der Verwendung des Begriffs „angemessen“ auch die eindeutige Verständlichkeit und Interpretation von „Handlungsspielraum“ fraglich. Handlungsspielraum wird auch in der Fachliteratur nicht nur synonym mit Autonomie, sondern auch mit (horizontaler) Variabilität verwendet (193). Abwechslungsreichtum der Tätigkeit ist jedoch bereits mit den Items Nr. 25 und Nr. 26 erfasst. Autonomie wiederum kann sich auf inhaltliche und zeitliche Aspekte beziehen. Letztere ist bereits mit Item Nr. 11 erfasst. Entsprechend war es für die Verbesserung der Messgenauigkeit notwendig das Item Nr. 30 so umzuformulieren, dass es von allen Probanden einheitlich im Sinne des „Grades der Entscheidungskompetenz“ (ebd.) in Bezug auf die eigene Tätigkeit verstanden und bewertet wird.

Drei Items zur Erfassung von Kooperations- und Abstimmungsstrukturen im Betrieb wurden neu im Fragebogen aufgenommen. Dies betrifft die Items Nr. 14 „Ich kann mich bei meiner Arbeit mit Kolleg*innen austauschen“, Nr. 15 „Wie beurteilen Sie den Abstimmungsbedarf mit Kolleg*innen?“ und Nr. 16 „Meine Aufgaben erledige ich in Zusammenarbeit mit Kolleg*innen“. Abstimmungsmöglichkeiten stellen laut BAuA einen wichtigen Gestaltungsaspekt der Dimension Arbeitsinhalt und -aufgabe sowie eine kritische Größe bei den arbeitsorganisatorisch Bedingten Kommunikations- und Kooperationserfordernissen dar (98). Da insbesondere im Rahmen der Expertenbefragung wiederholt darauf hingewiesen wurde, dass ein zu hoher Abstimmungsbedarf eine zu erfassende Gefährdung darstellen kann und auch die GDA eine zu hohe Interaktionsdichte als kritische Ausprägung beschreibt, wird dieser Aspekt der Kooperationsstrukturen nun explizit miterhoben. Da Alleinarbeit und Isolation durch die neuen technologischen Möglichkeiten der Automatisierung und Digitalisierung in Zukunft immer häufiger auftreten werden und eine psychische Gefährdung darstellen (228), sollte auch dieser Aspekt im AIGScreen 3.0 Berücksichtigung finden.

Ein weiteres Item, welches auf Hinweis der Zielgruppe aber auch im Sinne der Vollständigkeit der offiziellen Empfehlungen zur Gefährdungsbeurteilung aufgenommen wurde, ist das Item Nr. 20 („Über wichtige Dinge und Vorgänge in unserem Betrieb sind wir ausreichend informiert“), denn laut BAuA gehört es zur gesunden Arbeitsgestaltung, dass Beschäftigte alle für die Erledigung ihrer Aufgaben relevanten Informationen erhalten (98) und auch aus Sicht der Zielgruppe war dies ein relevanter, zu erhebender Aspekt.

Drei weitere Items wurden vor allem aufgrund der Befragungsergebnisse der Expertenbefragung im neuen Fragebogen aufgenommen, da sich aus Sicht mehrerer Experten noch zusätzlich erhoben werden sollten. Dies betrifft die Items Nr. 55, Nr. 56 und Nr. 59, welche Gefahrstoffexpositionen und Unfallgefahren erfragen. Damit wird auch die Erfassung aller relevanten Aspekte im Rahmen der Gefährdungsbeurteilung psychischer Belastungen gemäß GDA-Leitlinien komplettiert (196, 50). Denn gemäß DIN-Definition (DIN EN ISO 10075-1) zählen zu den psychischen Belastungen alle von außen auf den Menschen einwirkenden Einflüsse, die psychisch auf ihn Einwirken (229). Potenzielle chemische, physikalische oder biologische Gefährdungen wirken insofern auch auf psychischer Ebene, als dass sie Beschäftigte vor zusätzliche kognitive Herausforderungen stellen (z. B. Notwendigkeit erhöhter Achtsamkeit), aber auch emotional eine Belastung darstellen können.

Tabelle 17: Neuentwurf auf Basis der Fragebogenrevision (AIGScreen 3.0)

Folgt gefiltert auf	Arbeits-Nr.	Teilaspekt/Subskala	Item	Antwort-skala
Arbeitsorganisation				
CORE	10	Arbeitsintensität	Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit ausgewogen.	1
CORE	11	Arbeitszeit: Überstunden/Massierung	Meine Tätigkeit erfordert es, die Arbeitszeit zu überschreiten	2
CORE	12	Arbeitszeit: Selbstbestimmung / Flexibilität	Ich bin frei in der zeitlichen Einteilung meiner Arbeit.	1
CORE	13	Arbeitszeit: Pause	Ich habe jeden Tag die Möglichkeit, eine Pause von mind. 30 Minuten zu machen (bei mehr als 6 Arbeitsstunden bzw. mind. 45 Minuten bei mehr als 9 Arbeitsstunden)	1
CORE	14	Kommunikation / Kooperation: Abstimmung	Ich kann mich bei meiner Arbeit mit Kolleg*innen austauschen.	1
CORE	15	Kommunikation - Kooperation: Interaktionsdichte	Wie beurteilen Sie den Abstimmungsbedarf mit Kolleg*innen?	3
CORE	16	Kommunikation - Kooperation: Zusammenarbeit	Meine Aufgaben erledige ich in Zusammenarbeit mit Kolleg*innen.	2
CORE	17	Kommunikation: in-/direkte Kommunikation	Die Abstimmung mit Kolleg*innen erfolgt indirekt (z.B. E-Mail, Videotelefonie).	2
CORE	18	Arbeitsablauf: Störungen und Unterbrechungen	Störungen, die eine Unterbrechung oder einen Tätigkeitswechsel erzwingen, kommen bei meiner Arbeit nicht vor.	1
18	19	Arbeitsablauf: Störungen und Unterbrechungen	Störungen bei der Arbeit führen dazu, dass Resultate verloren gehen und ich neu beginnen muss.	2

Arbeitsinhalt/-aufgabe				
CORE	20	Kommunikation: Informationsangebot	Über wichtige Dinge und Vorgänge in unserem Betrieb sind wir ausreichend informiert.	1
CORE	21	Entwicklungsmöglichkeiten	Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit.	1
CORE	22	Qualifikation: Fachkompetenz	Wie würden Sie Ihre fachliche Qualifikation für Ihre Tätigkeit bewerten?	4
CORE	23	Qualifikation: Sozialkompetenz	Meine sozialen und kommunikativen Fähigkeiten werden den Anforderungen meiner Arbeit gerecht.	1
CORE	24	Qualifikation: Einarbeitung	In mein berufliches Tätigkeitsfeld wurde bzw. werde ich angemessen eingearbeitet, sodass ich gut arbeiten kann.	1
CORE	25	Variabilität: Aufgabe & hierarchische Vollständigkeit	Bei meiner Arbeit haben ich, insgesamt gesehen, häufig wechselnde, unterschiedliche Arbeitsaufgaben.	1
CORE	26	Variabilität: Anforderungen	Meine Aufgaben variieren in Anforderung und Komplexität.	1
CORE	27	Vollständigkeit: Aufgabengeschlossenheit (sequenzielle V. /Task Identity)	Meine Arbeit führe ich nicht nur aus, sondern plane, koordiniere und überprüfe sie auch selbst.	1
CORE	28	Verantwortung: Umfang	Die Verantwortung, die ich trage, ist...	3
CORE	29	Verantwortung: Transparenz/Rollenklarheit	Ich weiß genau, welche Dinge in meinen Verantwortungsbereich fallen.	1
CORE	30	Handlungsspielraum	Ich habe die Möglichkeit Arbeitsprozesse mitzubestimmen und zu beeinflussen.	1
CORE	31	Sinnhaftigkeit	Ich empfinde meine Arbeit als sinnvoll.	1

CORE	32	Emotionale Inanspruchnahme: Belastende Situationen	Ist Ihre Arbeit emotional fordernd?	5
CORE	33	Emotionale Inanspruchnahme: Gewalt/Gefahr	Kommt es bei Ihrer Arbeit zu Bedrohungen durch verbale oder körperliche Gewalt?	2
32	34	Emotionale Inanspruchnahme: Unterdrückung der eigenen Gefühle	Verlangt Ihre Arbeit von Ihnen, dass Sie Gefühle zeigen, die nicht mit Ihrem Empfinden übereinstimmen?	2

Soziale Aspekte

CORE	35	Führung: Feedback	Erhalten Sie von Ihren nächsten Vorgesetzten ausreichende Rückmeldung (Anerkennung, Kritik, Beurteilung) über die eigene Leistung?	5
CORE	36	Führung: Wertschätzung	Erfährt Ihre Arbeit Anerkennung und Wertschätzung durch Vorgesetzte / die Führung?	5
CORE	37	Führung: Unterstützung	Ich erfahre genügend Unterstützung von meinen unmittelbaren Vorgesetzten.	1
CORE	38	Vorgesetzt*	Wie häufig kommt es zu Konflikten mit Ihren unmittelbaren Vorgesetzten?	2
CORE	39	Führung: (Organisationale) Gerechtigkeit	Wird die Arbeit gerecht verteilt?	5
CORE	40	Kolleg*: Unterstützung	Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis.	1
CORE	41	Kolleg*: Konflikthäufigkeit	Wie häufig kommt es zu Konflikten mit Ihren Kolleg*innen?	2
CORE	42	Konfliktbewältigung	Konflikte [mit Kolleg*innen/Vorgesetzten] werden konstruktiv gelöst.	1
CORE	43	Kolleg*innen: Kontaktintensität	Die Anzahl sozialer Kontakte bei der Arbeit ist...	3

CORE	44	Untergebene	Ich erfahre genügend Unterstützung von den mir unterstellten Beschäftigten.	1
wenn bei 44 nicht "keine unterstellten Beschäftigten" ausgewählt	45	Untergebene	Wie häufig kommt es zu Konflikten mit den Ihnen unterstellten Beschäftigten?	2

Arbeitsumgebung

CORE	46	Äußere Faktoren	Die äußeren Bedingungen (Klima, Beleuchtung, Belüftung, Geräuschpegel etc.) an meinem Hauptarbeitsplatz sind angenehm.	1
46	47	Äußere Faktoren	Treffen eines oder mehrere der folgenden Merkmale auf Ihren Arbeitsplatz zu? (Liste mit potenziellen Gefährdungen, wie z.B. schlechte Beleuchtung (mangelnde oder defekte Lichtquellen, Flackern, starkes Gegenlicht), unangenehme Gerüche, Lärm, zu hohe Raumtemperaturen / zu niedrige Raumtemperaturen, fehlende Sozial- oder Pausenräume, mangelnde Belüftung, mangelhafte Hygiene, Platzmangel / zu geringe Arbeitsfläche etc.)	6
CORE	48	Physische Belastung	Wie häufig müssen Sie körperlich schwer arbeiten, z.B. schwer heben, tragen oder stemmen oder in ungünstiger Körperhaltung arbeiten?	2
48	49	Physische Belastung	Wie viele Male müssen Sie pro Arbeitstag körperlich schwer arbeiten?	7
48	50	Physische Belastung	Wie viele Minuten müssen Sie pro Arbeitstag körperlich schwer arbeiten?	7
48	51	Physische Belastung	Welche Gewichte müssen Sie in Kilogramm pro Umsatzvorhang bewegen (ohne das eigene Körpergewicht)?	7
48	52	Physische Belastung	Treffen eines oder mehrere der folgenden Merkmale auf Ihre Tätigkeit zu? (Liste mit potenziellen Gefährdungen, wie z.B. rutschiger oder verschmutzter Untergrund, hohe Unfallgefahr, schlechte Beleuchtung, Zwangshaltungen (lange gebeugt	6

			stehen, Arbeiten in gebückter oder verdrehter Haltung, Arbeiten über Kopf, Bewegungseinschränkungen, fehlende Standsicherheit etc.)	
CORE	53	Ergonomie	Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann (z. B. Arbeitshöhe, Sitzhöhe, Beinfreiheit, Durchgangsbreiten/Sicherheitsabstände).	1
CORE	54	Bewegung	Haben Sie die Möglichkeit, eine wechselnde Körperhaltung einzunehmen und / oder ausreichende Bewegung am Arbeitsplatz?	2
CORE	55	Gefahrstoffe	Arbeiten Sie mit Gefahrstoffen oder haben Sie im Rahmen Ihrer Tätigkeit Kontakt zu Gefahrstoffen?	2
			Hinweise auf Gefahrstoffe finden sich auf Produktverpackungen in Form von Gefahren-Piktogrammen sowie H- und P-Sätzen. Für die Einstufung und Kennzeichnung von Gefahrstoffen wird das weltweit vereinheitlichte System „Globally Harmonized System“ (GHS) verwendet.	
			Beispiele für Gefahrstoffe: Aceton, Blei, Bleichromat, Lösungsmittel in Klebern, p-Toluidin, Quecksilber, Benzol, Formaldehydlösung, weißer Phosphor	
55	56	Gefahrstoffe	Wenn ja, welche? (Liste)	6
CORE	57	Arbeitsmittel	Die nötigen Arbeitsmittel sind vorhanden.	1
CORE	58	Arbeitsmittel	Die vorhandenen Arbeitsmittel sind in gutem, funktionsfähigem Zustand.	1
CORE	59	Unfall-/Verletzungsgefahr	Meine Tätigkeit ist frei von erhöhten Unfall-, Verletzungs- oder Erkrankungsgefahren.	1

Antwortskalen: 1 = Stimme voll zu - stimme zu - teils/teils - stimme nicht zu - stimme überhaupt nicht zu; 2 = sehr häufig - regelmäßig - manchmal - selten - (fast) nie; 3 = zu hoch - etwas zu hoch - genau richtig - etwas zu niedrig - zu niedrig; 4 = unterqualifiziert - etwas unterqualifiziert - genau richtig - etwas überqualifiziert – überqualifiziert; 5 = in sehr hohem Maß, in hohem Maß, zum Teil, in geringem Maß, in sehr geringem Maß; 6 = Listenauswahl, dichotom (ja / nein); 7 = offen

5.6 Limitationen

Methodische Einschränkungen und Kritikpunkte an den hier eingesetzten Verfahren wurden bereits an mehreren Stellen diskutiert. In diesem Kapitel sollen noch einmal die Punkte zusammengefasst werden, welche sich einschränkend auf die Ergebnisinterpretation auswirken bzw. bei der Interpretation der Ergebnisse zumindest berücksichtigt werden sollten.

Bezüglich expertenbeteiligender Verfahren wurde bereits erläutert, dass Kritiker dieser Methoden hier die Gefahr von Verzerrungseffekten aufgrund subjektiver Einzelmeinungen sehen. Ferner ist bei einer freiwilligen Befragung immer die Möglichkeit der Verzerrung der Ergebnisse durch Selektionseffekte zu berücksichtigen. Da die Selbstselektion wahrscheinlich eher von Faktoren wie Interesse oder Berührungspunkten mit dem Thema bestimmt war und weniger von einer eindeutigen Ablehnung oder Befürwortung des Fragebogens, ist zu vermuten, dass die Bewertung der Fragebogeninhalte nicht oder zumindest nicht wesentlich in positive oder negative Richtung verzerrt ist. Auffallend war jedoch, dass die Untergruppe derjenigen, die als Beruf „Sonstiges“ angaben, ein tendenziell kritischeres Antwortverhalten aufwiesen. Für diese Gruppe muss eine Verzerrung durch Selbstselektionseffekte insofern angenommen werden, als dass hier vor allem Personen an der Befragung teilnahmen, die dem Fragebogen oder der Erhebung mittels AIGScreen im Rahmen des Betrieblichen Gesundheitsmanagements eher kritisch gegenüberstanden. Für diese These spricht, dass es sich dabei auch um Berufsgruppen handelt, welche laut Zielgruppendefinition gar nicht an der Befragung hätten teilnehmen sollen (z. B. Personen, die angaben ausschließlich BGM-Teilnehmer zu sein). Hier wäre eine stärkere Zugangskontrolle zur Befragung zu erwägen gewesen. Andererseits lässt sich diese Personengruppe anhand ihrer Berufsgruppenzugehörigkeit leicht herausfiltern, sodass auch eine Auswertung der Daten ohne diese Gruppe möglich ist. Zudem handelt es sich um eine sehr kleine Anzahl von Personen ($n = 11^{12}$), sodass insgesamt davon ausgegangen werden kann, dass die übergroße Mehrheit der Befragten der tatsächlichen Zielgruppe entstammt und mögliche Verzerrungen durch Selbstselektion sich ausgleichen. Allerdings unterliegt diese Annahme einer Einschränkung. Fehlende Angaben liegen vor allem für die Fragen vor, die am Ende des Fragebogens platziert waren. Dies lässt vermuten, dass die Befragten die Umfrage entweder zwischenspeicherten und nicht beendeten oder frühzeitig abbrachen. Beides spricht dafür, dass der Bearbeitungsaufwand für die Befragten zu hoch war. Diesem Problem könnte in Zukunft durch ein alternatives Erhebungsdesign begegnet werden. Möglich wäre z. B. den Fragebogen aufzuteilen und die einzelnen Befragten immer nur zufällig ausgewählte

¹² Diejenigen, die bei der Frage nach der Berufsgruppenzugehörigkeit „Sonstiges“ wählten und dann als Beruf eine zur Zielgruppe gehörende Position nannten wie z. B. Mitglied im Arbeitskreis Gesundheit zu sein, wurden nicht der Gruppe „Sonstige“ zugerechnet.

Abschnitte bearbeiten zu lassen, sodass nicht alle Experten alle Fragen beantworten müssen. Ist das nicht möglich, weil zum Beispiel die Gesamtzahl befragter Experten dafür zu klein ist oder für eine sinnvolle Bearbeitung der Fragen der vollständige Fragebogen den Probanden bekannt sein muss, könnte mit alternierenden Fragepositionen gearbeitet werden. Dadurch würden unterschiedliche Fragen oder Fragegruppen bei jeder einzelnen Befragung am Ende positioniert sein, sodass sich die Missingquote gleichmäßiger über alle Fragen verteilen würde. Dass keine dieser Möglichkeiten im Rahmen der vorliegenden Befragung eingesetzt wurde, ist bezüglich der Aussagekraft der Ergebnisse problematisch. Zum einen liegt so nur über circa die Hälfte der Stichprobe eine Information vor, zu welcher Berufsgruppe sie gehören und Verzerrung durch Selbstselektion können nicht sicher ausgeschlossen werden. Zum anderen stellt sich die Frage, ob alle wesentlichen Verbesserungsvorschläge für die am Ende platzierten Items gemacht werden konnten. Dies muss bei der Ergebnisinterpretation berücksichtigt werden, da weniger Anmerkungen seitens der Experten nicht zwangsläufig eine bessere Itemqualität implizieren, sondern durch frühzeitigen Abbruch und Ermüdung begünstigt sein können. Andererseits stellt der Merkmalsbereich mit den meisten Freitextkommentaren und Ergänzungsvorschlägen die Arbeitsumgebung dar. Diese wurde an letzter Stelle im Expertenfragebogen positioniert. Insofern ist ein allzu hoher Informationsverlust durch den Verzicht auf alternierende Fragepositionen wahrscheinlich nicht gegeben.

Es wird davon ausgegangen, dass die Ergebnisse der Expertenbefragung auf andere Arbeitssettings übertragbar sind, da Experten aus dem gesamten Geschäftsbereich BMVG teilgenommen haben, welcher ein vielfältiges Spektrum an Tätigkeitsfeldern und Standorten umfasst. Dennoch kann nicht endgültig ausgeschlossen werden, dass die Expertenmeinungen von der Zugehörigkeit zu ihrem Arbeitgeber beeinflusst sind und ein vergleichbares Studiendesign mit einer anderen Stichprobe zu anderen Ergebnissen kommen würde.

Weitere Selektionseffekte konnten durch Auswahl anhand mehrerer objektiver Kriterien weitgehend ausgeschlossen werden. Eine grundsätzliche Verzerrung der Ergebnisse aufgrund der geringen Objektivität wegen fehlender Standardisierung ist unwahrscheinlich, da die Teilnehmendenzahl um ein Vielfaches höher liegt als bei Expertenbefragungen üblich und somit mögliche extreme Einzelmeinungen relativiert werden und die Subjektivität der Urteile die Verlässlichkeit des Gesamtergebnisses nicht beeinträchtigt. Dies trifft jedoch nur auf die Expertenbefragung zu. Für das Fragebogenreview wurden die Experten zwar auch anhand mehrerer objektiver Kriterien ausgewählt, aber zum einen ist die Anzahl der Experten hier deutlich geringer ($n = 7$) und zum anderen kann nicht ausgeschlossen werden, dass sie aufgrund der berufsbedingten Nähe zum AIGScreen und Vorerfahrungen mit dem Instrument in ihren Einschätzungen und Urteilen beeinflusst sind und die Art der Beeinflussung für alle beteiligten Experten gleich oder zumindest ähnlich ist. Dieses Risiko wurde bewusst in Kauf

genommen, da der Vorteil andererseits gerade in diesem vertieften Wissen und den umfangreichen Erfahrungen zur Entwicklung, des Anwendungskontextes und der Datenauswertung des AIGScreen bestand. Nur so konnte der Empfehlung der europäischen Statistikbehörden zur Fragebogenentwicklung gefolgt werden, wonach die Experten nicht nur einzelne Items bewerten sollten, sondern diese Bewertung auch unter Berücksichtigung des Gesamtkontextes des Erhebungs- und Forschungsdesigns durchzuführen sei (70).

Eine weitere Einschränkung der Ergebnisinterpretation aus der Expertenbefragung ergibt sich aus dem Vorgehen bei der Abstrahierung und Quantifizierung der Freitexte. Die InterCoderübereinstimmung wurde lediglich als relativer Anteil der Codes, die übereinstimmend für eine Textpassage gegeben wurden, im Verhältnis zur Gesamtzahl der vergebenen Codes bemessen. Diese Maßzahl kann zu einer Überschätzung der InterCoderreliabilität führen (230), da hier keine Korrekturformel für zufällige Übereinstimmungen angewandt wird wie es z. B. beim alternativen Maß Kappa der Fall ist. Letzteres hat jedoch den Nachteil, dass es nur angewandt werden kann, wenn für Textpassagen genau ein Code verwendet werden kann (143). Da dies beim vorliegenden Datenmaterial nicht der Fall war und eine Segmentierung in eindeutige Abschnitte häufig auch die im Textmaterial enthaltenen Informationen nicht vollständig abgebildet hätte, was wiederum die Verlässlichkeit der Datenbasis negativ beeinflusst hätte (231), konnte Kappa in diesem Fall nicht als Maß der InterCoderübereinstimmung verwendet werden und eine niedrigere Reliabilität der Ergebnisse kann nicht ausgeschlossen werden.

Bezüglich der quantitativen Analysen ergaben sich Einschränkungen vor allem aufgrund der Datenstruktur. Die Mehrheit der Items wies eine linksschiefe Verteilung auf und aufgrund der vierstufigen Antwortskala sind Tests auf Normalverteilung nur begrenzt aussagekräftig. Gerade bei gesundheitsbezogenen Daten und Fragestellungen kann eine Normalverteilung jedoch oft nicht erreicht werden bzw. würde der Natur des Untersuchungsgegenstandes widersprechen (28). Um eine Verzerrung der Ergebnisse im Fall nicht-normalverteilter Daten zu vermeiden, wird die Anwendung nicht-parametrischer Tests oder polychorischer Verfahren von einigen Autoren empfohlen (232, 233). Da jedoch umfassend nachgewiesen ist, dass auch grobe Verletzungen der Normalverteilungsannahme insbesondere bei großen Stichproben kaum die Verlässlichkeit der Schätzer beeinflussen (213, 234), wurden parametrische Verfahren eingesetzt. Diese sind nicht-parametrischen Tests aufgrund ihrer geringen Teststärke wo immer möglich vorzuziehen (233). Die Ergebnisse der Faktorenanalyse konnten in einer polychorischen Analyse weitgehend repliziert werden, sodass von unverzerrten Ergebnissen ausgegangen werden kann. Eine umfassende Replikation der Analysen in nicht-parametrischen Verfahren wurde jedoch nicht durchgeführt und es existieren auch nicht für alle parametrischen Analysen nicht-parametrische Alternativen.

Korrelationen zwischen den Items verschiedener Skalen wirkten sich negativ auf die Ergebnisse der Faktorenanalyse und die Trennschärfe aus. Wie bereits in Abschnitt 5.3 erläutert, ist es für heterogene Konstrukte wie den AIGScreen allerdings plausibel, dass Merkmalsbereiche nicht vollkommen trennscharf gemessen werden können, da Zusammenhänge zwischen den einzelnen Aspekten theoretisch begründbar und zu erwarten sind. Insofern müssen Trennschärfekoeffizienten für solche heterogenen latenten Konstrukte nicht immer hoch sein und können nur bedingt zur Beurteilung der Item- und Skalenqualität herangezogen werden (71). Auch eine vollkommene Einfachstruktur im Rahmen der Faktorenanalyse ist bei der Erfassung eines heterogenen Konstrukts mit möglichst wenig Items kaum zu erwarten. Dies ist jedoch insofern problematisch, als dass eine fehlende Einfachstruktur und geringe interne Konsistenz grundsätzlich als problematisch hinsichtlich Reliabilität und Validität zu interpretieren sind (63, 78, 30, 176). Zudem tendieren die Modellfit-Maße, wie in Abschnitt 5.3.2 erläutert, bedingt durch vergleichsweise niedrige Primärladungen und viele Querladungen dazu, auch korrekt spezifizierte Modelle abzulehnen. Denkbar ist daher, dass die wissenschaftliche Güte des AIGScreen unterschätzt wird und bestimmte Anpassungen im Fragebogen, die durch statistische Analysen impliziert werden, negative Auswirkungen auf die Inhaltsvalidität haben könnten, da die Modifikationen theoretisch weniger plausibel sind.

Um die Ergebnisse der Datenanalyse gesichert auf die gesamte Zielgruppe übertragen zu können, muss die Stichprobe repräsentativ sein. Die Repräsentativität der Stichprobe konnte jedoch nur anhand weniger Merkmale, die bezüglich der Grundgesamtheit zur Verfügung standen, überprüft werden. Hierfür wurden das Alter, der Beschäftigungsstatus und das Geschlecht berücksichtigt. Für keines der Merkmale konnte bestätigt werden, dass die Stichprobe hinsichtlich der Verteilung dieser Merkmale für die Grundgesamtheit repräsentativ ist. Allerdings ist fraglich, inwiefern die fehlende Repräsentativität bezüglich dieser drei Merkmale einen Einfluss auf die Ergebnisinterpretation hat. Zum einen ist nicht festzustellen, inwieweit diese drei Parameter einen Einfluss auf die hier betrachteten Verteilungen haben. Zum anderen wäre selbst im Falle einer spezifischen Repräsentativität für diese drei Merkmale immer noch fraglich, ob die Stichprobe nicht bezüglich anderer, für die Einschätzung der Arbeitssituation relevanter Merkmale verzerrt ist. Beispielsweise kann argumentiert werden, dass ein positives Gesundheitsverhalten oder hohes Interesse am Thema BGM, die Teilnahmebereitschaft erhöhen. Somit wären jedoch vermehrt Personen in der Stichprobe, die potenziell über mehr Ressourcen verfügen und somit wahrscheinlich weniger Belastungen bzw. Beanspruchungen erleben. Die fehlende Repräsentativität bezüglich bestimmter soziodemographischer Merkmale muss insofern nicht problematisch für die Übertragung der Ergebnisse auf die gesamte Zielgruppe sein. Aber zukünftige Erhebungen sollten berücksichtigen, welche Merkmale die interessierenden Variablen beeinflussen können und

diese hinsichtlich einer repräsentativen Verteilung für die Grundgesamtheit überprüfen. Darüber hinaus sollte die Überprüfung der wissenschaftlichen Güte des AIGScreen auch mittels Stichproben aus anderen Arbeitssettings vorgenommen werden. Zwar wurde argumentiert, dass die Ergebnisse auch auf andere Kontexte übertragbar sind, aber der empirische Nachweis dieser Behauptung steht noch aus.

6 Zusammenfassung und Ausblick

Ziel dieser Arbeit war es, die wissenschaftliche Güte des AIGScreen zu beurteilen und daraus abzuleiten, inwiefern er als Screeninginstrument zur ressourceneffizienten Beurteilung von psychischen Gefährdungen am Arbeitsplatz eingesetzt werden kann. Da dieser Einsatz derzeit primär im Rahmen Betrieblichen Gesundheitsmanagements (BGM) erfolgt, sollten zudem insbesondere die Nebengütekriterien Testökonomie, Nützlichkeit, Vergleichbarkeit und Zumutbarkeit Berücksichtigung finden. Die Erhebung der notwendigen Daten erfolgte im Rahmen eines Forschungsprojekts zur Einführung von BGM im Geschäftsbereich des BMVg.

Zur vollständigen Beantwortung der Fragestellung war ein mixed methods-Ansatz notwendig, also die Kombination von qualitativen und quantitativen empirischen Methoden. Um zunächst die Eignung des Fragebogens für den beschriebenen Anwendungszweck und die Übereinstimmung mit allgemein anerkannten Qualitätskriterien für ein wissenschaftliches Testinstrument zu überprüfen, wurde ein mehrstufiges Expertenreview durchgeführt. Zunächst waren die Experten aufgefordert allein anhand standardisierter Checklisten mögliche Probleme beim AIGScreen zu identifizieren. In einer anschließenden Diskussionsrunde wurden diese Probleme mit allen Experten besprochen und ein Konsens darüber gefunden, wie dem Problem am besten begegnet werden sollte. Inwiefern der Fragebogen aus Sicht der Zielgruppe das Kriterium der Zumutbarkeit und der Nützlichkeit erfüllt und somit als allgemein akzeptiert bzw. augenscheinvalide gelten kann, wurden im Rahmen einer teilstandardisierten Online-Expertenbefragung ermittelt, deren primäres Ziel die Beurteilung der Inhaltsvalidität war. Aus den Umfrageergebnissen ließen sich Indikatoren zur Beurteilung der Relevanz einzelner Items und ganzer Itemskalen berechnen. Freitextantworten wurden im Rahmen einer qualitativen Inhaltsanalyse ausgewertet. Hieraus konnten insbesondere Informationen gewonnen werden, welche Aspekte zusätzlich erhoben werden sollten, um die Augenscheinvalidität und Akzeptanz bei der Zielgruppe zu erhöhen. Itemanalysen, explorative und konfirmatorische Faktorenanalysen dienten der Überprüfung von Reliabilität und Konstruktvalidität. Die Daten dafür wurden unter realen Feldbedingungen in der Zielgruppe erhoben. Fehlende Angaben wurden durch multiple Imputation ersetzt.

Das Fragebogenreview konnte die Objektivität des AIGScreen bestätigen, identifizierte jedoch einige Probleme bezüglich der Fragebogenqualität, für die jedoch praxistaugliche Lösungen gefunden werden konnten. Neben einzelnen Formulierungen, die aus Sicht der Experten die Reliabilität und Validität der Erhebung gefährden könnten, wurde insbesondere die verwendete Antwortskala kritisch bewertet. Die bisherige für alle Items verwendete vierstufige Likert-Skala sollte nach Experteneinschätzung durch eine mindestens fünfstufige Antwortskala ersetzt werden, deren Verbalisierung itemspezifisch sein sollte. Umfangreiche Forschungsergebnisse zur optimalen Gestaltung von Antwortskalen bestätigen, dass somit

eine Verbesserung der wissenschaftlichen Güte erzielt werden kann und dieses Format auch von den Probanden besser akzeptiert wird.

Die Inhaltsvalidität des AIGScreen konnte im Rahmen der Expertenbefragung bestätigt werden. Die Expertenbefragung diente zudem dem Zweck, zu eruieren, inwiefern eine mit den Empfehlungen der Deutschen Gemeinsamen Arbeitsschutzstrategie (GDA) zur Beurteilung psychischer Gefährdungen am Arbeitsplatz konforme inhaltliche Ausgestaltung des Fragebogens auch auf Akzeptanz bei der Zielgruppe trifft. Das Kriterium der Augenscheinvalidität stellt zwar kein wissenschaftliches Gütekriterium dar, ist aber dennoch insbesondere bei Erhebungen, die auf hohe Rücklaufquoten angewiesen sind von hoher Relevanz. Vergleicht man die Einschätzungen der Experten aus der Zielgruppe mit den Vorgaben der Leitlinien der GDA lässt sich insgesamt eine hohe Übereinstimmung feststellen. Bezüglich einzelner Aspekte wie z. B. Abwechslungsreichtum einer Tätigkeit oder Sinnhaftigkeit der eigenen Arbeit bestanden jedoch auch erhebliche Diskrepanzen, was die Bewertung der Relevanz seitens der Experten auf der einen Seite und den offiziellen Empfehlungen auch mit Blick auf den aktuellen Forschungsstand auf der anderen Seite betrifft. Hier konnte für einzelne Aspekte nicht abschließend geklärt werden, inwiefern diese Unterschiede auf settingspezifische Problemlagen oder eine mangelhafte Itemqualität zurückzuführen sind. Um die Akzeptanz des Fragebogens zu erhöhen, sollte jedoch unter Abwägung ökonomischer Kriterien grundsätzlich geprüft werden, ob eine Erweiterung des Instruments um die von den Experten vorgeschlagenen Aspekte zu ergänzen sind oder ob sie eventuell bereits an anderer Stelle erhoben werden. Bezüglich verschiedener Aspekte der Arbeitssicherheit in anderen Merkmalsbereichen wie z. B. Gefahrstoffen oder physikalischen Gefährdungen erschien eine Erweiterung des AIGScreen nicht sinnvoll, da sie einerseits die Probanden im Rahmen eines Grobscreenings zu psychischen Gefährdungen überfordern würden und andererseits auch dem Kriterium der Fragebogenökonomie widersprechen, da diese Informationen bereits an anderer Stelle (z. B. im Rahmen von Begehungen oder bei der Gefährdungsbeurteilung durch die Sicherheitsfachkraft) erhoben werden. Für andere Aspekte wurden in der überarbeiteten Fragebogenversion zusätzliche Items vorgeschlagen, auch wenn diese nicht Gegenstand der GDA-Empfehlungen sind. Damit soll auch die Bewertung der Relevanz auf Skalenebene gesteigert werden. Hier können die Ergebnisse aus der Expertenbefragung in Bezug auf die vier postulierten Hauptdimensionen noch nicht zufriedenstellen.

Vergleicht man die Ergebnisse der Reliabilitäts- und Konstruktvaliditätsanalysen, mit denen der Expertenbefragung in Bezug auf die vier Hauptdimensionen des AIGScreen muss vor allem festgestellt werden, dass der angewandte Methodenmix einen erheblichen Erkenntnisbeitrag geliefert hat. Obwohl die Inhaltsvalidität als gegeben betrachtet werden kann, zeigen die quantitativen Analysen einige Schwachstellen hinsichtlich Reliabilität und

Validität auf und die Ergebnisse der qualitativen Erhebungen konnten diese Mängel nur unzureichend identifizieren. Die größte Diskrepanz hinsichtlich Expertenbefragung und den Ergebnissen der statistischen Analyse konnte für die vierte Dimension, Arbeitsumgebung, festgestellt werden. Während diese seitens der Experten sehr positiv hinsichtlich Relevanz und inhaltlicher Vollständigkeit bewertet wurde, sprachen die Ergebnisse der Itemanalyse gegen eine ausreichende interne Konsistenz und Einfachstruktur dieser Skala. Allerdings muss bei der Ergebnisinterpretation auch berücksichtigt werden, dass die Skala nur vier Items enthielt, was sich negativ auf die Ergebnisse der verschiedenen statistischen Analysen auswirken kann. Die Skalen Arbeitsinhalt und Arbeitsaufgabe sowie Arbeitsorganisation können hingegen hinsichtlich ihrer internen Konsistenz als reliabel bewertet werden. Insbesondere für letztere Dimension lagen jedoch vergleichsweise schlechte Bewertungen aus der Expertenbefragung in Bezug auf Vollständigkeit und Relevanz vor. Die Skala Soziale Aspekte weist bezüglich Reliabilität und Inhaltsvalidität die besten Ergebnisse auf. Diese Skala kann als konsistent, inhaltlich vollständig und in Bezug auf die Zielgruppe als augenscheinvalide betrachtet werden. Um letzterem noch besser gerecht zu werden, wurde im überarbeiteten Fragebogenentwurf vorgeschlagen, den Fragebogen um die Aspekte Kontaktintensität, Konfliktbewältigung und Führung zu erweitern.

Die zunächst postulierte vierdimensionale Struktur des AIGScreen konnte im Rahmen der explorativen Faktorenanalyse nicht bestätigt werden. Mit Blick auf die Ladungsstruktur, inhaltliche Plausibilität und Modelparsamkeit wurde für die konfirmatorische Faktorenanalyse ein achtdimensionales Konstrukt ausgewählt, welches neben dem Ein-Faktor- und dem Vier-Faktoren-Modell hinsichtlich seiner Konstruktvalidität überprüft werden sollte. Obwohl acht Faktoren die Datenstruktur im Hinblick auf die Gütemaße besser abbildeten, als es im Modell mit vier oder einem Faktor der Fall war, konnte auch diese Lösung nicht als vollständig zufriedenstellend bewertet werden. Ein Grund für den schlechten Modellfit aller Modelle könnte im Fehlen einer Einfachstruktur und zum Teil niedrigen Faktorladungen liegen, da dies die χ^2 -basierten Modellfit-Statistiken negativ verzerrt. Multivariate Nicht-Normalverteilung könnte ebenfalls zu schlechteren Modellfit-Werten beigetragen haben. Letzteres könnte durch die Anpassung der Antwortskalen und die damit erwartete Erhöhung der Itemqualität verbessert werden. Aber auch dann deuten Studienergebnisse darauf hin, dass der Nachweis von Konstruktvalidität für heterogene latente Konstrukte trotz korrekter Modellspezifikation häufig scheitert, da die Modellfitmaße bei der konfirmatorischen Faktorenanalyse bei geringen Primärladungen und Verletzung der Einfachstruktur zur grundsätzlichen Ablehnung des Modells tendieren. Dieser Befund ist für zukünftige Validitätsanalysen des AIGScreen ernüchternd, da die einzige Möglichkeit zum Erreichen einer homogenen Skalenaufteilung darin bestünde, das Kriterium der Testökonomie zu verwerfen und die Anzahl der Items stark zu erhöhen.

Die im Rahmen dieser Arbeit gewonnen Erkenntnisse betreffen jedoch nicht nur die wissenschaftliche Güte eines bestimmten Fragebogens, sondern lassen auch Schlüsse für zukünftige Forschungsdesigns zu arbeitsbedingten Gefährdungen zu. So ist ein Ergebnis die offensichtliche Stärke der Mischung aus quantitativen und qualitativen Verfahren. Einerseits konnten so vertiefte Einblicke in die Bedarfe der Zielgruppe gewonnen werden und Verbesserungsmöglichkeiten entwickelt werden, die im Rahmen vollstandardisierter Erhebungen nicht hätten identifiziert werden können. Andererseits ist es mit Blick auf die Nebengütekriterien von Vorteil keine rein datengeleiteten Schlüsse zu ziehen, sondern diese auch sachlogisch und bezüglich ihrer Plausibilität von Fachexperten überprüfen zu lassen. Umgekehrt erlauben statistische Analysen zweifelsfrei objektive Bewertungen und gleichen somit die häufig formulierte Kritik aus, wonach Expertenurteile immer subjektiv seien und unverzerrte Ergebnisse entsprechend kaum möglich seien. Darüber hinaus stand die Frage im Raum, wer im Rahmen von Bedarfsanalysen als Experte gelten kann und inwiefern allgemeine Empfehlungen seitens der zuständigen Institutionen und der Wissenschaft sich mit der Wahrnehmung der betroffenen Personengruppen decken. Die vorliegende Erhebung hat gezeigt, dass Experten für den Gesundheitsschutz innerhalb eines bestimmten betrieblichen Settings verlässliche Informationsquellen für die Bedarfsanalyse hinsichtlich arbeitsbedingter Gefährdungen der individuellen Gesundheit sein können. Sie verbinden allgemeingültiges theoretisches Wissen mit praktischer Erfahrung und können frühzeitig neue Entwicklungen identifizieren. Sie verfügen außerdem über Kenntnisse der settingspezifischen Belastungen und können somit sinnvolle Ergänzungsvorschläge zum Erhebungsgegenstand machen. Dies gilt umso mehr, wenn auf eine interdisziplinäre Expertengruppe gesetzt wird, da hier nicht nur Anregungen zu wichtigen zusätzlichen Aspekten, sondern auch zu alternativen Informationsquellen innerhalb des Betriebes gegeben werden können, was den Ressourcenaufwand bei Bedarfsanalysen reduziert und sich positiv auf die Testökonomie auswirken kann. Allerdings hat sich auch gezeigt, dass die Einschätzung einzelner Experten nicht immer dem aktuellen Forschungsstand entspricht, sondern von persönlichen Meinungen und Erfahrungen am eigenen Arbeitsplatz bzw. dem individuellen beruflichen Hintergrund beeinflusst sein kann, weshalb eine Berücksichtigung der offiziellen Empfehlungen und dem aktuellen Forschungsstand auch in Zukunft bei der inhaltlichen Ausgestaltung unverzichtbar ist. Diese sollten auch in Zukunft als „Goldstandard“ bei der Gefährdungsanalyse gelten. Eine Ergänzung oder Modifizierung von Fragebogeninhalten durch die Experten aus der Zielgruppe ist an den Stellen, an denen es nicht konträr zu aktuellen wissenschaftlichen Erkenntnissen steht, jedoch sinnvoll und kann zu einer höheren Akzeptanz seitens der Zielgruppe sowie verlässlicheren Informationen beitragen, welche bedarfsangemessene und somit effiziente Interventionen ermöglichen. So kann eine breite Beteiligung verschiedener Akteure, welche die Zielgruppe möglichst gut repräsentieren, die Augenscheinvalidität erhöhen. Einerseits

kann es betriebsspezifische Themen geben, welche nicht im Rahmen der offiziellen Empfehlungen und Leitlinien genannt sind, aber deren Berücksichtigung für die Betroffenen ein großes Bedürfnis darstellt. In der vorliegenden Stichprobe waren das beispielsweise die Themen Hygiene und Sinnhaftigkeit der eigenen Tätigkeit. Je mehr die Beschäftigten das Gefühl haben, dass die Analyse die für sie relevanten Themen adressiert, desto höher ist die Wahrscheinlichkeit einer hohen Mitwirkungsbereitschaft und Beteiligung am Grobscreening und desto aussagekräftiger sind die damit gewonnen Daten. Gleichzeitig zeigte sich, dass die Einbeziehung alltäglicher Erfahrungen aus dem betrieblichen Kontext dazu führen kann, dass weitere wichtige Aspekte identifiziert werden, die ohne die Hinzuziehung von Experten unberücksichtigt blieben. So stellt beispielsweise Ergonomie sowohl in den offiziellen Empfehlungen als auch in der Einschätzung der befragten Experten einen wichtigen Aspekt dar. Was bisher in den Veröffentlichungen von GDA und BAuA jedoch nicht thematisiert wird, wird gleich mehrfach seitens der Experten gefordert: Die ergonomische Arbeitsgestaltung sollte auch darauf überprüft werden, ob sie adäquat für Menschen mit Behinderungen ist. Zudem kann die Beteiligung vielfältiger Experten aus den verschiedenen Bereichen, die sich mit der Sicherheit und Gesundheit bei der Arbeit befassen, dazu beitragen Redundanzen in der Datenerhebung und unnötige Mehrarbeit bei der Beurteilung von Arbeitsbedingungen zu vermeiden. So zeigte sich beispielsweise, dass zwar Experten aus der Gruppe der BGM-Beauftragten und BGM-Koordinatoren zurecht darauf hinweisen, dass auch sicherheitstechnische Aspekte und Gefährdungsfaktoren, welche nicht primär psychischer Art sind, aber eben auch auf die individuelle Psyche Einfluss nehmen können, im Rahmen eines Grobscreenings berücksichtigt werden sollten. Hingegen wurden diese Aspekte seitens der Gruppe der Arbeitssicherheitsfachkräfte nicht als zu erheben genannt. Hierbei handelt es sich um Gefährdungsfaktoren, welche bereits im Rahmen von Begehungen und der Gefährdungsbeurteilung physischer, chemischer, biologischer, mechanischer und sonstiger Faktoren erfasst werden. Im Sinne der Ressourceneffizienz und Fragebogenökonomie ist es daher angebracht, solche Aspekte zumindest nicht in hoher Detailtiefe zu erfassen.

Häufig wurden mehrere Gefährdungen in Kombination genannt. Zukünftige Gefährdungsanalysen sollten daher nicht ausschließlich das Vorliegen einer einzelnen Gefährdung, sondern vielmehr die Interdependenz verschiedener Faktoren berücksichtigen, welche einerseits erst in Kombination als besonders belastend wahrgenommen werden (z. B. hoher Abstimmungsbedarf bei gleichzeitig hoher Arbeitsintensität), aber andererseits auch negative Folgen neutralisieren können (z. B. stellt Autonomie eine Ressource dar, die dazu führt, dass hohe Arbeitsanforderungen als weniger belastend empfunden werden (187)). Dies sollte auch im Messmodell berücksichtigt werden, da es sich mit den empirischen Befunden dieser Arbeit deckt, wonach eine Einfachstruktur einzelner Gefährnungsdimensionen nicht nachweisbar ist. Ebenso scheint eine vollständige Trennung zwischen psychischen und

anderen Gefährdungen in der Analyse nicht möglich, da diese Bereiche interagieren bzw. sich überschneiden. Hier müssen Experten vor Ort einbezogen werden, um zu wissen, welche Informationen bereits an anderer Stelle erfasst werden und wo Redundanzen in der Datenerhebung oder aber sinnvolle Ergänzungsmöglichkeiten bestehen und wie verschieden Informationsquellen kombiniert ausgewertet werden können.

Der im Rahmen dieser Arbeit weiterzuentwickelnde und zu validierende AIGScreen, ein Screeninginstrument zur Messung der Arbeitssituation und des individuellen Gesundheitsempfindens, soll zu einer effizienten Ressourcenverwendung im Rahmen der arbeitsmedizinischen Intervention durch eine gezielte Bedarfsanalyse beitragen und somit auch dabei helfen, negative Effekte des Fachkräftemangels abzuschwächen. Erhebungsgegenstand des Fragebogens ist das subjektive Erleben von Arbeit und Gesundheit von Beschäftigten, woraus sich Befunde zur psychischen Belastungs- und Beanspruchungssituation im Arbeitskontext ableiten lassen. Für die Erstellung von Instrumenten, welche zur Gefährdungsbeurteilung psychischer Belastung eingesetzt werden, hat die GDA Qualitätsgrundsätze formuliert. Hierzu zählen die Eignung für die geplanten Einsatzbereich, die Beschreibung von Anwendungsvoraussetzungen, die methodische Qualität, die Erfassung aller Tätigkeiten und Belastungsfaktoren und die Partizipation der Beschäftigten. Zudem soll das Verfahren eine Beurteilung erlauben, ob Maßnahmen zur Reduktion von Gefährdungen notwendig sind. Diese Bedingungen werden vom AIGScreen erfüllt. So konnte die bisherige Datenerhebung aufzeigen, dass sich das Erhebungsdesign gut für verschiedenste Einsatzbereiche eignet. Die methodische Qualität war Hauptuntersuchungsgegenstand dieser Arbeit und kann insgesamt als gut bewertet werden, wobei sich durch die Revision des Fragebogens weitere Verbesserungen bei der wissenschaftlichen Güte erwartet werden. Der AIGScreen konnte nachweislich in allen Tätigkeitsbereichen eingesetzt werden, was auch eine Expertenbefragung in der Zielgruppe bestätigte. Die Fragebogenrevision diente dazu die bereits gute inhaltliche Vollständigkeit und Augenscheinvalidität aus Sicht der Zielgruppe weiter zu verbessern. Die Partizipation der Beschäftigten ist durch das Erhebungsdesign unmittelbar gegeben und anhand der Ergebnisse können Interventionsschwerpunkte identifiziert werden. Somit eignet sich der AIGScreen bereits jetzt als Instrument der Bedarfsanalyse und Maßnahmenplanung in Bezug auf arbeitsbedingte psychische Gefährdungen. Allerdings müssen bei der Bildung von Kennzahlen in Form von Summenindizes aus mehreren Items sowie bei Zeit- und Gruppenvergleichen die festgestellten Mängel hinsichtlich der wissenschaftlichen Güte berücksichtigt werden.

Aufgrund der festgestellten Modifikationsbedarfe wurde eine Revision des Fragebogens sowohl hinsichtlich des Inhalts als auch bezüglich Aspekten des Fragebogendesigns, insbesondere bezüglich der Antwortskalen erarbeitet. Mit der in Kapitel 5.5.1 vorgeschlagenen

Neufassung des Fragebogens, enthält der AIGScreen mit insgesamt 50 Items deutlich mehr als zuvor. Allerdings kann der Core-Fragebogen mit 41 Items im Vergleich zu anderen Screeninginstrumenten und aufgrund der geschlossenen Antwortformate immer noch als praxistauglich und ressourceneffizient bewertet werden. Mit Blick auf den Anwendungszweck des Fragebogens war zudem klar, dass die Erfüllung der wissenschaftlichen Gütekriterien dem Ziel der Befragungsökonomie überzuordnen war. Die Erweiterungen ergaben sich überwiegend aus den Ergebnissen der Expertenbefragung in der Zielgruppe. Sie dienen somit der Verbesserung der Inhaltsvalidität, aber auch der Augenscheinvalidität und Akzeptanz bei den Befragten. Damit können sie zu einer höheren Teilnahmebereitschaft bzw. einer höheren Rücklaufquote beitragen, was sich wiederum positiv auf die Verlässlichkeit und Repräsentativität der damit erhobenen Daten auswirkt und es ermöglicht, dass strategische Entscheidungen im Rahmen des BGM auf einer wissenschaftlich validierten Datenbasis getroffen werden können.

Da mit dem AIGScreen heterogene latente Konstrukte auf Basis subjektiver Einschätzungen mit möglichst wenig Items erhoben werden, lässt sich auch in Zukunft eine gewisse Unschärfe bzw. Unsicherheit bezüglich der Verlässlichkeit der Messung nicht ausschließen. Allerdings gibt es für diesen spezifischen Erhebungsgegenstand keine bessere Lösung und das Grobscreening hat sich bereits jetzt hinsichtlich vieler Kriterien als vorteilhaft erwiesen. So wird es dem Kriterium der Ökonomie gerecht, wobei die Online-Gefährdungsbeurteilung einen besonders ressourceneffizienten Ansatz darstellt, mit dem der Arbeitgeber seiner gesetzlichen Pflicht nachkommen kann. Gleichzeitig begegnet dieser Lösungsansatz dem bestehenden Fachkräftemangel im Bereich der Arbeitsmedizin und den sich daraus ergebenden Betreuungsdefiziten. Er ermöglicht, wie von der BAuA vorgeschlagen, so ressourceneffizient wie möglich einen Überblick über etwaige Interventionsschwerpunkte im Betrieb zu gewinnen und entlastet Sicherheitsfachkräfte und Betriebsärzte.

AIGScreen hat sich bereits jetzt als praktikables Instrument erwiesen, dass vielen Herausforderungen der Beurteilung psychischer Belastungen gerecht wird. Auch der Neuentwurf schafft eine gute Balance zwischen Fragebogenökonomie und den Anforderungen wissenschaftlicher Gütekriterien. So ist die Beteiligung der Beschäftigten sowie die für die Beurteilung psychischer Belastungen notwendige Erfassung von deren subjektiven Einschätzungen und Empfindungen mit dem AIGScreen möglich. Gleichzeitig kann das Instrument aufgrund der unterschiedlichen Erhebungsmodi niedrigschwellig in verschiedenen Tätigkeitsbereichen eingesetzt werden. Die Items sind so formuliert, dass verschiedene Berufsgruppen mit demselben Instrument befragt werden können, was einen Vergleich verschiedener Stichproben und Erhebungen möglich macht. Dies bietet sowohl Vorteile für die strategische Planung des Betrieblichen Gesundheitsmanagements als auch für den wissenschaftlichen Erkenntnisgewinn. Schließlich wurde eingangs erläutert, dass viele

Arbeitgeber Ihrer Pflicht zur Beurteilung psychischer Belastung aufgrund von Unwissenheit oder Unsicherheit bezüglich der Thematik nicht nachkommen können oder wollen. Der Einsatz eines Grobscreenings, wie dem hier präsentierten Fragebogen, ist ohne tiefere Kompetenzen möglich und verschafft jedem Arbeitgeber ohne größeren Personal- oder Zeitaufwand einen Überblick über die Gefährdungssituation und etwaige Interventionsschwerpunkte.

Da es keine allgemeinen Belastungsgrenzen bezüglich psychischer Belastung gibt, ist die durch den AIGScreen ermöglichte Quantifizierung und die sich daraus ergebende Vergleichbarkeit von Befragungsergebnissen ein hilfreicher Indikator für Arbeitgeber, wo Handlungsschwerpunkte in ihrem Betrieb zu identifizieren sind. Die inhaltliche Validität des Fragebogens wurde bereits durch eine Vielzahl von Experten bestätigt. Die Konsistenz und Eindimensionalität der Skalen im Neuentwurf des Fragebogens ist erneut zu überprüfen, um zu gewährleisten, dass daraus abgeleitete Kennzahlen verlässliche Schlüsse bezüglich der individuellen Arbeitssituation und des Gesundheitsempfindens ermöglichen. Die Fragebogenentwicklung und -validierung ist ein iterativer Prozess und die Wiederholung bestimmter Prüfschritte spricht nicht gegen das Instrument, sondern vielmehr für dessen kontinuierliche Pflege. Dass bisher die Kriterien der Reliabilität und Validität nicht vollständig erfüllt werden, spricht nicht grundsätzlich gegen die Auswertung und Interpretation von Daten, die mit dem AIGScreen erhoben wurden. Erstens tragen diese Erhebungen und daraus gewonnene Erkenntnisse wesentlich zur Verbesserung des Instruments bei. Zweitens ist die Beschäftigung mit der psychometrischen Güte insofern hilfreich, als dass auf diesem Weg für die Forschenden und das Publikum Transparenz geschaffen werden kann und die Grenzen der Interpretationsmöglichkeiten aufgezeigt werden. Daten, die mit einem nicht validen Instrument erhoben wurden, sind nicht unbrauchbar, die Möglichkeiten der Datenauswertung und -interpretation sind lediglich geringer. Aber es ist entscheidend zu wissen, ob ein Instrument valide misst, da die Anwender andernfalls Gefahr laufen, unzulässige Schlüsse zu ziehen. Die mit dem Instrument gewonnenen Ergebnisse können bereits jetzt als hilfreiche Orientierung bei der Arbeitsgestaltung und Maßnahmenplanung dienen. Der AIGScreen entlastet dabei Arbeitgeber und Arbeitsmediziner und leistet einen wichtigen Beitrag zur Erhaltung und Förderung der Gesundheit von Beschäftigten.

Literaturverzeichnis

1. Frank Faulbaum, Peter Prüfer, Margit Rexroth. Was ist eine gute Frage? Wiesbaden: VS Verlag für Sozialwissenschaften; 2009.
2. Arbeitsschutzgesetz vom 7. August 1996 (BGBl. I S. 1246), das zuletzt durch Artikel 12 des Gesetzes vom 22. November 2021 (BGBl. I S. 4906) geändert worden ist: ArbSchG [cited 2022 Feb 6]. Available from: URL: <https://www.gesetze-im-internet.de/arbschg/BJNR124610996.html#BJNR124610996BJNG000200000>.
3. Chapman LS. Meta-evaluation of worksite health promotion economic return studies: 2012 update. Am J Health Promot 2012; 26(4).
4. Goetzel RZ, Ozminkowski RJ. The health and cost benefits of work site health-promotion programs. Annu Rev Public Health 2008; 29:303–23.
5. Lück P, Eberle G, Bonitz D. Der Nutzen des betrieblichen Gesundheitsmanagements aus der Sicht von Unternehmen. In: Badura B, Bienert ML, editors. Betriebliches Gesundheitsmanagement: Kosten und Nutzen: Zahlen, Daten, Analysen aus allen Branchen der Wirtschaft. Heidelberg: Springer; 2009. p. 77–84 (Fehlzeiten-Report; vol. 2008).
6. Kramer I, Sockoll I, Bödeker W. Die Evidenzbasis für betriebliche Gesundheitsförderung und Prävention – Eine Synopse des wissenschaftlichen Kenntnisstandes. In: Badura B, Bienert ML, editors. Betriebliches Gesundheitsmanagement: Kosten und Nutzen: Zahlen, Daten, Analysen aus allen Branchen der Wirtschaft. Heidelberg: Springer; 2009. p. 65–76 (Fehlzeiten-Report; vol. 2008).
7. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Volkswirtschaftliche Kosten durch Arbeitsunfähigkeit 2018; 2020 [cited 2021 Apr 13]. Available from: URL: https://www.baua.de/DE/Themen/Arbeitswelt-und-Arbeitsschutz-im-Wandel/Arbeitsweltberichterstattung/Kosten-der-AU/pdf/Kosten-2018.pdf?__blob=publicationFile&v=3.
8. Thomas G. Grobe, Susanne Steinmann, Julia Gerr. Gesundheitsreport 2018. Siegburg: Müller Verlagsservice e. K; 2018. (Schriftenreihe zur Gesundheitsanalyse Band 9) [cited 2019 Jan 8]. Available from: URL: <https://www.barmer.de/blob/155284/c2ac6f9716e416c0b0d889a9a91ce9d8/data/dl-gesundheitsreport-bund.pdf>.
9. Techniker Krankenkasse. Gesundheitsreport 2018 Arbeitsunfähigkeiten [cited 2019 Jan 8]. Available from: URL: <https://www.tk.de/resource/blob/2034000/60cd049c105d066650f9867da5b4d7c1/gesundheit-sreport-au-2018-data.pdf>.

10. Jörg Marschall, Susanne Hildebrandt, Karsten Zich, Thorsten Tisch, Jelena Sörensen, Hans-Dieter Nolting. DAK-Gesundheitsreport 2018: DAK-Gesundheit; 2018 [cited 2019 Jan 8]. Available from: URL: <https://www.dak.de/dak/download/gesundheitsreport-2018-1970840.pdf>.
11. Baicker K, Cutler D, Song Z. Workplace wellness programs can generate savings. *Health Aff (Millwood)* 2010 [cited 2019 Feb 12]; 29(2):304–11.
12. Hochart C, Lang M. Impact of a comprehensive worksite wellness program on health risk, utilization, and health care costs. *Popul Health Manag* 2011 [cited 2019 Feb 12]; 14(3):111–6.
13. Treier M. Gefährdungsbeurteilung psychischer Belastungen: Begründung, Instrumente, Umsetzung. Wiesbaden: Springer; 2015. (Essentials).
14. European Agency for Safety and Health at Work. Expert forecast on emerging psychosocial risks related to occupational safety and health. Luxembourg: Office for Official Publ. of the Europ. Communities; 2007. (European risk observatory report; vol 5).
15. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. *J Occup Health Psychol* 2008; 13(1):69–93.
16. European Agency for Safety and Health at Work. Future occupational safety and health research needs and priorities in the member states of the European Union. Luxembourg: Off. for Off. Publ. of the Europ. Communities; 2000. (Research / European Agency for Safety and Health at Work).
17. Barth C, Hamacher W, Eickholt C. Arbeitsmedizinischer Betreuungsbedarf in Deutschland. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin; 2014 [cited 2017 Nov 7]. Available from: URL: https://www.baua.de/DE/Angebote/Publikationen/Berichte/F2326.pdf?__blob=publicationFile&v=5.
18. Dudenhöfer S, Schöne K, Letzel S, Rose D-M, editors. Gefährdungsbeurteilung in der Schule: Ergebnisse zur individuellen Arbeitssituation und dem Gesundheitsempfinden der Lehrkräfte.; 2013.
19. K. Schöne, T. Kurtz, S. Engler, K. Oehler, D.-M. Rose, S. Letzel. Onlinegefährdungsbeurteilung – Ein praktikabler Einstieg in ein bedarfsgerechtes Risiko- und Gesundheitsmanagement an staatlichen Schulen in Rheinland-Pfalz. *ASU Arbeitsmedizin Sozialmedizin Umweltmedizin - Zeitschrift für medizinische Prävention* 2017 [cited 2018 Jan 30]; (52):446–51. Available from: URL: <https://www.asu-arbeitsmedizin.com/article-767655->

30010/onlinegefaehrungsbeurteilung-ein-praktikabler-einstieg-in-ein-bedarfsgerechtes-risiko-und-gesundheitsmanagement-an-staatlichen-schulen-in-rheinland-pfalz-.html.

20. Schaarschmidt U. AVEM - ein persönlichkeitsdiagnostisches Instrument für die berufsbezogene Rehabilitation. In: Arbeitskreis Klinische Psychologie in der Rehabilitation BDP, editor. Psychologische Diagnostik - Weichenstellung für den Reha-Verlauf. Bonn: Deutscher Psychologen Verlag GmbH; 2006. p. 59–82.

21. Kieschke U. Arbeits-Bewertungs-Check für Lehrkräfte (ABC-L) – Ein Instrument für schulische Gestaltungsmaßnahmen. In: Verband für Bildung und Erziehung, editor. Fit für den Lehrerberuf – Pädagogische Profession im 21. Jahrhundert. Berlin: Wilke; 2007. p. 103–9.

22. Nübling M, Vomstein M, Haug A, Nübling T, Stöbel U, Hasselhorn HM, Hofmann F, Neuner R, Wirtz M, Krause A. Personenbezogene Gefährdungsbeurteilung an öffentlichen Schulen in Baden-Württemberg - Erhebung psychosozialer Faktoren bei der Arbeit. Freiburg; 2012.

23. Matthias Nübling, Ulrich Stöbel, Hans-Martin Hasselhorn, Martina Michaelis, Friedrich Hofmann. Measuring psychological stress and strain at work: Evaluation of the COPSOQ Questionnaire in Germany. GMS Psycho-Social-Medicine 2006 [cited 2019 Jan 18]; (3). Available from: URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2736502/pdf/PSM-03-05.pdf>.

24. Hoerger M. Participant dropout as a function of survey length in internet-mediated university studies: implications for study design and voluntary participation in psychological research. Cyberpsychol Behav Soc Netw 2010; 13(6):697–700.

25. Vicente P, Reis E. Using Questionnaire Design to Fight Nonresponse Bias in Web Surveys. Social Science Computer Review 2010; 28(2):251–67.

26. Sammito S, Schlattman A, Felfe J, Renner K-H, Kowalski J, Stein M et al. Betriebliches Gesundheitsmanagement im Geschäftsbereich des Bundesministeriums der Verteidigung - Wissenschaftliche Begleitung eines ehrgeizigen Projekts. Wehrmedizinische Monatszeitschrift 2015; 8(59):230–5.

27. Struhs-Wehr K. Betriebliches Gesundheitsmanagement und Führung: Gesundheitsorientierte Führung als Erfolgsfaktor im BGM; 2017.

28. Leonhart R. Einführung in die Testkonstruktion [cited 2017 Nov 6]. Available from: URL: http://www.reinhardt-verlag.de/pdf/Leonhart_Einf%C3%BChrung.pdf.

29. Moosbrugger H, Kelava A. Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In: Moosbrugger H, Kelava A, editors. Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 8–26.
30. Moosbrugger H, Kelava A, editors. Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012.
31. David Beck et al. Empfehlungen zur Umsetzung der Gefährdungsbeurteilung psychischer Belastung; 2017 [cited 2018 Sep 14]. Available from: URL: https://www.gda-psyche.de/SharedDocs/Publikationen/DE/broschuere-empfehlung-gefaehrungsbeurteilung.pdf;jsessionid=042C9EB305D37CBE38146D986E9F39DA?__blob=publicationFile&v=14.
32. Züll C, Menold N. Offene Fragen. In: Baur N, editor. Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS; 2014. p. 713–20 (Handbuch). Available from: URL: https://link.springer.com/content/pdf/10.1007%2F978-3-531-18939-0_52.pdf.
33. Sedlmeier P, Renkewitz F, editors. Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler. 2., aktualisierte und erweiterte Auflage. München: Pearson; 2013. (PS Psychologie).
34. Krebs D, Menold N. Gütekriterien quantitativer Sozialforschung. In: Baur N, editor. Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS; 2014. p. 425–38 (Handbuch).
35. Rohmert W. Das Belastungs-Beanspruchungs-Konzept. Zeitschrift für Arbeitswissenschaft; 1984(38):193–200.
36. Blättner B. Das Modell der Salutogenese. Präz Gesundheitsf 2007; 2(2):67–73.
37. Halbe-Haenschke B, Reck-Hog U. Die Erfolgsstrategie für Ihr BGM: Methoden und Umsetzung eines effektiven betrieblichen Gesundheitsmanagements. Wiesbaden: Springer Gabler; 2017.
38. Ruckenberg J. Gesundheit, Arbeit und Zusammenhalt. Herbolzheim: Centaurus Verlag & Media; 2013. (vol 45).
39. Schaper N. Wirkungen der Arbeit. In: Nerdinger FW, Blickle G, Schaper N, editors. Arbeits- und Organisationspsychologie: Mit 51 Tabellen. 3., vollst. überarb. Aufl. Berlin, Heidelberg: Springer; 2014. p. 517–38 (Springer-Lehrbuch).
40. Neuner R. Psychische Gesundheit bei der Arbeit: Betriebliches Gesundheitsmanagement und Gefährdungsbeurteilung psychischer Belastung. 2., überarb. Aufl. 2016; 2016.

41. Kauffeld S. Arbeits-, Organisations- und Personalpsychologie für Bachelor: Mit 34 Tabellen ; [lesen, hören, lernen im Web]. Berlin u.a.: Springer; 2011. (Springer-Lehrbuch).
42. Antonovsky A. Salutogenese: Zur Entmystifizierung der Gesundheit. Tübingen: dgvt-Verlag; 1997.
43. Faltermaier T. Salutogenese; 2020. Available from: URL:
<https://leitbegriffe.bzga.de/alphabetisches-verzeichnis/salutogenese/>.
44. baua. Aktuell - Ausgabe 2/2012. Amtliche Mitteilungen der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Dortmund; 2012 [cited 2018 Feb 8]. Available from: URL:
<https://www.baua.de/DE/Angebote/Publikationen/Aktuell/2-2012.html>.
45. Oppolzer R. Psychische Belastungen in der Arbeitswelt. 3.th ed. Mannheim; 2009.
46. Kirchner J-H. Arbeitswissenschaft Entwicklung eines Grundkonzeptes. Zeitschrift für Arbeitswissenschaft 1993; (57):85–92.
47. Metz A-M, Rothe H-J. Screening psychischer Arbeitsbelastung. Wiesbaden: Springer Fachmedien Wiesbaden; 2017.
48. Hurrelmann K, Franzkowiak P. Gesundheit; 2018. Available from: URL:
<https://leitbegriffe.bzga.de/alphabetisches-verzeichnis/gesundheit/>.
49. Weltgesundheitsorganisation. Ottawa Charter for Health Promotion; 1986 [cited 2022 Feb 1]. Available from: URL:
https://www.euro.who.int/__data/assets/pdf_file/0006/129534/Ottawa_Charter_G.pdf.
50. Gemeinsame Deutsche Arbeitsschutzstrategie. Leitlinie Beratung und Überwachung bei psychischer Belastung am Arbeitsplatz. Berlin; 2018.
51. Gemeinsame Deutsche Arbeitsschutzstrategie. Träger der GDA; 2022 [cited 2022 Feb 7]. Available from: URL: https://www.gda-portal.de/DE/GDA/Traeger/Traeger_node.html.
52. Eichhorn D, Schuller K. Gefährdungsbeurteilung psychischer Belastung – Reine Pflichterfüllung oder Nutzen für die Betriebe? sis 2017 [cited 2022 Feb 6]; (10):428–33. Available from: URL:
<https://www.baua.de/DE/Angebote/Publikationen/Aufsaetze/artikel1366.html>.
53. Beck D, Richter G, Ertel M, Morschhäuser M. Gefährdungsbeurteilung bei psychischen Belastungen in Deutschland. Präz Gesundheitsf 2012 [cited 2022 Feb 6]; 7(2):115–9. Available from: URL:
https://www.baua.de/DE/Angebote/Publikationen/Aufsaetze/artikel28.pdf?__blob=publicationFile&v=2.

54. Kaempf, S., Krause, A. Gefährdungsbeurteilung zur Analyse psychischer Belastungen am Arbeitsplatz Schule. In: Bungard W, Koop B, Liebig C, editors. Psychologie und Wirtschaft leben: Aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis ; [Dritte Tagung der Fachgruppe für Arbeits- und Organisationspsychologie der Deutschen Gesellschaft für Psychologie im September 2003 in Mannheim]. 1. Aufl. München: Hampp; 2004. p. 314–9.
55. Kristensen, Tage S, Harald Hannerz, Annie Høgh, and Vilhelm Borg. The Copenhagen Psychosocial Questionnaire-a Tool for the Assessment and Improvement of the Psychosocial Work Environment. *Scandinavian Journal of Work, Environment & Health* 2005 [cited 2018 Oct 16]; 31(6):438–49. Available from: URL: <https://www.jstor.org/stable/40967527>.
56. Liebl A, Kittel M. Psychische Gesundheit in der Arbeitswelt. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA).
57. Beck D, Schuller K. Gefährdungsbeurteilung psychischer Belastung in der betrieblichen Praxis. baua: Bericht kompakt 2020 [cited 2022 Feb 6]. Available from: URL: <https://www.baua.de/DE/Angebote/Publikationen/Bericht-kompakt/F2358.html>.
58. Beck D, Berger S, Breutmann N, Fergen A, Gregersen S, Morschhäuser M et al. Empfehlungen zur Umsetzung der Gefährdungsbeurteilung psychischer Belastung; 2018 [cited 2022 Feb 7]. Available from: URL: <https://www.gda-psyche.de/SharedDocs/Publikationen/DE/broschuere-empfehlung-gefaehrungsbeurteilung.html;jsessionid=AA1B24BCE1B4872B29ECEDC42074CC77>.
59. Reichertz J. Empirische Sozialforschung und soziologische Theorie. In: Baur N, editor. *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS; 2014. p. 65–80 (Handbuch).
60. Döring N, Bortz J. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5. vollständig überarbeitete, aktualisierte und erweiterte Auflage. Berlin, Heidelberg: Springer; 2016. (Springer-Lehrbuch).
61. Bortz J, Schuster C. *Statistik für Human- und Sozialwissenschaftler*. 7., vollständig überarbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg; 2010. (Springer-Lehrbuch).
62. Baur N, Blasius J. Methoden der empirischen Sozialforschung: Ein Überblick. In: Baur N, editor. *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS; 2014. p. 41–64 (Handbuch).

63. Rammstedt B. Zur Bestimmung der Güte von Multi-Item-Skalen: Eine Einführung. ZUMA How-to-Reihe 2004; 12.
64. Bühner M. Einführung in die Test- und Fragebogenkonstruktion. 2., aktualisierte und erw. Aufl. München u.a.: Pearson Studium; 2006. (PS Psychologie).
65. Breyer B, Bluemke M, Danner D. ZIS Publication Guide: Leitfaden für die Veröffentlichung von Erhebungsinstrumenten in der Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS).; 2016.
66. Peter JP. Construct Validity: A Review of Basic Issues and Marketing Practices. Journal of Marketing Research 1981; 18(2):133.
67. Karin Schermelleh-Engel, Helfried Moosbrugger. Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. Methods of Psychological Research Online 2003 [cited 2018 Oct 9]; (8):23–74. Available from: URL: <http://www.dgps.de/fachgruppen/methoden/mpr-online/>.
68. Lawshe CH. A quantitative approach to content validity. Personnel Psychology 1975; (28):563–75.
69. Hartig J, Frey A, Jude N. Validität. In: Moosbrugger H, Kelava A, editors. Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 143–72.
70. G. Brancato, S. Macchia, M. Murgia, M. Signore, G. Simeoni, K. Blanke, T. Körner, A. Nimmergut, P. Lima, R. Paulino, J.H.P. Hoffmeyer-Zlotnik. Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System; 2006.
71. Schmidt-Atzert L, Amelang M, Fydrich T. Psychologische Diagnostik: Mit 82 Tabellen. 5., vollständig überarbeitete und erweiterte Auflage. Berlin: Springer; 2012. (Springer-Lehrbuch).
72. Coltman T, Devinney TM, Midgley DF, Venaik F. Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research 2008 [cited 2018 Apr 4]; 61(12):1250–62.
73. Meinold T. Thielsch, Timo Lenzner & Torsten Melles. Wie gestalte ich gute Items und Interviewfragen? In: Thielsch, Mainold T. & Brandenburg, Torsten, editor. Praxis der Wirtschaftspsychologie II: Praxis und Fallbeispiele für Studium und Praxis. Münster: MV Wissenschaft; 2012. p. 221–40 [cited 2018 May 30]. Available from: URL: http://www.thielsch.org/download/wirtschaftspsychologie/Thielsch_2012b.pdf.
74. Kelle U, Langfeldt B, Metje B. Qualitätssicherung von Einstellungsskalen mit Hilfe qualitativer Methoden und von „Mixed-Methods-Designs“ – die Messung religiöser Überzeugungen. In: p. 225–60.

75. Häder M. Empirische Sozialforschung. Wiesbaden: Springer Fachmedien Wiesbaden; 2019.
76. Taylor CS. Validity and Validation. Oxford: Oxford University Press USA; 2013. (Understanding Statistics).
77. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007; 22(11):1596–602.
78. Edwards JR. The Fallacy of Formative Measurement. *Organizational Research Methods* 2011; 14(2):370–88.
79. Christophersen T, Grape C. Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle. In: p. 103–18.
80. Jarvis CB, MacKenzie SB, Podsakoff PM. A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *J CONSUM RES* 2003; 30(2):199–218.
81. G. Brancato, S. Macchia, M. Murgia, M. Signore, G. Simeoni, K. Blanke, T. Körner, A. Nimmergut, P. Lima, R. Paulino, J.H.P. Hoffmeyer-Zlotnik. Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System; 2006.
82. Richter G. Toolbox Version 1.2: Instrumente zur Erfassung psychischer Belastungen; Forschung Projekt F 1965. 2., unveränd. Aufl. Dortmund, Berlin, Dresden: Baua; 2011.
83. Kristensen, Tage S, Harald Hannerz, Annie Høgh, and Vilhelm Borg. The Copenhagen Psychosocial Questionnaire-a Tool for the Assessment and Improvement of the Psychosocial Work Environment. *Scandinavian Journal of Work, Environment & Health* 2005 [cited 2018 Oct 16]; 31(6):438–49. Available from: URL: <https://www.jstor.org/stable/40967527>.
84. Semmer, N. K., Zapf, D., & Dunckel, H. Assessing stress at work: A framework and an instrument. In: Dunckel H, editor. Handbuch psychologischer Arbeitsanalyseverfahren. 2., [omarb] uppl. Zürich: Vdf-Hochschulverl.; 1999. p. 105–13 (Mensch, Technik, Organisation; vol. 14).
85. Morgeson FP, Humphrey SE. The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. *J Appl Psychol* 2006; 91(6):1321–39.
86. Jochen Prümper, Klaus Hartmannsgruber, Michael Frese. KFZA - Kurz-Fragebogen zur Arbeitsanalyse. In: Sarges W, Wottawa H, Roos C, editors. Handbuch

wirtschaftspsychologischer Testverfahren: Band II: Organisationspsychologische Instrumente. 1. Auflage. Lengerich: Pabst Science Publishers; 2010. p. 157–64.

87. David Beck et al. Empfehlungen zur Umsetzung der Gefährdungsbeurteilung psychischer Belastung; 2017 [cited 2018 Sep 14]. Available from: URL: https://www.gda-psyche.de/SharedDocs/Publikationen/DE/broschuere-empfehlung-gefaehrungsbeurteilung.pdf;jsessionid=042C9EB305D37CBE38146D986E9F39DA?__blob=publicationFile&v=14.

88. Kaempfer, S., Krause, A. Gefährdungsbeurteilung zur Analyse psychischer Belastungen am Arbeitsplatz Schule. In: Bungard W, Koop B, Liebig C, editors. Psychologie und Wirtschaft leben: Aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis ; [Dritte Tagung der Fachgruppe für Arbeits- und Organisationspsychologie der Deutschen Gesellschaft für Psychologie im September 2003 in Mannheim]. 1. Aufl. München: Hampp; 2004. p. 314–9.

89. Nickel S, Kersten M. Psychometrische Prüfung des Fragebogens zur psychischen Belastung in der stationären Krankenpflege („miab“). Hamburg; 2014 [cited 2022 Mar 17]. Available from: URL: <https://www.bgw-online.de/resource/blob/22130/7d747db5629a9909761bcbf48790d2ef/bgw08-00-111-bericht-miab-stat-krankenpflege-download-data.pdf>.

90. Schöne Klaus, Kurtz Tanja, Engler Sarah, Oehler Kathri, Rose Dirk-Matthias, Letzel Stephan. Onlinegefährdungsbeurteilung — Ein praktikabler Einstieg in ein bedarfsgerechtes Risiko- und Gesundheitsmanagement an staatlichen Schulen in Rheinland-Pfalz. ASU 2017; 52(06):446–51.

91. Nübling M. Erprobung eines Messinstrumentes (COPSOQ) ; [Abschlussbericht zum Projekt "Methoden zur Erfassung psychischer Belastungen - Erprobung eines Messinstrumentes (COPSOQ)" - Projekt F 1885. Bremerhaven: Wirtschaftsverl. NW Verl. für Neue Wiss; 2005. (Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Forschung; vol 1058) [cited 2019 Jan 18].

92. DIN Deutsches Institut für Normung e. V. Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung - Teil 1: Allgemeine Aspekte und Konzepte und Begriffe: Beuth Verlag GmbH; 2017 2017. Available from: URL: <https://www.din.de/de/meta/suche/62730!search?query=DIN+EN+ISO+10075&submit-btn=Submit> [cited 2022 Jul 8].

93. Roßbach B, Löffler KI, Mayer-Popken O, Konietzko J, Dupuis H. Belastungs- und Beanspruchungskonzept. In: Letzel S, Nowak D, Konietzko J, Dupuis H, editor. Handbuch

- der Arbeitsmedizin: Arbeitsphysiologie, Arbeitspsychologie, klinische Arbeitsmedizin, Prävention und Gesundheitsförderung. 3rd ed. Landsberg: ecomed Medizin; 2007 (A II-1).
94. Joiko K, Schmauder M, Wolff G. Psychische Belastung und Beanspruchung im Berufsleben: Erkennen - gestalten. 5. Aufl. Dortmund: Baua; 2010.
95. Wöhrmann AM. Psychische Gesundheit in der Arbeitswelt. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA).
96. Bradtke E, Melzer M. Psychische Gesundheit in der Arbeitswelt: Vollständigkeit. Dortmund/Berlin/Dresden; 2016 [cited 2020 Oct 19]. Available from: URL: <https://www.baua.de/DE/Angebote/Publikationen/Berichte/F2353-1c.html>.
97. Rothe I, Adolph L, Beermann B, Schütte M, Windel A, Grever A et al. Psychische Gesundheit in der Arbeitswelt; 2017.
98. Adolph L, Beermann B, Brenscheidt F, Otto J, Barth C, Grzech-Sukalo H et al. Ratgeber zur Gefährdungsbeurteilung Handbuch für Arbeitsschutzfachleute. 3. Auflage. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin; 2016.
99. Paridon H. Gefährdungsbeurteilung psychischer Belastungen: Tipps zum Einstieg. [Stand]: Juni 2013. Berlin: Dt. Gesetzliche Unfallversicherung; 2013. (IAG-Report; vol 2013,1).
100. Rainer Schnell, Paul B. Hill, Elke Esser. Methoden der empirischen Sozialforschung. 7. völlig überarbeitete und erweiterte Auflage. München: Oldenbourg Wissenschaftsverlag GmbH; 2005.
101. Krosnick JA. Questionnaire Design. In: Vannette DL, Krosnick JA, editors. The Palgrave Handbook of Survey Research. Cham: Springer International Publishing; 2018. p. 439–55.
102. Porst R. Fragebogen. Wiesbaden: Springer Fachmedien Wiesbaden; 2014.
103. Wolter F, Junkermann J. Antwortvalidität in Survey-Interviews: Meinungsäußerungen zu fiktiven Dingen. In: Menold N, Wolbring T, editors. Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente. Wiesbaden: Springer Fachmedien Wiesbaden; 2019. p. 339–68.
104. Thielsch, Mainald T. & Brandenburg, Torsten, editor. Praxis der Wirtschaftspsychologie II: Praxis und Fallbeispiele für Studium und Praxis. Münster: MV Wissenschaft; 2012.
105. Martin E. Survey questionnaire construction. Survey methodology 2006; (13).
106. Thielsch MT, Lenzner T, Melles T. Wie gestalte ich gute Items und Interviewfragen? In: Brandenburg T, Thielsch MT, editors. Praxis der Wirtschaftspsychologie: Themen und

Fallbeispiele für Studium und Anwendung. Münster: Monsenstein und Vannerdat; 2012. 221-240 (MV Wissenschaft).

107. Converse JM, Presser S. Survey questions: Handcrafting the standardized questionnaire. Thousand Oaks; 1986. Quantitative Applications in the Social Sciences 63.

108. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl. Cognit. Psychol. 1991; 5(3):213–36.

109. Krosnick J, Presser S. Question and Questionnaire Design. In: Marsden PV, Wright JD, editors. Handbook of Survey Research. 2nd ed. Bingley, UK: Emerald Group Publishing; 2010. p. 263–313.

110. Weiss DJ, Shanteau J. Empirical Assessment of Expertise. Human Factors 2003; 45(1):104–14.

111. Mauksch S, Gracht HA von der, Gordon TJ. Who is an expert for foresight? A review of identification methods. Technological Forecasting and Social Change 2020; 154:119982.

112. Shanteau J, Weiss DJ, Thomas RP, Pounds J. How Can You Tell If Someone Is an Expert? Performance-Based Assessment of Expertise. In: Sandra L. Schneider, S. L., & Shanteau, J., editor. Emerging perspectives on decision research. Cambridge: Cambridge University Press; 2003. p. 620–40.

113. Sandra L. Schneider, S. L., & Shanteau, J., editor. Emerging perspectives on decision research. Cambridge: Cambridge University Press; 2003.

114. Bogner A, Littig B, Menz W. Interviews mit Experten: Eine praxisorientierte Einführung. Wiesbaden: Springer VS; 2014. (Qualitative Sozialforschung).

115. Renate Liebold, Rainer Trinczek. Experteninterview. In: Kühl S, editor. Handbuch Methoden der Organisationsforschung: Quantitative und qualitative Methoden. 1. Aufl. Wiesbaden: Verl. für Sozialwiss. / GWV Fachverl.; 2009. p. 32–56.

116. Olson K. An Examination of Questionnaire Evaluation by Expert Reviewers. Field Methods 2010; 22(4):295–318.

117. Klein G, Shneiderman B, Hoffman RR, Ford KM. Why Expertise Matters: A Response to the Challenges. IEEE Intell. Syst. 2017; 32(6):67–73.

118. Presser S, Blair J. Survey pretesting: Do different methods produce different results? Sociological methodology; (24):73–104. Available from: URL: https://www.jstor.org/stable/pdf/270979.pdf?casa_token=PQOj8dkGgzQAAAAA:9ebWHb4sJmpr1M3_VTx8gn7nSHkd1FuE8mOknqwrGTiOCizVuPNli02Gymj9i-RtTINGxkjvYIYPf0kWdjrCIS_Zvjci6f9Ei2RMcb7sgFdLKsNIUJdU.

119. Moosbrugger H, Kelava A. Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In: Moosbrugger H, Kelava A, editors. Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 8–26.
120. Menold N, Wolbring T, editors. Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente. Wiesbaden: Springer Fachmedien Wiesbaden; 2019. (Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute).
121. Kelle U, Langfeldt B, Metje B. Qualitätssicherung von Einstellungsskalen mit Hilfe qualitativer Methoden und von „Mixed-Methods-Designs“ – die Messung religiöser Überzeugungen. In: Menold N, Wolbring T, editors. Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente. Wiesbaden: Springer Fachmedien Wiesbaden; 2019. p. 225–60 (Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute).
122. Shanteau J, Weiss DJ, Thomas RP, Pounds JC. Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research* 2002; 136(2):253–63. Available from: URL: [https://doi.org/10.1016/S0377-2217\(01\)00113-8](https://doi.org/10.1016/S0377-2217(01)00113-8).
123. Leonhart R. Einführung in die Testkonstruktion [cited 2017 Nov 6]. Available from: URL: http://www.reinhardt-verlag.de/pdf/Leonhart_Einf%C3%BChrung.pdf.
124. Meinold T. Thielsch, Timo Lenzner & Torsten Melles. Wie gestalte ich gute Items und Interviewfragen? In: Thielsch, Mainold T. & Brandenburg, Torsten, editor. *Praxis der Wirtschaftspsychologie II: Praxis und Fallbeispiele für Studium und Praxis*. Münster: MV Wissenschaft; 2012. p. 221–40 [cited 2018 May 30]. Available from: URL: http://www.thielsch.org/download/wirtschaftspsychologie/Thielsch_2012b.pdf.
125. Natalja Menold, Kathrin Bogner. Gestaltung von Ratingskalen in Fragebögen: Version 1.0: GESIS – Leibniz-Institut für Sozialwissenschaften; 2015 [cited 2018 Jun 14]. Available from: URL: https://www.gesis.org/fileadmin/upload/SDMwiki/Archiv/Ratingskalen_MenoldBogner_012015_1.0.pdf.
126. Renate Liebold, Rainer Trinczek. Experteninterview. In: Kühl S, editor. *Handbuch Methoden der Organisationsforschung: Quantitative und qualitative Methoden*. 1. Aufl. Wiesbaden: Verl. für Sozialwiss. / GWV Fachverl.; 2009. p. 32–56.
127. Ernst S. Prozessorientierte Methoden in der Arbeits- und Organisationsforschung: Eine Einführung. Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden; 2010.

128. G. Brancato, S. Macchia, M. Murgia, M. Signore, G. Simeoni, K. Blanke, T. Körner, A. Nimmergut, P. Lima, R. Paulino, J.H.P. Hoffmeyer-Zlotnik. Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System; 2006.
129. Beck D, Lenhardt U. Betriebliche Gesundheitsförderung in Deutschland: Verbreitung und Inanspruchnahme. Ergebnisse der BIBB/BAuA-Erwerbstätigenbefragungen 2006 und 2012. Gesundheitswesen 2016; 78(1):56–62.
130. David Beck et al. Empfehlungen zur Umsetzung der Gefährdungsbeurteilung psychischer Belastung; 2017 [cited 2018 Sep 14]. Available from: URL: https://www.gda-psyche.de/SharedDocs/Publikationen/DE/broschuere-empfehlung-gefaehrdungsbeurteilung.pdf;jsessionid=042C9EB305D37CBE38146D986E9F39DA?__blob=publicationFile&v=14.
131. Lawshe CH. A quantitative approach to content validity. Personnel Psychology 1975; (28):563–75.
132. Gilbert GE, Prion S. Making Sense of Methods and Measurement: Lawshe's Content Validity Index. Clinical Simulation in Nursing 2016; 12(12):530–1.
133. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. West J Nurs Res 2003; 25(5):508–18.
134. Yusoff MSB. ABC of Content Validation and Content Validity Index Calculation. EIMJ 2019; 11(2):49–54.
135. Yaghmaei F. Content Validity And Its Estimation. JOURNAL OF MEDICAL EDUCATION 2003; 3(1):25–7.
136. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychologica 2000; (104):1–15.
137. Bogner K, Landrock U. Antworttendenzen in standardisierten Umfragen.
138. Ayre C, Scally AJ. Critical Values for Lawshe's Content Validity Ratio. Measurement and Evaluation in Counseling and Development 2014; 47(1):79–86.
139. Früh W. Inhaltsanalyse: Theorie und Praxis. 9. überarbeitete Auflage. Konstanz, München: UVK Verlagsgesellschaft mbH mit UVK/Lucius; 2017. (utb-studi-e-book; vol 2501).
140. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007; 30(4):459–67.

141. Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken. Neuausgabe. s.l.: Beltz Verlagsgruppe; 2010.
142. Züll C. Offene Fragen. Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften; 2015. GESIS Survey Guidelines.
143. Burla L, Knierim B, Barth J, Liewald K, Duetz M, Abel T. From text to codings: intercoder reliability assessment in qualitative content analysis. *Nurs Res* 2008; 57(2):113–7.
144. Mayring P, Fenzl T. Qualitative Inhaltsanalyse. In: Baur N, editor. *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS; 2014. p. 633–48 (Handbuch).
145. Lüdtke O, Robitzsch A, Trautwein U, Köller O. Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau* 2007; 58(2):103–17.
146. Rentzsch K, Schütz A. *Psychologische Diagnostik: Grundlagen und Anwendungsperspektiven*. 1. Auflage; 2009. (Grundriss der Psychologie).
147. Eye A von, Bogat AG. Testing the assumption of multivariate normality. *Psychology Science* 2004; (46):243–58.
148. Mardia KV. Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika* 1970 [cited 2018 Oct 9]; 57(3):519–30. Available from: URL: <https://www.jstor.org/stable/2334770>.
149. Yiğit S, Mendeş M. Usage of Multidimensional Scaling Technique for Evaluating Performances of Multivariate Normality Tests. *BJAST* 2016; 16(1):1–8.
150. Doornik JA, Hansen H. An Omnibus Test for Univariate and Multivariate Normality*. *Oxford Bulletin of Economics and Statistics* 2008; 70:927–39.
151. Mayer SJ. Hauptkomponentenanalyse und explorative Faktorenanalyse. In: Wagemann C, Goerres A, Siewert M, editors. *Handbuch Methoden der Politikwissenschaft*. Wiesbaden: Springer Fachmedien Wiesbaden; 2018. p. 1–26.
152. Cleff T. *Deskriptive Statistik und moderne Datenanalyse: Eine computergestützte Einführung mit Excel, PASW (SPSS) und STATA*. 2., überarb. und erw. Aufl.; 2011. (Lehrbuch).
153. StataCorp LP. *STATA User's Guide*. Release 13: Alpha. College Station, Texas: Stata Press; 2013 [cited 2022 Jan 19]. Available from: URL: <https://www.stata.com/manuals/mvalpha.pdf>.
154. Acock AC. *A gentle introduction to Stata*. 6th edition. College Station, Texas: Stata Press; 2018. (A Stata Press publication).

155. Aichholzer J. Einführung in lineare Strukturgleichungsmodelle mit Stata. Wiesbaden: Springer Fachmedien Wiesbaden; 2017.
156. Backhaus K, Erichson B, Plinke W, Weiber R. Multivariate Analysemethoden: Eine anwendungsorientierte Einführung ; [Extras im Web. 13., überarbeitete Auflage. Berlin, Dordrecht, London, New York: Springer; 2011. (Springer-Lehrbuch).
157. Zygmunt C, Smith MR. Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. TQMP 2014 [cited 2019 Feb 5]; 10(1):40–55.
158. Nassiri V, Lovik A, Molenberghs G, Verbeke G. On using multiple imputation for exploratory factor analysis of incomplete data. Behav Res Methods 2018 [cited 2019 May 15]; 50(2):501–17.
159. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res 2011; 20(1):40–9.
160. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009; 60:549–76.
161. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods 2002; 7(2):147–77.
162. Böwing-Schmalenbrock M, Jurczok A. Multiple Imputation in der Praxis: Ein sozialwissenschaftliches Anwendungsbeispiel: Universität Potsdam; 2011. Available from: URL: <http://opus.kobv.de/ubp/volltexte/2012/5811/>.
163. Medeiros R. Handling missing data in Stata: Imputation and likelihood-based approaches; 2016 [cited 2019 Jan 11]. Available from: URL: <https://www.stata.com/meeting/switzerland16/slides/medeiros-switzerland16.pdf>.
164. Royston P, White I. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. J. Stat. Soft. 2011; 45(4).
165. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. Am J Epidemiol 2010; 171(5):624–32.
166. Schwartz-Soicher O. Multiple Imputation of Missing Data Using Stata: Princeton University; 2017 [cited 2022 Sep 19]. Available from: URL: <https://dss.princeton.edu/training/MlStata.pdf>.
167. Carlin JB, Li N, Greenwood P, Coffey C. Tools for analyzing multiple imputed datasets. The Stata Journal 2003; 3(3):226–44.

168. Royston P. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. *The Stata Journal* 2009; 9(3):466–77.
169. Schermelleh-Engel K, Werner C, Moosbrugger H. Exploratorische Faktorenanalyse: Hauptachsenanalyse und Hauptkomponentenanalyse: SPSS-Beispiel zu Kapitel 13. In: Moosbrugger H, Kelava A, editors. *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer Medizin Verlag; 2007.
170. Yong AG, Pearce S. A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *TQMP* 2013; 9(2):79–94.
171. Williams B, Onsmann A, Brown T. Exploratory factor analysis: A five-step guide for novices. *Journal of Emergency Primary Health Care (JEPHC)* 2010; 8(3):1–13.
172. Moosbrugger H, Schermelleh-Engel K. Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In: Moosbrugger H, Kelava A, editors. *Testtheorie und Fragebogenkonstruktion*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 326–42.
173. Anna B. Costello, Jon W. Osborne. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. - *Practical Assessment, Research & Evaluation*. *Practical Assessment, Research & Evaluation* 2005 [cited 2019 Feb 5]; 10(7). Available from: URL: <https://pareonline.net/pdf/v10n7.pdf>.
174. Revelle W. psych: Procedures for Personality and Psychological Research; 2018 [cited 2019 Nov 4]. Available from: URL: <https://CRAN.R-project.org/package=psych>.
175. Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin* 1986; 99(3):432–42.
176. Karin Schermelleh-Engel, Helfried Moosbrugger. Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online* 2003 [cited 2018 Oct 9]; (8):23–74. Available from: URL: <http://www.dgps.de/fachgruppen/methoden/mpr-online/>.
177. Maeda H. Response option configuration of online administered Likert scales. *International Journal of Social Research Methodology* 2015; 18(1):15–26.
178. Krebs D, Menold N. Gütekriterien quantitativer Sozialforschung. In: Baur N, editor. *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS; 2014. p. 425–38 (Handbuch).
179. Karin Schermelleh-Engel, Helfried Moosbrugger. Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online* 2003 [cited 2018 Oct 9]; (8):23–74. Available from: URL: <http://www.dgps.de/fachgruppen/methoden/mpr-online/>.

180. Bentler PM, Bonett DG. Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychological Bulletin* 1980; 88(3):588–606.
181. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 1999; 6(1):1–55.
182. Satorra A, Bentler PM. Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis. In: Eye A von, editor. *Latent variables analysis: Applications for developmental research*. Thousand Oaks: Sage; 1994. p. 399–419.
183. Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 2001; 66(4):507–14.
184. StataCorp LLC. Satorra–Bentler adjustments [cited 2022 Sep 23]. Available from: URL: <https://www.stata.com/features/overview/sem-satorra-bentler/>.
185. Amlinger-Chatterjee M. *Psychische Gesundheit in der Arbeitswelt*. Berlin, Dresden; 2016.
186. *Gefährdungsfaktoren: Ein Ratgeber*. Dortmund; 2019 [cited 2020 Sep 9]. Available from: URL: <https://www.baua.de/DE/Themen/Arbeitsgestaltung-im-Betrieb/Gefahrungsbeurteilung/Expertenwissen/Expertenwissen.html?view=pdfViewExt>.
187. Drössler S, Steputat A, Schubert M, Euler U, Seidler A. *Psychische Gesundheit in der Arbeitswelt: Soziale Beziehungen*. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA).
188. Montano D, Reeske-Behrens A, Franke F. *Psychische Gesundheit in der Arbeitswelt*. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA); 2016 [cited 2020 Oct 19]. Available from: URL: https://www.baua.de/DE/Angebote/Publikationen/Berichte/F2353-2a.pdf?__blob=publicationFile&v=9.
189. Richter G, Henkel D, Rau R, Schütte M. Infoteil A: Beschreibung psychischer Belastungsfaktoren bei der Arbeit: Gefährdungsbeurteilung psychischer Belastung. Berlin: Erich Schmidt, Berlin; 2014.
190. *Gemeinsame Deutsche Arbeitsschutzstrategie. Leitlinie Beratung und Überwachung bei psychischer Belastung am Arbeitsplatz*. Berlin; 2018.
191. Kauffeld S, Hoppe, Diana. *Arbeit und Gesundheit*. In: Kauffeld S, editor. *Arbeits-, Organisations- und Personalpsychologie: Für Bachelor ; mit 34 Tabellen*. Berlin, Heidelberg: Springer; 2011. p. 223–44 (Springer-Lehrbuch).

192. Blickle G. Anforderungsanalyse. In: Nerdinger FW, Blickle G, Schaper N, editors. Arbeits- und Organisationspsychologie. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 207–22.
193. Rosen, Patricia Helen. Psychische Gesundheit in der Arbeitswelt. Berlin, Dresden; 2016.
194. P. H. Rosen. Tätigkeitsspielräume in Produktionsaufgaben — Arbeitswissenschaftliche Erkenntnisse und Gestaltungsoptionen. ASU Arbeitsmedizin Sozialmedizin Umweltmedizin - Zeitschrift für medizinische Prävention 2018 [cited 2020 Sep 8]; (Sonderheft 2018). Available from: URL: <https://www.asu-arbeitsmedizin.com/taetigkeitsspielraeume-produktionsaufgaben/taetigkeitsspielraeume-produktionsaufgaben>.
195. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Flexible Arbeitszeitmodelle - Überblick und Umsetzung. 2nd ed. Bönen: Druck & Verlag Kettler GmbH; 2019.
196. Leitlinie Gefährdungsbeurteilung und Dokumentation. Berlin; 2017 May 22 [cited 2020 Sep 8]. Available from: URL: https://www.gda-portal.de/DE/Downloads/pdf/Leitlinie-Gefaehrdungsbeurteilung.pdf?__blob=publicationFile&v=2.
197. Joiko K, Schmauder M, Wolff G. Psychische Belastung und Beanspruchung im Berufsleben: Erkennen - Gestalten. 5., Aufl. Dortmund; 2010.
198. Liebers F, Schust M. Gefährdungen durch physische Belastung [cited 2022 Nov 17]. Available from: URL: https://www.baua.de/DE/Themen/Arbeitsgestaltung-im-Betrieb/Gefaehrdungsbeurteilung/Expertenwissen/Physische-Belastung/Physische-Belastung_node.html.
199. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. MEGAPHYS - Mehrstufige Gefährdungsanalyse physischer Belastungen am Arbeitsplatz: Gemeinsamer Abschlussbericht der BAuA und der DGUV; 2019 Band 1.
200. Wiessmann, Fritz. Unfallrisiko Betriebsklima. In: Windemuth D, Jung D, Petermann O, editors. Praxishandbuch psychische Belastungen im Beruf: Vorbeugen - erkennen - handeln. Lizenzausg., 1. Aufl. Stuttgart: Gentner; 2010. p. 293–304.
201. Rigotti T, Otto K. Arbeitsplatzunsicherheit als Risikofaktor für Arbeitsunfälle. In: Windemuth D, Jung D, Petermann O, editors. Praxishandbuch psychische Belastungen im Beruf: Vorbeugen - erkennen - handeln. Lizenzausg., 1. Aufl. Stuttgart: Gentner; 2010. p. 228–36.
202. Haase E, Wegge J, Wach D, Mavius F. Führung als Belastungsfaktor für Mitarbeitergesundheit und Arbeitsunfälle. In: Windemuth D, Jung D, Petermann O, editors.

Praxishandbuch psychische Belastungen im Beruf: Vorbeugen - erkennen - handeln.
Lizenzausg., 1. Aufl. Stuttgart: Gentner; 2010. p. 182–98.

203. Cosmar M. Manipulation von Schutzeinrichtungen von Maschinen - Motivation und Prävention aus psychologischer Sicht. In: Windemuth D, Jung D, Petermann O, editors. Praxishandbuch psychische Belastungen im Beruf: Vorbeugen - erkennen - handeln. Lizenzausg., 1. Aufl. Stuttgart: Gentner; 2010. p. 211–27.

204. Arbeitsstättenverordnung vom 12. August 2004 (BGBl. I S. 2179), die zuletzt durch Artikel 4 des Gesetzes vom 22. Dezember 2020 (BGBl. I S. 3334) geändert worden ist: ArbStättV; 2004 [cited 2022 Nov 18]. Available from: URL: https://www.gesetze-im-internet.de/arbst_ttv_2004/BJNR217910004.html.

205. Eichhorst W, Buhlmann F. Die Zukunft der Arbeit und der Wandel der Arbeitswelt. IZA Standpunkte 2015 [cited 2020 Aug 21]; (77). Available from: URL: <http://hdl.handle.net/10419/121271>.

206. Sarges W, Wottawa H, Roos C, editors. Handbuch wirtschaftspsychologischer Testverfahren: Band II: Organisationspsychologische Instrumente. 1. Auflage. Lengerich: Pabst Science Publishers; 2010.

207. Kauffeld S, Ianiro, Patrizia M, Sauer Nils C. Führung. In: Kauffeld S, editor. Arbeits-, Organisations- und Personalpsychologie: Für Bachelor ; mit 34 Tabellen. Berlin, Heidelberg: Springer; 2011. p. 67–92 (Springer-Lehrbuch).

208. Nerdinger FW, Schaper N. Formen des Arbeitsverhaltens. In: Nerdinger FW, Blickle G, Schaper N, editors. Arbeits- und Organisationspsychologie. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 441–60.

209. Schaper N. Psychologie der Arbeitssicherheit. In: Nerdinger FW, Blickle G, Schaper N, editors. Arbeits- und Organisationspsychologie. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 489–516.

210. Beauducel A, Wittmann WW. Simulation Study on Fit Indexes in CFA Based on Data With Slightly Distorted Simple Structure. Structural Equation Modeling: A Multidisciplinary Journal 2005; 12(1):41–75.

211. Wilcox RR. Introduction to Robust Estimation and Hypothesis Testing. 3rd ed. St. Louis: Elsevier Science & Technology; 2013. (Statistical Modeling and Decision Science Ser).

212. Havlicek LL, Peterson NL. Robustness of the Pearson Correlation against Violations of Assumptions. Percept Mot Skills 1976; 43(3_suppl):1319–34.

213. Pearson ES. The Analysis of Variance in Cases of Non-Normal Variation. Biometrika 1931; 23(1/2):114–33.

214. Boneau CA. The effects of violations of assumptions underlying the test. *Psychological Bulletin* 1960; 57:49–64.
215. Carifio J, Perla R. Resolving the 50-year debate around using and misusing Likert scales. *Med Educ* 2008; 42(12):1150–2.
216. Rasch D, Guiard V. The robustness of parametric statistical methods. *Psychology Science* 2004; 46(2):175–208.
217. Gregoire TG, Driver BL. Analysis of ordinal data to detect population differences. *Psychological Bulletin* 1987; 101(1):159–65.
218. Cohen ME. Analysis of ordinal dental data: evaluation of conflicting recommendations. *J Dent Res* 2001; 80(1):309–13.
219. Li C-H. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods* 2016; 48(3):936–49.
220. Fayers P. Alphas, betas and skewy distributions: two ways of getting the wrong answer. *Adv Health Sci Educ Theory Pract* 2011; 16(3):291–6.
221. Revilla MA, Saris WE, Krosnick JA. Choosing the Number of Categories in Agree–Disagree Scales. *Sociological Methods & Research* 2014; 43(1):73–97.
222. Saris WE, Revilla M, Krosnick JA, Shaeffer EM. Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options. *Survey Research Methods* 2010; 4(1):61–79.
223. Krebs D. In der Mitte ist Platz für mehrere Meinungen: Vergleich von partiell- und vollverbalisierten Skalen mit unterschiedlicher Formulierung der Skalenmitte. In: Menold N, Wolbring T, editors. *Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente*. Wiesbaden: Springer Fachmedien Wiesbaden; 2019. p. 133–58 (Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute).
224. Rosebrock A, Schlosser S, Höhne JK, Kühnel S. Einflüsse unterschiedlicher Formen der Verbalisierung von Antwortskalen auf das Antwortverhalten von Befragungspersonen. In: Menold N, Wolbring T, editors. *Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente*. Wiesbaden: Springer Fachmedien Wiesbaden; 2019. p. 65–102 (Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute).
225. Dawes J. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International journal of market research* 2008; 50(1):61–104.

226. Badura B. Sozialkapital und Gesundheit. In: Badura B, editor. Arbeit und Gesundheit im 21. Jahrhundert. Berlin, Heidelberg: Springer Berlin Heidelberg; 2017. p. 37–70.
227. Grzech-Sukalo H, Hänecke K. Entwicklung einer Gefährdungsbeurteilung im Hinblick auf die Arbeitszeit. Berlin, Dresden; 2008.
228. Nerdinger FW, Blickle G, Schaper N, editors. Arbeits- und Organisationspsychologie: Mit 51 Tabellen. 3., vollst. überarb. Aufl. Berlin, Heidelberg: Springer; 2014. (Springer-Lehrbuch).
229. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA). Psychische Faktoren; 2022 [cited 2022 Nov 17]. Available from: URL: https://www.baua.de/DE/Themen/Arbeitsgestaltung-im-Betrieb/Gefaehrdungsbeurteilung/Expertenwissen/Psychische-Faktoren/Psychische-Faktoren_node.html.
230. Roland T Rust, Bruce Cooil. Reliability Measures for Qualitative Data: Theory and Implications. Journal of Marketing Research 1994; 31(1):1–14.
231. Kurasaki KS. Intercoder Reliability for Validating Conclusions Drawn from Open-Ended Interview Data. Field Methods 2000; 12(3):179–94.
232. Kolenikov, Stanislav, und Gustavo Angeles. The use of discrete data in PCA: Theory, simulations, and applications to socioeconomic indices [CPC/MEASURE Working paper.]: Carolina Population Center, University of North Carolina; 2004 [cited 2018 Sep 18]. Available from: URL: https://s3.amazonaws.com/academia.edu.documents/30773824/Kolenikov_jm.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1537274773&Signature=ntAXwyq9j6CnYo5g59O9H0MNtfQ%3D&response-content-disposition=inline%3B%20filename%3DThe_use_of_discrete_data_in_PCA_theory_s.pdf.
233. Schäfer T. Statistik II. Wiesbaden: VS Verlag für Sozialwissenschaften; 2011.
234. Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract 2010; 15(5):625–32.

Anlagen

Anlage 1: AIGScreen 2.0

Beurteilung Arbeitssituation, Gesundheitsempfinden und –verhalten

(Zielgruppe: Angehörige im Geschäftsbereich des Bundesministeriums der Verteidigung)

Nachfolgend finden Sie unter anderem Fragen zu Ihrer Arbeitssituation sowie zu Ihrem Gesundheitsempfinden. Ihre Antworten dienen uns dazu, Ihre Ressourcen, Belastungen und Beschwerden zu erfassen, um anschließend Maßnahmen abzuleiten. Klicken Sie bitte die Antwort an, die Ihrer Meinung am ehesten entspricht. Da es um Ihr persönliches Empfinden geht, gibt es keine richtigen oder falschen Antworten. Des Weiteren beinhaltet der Fragebogen Fragen zu Ihrem gesundheitlichen Zustand sowie Ihrem gesundheitlichen Verhalten, deren wahrheitsgemäße Beantwortung im Sinne einer verlässlichen Datenbasis und einer bedarfsgerechten Angebotsgestaltung von BGM wünschenswert ist.

Bitte beachten Sie die folgenden Hinweise, damit Ihre Angaben in die Auswertung eingehen können:

Wie wird der Fragebogen ausgefüllt?

Bitte füllen Sie den Fragebogen möglichst mit einem Kugelschreiber aus, indem Sie:

- ein Kreuz in die zutreffenden Kästchen setzen

Beispiel: Welches Geschlecht haben Sie?

☐ männlich ☐ weiblich

- oder die entsprechenden Zahlen in die weißen Felder eintragen.

Beispiel: Welches Gewicht haben Sie? kg

- Wenn Sie sich verschrieben haben, korrigieren Sie die Antworten bitte wie hier gezeigt:

☐	☒	☐	☐	☐	Markierung	→	☐	☒	☐	☐	☐	Korrektur
--------------------------	-------------------------------------	--------------------------	--------------------------	--------------------------	------------	---	--------------------------	-------------------------------------	--------------------------	--------------------------	--------------------------	-----------

- Handschriftliche Ergänzungen werden nur in den dafür vorgesehenen Freitextfeldern registriert.

Beispiel: sonstigen Beschwerden/Erkrankungen:

Den Fragebogen bitte NICHT falten oder knicken!!

Um Ihre Angaben für wissenschaftliche Auswertungen mit späteren Befragungen verknüpfen zu können und dabei gleichzeitig Ihre Anonymität zu wahren, bitten wir Sie, den „individuellen Code“ zu erstellen. Dieser Code ist nur Ihnen persönlich bekannt. Er wird lediglich benutzt, um die Daten aus der ersten Befragung mit den Daten aus späteren Befragungen zu verknüpfen.

Ihr individueller Code setzt sich aus folgenden Zeichen zusammen:							
Die ersten beiden Buchstaben Ihres Geburtsortes		Die ersten beiden Buchstaben des Vornamens Ihrer Mutter		Datum Ihres Geburtstages		Die ersten beiden Buchstaben des Vornamens Ihres Vaters	
Beispiel: B onn		Beispiel: B arbara		Beispiel: 07 .11.1991		Beispiel: W erner	
B	o	B	a	0	7	W	e

Bitte geben Sie hier Ihren individuellen Teilnehmercode ein (persönlicher Schlüssel)

--	--	--	--	--	--

Damit wir Ihre Fragebogenantworten Ihrem Standort/Ihrer Einheit zuordnen können, bitte wir Sie hier Ihre Organisations-ID anzugeben. Diese finden Sie auf dem Teilnahme-Informationsschreiben.

--

Sollte eine Aussage für Sie nicht zutreffen, kreuzen Sie bitte die Kategorie „unzutreffend“ an.

Bitte geben Sie an, inwiefern die folgenden Aussagen auf Ihr Ernährungs- und Bewegungsverhalten zutreffen.		nein	eher nein	eher ja	ja	unzutreffend
1	Ich esse normalerweise vielseitig und ausgewogen.	☐	☐	☐	☐	☐
2	Ich esse normalerweise jeden Tag Obst und Gemüse/Salat.	☐	☐	☐	☐	☐
3	Ich esse normalerweise mehrmals in der Woche Vollkornprodukte (z. B. Vollkornbrot, Vollkornbrötchen, Vollkornreis, Vollkornnudeln).	☐	☐	☐	☐	☐
4	Ich trinke normalerweise mindestens 1,5 l am Tag.	☐	☐	☐	☐	☐
5	Die körperliche Leistungsfähigkeit ist wichtig, damit ich die Anforderungen meines Dienstpostens erfüllen kann.	☐	☐	☐	☐	☐
6	Ich kann während meiner Dienst- und Arbeitszeit ausreichend Sport treiben.	☐	☐	☐	☐	☐
7	Ich versuche bewusst, in meinen Alltag Bewegungsaktivitäten einzubauen (z.B. möglichst Fahrrad statt Auto oder Treppe statt Aufzug).	☐	☐	☐	☐	☐
8	Die Anzahl der im Dienstsport/militärischen Fitnesstraining angebotenen Maßnahmen ist genau richtig.	☐	☐	☐	☐	☐
9	Die Qualität des Dienstsports/militärischen Fitnesstrainings ist genau richtig.	☐	☐	☐	☐	☐

Die folgenden Aussagen beziehen sich auf Ihre Arbeitsmerkmale.

Bitte schätzen Sie ein, inwiefern die folgenden Arbeitsmerkmale an Ihrem Hauptarbeitsplatz vorhanden sind.

nein eher
nein eher
ja unzutreff
end

Kategorie: Arbeitsmenge, zeitliche Aspekte, Arbeitspausen

10	Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit ausgewogen.	☐	☐	☐	☐	☐
11	Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen.	☐	☐	☐	☐	☐
12	Die Anzahl der von mir verlangten Überstunden ist angemessen.	☐	☐	☐	☐	☐
13	Meine Arbeitszeitverteilung ist mir angenehm.	☐	☐	☐	☐	☐
14	Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen.	☐	☐	☐	☐	☐
15	Meine Arbeitspausen sind ausreichend.	☐	☐	☐	☐	☐
16	Meine Arbeitspausen sind störungsfrei.	☐	☐	☐	☐	☐

Kategorie: Arbeitsanforderungen, Kontrolle, organisatorische Arbeitsbedingungen

17	Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet.	☐	☐	☐	☐	☐
18	Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit.	☐	☐	☐	☐	☐
19	Meine Arbeit ist in angenehmem Maße abwechslungsreich.	☐	☐	☐	☐	☐
20	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z. B. Personalverantwortung).	☐	☐	☐	☐	☐
21	Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material).	☐	☐	☐	☐	☐
22	Ich kann erkennen, ob ich meine Arbeit gut gemacht habe.	☐	☐	☐	☐	☐
23	Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit.	☐	☐	☐	☐	☐
24	Ich habe einen angemessenen Handlungsspielraum (z. B. Entscheidungsfreiheit) bei meiner Arbeit.	☐	☐	☐	☐	☐
25	Ich werde bei meiner Arbeit selten unterbrochen.	☐	☐	☐	☐	☐
26	Die nötigen Arbeitsmittel sind vorhanden.	☐	☐	☐	☐	☐
27	Arbeitsräume sind in ausreichender Anzahl vorhanden.	☐	☐	☐	☐	☐
28	Die Einrichtung an meinem Hauptarbeitsplatz ist angemessen.	☐	☐	☐	☐	☐

Kategorie: Soziale Aspekte der Arbeit

29	Mein Kundenkreis für interne Dienstleistungen (z. B. auch Patienten, Leistungsempfänger) verhält sich angemessen.	☐	☐	☐	☐	☐
----	---	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Die folgenden Aussagen beziehen sich auf Ihre Arbeitsmerkmale.

Bitte schätzen Sie ein, inwiefern die folgenden Arbeitsmerkmale an Ihrem Hauptarbeitsplatz vorhanden sind.

		nein	eher nein	eher ja	ja	unzutreff end
30	Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.	☐	☐	☐	☐	☐
31	Die Anerkennung meiner Arbeit durch meinen Mitarbeiterkreis ist angemessen.	☐	☐	☐	☐	☐
32	Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis.	☐	☐	☐	☐	☐
33	Mit meinen Mitarbeitenden gibt es keine Konflikte.	☐	☐	☐	☐	☐
34	Die Anerkennung meiner Arbeit durch meine unmittelbaren Vorgesetzten ist angemessen.	☐	☐	☐	☐	☐
35	Ich erfahre genügend Unterstützung von meinen unmittelbaren Vorgesetzten.	☐	☐	☐	☐	☐
36	Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte.	☐	☐	☐	☐	☐
37	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Beschäftigten ist angemessen.	☐	☐	☐	☐	☐
38	Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Beschäftigten.	☐	☐	☐	☐	☐
39	Mit den Untergebenen/den mir unterstellten Beschäftigten gibt es keine Konflikte.	☐	☐	☐	☐	☐
40	Konflikte werden offen besprochen.	☐	☐	☐	☐	☐

Kategorie: Umgebungsbedingungen am Hauptarbeitsplatz

41	Den Geräuschpegel an meinen Hauptarbeitsplatz empfinde ich als angenehm.	☐	☐	☐	☐	☐
42	Die äußeren Bedingungen (Klima, Beleuchtung, Belüftung etc.) an meinem Hauptarbeitsplatz sind angenehm.	☐	☐	☐	☐	☐
43	Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen.	☐	☐	☐	☐	☐
44	Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann (z. B. Arbeitshöhe, Sitzhöhe, Beinfreiheit, Durchgangsbreiten/Sicherheitsabstände).	☐	☐	☐	☐	☐
45	Sonstiges:					

Die folgenden Aussagen beziehen sich auf Ihre Gesundheit. Dabei geht es nicht nur um Beschwerden, welche auf den Dienst zurückzuführen sind, sondern um Ihren Gesundheitszustand im Allgemeinen. Bitte geben Sie an, an welchen Beschwerden Sie in den letzten 4 Wochen gelitten haben.

nein eher
nein ja ich möchte
keine
Aussage
machen

Ich litt in den letzten 4 Wochen unter...

46	...Kopfschmerzen (einschließlich Migräne)	☐	☐	☐	☐	☐
47	...Augenbeschwerden	☐	☐	☐	☐	☐
48	...Atemwegserkrankungen (z.B. Erkältungserkrankungen)	☐	☐	☐	☐	☐
48	...Hals-, Nasen-, Ohrenbeschwerden	☐	☐	☐	☐	☐
50	...Übermüdung/Erschöpfung	☐	☐	☐	☐	☐
51	...Leistungsabfall/Konzentrationsproblemen	☐	☐	☐	☐	☐
52	...Schlafstörungen	☐	☐	☐	☐	☐
53	...Psychischen Beschwerden (z. B. Depressionen, Ängste)	☐	☐	☐	☐	☐
54	...Stimm- und Sprechbeschwerden (durch ständiges lautes Sprechen)	☐	☐	☐	☐	☐
55	...Rückenbeschwerden	☐	☐	☐	☐	☐
56	...Nackenbeschwerden	☐	☐	☐	☐	☐
57	...Beschwerden in Schultern, Ellenbogen, Händen	☐	☐	☐	☐	☐
58	...Beschwerden an Hüften, Knien, Füßen	☐	☐	☐	☐	☐
59	...Herz-/Kreislaufbeschwerden	☐	☐	☐	☐	☐
60	...Bluthochdruck	☐	☐	☐	☐	☐
61	...Magen-Darm-Beschwerden	☐	☐	☐	☐	☐
62	...Hautirritationen/Hauterkrankungen	☐	☐	☐	☐	☐
63	...Allergien	☐	☐	☐	☐	☐
64	...Zuckerkrankheit (Diabetes Typ 1 oder 2, behandelt oder unbehandelt)	☐	☐	☐	☐	☐
65	...sonstigen Beschwerden/Erkrankungen:					

Die folgenden Aussagen beziehen sich auf ihren Alkoholkonsum und ihr Rauchverhalten.

Bitte geben Sie jeweils an, wie hoch ihr Konsum / ihr Verhalten ist.

- 66 An wie vielen Tagen in der Woche trinken Sie alkoholhaltige Getränke? (Anzahl ganze Tage)
- 67 Wie viele 0,3l-Flaschen Bier trinken sie im Durchschnitt pro Woche? (Anzahl Flaschen)
- 68 Wie viele Gläser Wein (200ml-Gläser) trinken Sie im Durchschnitt pro Woche? (Anzahl Gläser)
- 69 Wie viele Gläser Schnaps (2cl-Gläser) trinken Sie im Durchschnitt pro Woche? (Anzahl Gläser)
- 70 Wie viele Zigaretten konsumieren Sie durchschnittlich pro Tag? (Anzahl, Nichtraucher bitte „0“ eintragen)
- 71 Wie viele Zigarren/Zigarillos konsumieren Sie durchschnittlich pro Tag? (Anzahl, Nichtraucher bitte „0“ eintragen)
- 72 Bei E-Zigarten: Wie hoch ist Ihr durchschnittlicher Konsum pro Tag in ml? (Menge in ml, Nichtraucher bitte „0“ eintragen)

Die folgenden Aussagen beziehen sich auf Ihre Belastungen und Beanspruchungen am Arbeitsplatz. Bitte schätzen Sie ein, inwiefern die folgenden Aussagen bei Ihnen zutreffen.

nein eher nein eher ja unzutreffend

- 73 ...häufigen Krankheitstagen (> 14 Tage in den letzten 12 Monaten). ☐ ☐ ☐ ☐ ☐
- 74 ...Präsentismus, d. h. trotz Krankheit fehle ich nicht (mind. zweimal in den letzten 12 Monaten). ☐ ☐ ☐ ☐ ☐
- 75 ...Rückgang der Arbeitszufriedenheit. ☐ ☐ ☐ ☐ ☐
- 76 ...Dienst nach Vorschrift (d. h. ich mache nur das Nötigste). ☐ ☐ ☐ ☐ ☐
- 77 ...«Innerer Kündigung». ☐ ☐ ☐ ☐ ☐
- 78 ...zu körperlichen Beschwerden. ☐ ☐ ☐ ☐ ☐
- 79 ...zu psychischen Beschwerden (inkl. Schlafstörungen). ☐ ☐ ☐ ☐ ☐
- 80 ... zu sonstigem:

Die folgenden Aussagen beziehen sich auf Ihren Umgang mit Ihrer Gesundheit und auf Ihr Stresserleben. Bitte geben Sie an, inwiefern die folgenden Aussagen auf Sie zutreffen.

nein eher nein eher ja Unzu-
treffend

- 81 Wichtige Entscheidungen mache ich vor allem davon abhängig, was sie für meine Gesundheit bedeuten. ☐ ☐ ☐ ☐ ☐
- 82 Meine Gesundheit hat für mich erste Priorität. ☐ ☐ ☐ ☐ ☐

Die folgenden Aussagen beziehen sich auf Ihren Umgang mit Ihrer Gesundheit und auf Ihr Stresserleben. Bitte geben Sie an, inwiefern die folgenden Aussagen auf Sie zutreffen.		nein	eher nein	eher ja	ja	Unzu- treffend
83	Ich achte bewusst auf gesundheitliche Warnsignale.	☐	☐	☐	☐	☐
84	Ich achte auf eine gesunde Lebensweise (z. B. gesunde Ernährung, nicht rauchen, Sport, Schlaf und Erholung).	☐	☐	☐	☐	☐
85	In meinem Job kommt es immer wieder vor, dass ich zu wenig Rücksicht auf meine Gesundheit nehmen kann.	☐	☐	☐	☐	☐
86	Es fällt mir schwer, nach der Arbeit abzuschalten.	☐	☐	☐	☐	☐
87	Ich fühle mich ab und zu wie jemand, den man als Nervenbündel bezeichnet.	☐	☐	☐	☐	☐
88	Ich reagiere gereizt, obwohl ich es gar nicht will.	☐	☐	☐	☐	☐

Wie wichtig ist Ihnen, dass, ...		nicht wichtig	Eher nicht wichtig	Eher wichtig	wichtig	Unzu- treffend
89	... sich die Bundeswehr um die Gesundheit ihrer Beschäftigten kümmert?	☐	☐	☐	☐	☐
90	... Arbeits- und Gesundheitsschutz bei der Bundeswehr sehr ernst genommen werden?	☐	☐	☐	☐	☐
91	... die Bundeswehr sich darum kümmert, gesundheitliche Risiken in der Arbeit abzubauen?	☐	☐	☐	☐	☐
92	... die Bundeswehr attraktive Angebote im Rahmen der Betrieblichen Gesundheitsförderung macht?	☐	☐	☐	☐	☐

Im Folgenden geht es um Ihr generelles Interesse an Angeboten des Betrieblichen Gesundheitsmanagements in der Bundeswehr, wie z.B. Sport, Ernährung, Stressprävention etc. Bitte kreuzen Sie an, wie sehr die folgenden Aussagen auf Sie persönlich zutreffen.		nein	eher nein	eher ja	ja	Unzu- treffen d
93	Ich würde gerne mehr für meine Gesundheit tun und entsprechende Angebote in der Bundeswehr nutzen.	☐	☐	☐	☐	☐
94	Ich plane fest, mich dauerhaft an möglichen Angeboten des Betrieblichen Gesundheitsmanagements in der Bundeswehr zu beteiligen.	☐	☐	☐	☐	☐
95	Mein täglicher Dienstbetrieb lässt mir ausreichend zeitlichen Spielraum, um an Angeboten des Betrieblichen Gesundheitsmanagements der Bundeswehr teilzunehmen.	☐	☐	☐	☐	☐
96	Die Teilnahme an Angeboten des Betrieblichen Gesundheitsmanagements wird an meiner Dienststelle unterstützt.	☐	☐	☐	☐	☐
97	Ich würde an BGM-Maßnahmen teilnehmen, wenn sie außerhalb der Dienstzeit stattfinden würden.	☐	☐	☐	☐	☐
98	Ich würde an BGM-Maßnahmen teilnehmen, wenn sie zwar während der Dienstzeit stattfinden würden, ich aber diese Zeit nacharbeiten müsste.	☐	☐	☐	☐	☐

Im Folgenden geht es um Ihr generelles Interesse an Angeboten des Betrieblichen Gesundheitsmanagements in der Bundeswehr, wie z.B. Sport, Ernährung, Stressprävention etc. Bitte kreuzen Sie an, wie sehr die folgenden Aussagen auf Sie persönlich zutreffen.		nein	eher nein	eher ja	ja	Unzu- treffen d
99	Ich würde an BGM-Maßnahmen teilnehmen, wenn sie auf die Arbeitszeit angerechnet werden.	☐	☐	☐	☐	☐

Im Folgenden geht es um allgemeine Aspekte Ihrer Arbeit. Bitte beantworten Sie die Fragen auf einer Skala von (1) „stimme voll zu“ bis (6) „stimme überhaupt nicht zu“.		Stimme voll zu					stimme überhaupt nicht zu
		1	2	3	4	5	6
100	Alle meine Aufgaben bearbeite ich ausdauernd und zielgerichtet.	☐	☐	☐	☐	☐	☐
101	Ich habe einen starken Antrieb, Arbeitsergebnisse von hoher Qualität zu erzielen.	☐	☐	☐	☐	☐	☐
102	Mein Mitarbeiterkreis kann sich immer auf meine Unterstützung verlassen.	☐	☐	☐	☐	☐	☐
103	Ich bin überzeugt, dass ich meine Arbeit gut mache.	☐	☐	☐	☐	☐	☐
104	Ich bin bereit, für das Erreichen der Teamziele mein Bestes zu geben.	☐	☐	☐	☐	☐	☐
105	Ich bringe mich aktiv in Besprechungen ein.	☐	☐	☐	☐	☐	☐
106	Im Arbeitsalltag gelingt es mir, meinen Mitarbeiterkreis für Ideen zu begeistern.	☐	☐	☐	☐	☐	☐
107	Ich trage dazu bei, dass der Informationsaustausch im Team gelingt.	☐	☐	☐	☐	☐	☐
108	Ich ermuntere meinen Mitarbeiterkreis, sich für die Teamziele mit ganzer Kraft einzusetzen.	☐	☐	☐	☐	☐	☐
109	Ich bin stolz darauf, am Auftrag der Bundeswehr mitzuarbeiten.	☐	☐	☐	☐	☐	☐
110	Reformen in der Bundeswehr unterstütze ich nach meinen Möglichkeiten.	☐	☐	☐	☐	☐	☐
111	Ich richte mein tägliches Handeln nach den Zielen der Dienststelle aus.	☐	☐	☐	☐	☐	☐
112	Wenn ich heute noch einmal die Wahl hätte, würde ich mich wieder für die Bundeswehr als Arbeitgeber entscheiden.	☐	☐	☐	☐	☐	☐
113	Ich kann mein Arbeitspensum insgesamt gut bewältigen.	☐	☐	☐	☐	☐	☐
114	Ich bin überzeugt, meinen Arbeitsanforderungen dauerhaft gewachsen zu sein.	☐	☐	☐	☐	☐	☐
115	Ich Sorge dafür, dass meine beruflichen und privaten Verpflichtungen miteinander vereinbar sind.	☐	☐	☐	☐	☐	☐
116	Ich habe bei meiner Arbeit die Möglichkeit, das zu tun, was ich am besten kann.	☐	☐	☐	☐	☐	☐

Im Folgenden geht es um allgemeine Aspekte Ihrer Arbeit. Bitte beantworten Sie die Fragen auf einer Skala von (1) „stimme voll zu“ bis (6) „stimme überhaupt nicht zu“.		Stimme voll zu						stimme überhaupt nicht zu
		1	2	3	4	5	6	
117	Ich kann meine Bedürfnisse angemessen in die Arbeitsbeziehung einbringen.	☐	☐	☐	☐	☐	☐	
118	Meine Arbeit ist für mich mehr als nur ein Job.	☐	☐	☐	☐	☐	☐	

Alles in allem...	sehr schlecht	schlecht	nicht ganz zufrieden stellend	zufrieden stellend	gut	sehr gut
119 Wie bewerten Sie Ihre Arbeitssituation im Allgemeinen?	☐	☐	☐	☐	☐	☐
120 Wie bewerten Sie Ihren Gesundheitszustand im Allgemeinen?	☐	☐	☐	☐	☐	☐

Hier möchten wir Sie bitten, einige Fragen zu Ihrer körperlichen Aktivität zu beantworten.

- 121 Die folgende Frage bezieht sich auf die letzten 3 Monate:
An wie vielen Tagen in der Woche sind Sie körperlich so aktiv, dass Sie ins Schwitzen oder außer Atem geraten?
Es geht um eine durchschnittliche Woche. Tage (bei 0 Tagen bitte weiter mit Frage 121)
- 122 Und wie lange sind Sie an diesen Tagen, an denen Sie durch Ihre körperliche Aktivität ins Schwitzen oder außer Atem geraten, durchschnittlich körperlich aktiv?
- ☐ weniger als 10 min
- ☐ 10 bis unter 30 min
- ☐ 30 bis unter 60 min
- ☐ mehr als 60 min
- 123 Wie oft treiben Sie Sport?
- ☐ keine sportliche Betätigung
- ☐ weniger als 1 Stunde in der Woche
- ☐ regelmäßig 1–2 Stunden in der Woche
- ☐ regelmäßig 2–4 Stunden in der Woche
- ☐ regelmäßig mehr als 4 Stunden in der Woche

Am Ende möchten wir Sie bitten, einige Fragen zu Ihrer Person und Ihrem Beschäftigungsverhältnis zu beantworten.

- Welches Geschlecht haben Sie?
- 124 ☐ männlich ☐ weiblich

Am Ende möchten wir Sie bitten, einige Fragen zu Ihrer Person und Ihrem Beschäftigungsverhältnis zu beantworten.

Wie alt sind Sie?

☐ bis einschließlich 29 Jahre

☐ 30-39 Jahre

☐ 40-49 Jahre

125

☐ 50 Jahre und älter

126 Wie groß sind Sie? cm

127 Welches Gewicht haben Sie? kg (Bitte nur ganze Zahlen eintragen, ggf. runden)

Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?

☐ ohne Schulabschluss

☐ Hauptschulabschluss

☐ Realschulabschluss / Mittlere Reife / Polytechnische Oberschule

☐ Hochschulreife / Erweiterte polytechnische Oberschule

128

☐ anderer Schulabschluss

Leben Sie in einer festen Partnerschaft?

129 ☐ Ja

☐ Nein

130 Wie lange sind Sie schon für die Bundeswehr tätig?

(in ganzen Jahren, ggf. bitte runden)

Sie arbeiten derzeit in Vollzeit/Teilzeit?

☐ 100%

131 ☐ weniger als 100% und mehr als 75%

☐ zwischen 75% und 50%

☐ weniger als 50%

In welchem Beschäftigungsverhältnis stehen Sie?

Für militärisches Personal:

☐ Berufssoldat/in

☐ Zeitsoldat/in (SaZ): Jahre (Bitte nur ganze Zahlen eintragen, ggf. runden)

☐ Freiwillig Wehrdienstleistende/r (FWDL)

☐ Reservist/in

132

☐ Sonstiges:

Für ziviles Personal

☐ Beamte

☐ Beschäftigte mit unbefristetem Vertrag

☐ Beschäftigte mit befristetem Vertrag: Jahre (bitte geben sie die Gesamtdauer Ihres befristeten Vertrages

Am Ende möchten wir Sie bitten, einige Fragen zu Ihrer Person und Ihrem Beschäftigungsverhältnis zu beantworten.

- 133 Haben Sie eine Führungsposition? (militärisch/zivil)
- ☐ Ja, mit Personalverantwortung
- ☐ Ja, ohne Personalverantwortung
- ☐ Nein

-
- 134 Wurden Sie beruflich bereits im Ausland eingesetzt?
- ☐ Ja
- ☐ Nein

Vielen Dank für das Ausfüllen unseres Fragebogens.

Möchten Sie uns noch etwas zu dem Fragebogen mitteilen?

- 135 Ja, folgendes:

Anlage 2 – Checkliste für Fragebogenreview

Sehr geehrte Damen und Herren,

herzlichen Dank für Ihre Teilnahme!

Bevor Sie mit der Bewertung des Fragebogens beginnen, möchten wir Ihnen noch ein paar Hinweise zum Gebrauch der beiliegenden Checklisten geben.

1. Die Liste stellt keine vollständige Aufstellung möglicher Probleme und Fehler dar. Sie soll Ihnen lediglich die Arbeit erleichtern, indem Sie für die zutreffenden Kategorien die jeweilige Nummer eintragen können und das Problem nicht ausformulieren müssen.
2. Es ist daher nicht notwendig jedes Item auf Zutreffen aller möglichen hier aufgelisteten Probleme zu überprüfen. Machen Sie einfach dort Anmerkungen, wo Ihnen etwas auffällt.
3. Machen Sie nur bei den Items Angaben, die Ihnen als problematisch auffallen bzw. zu denen Sie eine Anmerkung machen möchten. Alle anderen Felder können einfach leer gelassen werden.
4. Um Ihnen ein möglichst realitätsnahes Bild zu vermitteln, sind alle Fragen genau so dargestellt, wie sie auch eingesetzt werden. Es ist aber selbstverständlich nicht notwendig, diese Fragen zu beantworten.

Beachten Sie, dass die Checkliste jeweils für nur eine Frage verwendet wird. Machen Sie nur bei den Problembereichen Angaben, die Ihnen als problematisch auffallen bzw. zu denen Sie eine Anmerkung machen möchten. Kreuzen Sie jeweils an, wenn das genannte Problem/der Fehler in der zu untersuchenden Frage vorhanden ist. Nutzen Sie bei Bedarf die Freitextfelder, um das Problem zu erläutern. Nicht relevante Problembereiche lassen Sie einfach unbeantwortet.

Die Checkliste muss nur für die Items bearbeitet werden, bei denen Probleme identifiziert werden. Tragen Sie bitte im untenstehenden Feld die Fragennummer ein. Für Fragen, die keine Probleme aufweisen, muss die Checkliste nicht ausgefüllt werden.

Problembereich 1: Frageverständnis/Wortlaut

- ☐ Der Fragetext enthält Worte/Formulierungen mit vager/unklarer Bedeutung
 - ☐ Unpräzise Worte/Formulierungen
 - ☐ Frage oder bestimmte Wörter lassen mehr als eine Interpretation zu
 - ☐ Ungenaue Bezeichnungen von Zeiträumen oder Zeitpunkten
 - ☐ Ungenaue Häufigkeitsbezeichnungen
 - ☐ Ungenaue Intensitätsbezeichnungen
 - ☐ Ungenaue Mengenbezeichnungen
- ☐ In der Frage werden mehrere inhaltlich unterschiedliche Sachverhalte angesprochen
- ☐ Der inhaltliche Fokus des Einleitungstextes entspricht nicht dem der Frage/Aussage
- ☐ Die Frage ist hypothetisch formuliert
- ☐ Der sprachliche Ausdruck kann vereinfacht werden

- ☐ Es werden Fremdworte oder Fachausdrücke verwendet, die für Befragte vielleicht unbekannt sind
- ☐ Grammatikalische/Orthographische Fehler
- ☐ Sonstiges

Kommentar: _____

Problembereich 2: Unzutreffende Annahmen über Befragte

- ☐ Die Frage geht von Annahmen über Merkmale/ Verhaltensweisen von Befragten aus, die unter Umständen nicht zutreffen
- ☐ Sonstiges

Kommentar: _____

Problembereich 3: Komplexität der mentalen Leistung

- ☐ Eine mentale Leistung ist schwer zu erbringen
- ☐ Eine mentale Leistung ist für einen Teil der Zielgruppe schwer zu erbringen/Verzerrungseffekt aufgrund unterschiedlicher Kompetenz
- ☐ Sonstiges

Kommentar: _____

Problembereich 4: Fragekontext und Nebeneffekte

- ☐ Es besteht die Gefahr, dass die Frage auf Grund vorangegangener Fragen nicht in der intendierten Weise interpretiert wird
- ☐ Es besteht die Gefahr, dass die Beantwortung der Frage von der Beantwortung vorangegangener Fragen beeinflusst wird
- ☐ Es besteht die Gefahr, dass die Frage gar nicht oder nicht ehrlich beantwortet wird (z.B., weil sie besonders sensible Aspekte thematisieren oder Tendenz zu sozial erwünschten Antworten)
- ☐ Es besteht die Gefahr suggestiver Nebeneffekte
- ☐ Es besteht die Gefahr von Ermüdungs-/Demotivationseffekten (z.B. aufgrund der Fragebogenlänge)
- ☐ Sonstiges

Kommentar: _____

Problembereich 5: „Unzutreffend“-Kategorie bei Einstellungsfragen

- ☐ Probleme mit der expliziten Vorgabe einer „unzutreffend“-Kategorie bei Einstellungsfragen

☐ Sonstiges

Kommentar: _____

Problembereich 6: Antwortvorgaben/Antwortskalen

☐ Die Antwortvorgaben passen nicht zur Frage

☐ Antwortvorgaben überschneiden sich

☐ Antwortvorgaben fehlen

☐ Antwortvorgaben sind unlogisch/unsystematisch angeordnet

☐ Anzahl der Skalenpunkte ist zu hoch/zu niedrig

☐ Die Abstände zwischen Skalenpunkten sind nicht gleich/die Skala ist nicht ausbalanciert

☐ Eine negativ formulierte Aussage soll mittels einer bipolaren Antwortskala bewertet werden

☐ Es besteht die Gefahr der Antwortverzerrung durch Antwortvorgaben/Antwortskalen

☐ Sonstiges

Kommentar: _____

Problembereich 7: Probleme im Zusammenhang mit der Aufgabenbeschreibung

☐ Es fehlen Anweisungen zur Vorgehensweise

☐ Mehrere Fragen müssen vom Befragten in einem komplizierten Antwortschema beantwortet werden

☐ Wichtige Zusatzinformationen fehlen oder sind an der falschen Stelle aufgeführt

☐ Der Fragetext ist unvollständig

☐ Sonstiges

Kommentar: _____

Anlage 3 – Fragebogen für die Expertenbefragung

Sehr geehrte Damen und Herren,

vielen Dank für Ihre Unterstützung!

Bevor Sie mit der Bewertung des Fragebogens beginnen, möchten wir Sie kurz über den Hintergrund der Studie informieren.

1. Zweck des Fragebogens und Anforderungen an den Fragebogen

Der Fragebogen (AIGScreenBw) soll:

- a) zur *Bedarfsanalyse* im Rahmen von BGM dienen,
- b) Handlungsschwerpunkte identifizieren, indem er die *Arbeits- und Gesundheitssituation der Beschäftigten* erfasst,
- c) wesentliche Elemente der *psychischen Gefährdungsbeurteilung* abdecken,
- d) *individuelle Reaktionen auf die Beanspruchung am Arbeitsplatz* aufzeigen,
- e) den *speziellen Erfordernissen* des Geschäftsbereichs BMVg angepasst sein.

2. Zweck dieser Befragung

Wir haben Sie zur Teilnahme an dieser Expertenbefragung eingeladen, weil Sie sowohl Kenntnisse und Erfahrungen in den Bereichen Arbeitssicherheit und/oder BGM haben und weil Sie zudem Experten für Ihre Dienststelle sind. Daher können Sie am besten beurteilen, ob

- a) der Fragebogen zu den Gegebenheiten Ihrer Dienststelle passt,
- b) der Fragebogen alle wesentlichen Aspekte im Rahmen der Bedarfsanalyse für ein BGM erfasst,
- c) der Fragebogen Ihnen als ein geeignetes Instrument für die psychische Gefährdungsbeurteilung erscheint und aus Ihrer Sicht alle wesentlichen Aspekte abbildet.

In der folgenden Befragung geht es daher um Ihre persönliche Einschätzung. Es gibt keine richtigen oder falschen Antworten. Die Befragung gliedert sich in zwei Teile.

Im ersten Teil werden Sie gebeten, alle Aspekte zu nennen, die ein Fragebogen Ihrer Meinung nach enthalten sollte, um den oben genannten Anforderungen gerecht zu werden.

Im zweiten Teil werden Sie gebeten Fragen aus dem Fragebogen zu vier Bereichen der Arbeitssituation (Arbeitsorganisation, Arbeitsinhalt und Arbeitsaufgabe, soziale Beziehungen und Arbeitsumgebung) und zu Beanspruchungsreaktionen zu bewerten. Die Fragen zu jedem der fünf Bereiche sind gleich und werden sich daher fünfmal wiederholen.

Uns ist bewusst, dass die Umfrage sehr umfangreich ist und wir bedanken uns im Vorfeld für Ihre Ausdauer! Sie haben jederzeit die Möglichkeit die Befragung zwischenzuspeichern und zu einem späteren Zeitpunkt fortzusetzen.

Wenn Sie Fragen haben oder Hilfe bei technischen Problemen benötigen, wenden Sie sich an:

Annika Claus (06131/17 88 -56 oder annika.claus@unimedizin-mainz.de)

Vielen Dank!

Hatten Sie bereits Kontakt mit dem AIGScreenBw, d. h. wurde die Befragung in Ihrem Tätigkeitsbereich durchgeführt oder ist eine Befragung mit dem AIGScreenBw in Ihrem Tätigkeitsbereich geplant und Sie sind aktiv in die Vorbereitung involviert?

☐	Nein
☐	Ja
☐	Keine Antwort

Teil 1: Offene Befragung zu relevanten Inhalten

Bitte notieren Sie, welche Aspekte Ihrer Meinung nach bezüglich der beschriebenen Anforderungen im Fragebogen enthalten sein müssen:

Freitext

Teil 2: Bewertung des vorliegenden Fragebogens

A. Folgende Items sollen den Merkmalsbereich Arbeitsorganisation abbilden.

Arbeitsorganisation umfasst

- die Einhaltung gesetzlicher Bestimmungen zur Arbeitszeit,
- die Organisation und Planung von Arbeitszeiten und Pausen,
- die Anpassung der Arbeitszeit an die Arbeitsbelastung-/menge/-komplexität,
- die Strukturierung von Arbeitsabläufen und die Häufigkeit von Arbeitsstörungen und -unterbrechungen,
- Abstimmungs- und Koordinierungsmöglichkeiten im Kollegium sowie
- die Art, Dauer und Inhalte der Kommunikations- und Informationserfordernisse und die Anzahl der Kommunikationspartnerinnen und -partner.

F A.1 Bitte bewerten Sie die Relevanz des jeweiligen Items für diesen Merkmalsbereich.

		1 - Nicht relevant	2 - Wenig relevant	3 - Etwas relevant	4 - Ziemlich relevant	5 - Sehr relevant	6 - Höchst relevant
1	Ich empfinde das Verhältnis von Arbeitsmenge zu Arbeitszeit ausgewogen.	☐	☐	☐	☐	☐	☐
2	Die Anzahl der von mir verlangten Vertretungen/Bereitschaftsdienste ist angemessen.	☐	☐	☐	☐	☐	☐
3	Die Anzahl der von mir verlangten Überstunden ist angemessen.	☐	☐	☐	☐	☐	☐

4	Meine Arbeitszeitverteilung ist mir angenehm.	☐	☐	☐	☐	☐	☐
5	Ich empfinde das Verhältnis von Arbeitszeit zu Freizeit als ausgewogen.	☐	☐	☐	☐	☐	☐
6	Meine Arbeitspausen sind ausreichend.	☐	☐	☐	☐	☐	☐
7	Meine Arbeitspausen sind störungsfrei.	☐	☐	☐	☐	☐	☐
8	Ich werde bei meiner Arbeit selten unterbrochen.	☐	☐	☐	☐	☐	☐

F A.2 Bitte bewerten Sie, inwiefern folgende Aussage zutrifft:

Der Merkmalsbereich Arbeitsorganisation ist vollständig repräsentiert.

1 – Stimme überhaupt nicht zu	2- Stimme nicht zu	3 - Stimme eher nicht zu	4 - Stimme eher zu	5 – Stimme zu	6 – Stimme vollkommen zu
☐	☐	☐	☐	☐	☐

F A.3 Würden Sie weitere Aspekte (Items) zum Merkmalsbereich erfassen?

☐	Nein	
☐	Ja	→ Welche weiteren Items würden Sie erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich relevant?
		Freitext

F A.4 Würden Sie von den genannten Aspekten (Items) eines oder mehrere nicht erfassen?

☐	Nein	
☐	Ja	→ Welche Items würden Sie nicht erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich nicht relevant?
		Freitext

F A.5 Ist die Definition des Merkmalsbereichs Ihrer Meinung nach vollständig?

☐	Ja	
☐	Nein*	→ Was würden Sie bei der Definition des Merkmalsbereichs Arbeitsorganisation ergänzen oder ändern?
		Freitext

* An dieser Stelle wird die Definition des Merkmalsbereichs in der Online-Befragung erneut unter dem Fragetext eingeblendet.

F A.6 Ist die Definition des Merkmalsbereichs Ihrer Meinung nach zu weit gefasst?

☐	Ja	
☐	Nein*	→ Was würden Sie aus der Definition des Merkmalsbereichs Arbeitsorganisation ausschließen?

	Freitext
--	----------

F A.7 Haben Sie weitere Anmerkungen zum Merkmalsbereich Arbeitsorganisation?

Freitext

B. Folgende Items sollen den Merkmalsbereich **Arbeitsinhalt und Arbeitsaufgabe** abbilden.

Arbeitsinhalt und Arbeitsaufgabe umfassen

- den Grad der Kontrolle und der Verantwortung, über den Beschäftigte bei der Ausführung ihrer Aufgaben bzgl. Planung, Organisation und Durchführung verfügen (Vollständigkeit)
- den Grad der Abwechslung bzgl. der kognitiven, informatorischen und emotionalen Anforderungen und deren Übereinstimmung mit den individuellen Kompetenzen. (Variabilität)
- den Grad der Entscheidungsfreiheiten bezüglich der inhaltlichen und zeitlichen Arbeitsgestaltung hinsichtlich Arbeitsinhalten, Arbeitspensum, Arbeitsmethode/-mittel und Reihenfolge der Tätigkeiten (Handlungsspielraum)
- die Vollständigkeit, Relevanz, Aktualität, Verständlichkeit, Zugänglichkeit und Eindeutigkeit der arbeitsbezogenen Informationen
- den Umfang und die Transparenz der Pflichten (Verantwortung)
- die Anforderungen an die Kompetenzen und Kenntnisse und die Einweisung in die Tätigkeit (Qualifikation)
- die emotionale Inanspruchnahme durch die Ausübung der Tätigkeit

F B.1 Bitte bewerten Sie die Relevanz des jeweiligen Items für diesen Merkmalsbereich

		1 - Nicht relevant	2 - Wenig relevant	3 - Etwas relevant	4 - Ziemlich relevant	5 - Sehr relevant	6 - Höchst relevant
8	Ich fühle mich auf meine beruflichen Anforderungen gut vorbereitet.	☐	☐	☐	☐	☐	☐
9	Ich habe ausreichende Fort- und Weiterbildungsmöglichkeiten für meine Tätigkeit.	☐	☐	☐	☐	☐	☐
10	Meine Arbeit ist in angenehmem Maße abwechslungsreich.	☐	☐	☐	☐	☐	☐
11	Das Ausmaß an Verantwortung, das ich für andere Personen trage, ist angemessen (z. B. Personalverantwortung).	☐	☐	☐	☐	☐	☐
12	Das Ausmaß an Verantwortung, das ich für Objekte trage, ist angemessen (z. B. Infrastruktur, Material).	☐	☐	☐	☐	☐	☐
13	Ich habe einen angemessenen Handlungsspielraum (z. B. Entscheidungsfreiheit) bei meiner Arbeit.	☐	☐	☐	☐	☐	☐
14	Die körperliche Leistungsfähigkeit ist wichtig, damit ich die Anforderungen meines Dienstpostens erfüllen kann.	☐	☐	☐	☐	☐	☐
15	Ich kann erkennen, ob ich meine Arbeit gut gemacht habe.	☐	☐	☐	☐	☐	☐

F B.2 Bitte bewerten Sie, inwiefern folgende Aussage zutrifft:

Der Merkmalsbereich Arbeitsinhalt und Arbeitsaufgabe ist vollständig repräsentiert.

1 – Stimme überhaupt nicht zu	2- Stimme nicht zu	3 - Stimme eher nicht zu	4 - Stimme eher zu	5 – Stimme zu	6 – Stimme vollkommen zu
☐	☐	☐	☐	☐	☐

F A.3 Würden Sie weitere Aspekte (Items) zum Merkmalsbereich erfassen?

☐	Nein	
☐	Ja	→ Welche weiteren Items würden Sie erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich relevant?
		Freitext

F A.4 Würden Sie von den genannten Aspekten (Items) eines oder mehrere nicht erfassen?

☐	Nein	
☐	Ja	→ Welche Items würden Sie nicht erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich nicht relevant?
		Freitext

F A.5 Ist die Definition des Merkmalsbereichs Ihrer Meinung nach vollständig?

☐	Ja	
☐	Nein*	→ Was würden Sie bei der Definition des Merkmalsbereichs Arbeitsinhalt und -aufgabe ergänzen oder ändern?
		Freitext

* An dieser Stelle wird die Definition des Merkmalsbereichs in der Online-Befragung erneut unter dem Fragetext eingeblendet.

F A.6 Ist die Definition des Merkmalsbereichs Ihrer Meinung nach zu weit gefasst?

☐	Nein	
☐	Ja	→ Was würden Sie aus der Definition des Merkmalsbereichs Arbeitsinhalt und -aufgabe ausschließen?
		Freitext

F B.7 Haben Sie weitere Anmerkungen zum Merkmalsbereich Arbeitsinhalt und -aufgabe?

Freitext

C. Folgende Items sollen den Merkmalsbereich soziale Beziehungen abbilden.

Soziale Beziehungen umfassen

- die Häufigkeit und die Qualität arbeitsbedingter und arbeitsbezogener Interaktionen mit dem Kollegium und Vorgesetzten
- die organisatorische Einbindung der eigenen Arbeit in das Arbeitssystem
- der Grad sozialer Unterstützung durch das Kollegium und Vorgesetzte
- das Feedback der Vorgesetzten
- den Führungsstil und die Qualifikation der Führungskräfte
- Streitigkeiten und Konflikte mit dem Kollegium und Führungskräften

F C.1 Bitte bewerten Sie die Relevanz des jeweiligen Items für diesen Merkmalsbereich

		1 - Nicht relevant	2- Wenig relevant	3 - Etwas relevant	4 - Ziemlich relevant	5 - Sehr relevant	6 - Höchst relevant
16	Ich erhalte angemessene Rückmeldung von meinen nächsten Vorgesetzten bezüglich meiner Arbeit.	☐	☐	☐	☐	☐	☐
17	Mein Kundenkreis für interne Dienstleistungen (z. B. auch Patienten, Leistungsempfänger) verhält sich angemessen.	☐	☐	☐	☐	☐	☐
18	Ich werde von dem internen Kundenkreis (Patienten, Leistungsempfänger) respektiert und in meiner Arbeit unterstützt.	☐	☐	☐	☐	☐	☐
19	Die Anerkennung meiner Arbeit durch meinen Mitarbeiterkreis ist angemessen.	☐	☐	☐	☐	☐	☐
20	Ich erfahre genügend Unterstützung durch meinen Mitarbeiterkreis.	☐	☐	☐	☐	☐	☐
21	Mit meinen Mitarbeitenden gibt es keine Konflikte.	☐	☐	☐	☐	☐	☐
22	Die Anerkennung meiner Arbeit durch meine unmittelbaren Vorgesetzten ist angemessen.	☐	☐	☐	☐	☐	☐
23	Ich erfahre genügend Unterstützung von meinen unmittelbaren Vorgesetzten.	☐	☐	☐	☐	☐	☐
24	Mit meinen unmittelbaren Vorgesetzten gibt es keine Konflikte.	☐	☐	☐	☐	☐	☐
25	Die Anerkennung meiner Arbeit durch Untergebene/die mir unterstellten Beschäftigten ist angemessen.	☐	☐	☐	☐	☐	☐
26	Ich erfahre genügend Unterstützung von Untergebenen/den mir unterstellten Beschäftigten.	☐	☐	☐	☐	☐	☐

27	Mit den Untergebenen/den mir unterstellten Beschäftigten gibt es keine Konflikte.	☐	☐	☐	☐	☐	☐
28	Konflikte werden offen besprochen.	☐	☐	☐	☐	☐	☐

F C.2 Bitte bewerten Sie, inwiefern folgende Aussage zutrifft:

Der Merkmalsbereich soziale Beziehungen ist vollständig repräsentiert.

1 – Stimme überhaupt nicht zu	2- Stimme nicht zu	3 - Stimme eher nicht zu	4 - Stimme eher zu	5 – Stimme zu	6 – Stimme vollkommen zu
☐	☐	☐	☐	☐	☐

F C.3 Würden Sie weitere Aspekte (Items) zum Merkmalsbereich erfassen?

☐	Nein	
☐	Ja	→ Welche weiteren Items würden Sie erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich relevant?
		Freitext

F C.4 Würden Sie von den genannten Aspekten (Items) eines oder mehrere nicht erfassen?

☐	Nein	
☐	Ja	→ Welche Items würden Sie nicht erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich nicht relevant?
		Freitext

F C.5 Ist die Definition des Merkmalsbereichs Ihrer Meinung nach vollständig?

☐	Ja	
☐	Nein*	→ Was würden Sie bei der Definition des Merkmalsbereichs soziale Beziehungen ergänzen oder ändern?
		Freitext

* An dieser Stelle wird die Definition des Merkmalsbereichs in der Online-Befragung erneut unter dem Fragetext eingeblendet.

F C.6 Ist die Definition des Merkmalsbereichs Ihrer Meinung nach zu weit gefasst?

☐	Nein	
☐	Ja	→ Was würden Sie aus der Definition des Merkmalsbereichs soziale Beziehungen ausschließen?
		Freitext

F C.7 Haben Sie weitere Anmerkungen zum Merkmalsbereich soziale Beziehungen?

Freitext

D. Folgende Items sollen den Merkmalsbereich Arbeitsumgebung abbilden.

Arbeitsumgebung umfasst

- Unfallgefahren und Fluchtwege am Arbeitsplatz sowie Sicherheits- und Gesundheitsschutzkennzeichnung
- physikalische und chemische Faktoren die während der Ausübung der Arbeitstätigkeit auf Arbeitskräfte einwirken
- die ergonomische Arbeitsplatzgestaltung
- physische Arbeitsbedingungen
- die Anordnung des Arbeitsplatzes und der Pausen-/Sanitärräume
- die Arbeitsmittel

F D.1 Bitte bewerten Sie die Relevanz des jeweiligen Items für diesen Merkmalsbereich

		1 - Nicht relevant	2 - Wenig relevant	3 - Etwas relevant	4 - Ziemlich relevant	5 - Sehr relevant	6 - Höchst relevant
29	Die nötigen Arbeitsmittel sind vorhanden.	☐	☐	☐	☐	☐	☐
30	Arbeitsräume sind in ausreichender Anzahl vorhanden.	☐	☐	☐	☐	☐	☐
31	Die Einrichtung an meinem Hauptarbeitsplatz ist angemessen.	☐	☐	☐	☐	☐	☐
32	Den Geräuschpegel an meinen Hauptarbeitsplatz empfinde ich als angenehm.	☐	☐	☐	☐	☐	☐
33	Die äußeren Bedingungen (Klima, Beleuchtung, Belüftung etc.) an meinem Hauptarbeitsplatz sind angenehm.	☐	☐	☐	☐	☐	☐
34	Die körperlichen Anforderungen meiner beruflichen Tätigkeit sind angemessen.	☐	☐	☐	☐	☐	☐
35	Mein Hauptarbeitsplatz ist so gestaltet, dass ich gut arbeiten kann (z. B. Arbeitshöhe, Sitzhöhe, Beinfreiheit, Durchgangsbreiten/Sicherheitsabstände).	☐	☐	☐	☐	☐	☐

F D.2 Bitte bewerten Sie, inwiefern folgende Aussage zutrifft: Der Merkmalsbereich

Arbeitsumgebung ist vollständig repräsentiert.

1 – Stimme überhaupt nicht zu	2- Stimme nicht zu	3 - Stimme eher nicht zu	4 - Stimme eher zu	5 – Stimme zu	6 – Stimme vollkommen zu
☐	☐	☐	☐	☐	☐

F D.3 Würden Sie weitere Aspekte (Items) zum Merkmalsbereich erfassen?

☐	Nein	
☐	Ja	→ Welche weiteren Items würden Sie erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich relevant?
		Freitext

F D.4 Würden Sie von den genannten Aspekten (Items) eines oder mehrere nicht erfassen?

☐	Nein	
☐	Ja	→ Welche Items würden Sie nicht erfassen? Warum sind diese Aspekte Ihrer Meinung nach für den Merkmalsbereich nicht relevant?
		Freitext

F D.5 Ist die Definition des Merkmalsbereichs Ihrer Meinung nach vollständig?

☐	Ja	
☐	Nein*	→ Was würden Sie bei der Definition des Merkmalsbereichs Arbeitsumgebung ergänzen oder ändern?
		Freitext

* An dieser Stelle wird die Definition des Merkmalsbereichs in der Online-Befragung erneut unter dem Fragetext eingeblendet.

F D.6 Ist die Definition des Merkmalsbereichs Ihrer Meinung nach zu weit gefasst?

☐	Nein	
☐	Ja	→ Was würden Sie aus der Definition des Merkmalsbereichs Arbeitsumgebung ausschließen?
		Freitext

F D.7 Haben Sie weitere Anmerkungen zum Merkmalsbereich Arbeitsumgebung?

Freitext

Teil 3 – Ihre Verwendung

1. Welche Funktion nehmen Sie wahr?

☐	BGM-Koordinierende
☐	BGM-Beauftragte
☐	Fachkraft für Arbeitssicherheit / Sicherheitsingenieur
☐	Betriebsmediziner
☐	Dienststellenleitung
☐	Beteiligungsgremium und Arbeitskreis Gesundheit
☐	Sonstige: Freitext

2. Welche Dienststelleneigenschaft trifft auf Ihren aktuellen Einsatzbereich zu?

☐	Einsatzverband (z.B. Regimente, Luftwaffensicherung)
☐	Kommandobehörde
☐	Sanitätseinrichtung (z.B. Krankenhaus, SanVersZ)
☐	Seegehende Einheit inkl. Unterstützungskräfte
☐	Fliegende Einheit inkl. Unterstützungskräfte und Flugsicherung
☐	Forschungseinrichtung (z.B. Resortforschungseinrichtung, Institute)
☐	Ausbildungseinrichtung (z. B. Sanitätsakademie, ZAW, Feldjägerdienstschule)
☐	Verwaltung (z.B. BwDLZ, BMVg, BAAINBw)
☐	Sonstige: Freitext

3. Die Befragung ist hiermit beendet. Wenn Ihnen noch etwas aufgefallen ist, dass Sie nicht anhand der Fragen anmerken können oder wenn Sie zusätzlich etwas ergänzen möchten, können Sie dies hier tun.

Freitext

Anlage 4: Auswertung der Freitextantworten der Expertenbefragung auf die Frage „Bitte notieren Sie, welche Aspekte Ihrer Meinung nach bezüglich der beschriebenen Anforderungen im Fragebogen enthalten sein müssen“ (n = 376)

Merkmalsbereich	Kategorie	Nennungen absolut	Anteil im Merkmalsbereich	Beispiel
Bedarfs- /Wirksamkeitsanalyse/ Evaluation und Umsetzung BGM (n = 70)	Handlungsschwerpunkte identifizieren	14	20,0%	„Der Schwerpunkt sollte jedoch darauf gelegt werden entsprechende Handlungsschwerpunkte in der Arbeits- und Gesamtsituation inklusive des medizinischen Arbeitsschutzes sowie der psychischen Gesundheit zu identifizieren, auszuwerten und dann ein individuellen Konzept zu erarbeiten für die entsprechende Dienststelle.“
	Bedarfs-/Istanalyse	41	58,6%	„IST-Zustandsanalyse, Krankenstand, Gesundheitszustand, Arbeitszufriedenheit, Arbeitsplatzmerkmale“
	Umsetzung BGM an DS evaluieren	7	10,0%	„BGM + Vorkenntnisse/Informationen über BGM? + Ansprechpartner bekannt?“
	BGM-Compliance erfassen	5	7,1%	„Bereitschaft zur Teilnahme? Bereitschaft zur Unterstützung (z.B. Leitung von einzelnen Angeboten)? Teilnahme während/nach Regelarbeitszeit? Bereitschaft zu finanziellem Eigenbeitrag für bestimmte Angebote? gewünschte Angebote (Info-Vorträge, Beratung, Sport, "Gesundheitssport" etc.)“
	BGM fördern und bewerben	3	4,3%	„wird es Schichtbegleitend angeboten, wurde im Vorfeld genug darüber informiert“
Arbeitssituation	Arbeitsorganisation	14	16,9%	„Transparenz und Informationsfluss in der Organisation / im jeweiligen Bereich“

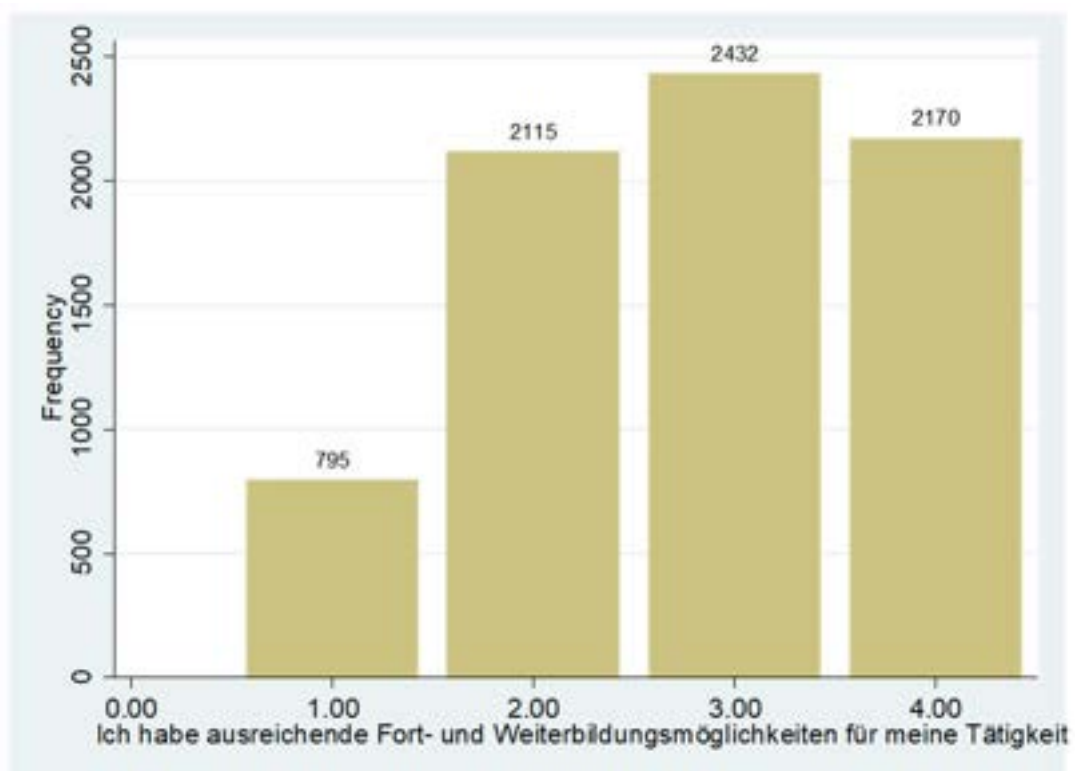
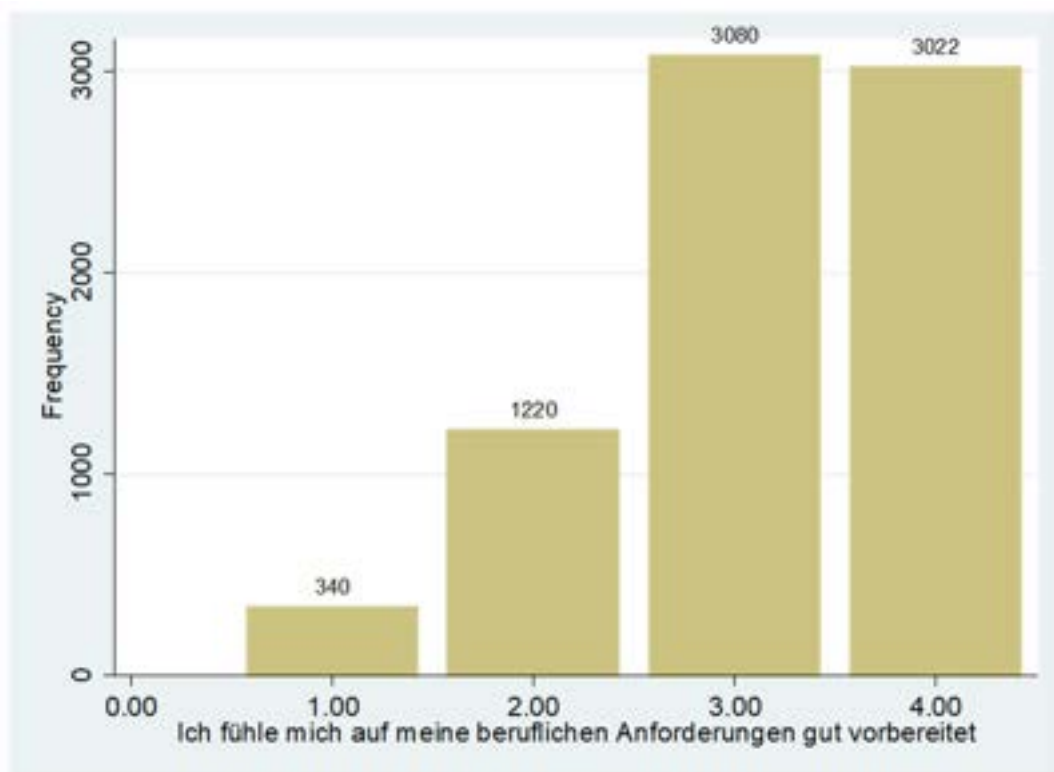
(n = 83)	Arbeitssituation allgemein	34	41,0%	„Situation am Arbeitsplatz“
	Arbeitsmittel und Arbeitsplatzumgebung	13	15,7%	„Bewertung der Standardbüroausstattung durch die Beschäftigten als angemessen oder nicht, falls Sonderausstattung vorhanden Bewertung als angemessen oder nicht, Bewertung der sonstigen Arbeitsmittel; Aussage zu Raumgröße bzw. Größe des Arbeitsplatzes zu klein ...; Einzelbüro vorhanden, erwünscht?“
	Arbeitsinhalt	22	26,5%	„Einsatz beim Dienstherrn gemäß der eigenen Fähigkeiten und entsprechende Dienstgrad / Dienststellung? (Bsp. zur Erläuterung: Soldaten mit abgeschlossenem Studium dienen als Mannschafter auf Fregatten --> Deutliche Diskrepanz zwischen Ausbildung und Verwendung)“
Psychische Belastungen und Gefährdungen (n = 40)	Psychische Gefährdungsbeurteilung	31	77,5%	„Als besonders wichtig erscheint u.a. das umfängliche individuelle Erfassen der Arbeits- und Gesundheitssituation der Beschäftigten, gerade um BGM-Maßnahmen besser am Bedarf ausrichten zu können. Dies schließt die Beurteilung einer psychischen Gefährdung mit ein.“
	Stress	6	15,0%	„Erfassung der Stressoren, [...], Individuelles Stresserleben, Möglichkeiten des Stressabbaus“
	Emotionales Befinden	1	2,5%	„emotionalen Befindes eines Mitarbeites“
	psychische Belastungen	2	5,0%	„Arbeitsschutz: psychische Belastung“
individuelles Gesundheitsverhalten (n = 15)	Bewegung	5	33,3%	„Bewegung/Sport, wie oft wird in der Woche Sport gemacht (inklusive Freizeit), welche Sportarten: Ballsportarten, Ausdauer, Kraftsport, Mannschaftssportarten“
	Sucht/Genussmittel	2	13,3%	„private Gesundheitsvorsorge: [...] Alkohol, Rauchen etc?“
	Ernährung	4	26,7%	„Individuelles Risikoverhalten im puncto [...] Ernährung“

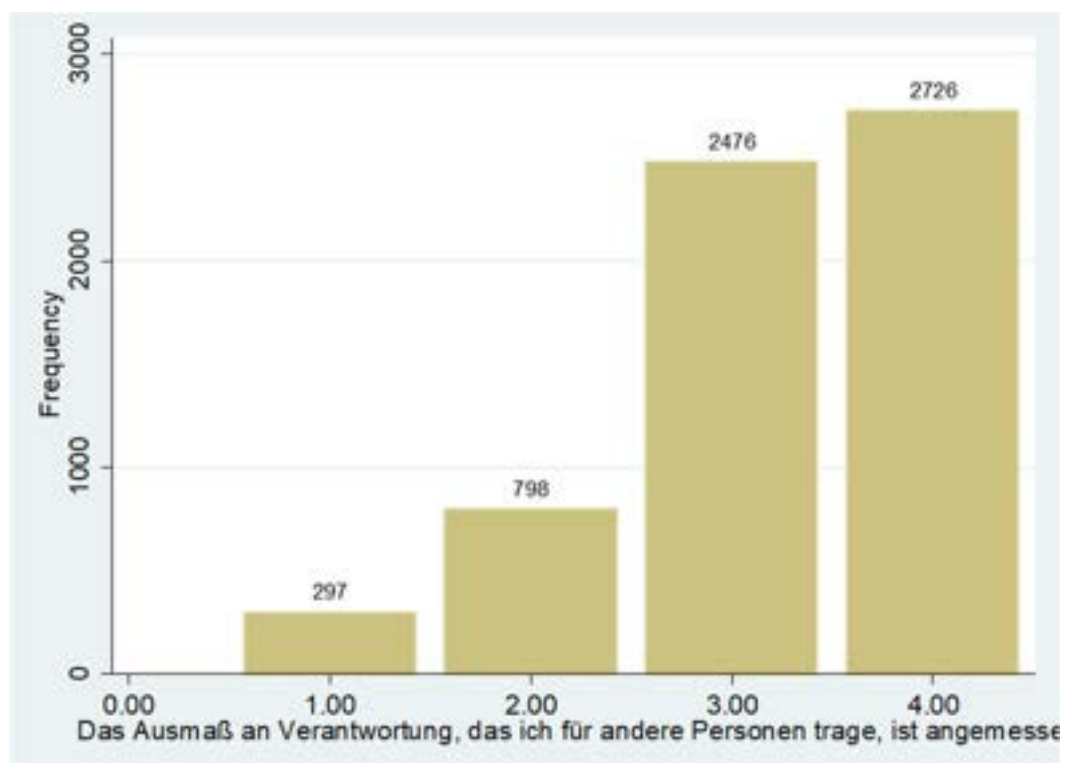
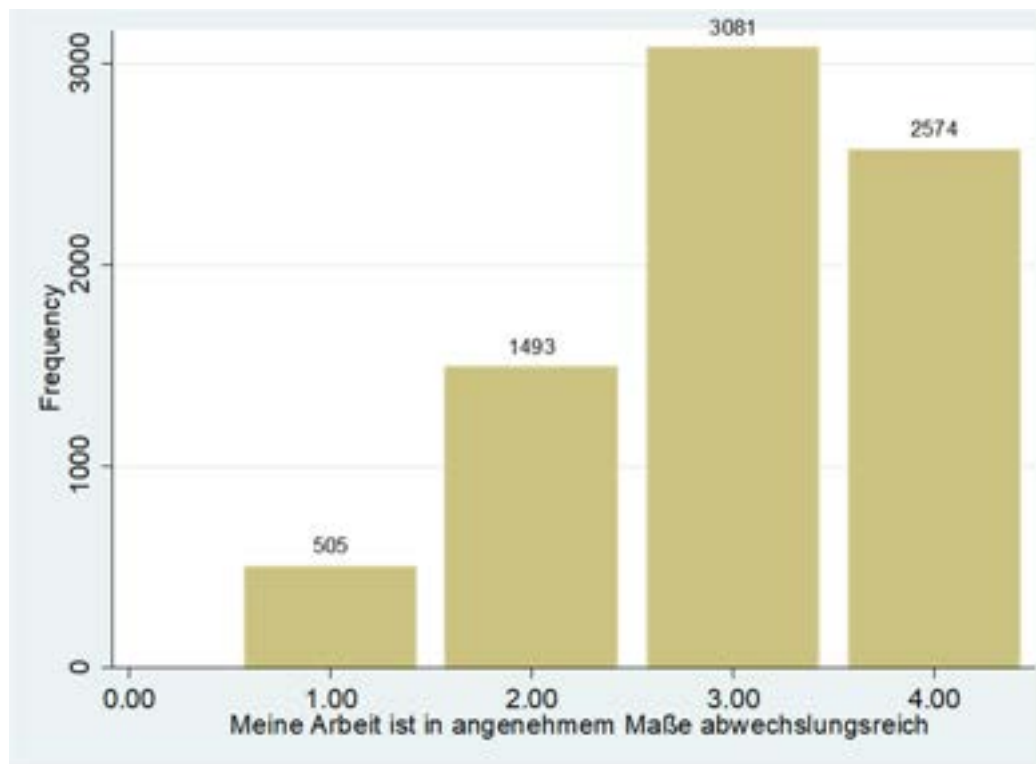
	Gesundheitsverhalten allgemein	4	26,7%	„Gesundheitsverhalten“
Gesundheits- situation (n = 41)	keine Unterkategorien	-	100,0%	„Gesundheitssituation der Beschäftigten“
Individuelle Beanspruchungs- reaktionen (n = 26)	keine Unterkategorien	-	100,0%	„Individuelle Reaktionen auf die Beanspruchung am Arbeitsplatz aufzeigen (Kdo Ebene also Bürotätigkeiten)“
	zusätzliche Fragebogeninhalte	23	33,3%	„Auflistung effektiver Gegenmaßnahmen und Möglichkeit, Wünsche und Vorschläge zu äußern“
	Differenzierte Erfassung des Tätigkeitsprofils	22	31,9%	„gezielter auf die jeweiligen Arbeitsplätze eingehen, konkreter gestellt werden“
Erhebungs- und Auswertungsdesign (n = 69)	Auswertungsdesign	6	8,7%	„Codiert sein, für eine erneute Befragung“
	Fragebogen- gestaltung	13	18,8%	„Abstufungen der Antwortmöglichkeiten beinhalten“
	Zielsetzung der Befragung	3	4,3%	„Wer soll erreicht werden? „
	Rahmenbedingungen	2	2,9%	„wird es Schichtbegleitend angeboten, wurde im Vorfeld genug darüber informiert“
Soziales (n = 19)	Arbeitsklima	7	36,8%	„Wie ist das Arbeitsklima/Miteinander?“
	soziale Beziehungen	5	26,3%	„Unterstützung durch Kollegen, Vorgesetzte etc.“

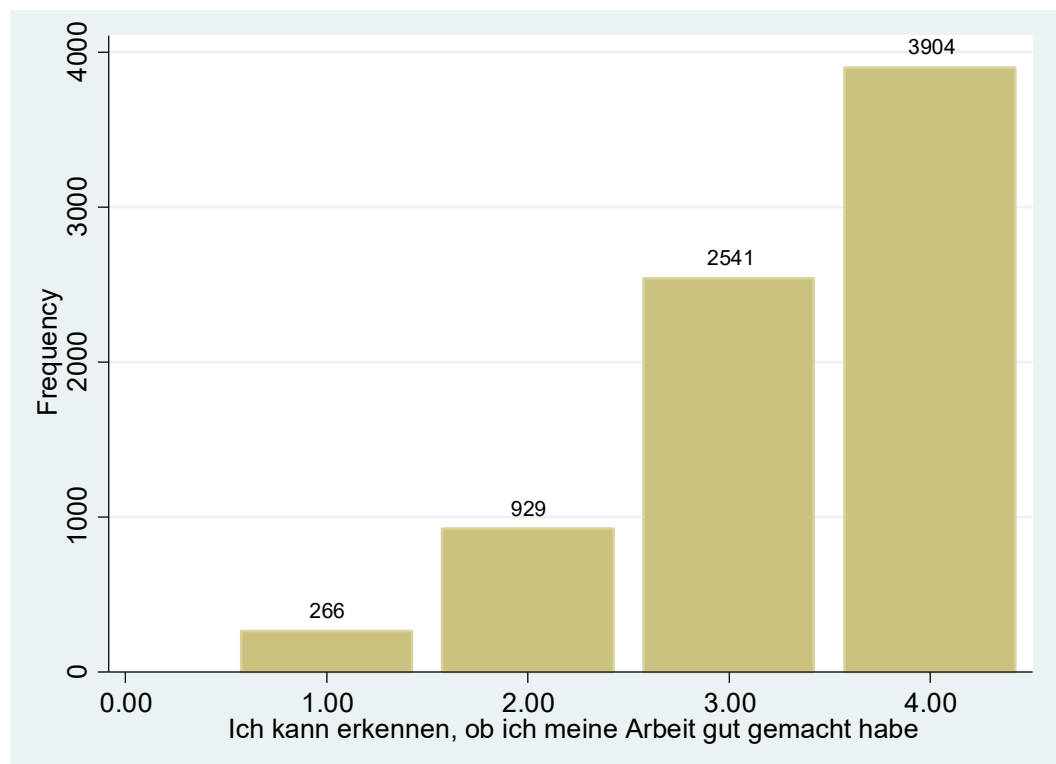
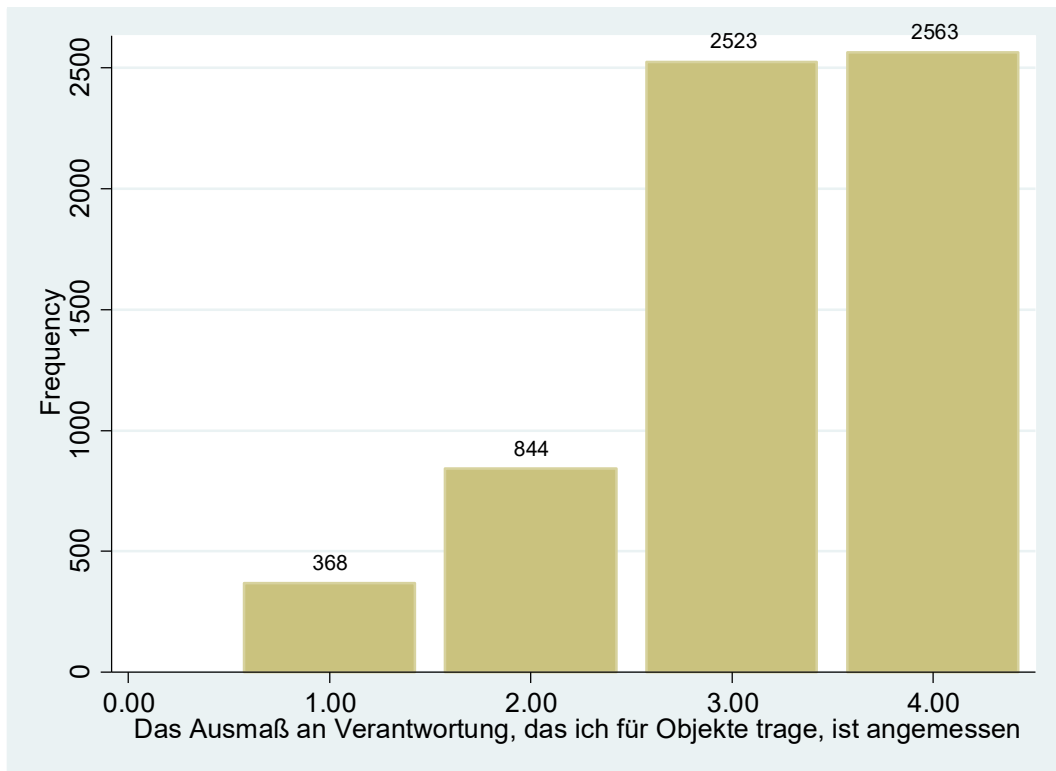
Führung (n = 13)	Negative soziale Interaktionen	2	10,5%	„Mobbing am Arbeitsplatz? Sexuelle Belästigung am Arbeitsplatz?“
	Fehler-/Konfliktkultur	2	10,5%	„Umgang mit Fehlern/Fehlerkultur“
	Sonstiges	3	15,8%	„Soziale Nachhaltigkeit“
	Unterstützung durch Vorgesetzte	2	15,4%	„Erlebte/erfahrene Unterstützung durch [...] die unmittelbare und die nächst höhere Führungskraft“
	Feedback / Kommunikation	2	15,4%	„Werden meine Arbeitsergebnisse in meinen Augen richtig bewertet, erkennt der Vorgesetzte besondere Härten“
	Führungsverhalten	5	38,5%	„tradierte aber obsolekte Führungs- und Verhaltensweisen“
	allgemein	4	30,8%	„eigene Einschätzung [...] die Vorgesetzten und den Dienstherrn“

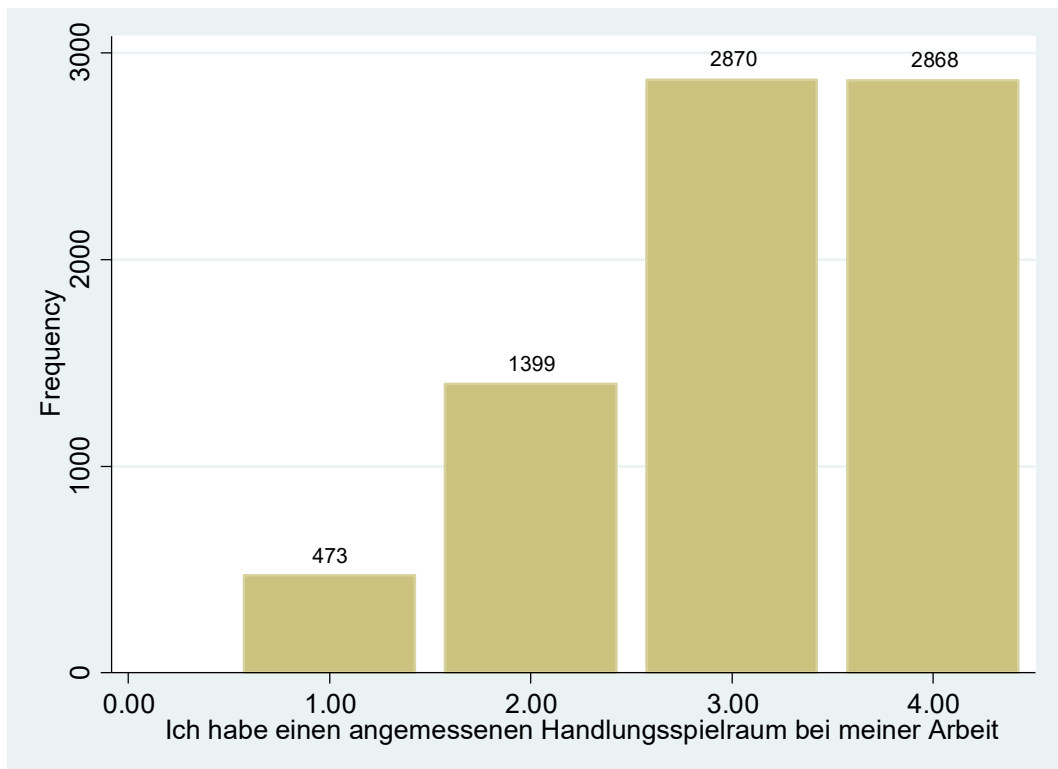
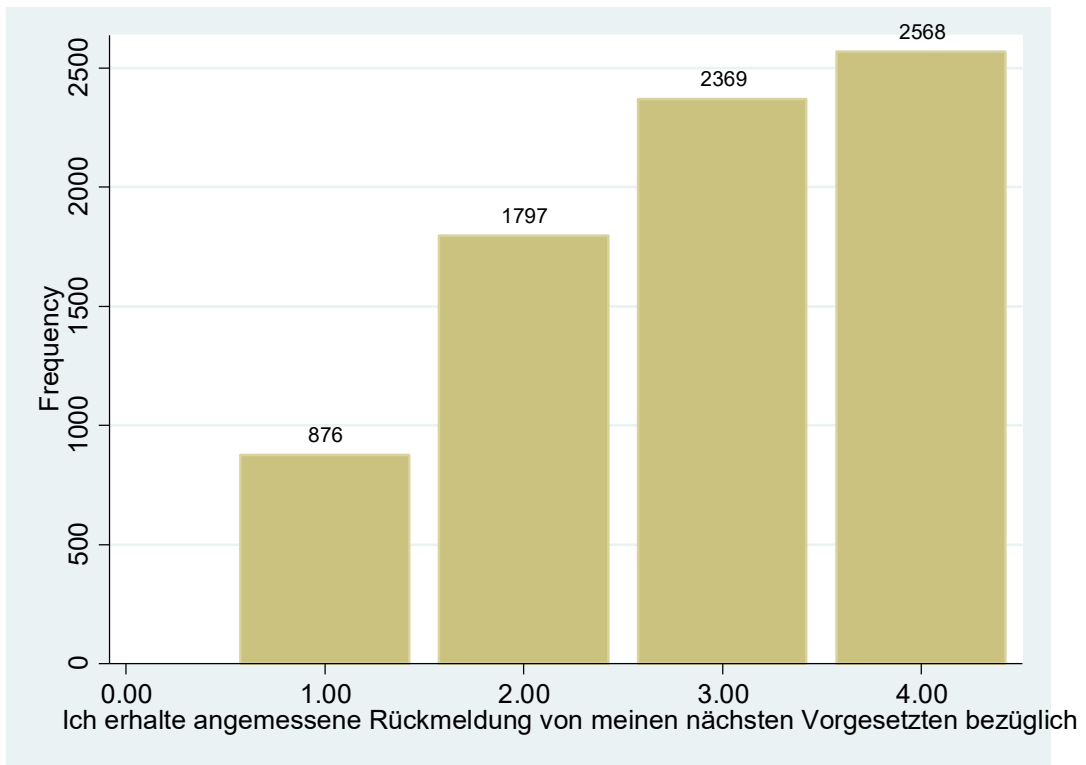
Hinweis: Die Beispiele geben Originalformulierungen wieder, weshalb Orthografie und Grammatik unkorrigiert übernommen wurden

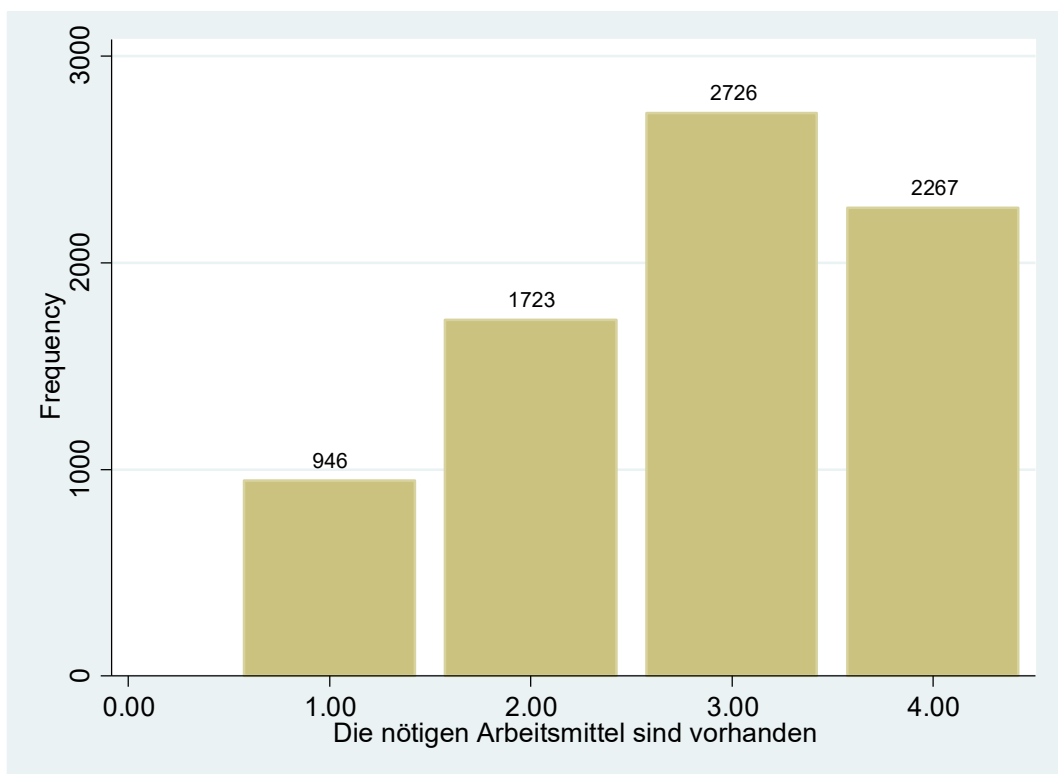
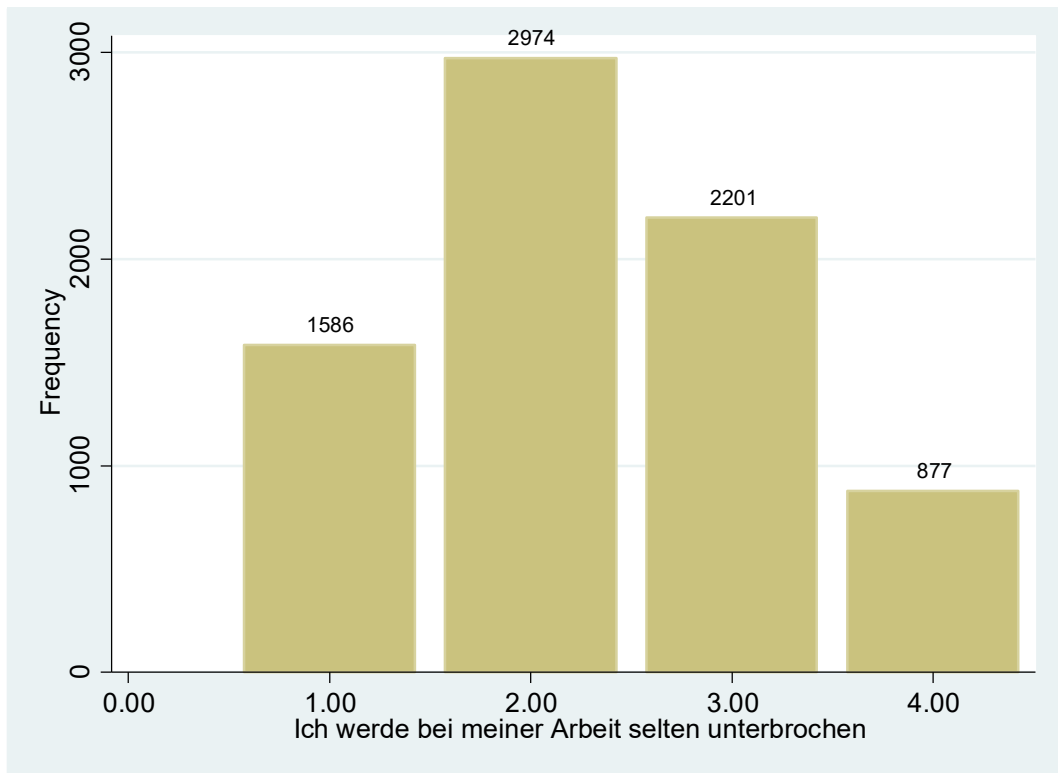
Anlage 5: Ergebnisse der Itemanalyse – Häufigkeitsverteilungen der Antworten auf alle 34 Items als Histogramm

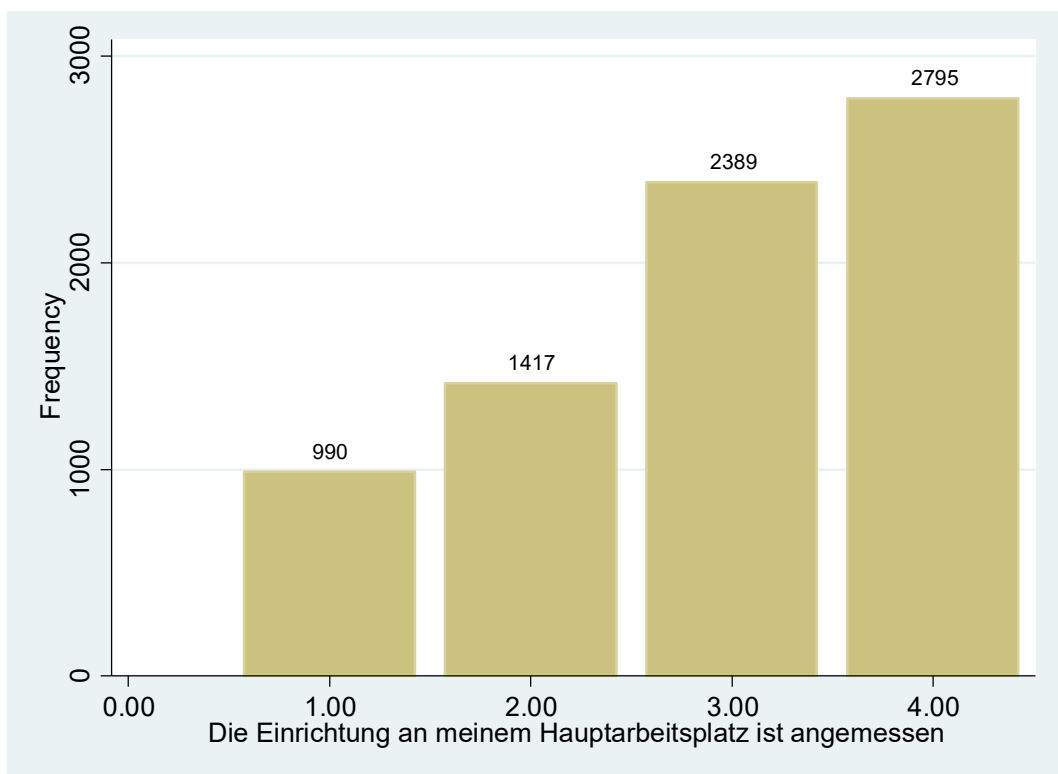
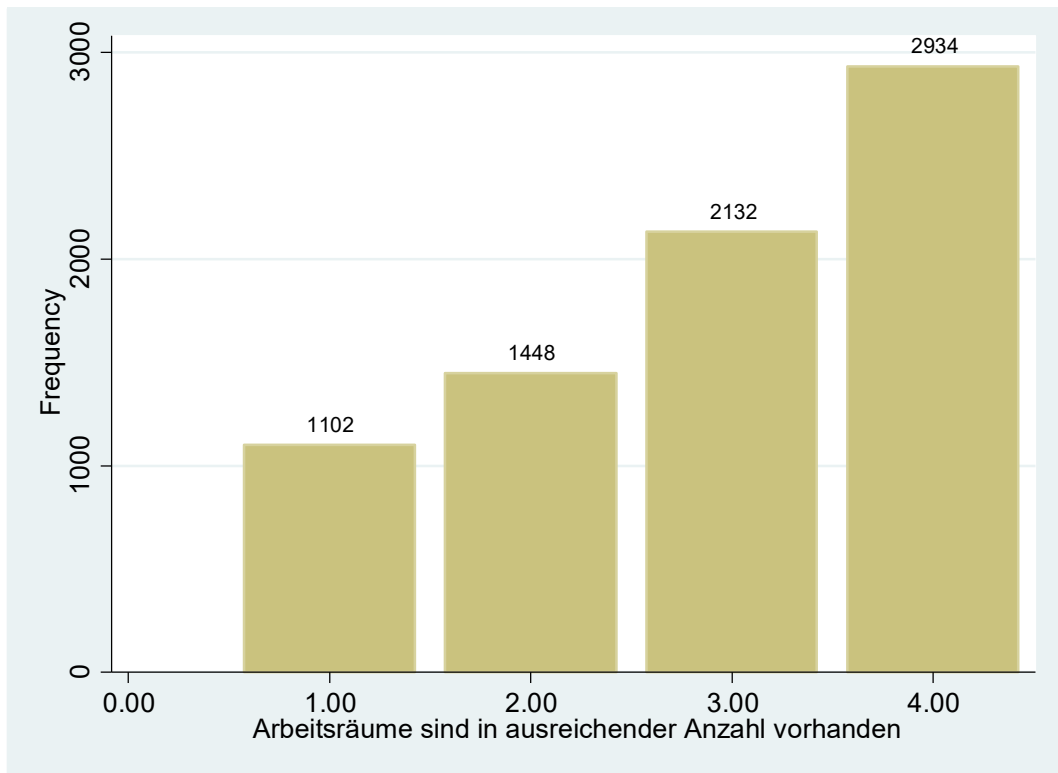


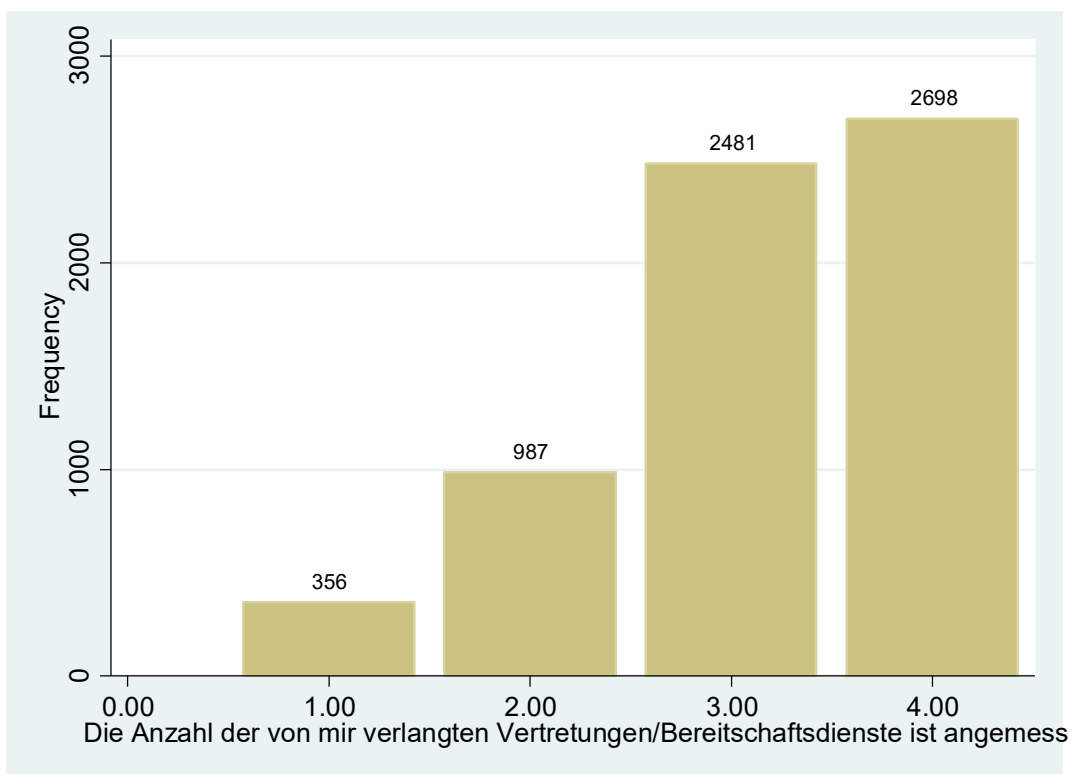
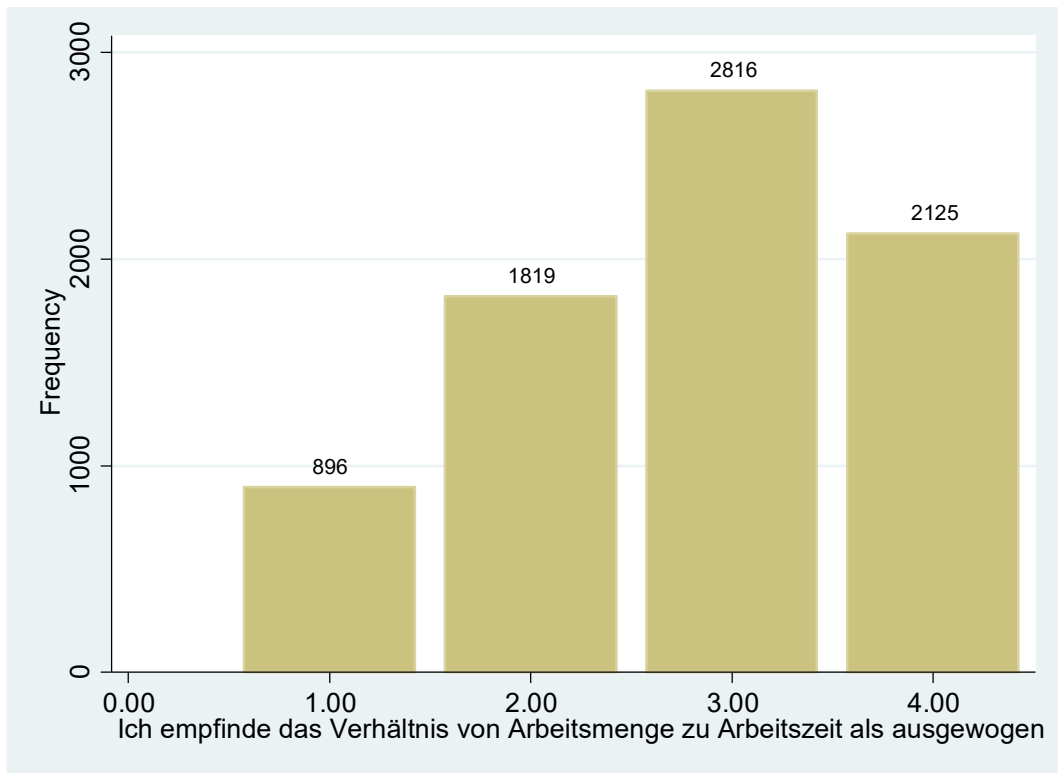


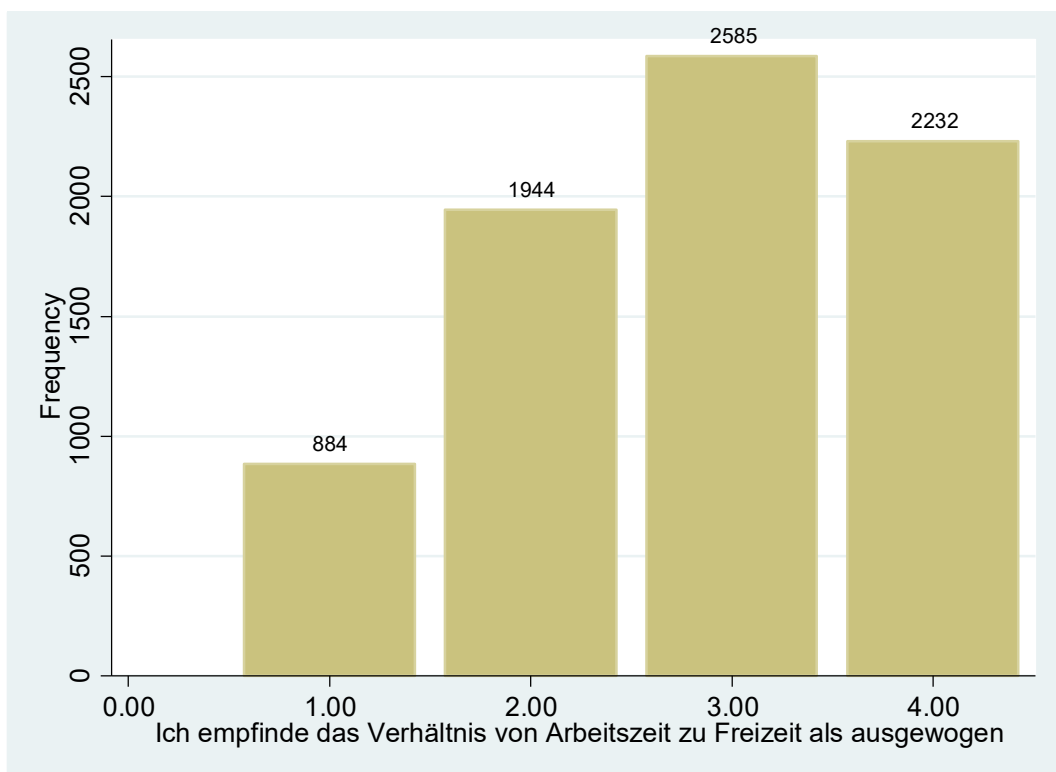
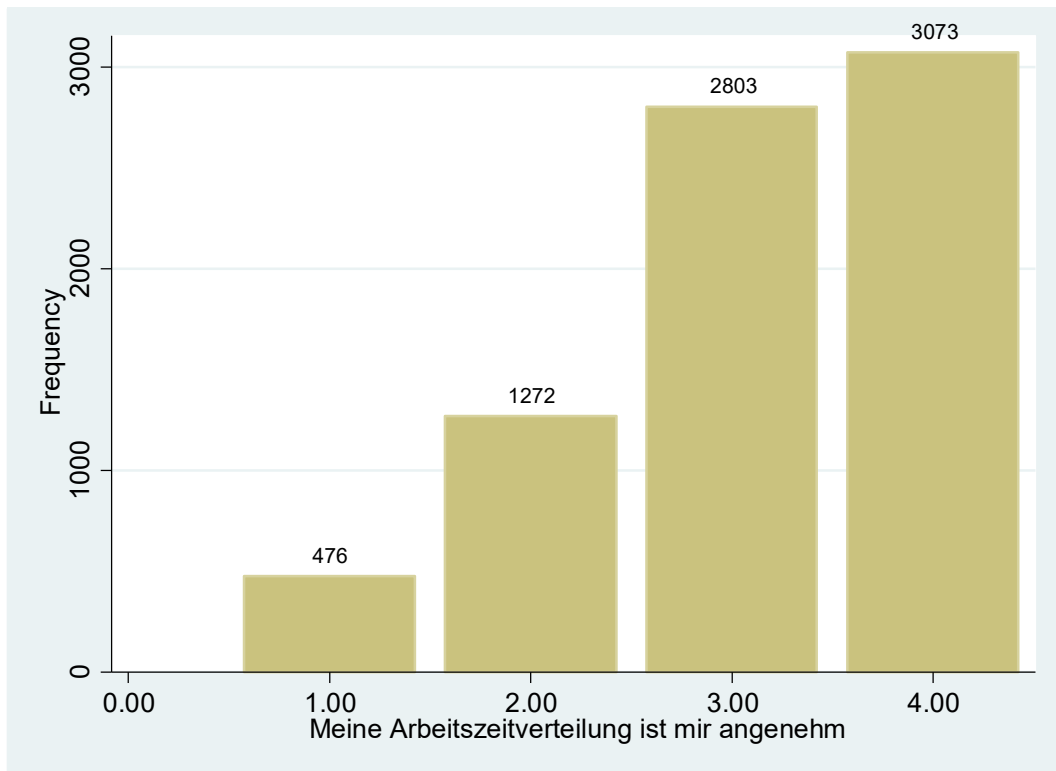


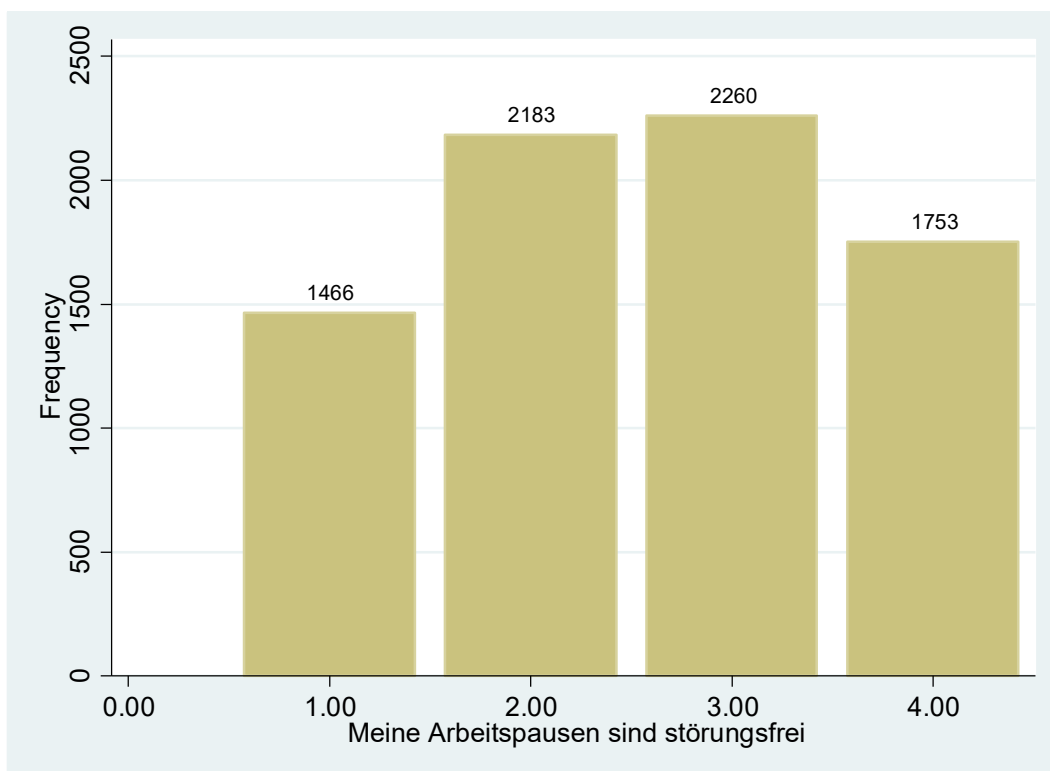
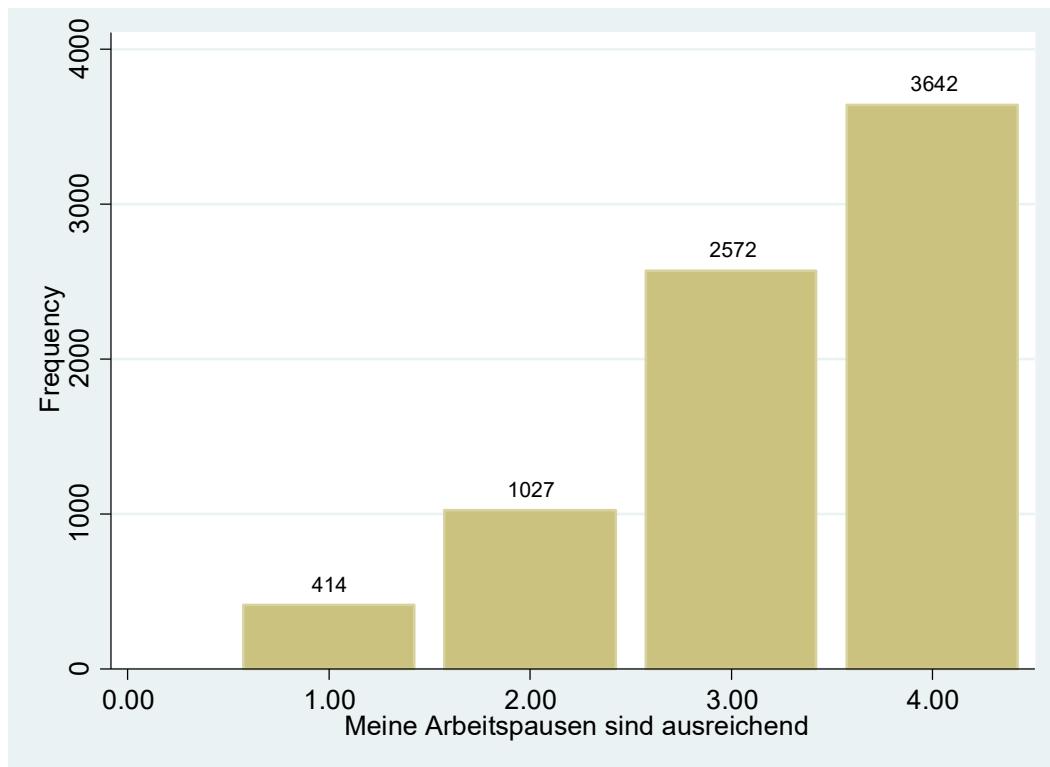


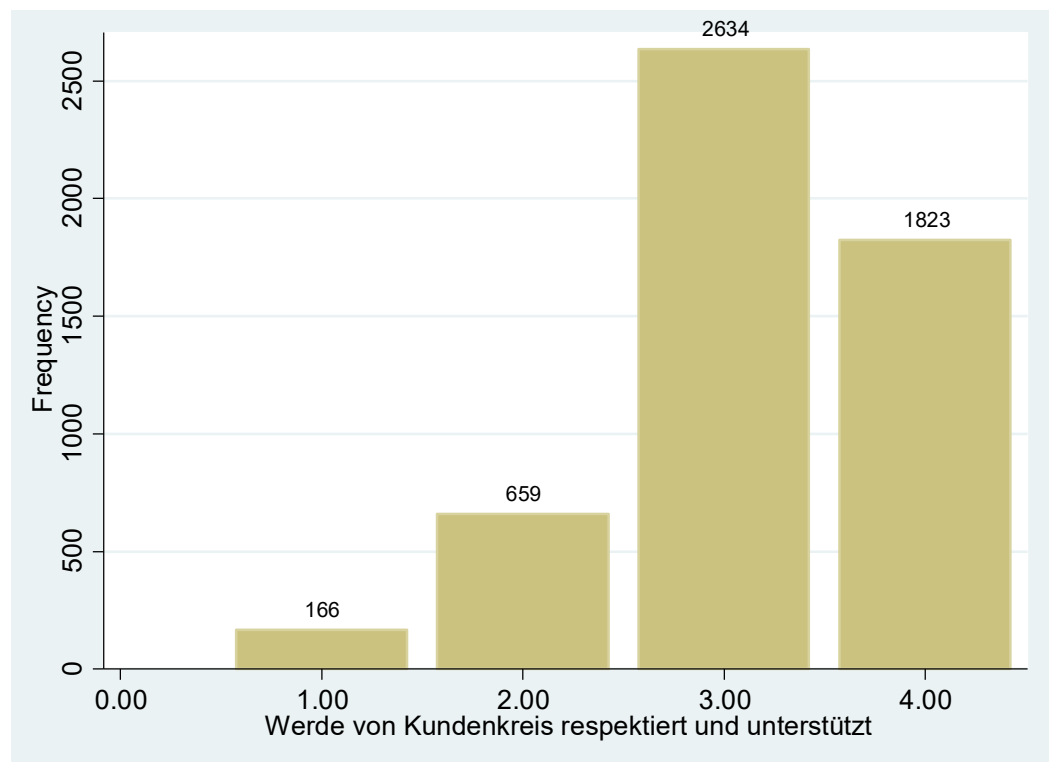
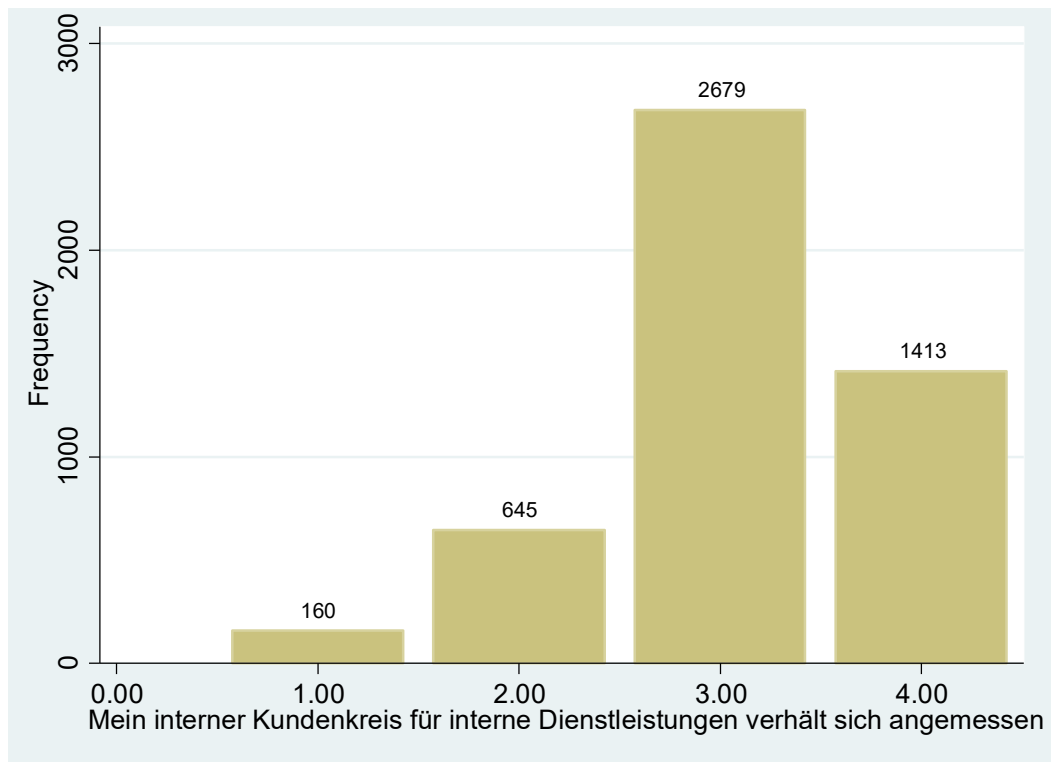


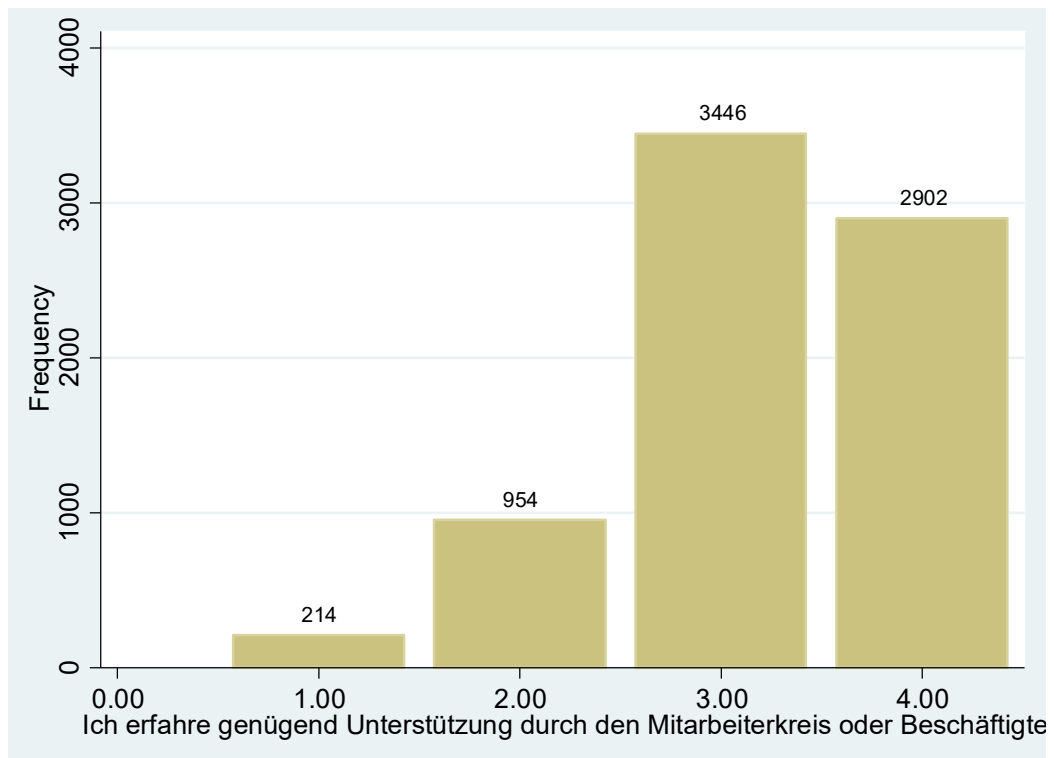
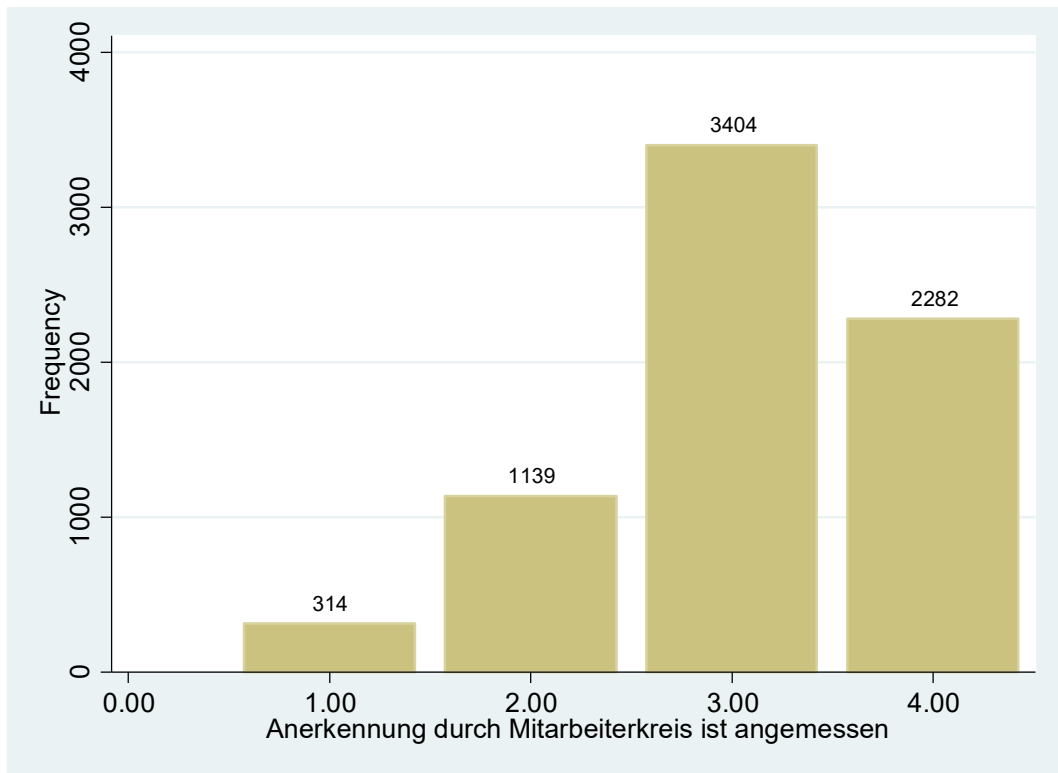


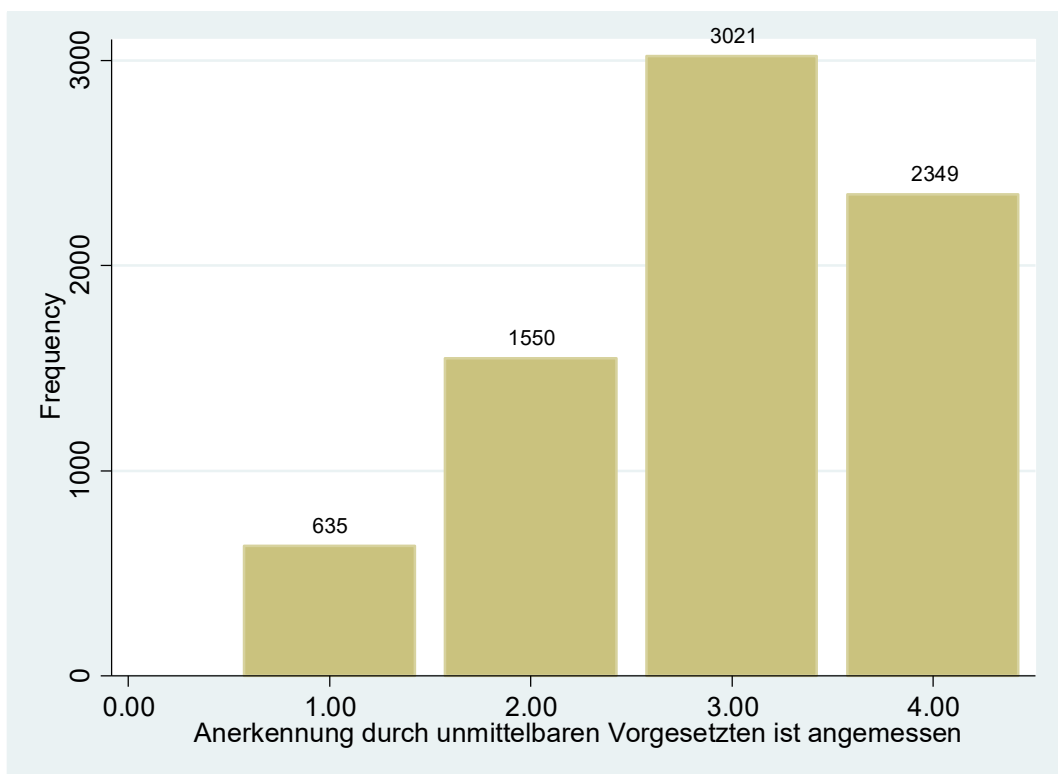
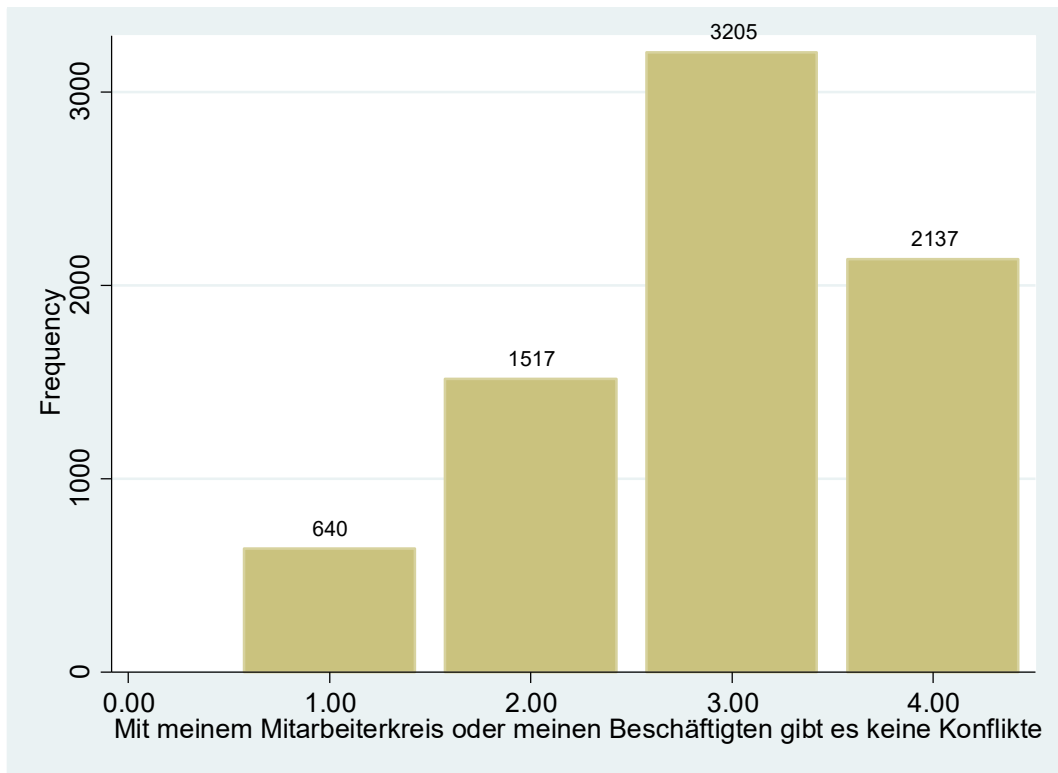


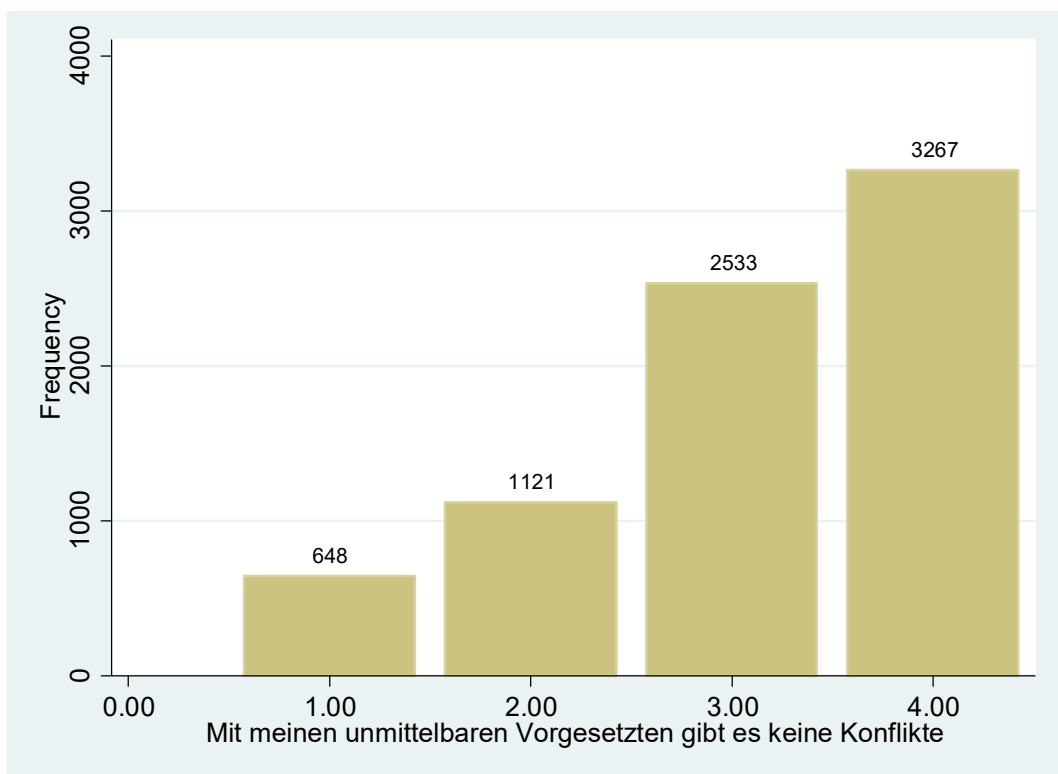
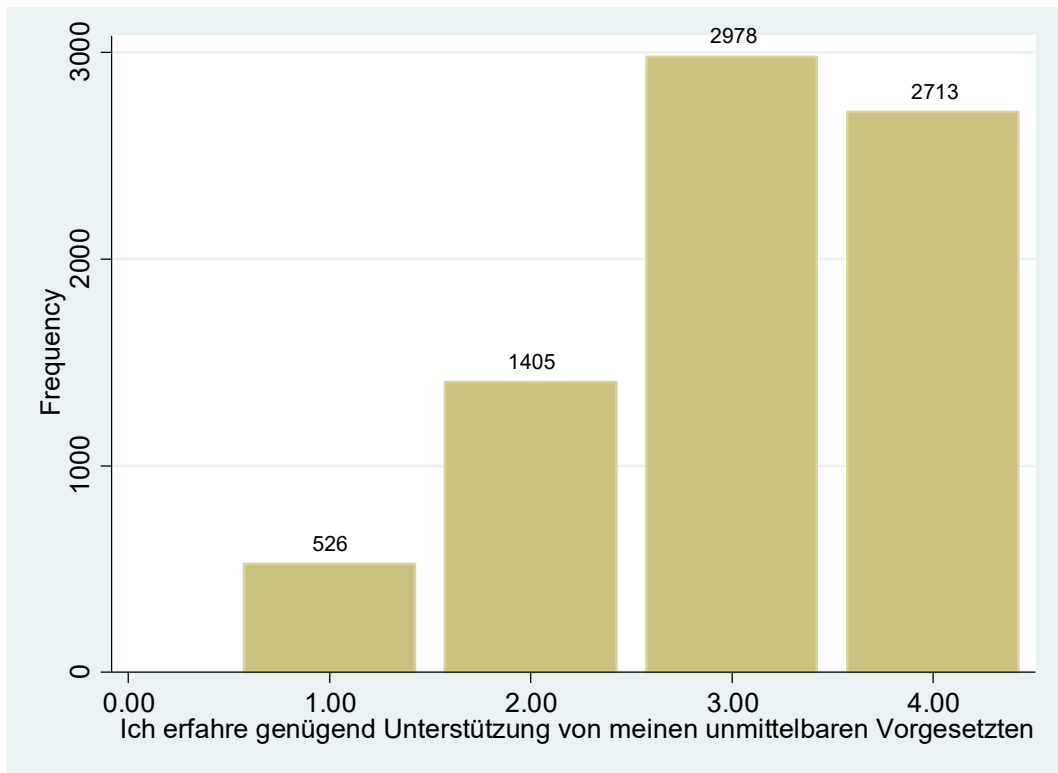


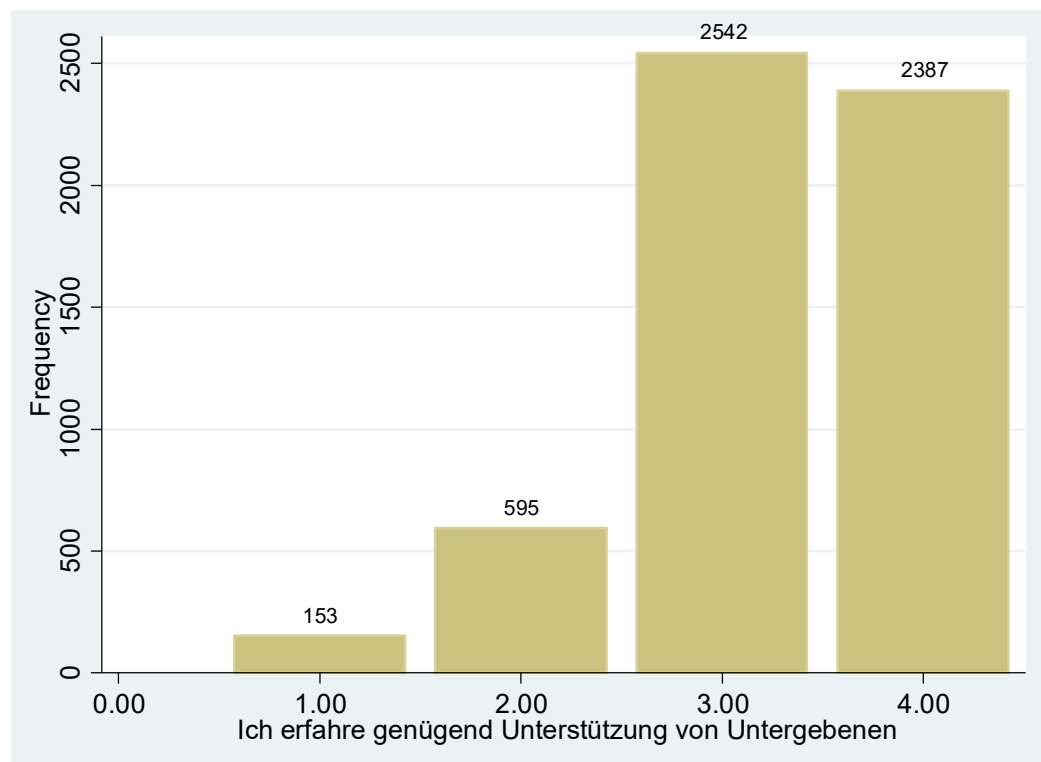
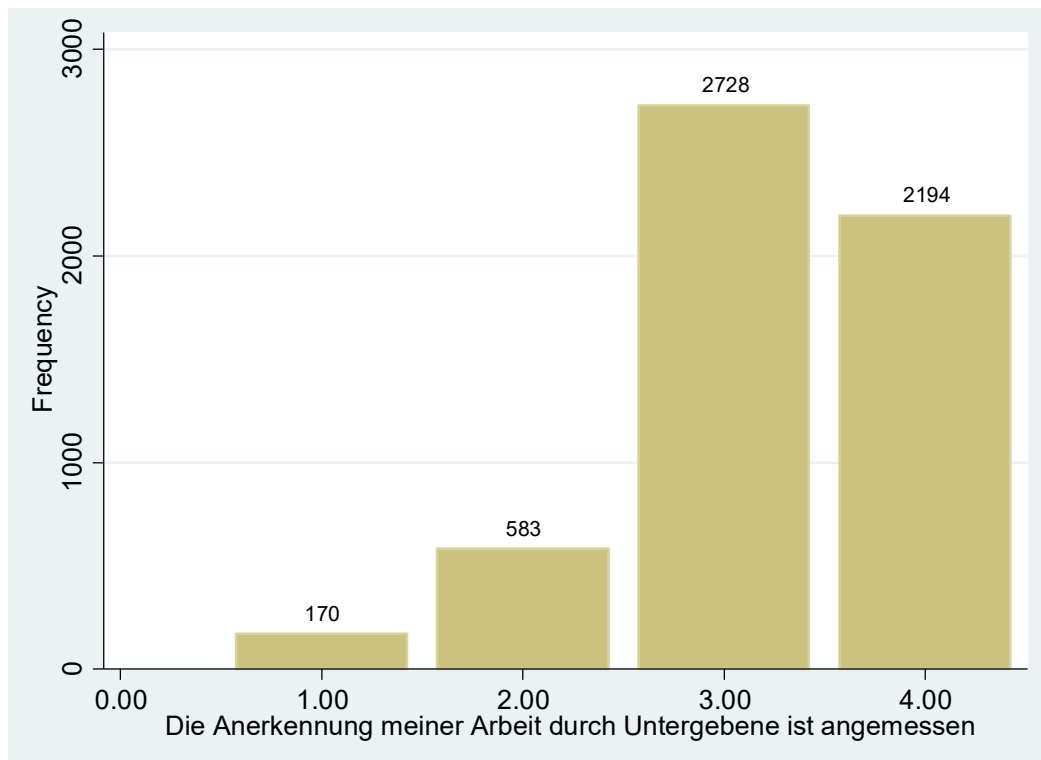


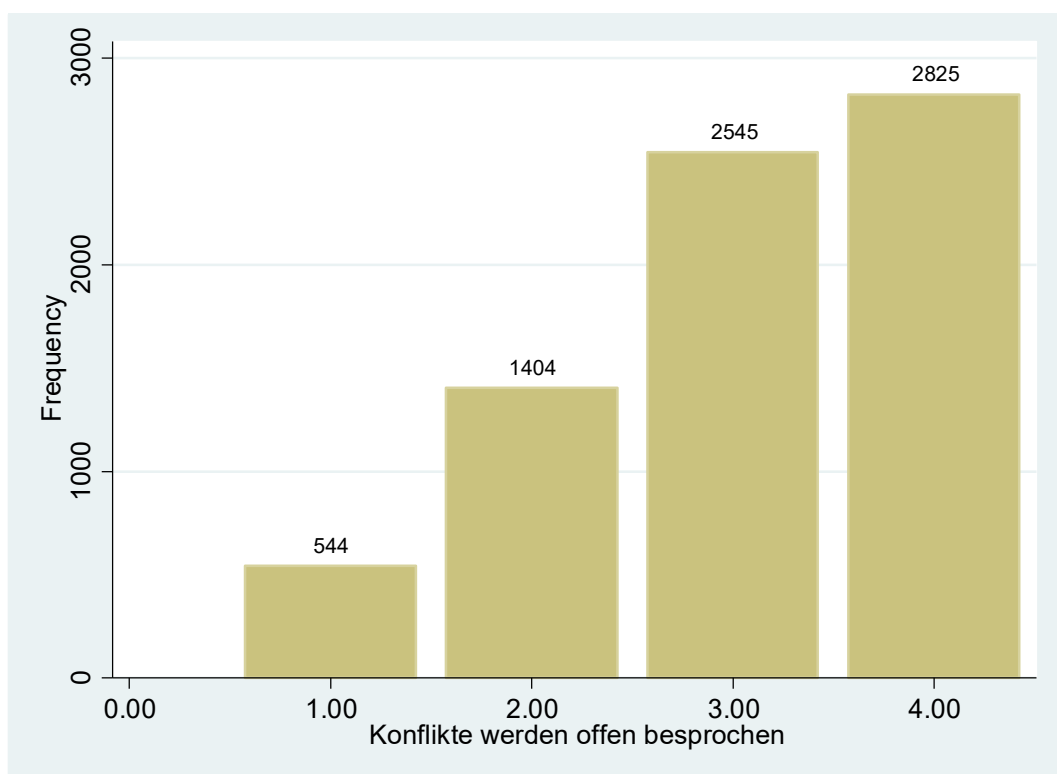
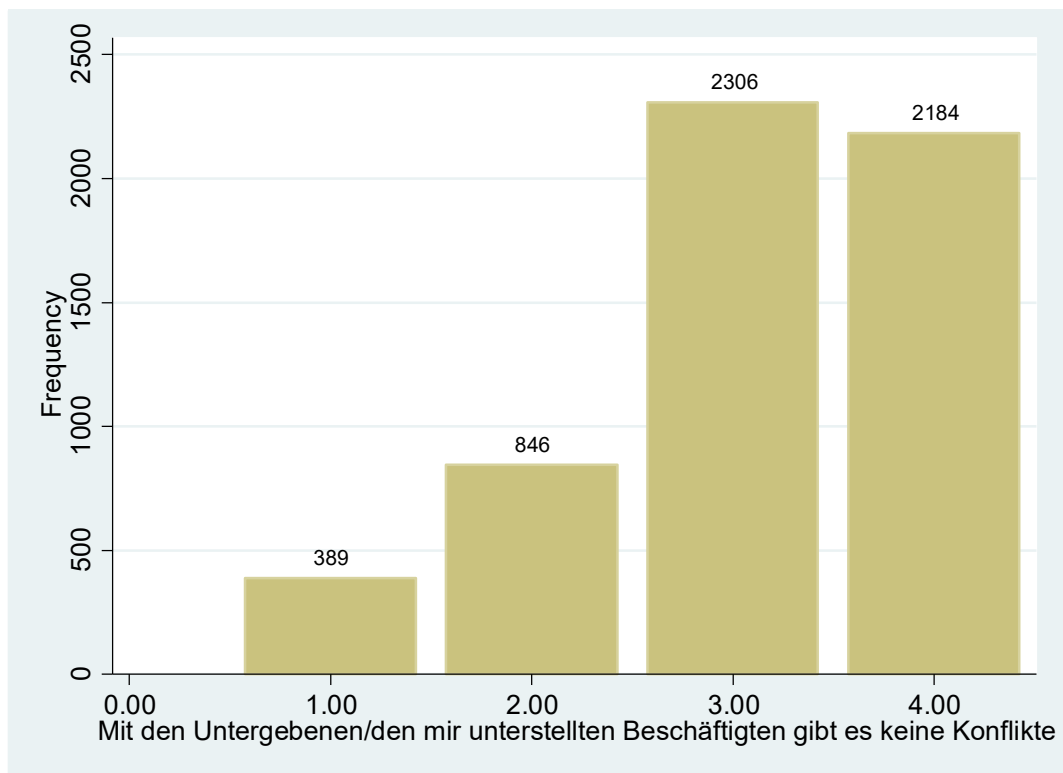


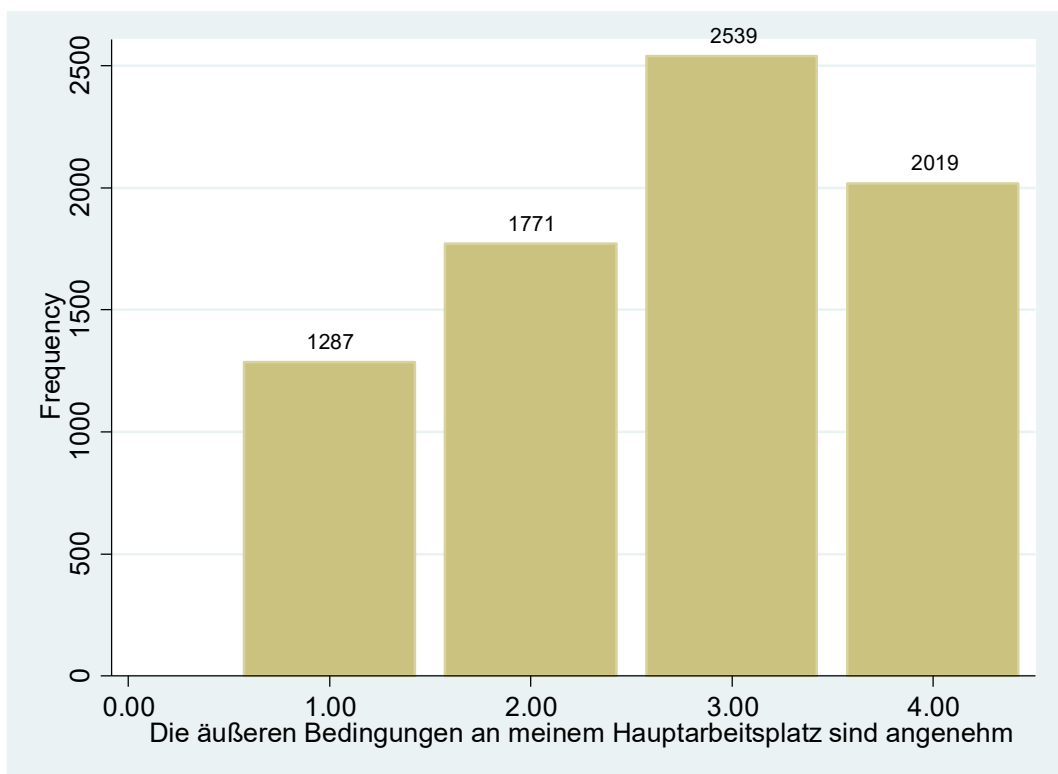
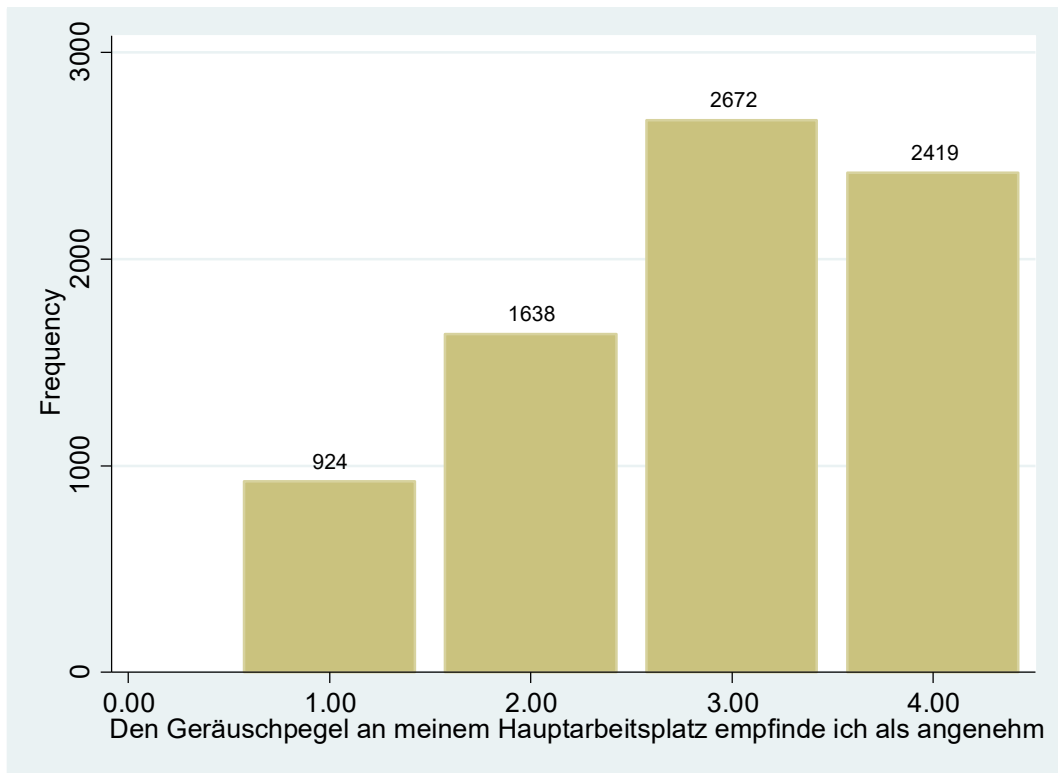


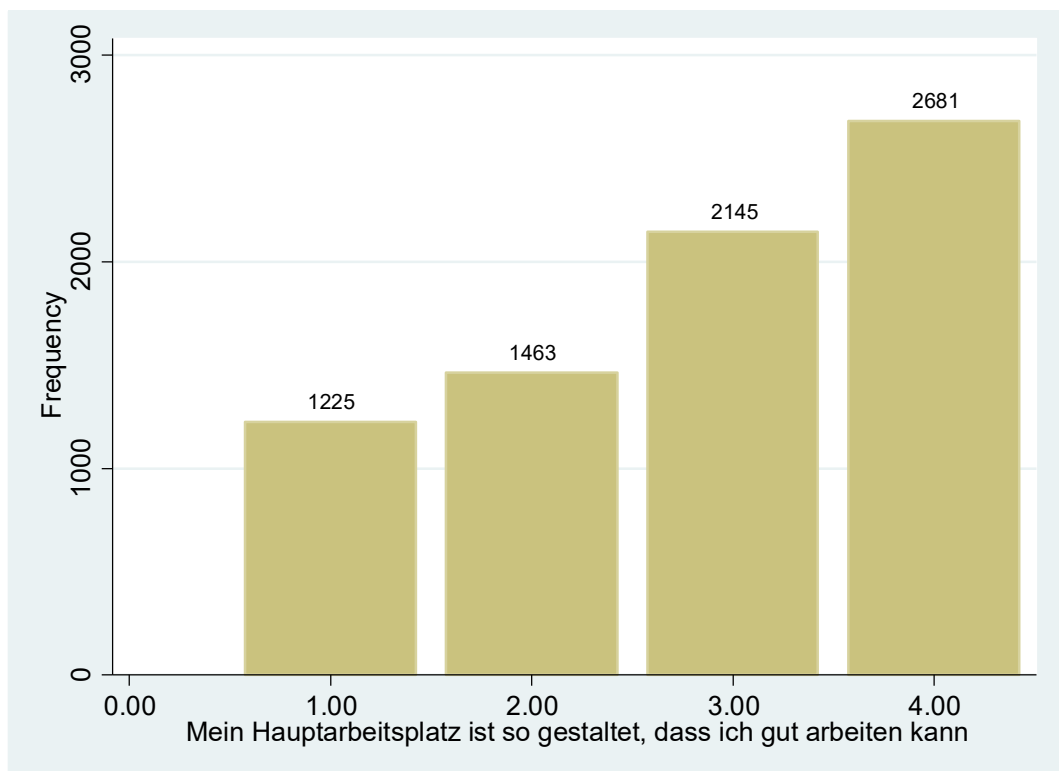
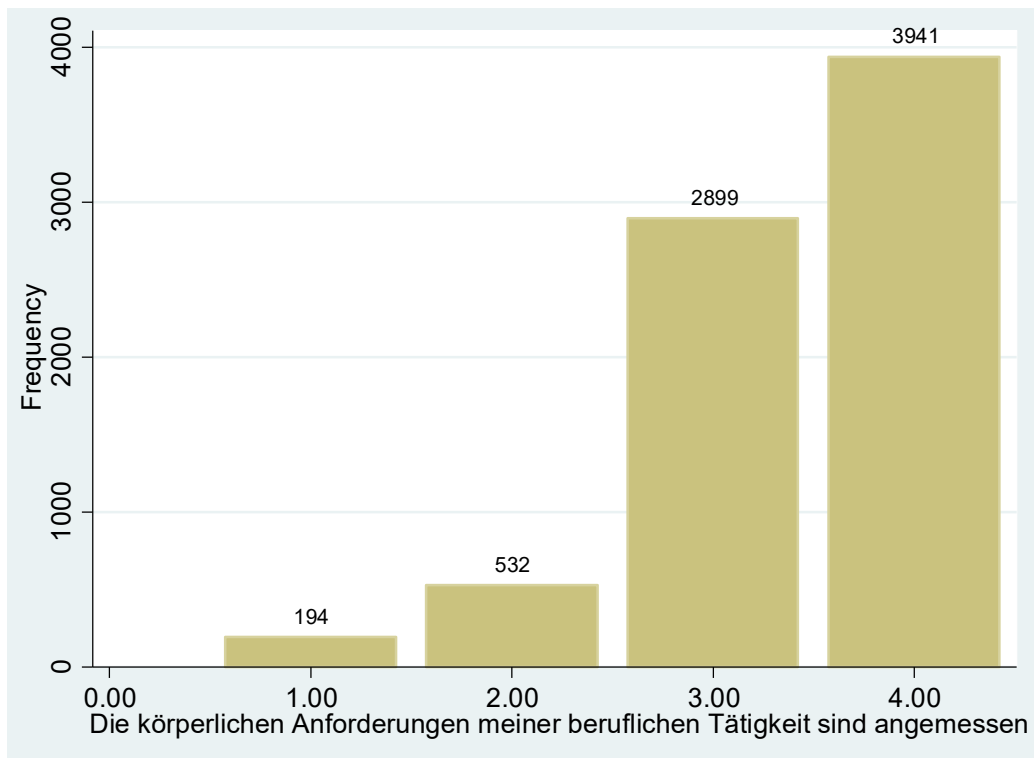












Anlage 6: Itemkorrelationen nach Pearson und Spearman

Item		aa1	aa2	aa3	aa4	aa5	aa6	am1	am2	am3	am4
aa1	r	10000									
	N	7656									
aa2	r	0.4776*	10000								
	rho	0.4722*									
	p (r)	0.0000									
	p (rho)	0.0000									
	N	6472	6522								
aa3	r	0.5157*	0.4615*	10000							
	rho	0.5155*	0.4644*	10000							
	p (r)	0.0000	0.0000								
	p (rho)	0.0000	0.0000								
	N	7547	6461	7624							
aa4	r	0.4443*	0.3834*	0.5671*	10000						
	rho	0.4408*	0.3784*	0.5666*	10000						
	p (r)	0.0000	0.0000	0.0000							
	p (rho)	0.0000	0.0000	0.0000							
	N	7574	6466	7546	7645						
aa5	r	0.4052*	0.3911*	0.4384*	0.4951*	10000					
	rho	0.4022*	0.3992*	0.4510*	0.4997*	10000					
	p (r)	0.0000	0.0000	0.0000	0.0000						
	p (rho)	0.0000	0.0000	0.0000	0.0000						
	N	7575	6464	7546	7575	7655					

aa6	r	0.3883*	0.3727*	0.3909*	0.4019*	0.5231*	10000				
	rho	0.3850*	0.3659*	0.3916*	0.4015*	0.5196*	10000				
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000					
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000					
	N	7581	6468	7552	7579	7624	7662				
am1	r	0.2401*	0.2183*	0.2635*	0.2146*	0.2083*	0.1977*	10000			
	rho	0.2411*	0.2188*	0.2679*	0.2147*	0.2245*	0.1938*	10000			
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
	N	7576	6473	7543	7564	7573	7584	7662			
am2	r	0.2048*	0.1810*	0.2365*	0.2065*	0.1975*	0.2303*	0.4376*	10000		
	rho	0.2042*	0.1779*	0.2369*	0.2060*	0.1961*	0.2281*	0.4335*	10000		
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
	N	7425	6388	7397	7420	7431	7438	7463	7512		
am3	r	0.1738*	0.1126*	0.2119*	0.1789*	0.1448*	0.1174*	0.3055*	0.3405*	10000	
	rho	0.1647*	0.1105*	0.2077*	0.1796*	0.1535*	0.1150*	0.3016*	0.3399*	10000	
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
	N	7563	6452	7537	7554	7567	7575	7587	7444	7653	
am4	r	0.1996*	0.2360*	0.2020*	0.1771*	0.1802*	0.1326*	0.2987*	0.2861*	0.3994*	10000
	rho	0.2002*	0.2455*	0.2183*	0.1816*	0.2019*	0.1320*	0.3095*	0.2873*	0.4068*	10000
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	N	6235	5574	6215	6221	6233	6237	6251	6186	6245	6297

am5	r	0.2400*	0.2847*	0.2431*	0.1996*	0.1973*	0.1834*	0.2805*	0.2536*	0.2979*	0.5973*
	rho	0.2410*	0.2925*	0.2625*	0.2063*	0.2150*	0.1785*	0.2819*	0.2567*	0.3059*	0.6054*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	6235	5539	6209	6221	6236	6240	6256	6168	6244	5650
am6	r	0.1683*	0.1286*	0.1793*	0.1495*	0.1555*	0.1155*	0.2972*	0.1911*	0.2742*	0.2729*
	rho	0.1606*	0.1297*	0.1764*	0.1471*	0.1625*	0.1040*	0.2987*	0.1815*	0.2618*	0.2767*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7554	6445	7524	7542	7556	7562	7573	7428	7565	6243
am7	r	0.1711*	0.1828*	0.2183*	0.1897*	0.1909*	0.2221*	0.2164*	0.2565*	0.2589*	0.2427*
	rho	0.1686*	0.1736*	0.2175*	0.1873*	0.1949*	0.2156*	0.2165*	0.2522*	0.2583*	0.2439*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7524	6449	7499	7514	7524	7535	7544	7401	7542	6222
am8	r	0.1355*	0.1350*	0.2044*	0.1714*	0.1811*	0.1428*	0.2738*	0.2467*	0.3278*	0.3209*
	rho	0.1295*	0.1348*	0.2027*	0.1625*	0.1765*	0.1319*	0.2735*	0.2468*	0.3217*	0.3109*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7524	6440	7498	7516	7525	7534	7546	7407	7540	6247
am9	r	0.3602*	0.2757*	0.3112*	0.3060*	0.3294*	0.5027*	0.1920*	0.2046*	0.0925*	0.1253*
	rho	0.3634*	0.2701*	0.3139*	0.3097*	0.3247*	0.5054*	0.1837*	0.2040*	0.0865*	0.1190*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7555	6449	7527	7550	7555	7563	7571	7425	7566	6233
am10	r	0.2229*	0.2209*	0.2835*	0.2221*	0.2080*	0.2656*	0.2467*	0.2312*	0.1283*	0.1642*

	rho	0.2214*	0.2170*	0.2853*	0.2198*	0.2157*	0.2654*	0.2360*	0.2281*	0.1232*	0.1705*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7574	6464	7544	7563	7570	7582	7589	7438	7585	6249
am11	r	0.2262*	0.1992*	0.2687*	0.2402*	0.2226*	0.2693*	0.2027*	0.1863*	0.1566*	0.1539*
	rho	0.2209*	0.1983*	0.2722*	0.2392*	0.2369*	0.2710*	0.2042*	0.1870*	0.1627*	0.1740*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7530	6434	7502	7519	7530	7540	7543	7397	7542	6219
am12	r	0.2114*	0.2068*	0.2881*	0.2528*	0.2124*	0.2628*	0.2371*	0.2161*	0.1843*	0.1586*
	rho	0.2084*	0.2106*	0.2894*	0.2519*	0.2247*	0.2642*	0.2366*	0.2177*	0.1854*	0.1781*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7509	6415	7478	7500	7506	7517	7522	7379	7518	6204
sa1	r	0.2788*	0.2805*	0.2879*	0.2433*	0.2597*	0.2729*	0.2741*	0.2288*	0.2467*	0.3050*
	rho	0.2838*	0.2809*	0.2908*	0.2469*	0.2731*	0.2741*	0.2720*	0.2261*	0.2414*	0.3123*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	4845	4319	4833	4840	4842	4834	4866	4806	4850	4257
sa2	r	0.2227*	0.2134*	0.2478*	0.2217*	0.2139*	0.2162*	0.2910*	0.2308*	0.2950*	0.3142*
	rho	0.2251*	0.2106*	0.2524*	0.2235*	0.2280*	0.2164*	0.2892*	0.2285*	0.2888*	0.3128*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	5225	4616	5204	5216	5222	5219	5246	5172	5231	4556
sa3	r	0.1895*	0.1867*	0.2393*	0.2178*	0.1950*	0.1947*	0.2763*	0.2313*	0.3154*	0.3070*
	rho	0.1872*	0.1860*	0.2424*	0.2184*	0.2062*	0.1913*	0.2813*	0.2287*	0.3042*	0.3094*

	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7062	6076	7038	7059	7060	7063	7082	6957	7079	5869
sa4	r	0.2324*	0.2264*	0.2645*	0.2298*	0.2436*	0.2474*	0.2494*	0.2412*	0.2720*	0.2851*
	rho	0.2332*	0.2270*	0.2684*	0.2323*	0.2495*	0.2447*	0.2478*	0.2360*	0.2694*	0.2833*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7427	6366	7396	7418	7425	7435	7446	7301	7440	6150
sa5	r	0.1738*	0.1640*	0.1947*	0.1450*	0.1610*	0.2224*	0.1648*	0.1671*	0.1505*	0.1589*
	rho	0.1806*	0.1677*	0.1984*	0.1470*	0.1679*	0.2287*	0.1657*	0.1668*	0.1458*	0.1598*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7411	6339	7377	7401	7402	7413	7426	7282	7417	6134
sa6	r	0.2183*	0.2153*	0.2606*	0.2315*	0.2394*	0.2442*	0.2412*	0.2476*	0.2732*	0.2749*
	rho	0.2157*	0.2055*	0.2582*	0.2281*	0.2396*	0.2416*	0.2427*	0.2436*	0.2718*	0.2696*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7466	6394	7439	7457	7466	7474	7485	7343	7481	6173
sa7	r	0.2483*	0.2429*	0.2683*	0.2339*	0.2741*	0.2735*	0.2441*	0.2560*	0.2547*	0.2514*
	rho	0.2449*	0.2372*	0.2714*	0.2360*	0.2780*	0.2733*	0.2405*	0.2528*	0.2533*	0.2517*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7530	6435	7501	7522	7529	7537	7551	7406	7540	6220
sa8	r	0.1831*	0.1909*	0.1935*	0.1581*	0.1957*	0.2077*	0.1688*	0.1964*	0.1527*	0.1863*
	rho	0.1965*	0.2042*	0.2103*	0.1662*	0.2069*	0.2191*	0.1781*	0.2003*	0.1557*	0.1925*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7481	6386	7448	7472	7478	7485	7499	7353	7491	6165
sa9	r	0.1724*	0.1703*	0.1978*	0.1904*	0.1928*	0.1660*	0.2251*	0.1948*	0.2486*	0.3057*
	rho	0.1701*	0.1754*	0.2084*	0.1895*	0.2030*	0.1612*	0.2322*	0.1930*	0.2436*	0.3016*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	5611	4994	5595	5608	5609	5613	5623	5562	5618	5227
sa10	r	0.1642*	0.1841*	0.1922*	0.1741*	0.2020*	0.1758*	0.1925*	0.1964*	0.2270*	0.2971*
	rho	0.1655*	0.1905*	0.2026*	0.1794*	0.2133*	0.1683*	0.1997*	0.1920*	0.2225*	0.2943*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	5613	4995	5597	5608	5607	5614	5624	5570	5618	5235
sa11	r	0.1349*	0.1427*	0.1431*	0.1290*	0.1429*	0.1766*	0.1199*	0.1326*	0.1065*	0.1611*
	rho	0.1460*	0.1524*	0.1621*	0.1389*	0.1582*	0.1801*	0.1374*	0.1357*	0.1093*	0.1642*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	5658	5031	5644	5656	5658	5663	5676	5615	5666	5257
sa12	r	0.1301*	0.1265*	0.1713*	0.1260*	0.1497*	0.1398*	0.2132*	0.2024*	0.2316*	0.2500*
	rho	0.1274*	0.1168*	0.1686*	0.1233*	0.1521*	0.1296*	0.2078*	0.1954*	0.2223*	0.2474*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	7227	6209	7199	7221	7229	7236	7249	7121	7244	6020
um1	r	0.2288*	0.2560*	0.3019*	0.2709*	0.2225*	0.3511*	0.1898*	0.1897*	0.1731*	0.1700*
	rho	0.2289*	0.2539*	0.2962*	0.2632*	0.2233*	0.3465*	0.1895*	0.1921*	0.1744*	0.1803*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	<i>N</i>	7559	6447	7529	7546	7558	7565	7577	7423	7566	6230
um2	r	0.2046*	0.2006*	0.2911*	0.2757*	0.2130*	0.3146*	0.1889*	0.2037*	0.1907*	0.1538*
	rho	0.2078*	0.2004*	0.2887*	0.2741*	0.2164*	0.3128*	0.1874*	0.2029*	0.1866*	0.1623*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>N</i>	7524	6415	7491	7511	7521	7530	7540	7390	7529	6203
um3	r	0.2244*	0.2818*	0.2865*	0.2462*	0.2451*	0.2228*	0.2489*	0.1942*	0.2372*	0.2612*
	rho	0.2259*	0.3087*	0.3026*	0.2517*	0.2724*	0.2281*	0.2533*	0.1902*	0.2274*	0.2854*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>N</i>	7474	6391	7445	7461	7477	7480	7489	7346	7485	6188
um4	r	0.2086*	0.1971*	0.2985*	0.2703*	0.2202*	0.3041*	0.2146*	0.2082*	0.1781*	0.1525*
	rho	0.2092*	0.1993*	0.3013*	0.2690*	0.2269*	0.3004*	0.2112*	0.2097*	0.1852*	0.1716*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>N</i>	7427	6361	7397	7410	7425	7431	7444	7297	7431	6140

Item		am5	am6	am7	am8	am9	am10	am11	am12	sa1	sa2	sa3	sa4	sa5
am6	r	0.2641*	10000											
	rho	0.2728*	10000											
	p (r)	0.0000												
	p (rho)	0.0000												
	<i>N</i>	6253	7640											
am7	r	0.2472*	0.3691*	10000										
	rho	0.2482*	0.3525*	10000										
	p (r)	0.0000	0.0000											

	p (rho)	0.0000	0.0000					
	N	6232	7547	7610				
am8	r	0.2802*	0.3014*	0.3738*	10000			
	rho	0.2733*	0.2908*	0.3774*	10000			
	p (r)	0.0000	0.0000	0.0000				
	p (rho)	0.0000	0.0000	0.0000				
	N	6244	7537	7521	7610			
am9	r	0.1675*	0.1368*	0.1947*	0.1842*	10000		
	rho	0.1585*	0.1196*	0.1865*	0.1731*	10000		
	p (r)	0.0000	0.0000	0.0000	0.0000			
	p (rho)	0.0000	0.0000	0.0000	0.0000			
	N	6237	7550	7531	7539	7638		
am10	r	0.2066*	0.1234*	0.1932*	0.2050*	0.2679*	10000	
	rho	0.2068*	0.1211*	0.1890*	0.2007*	0.2640*	10000	
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000		
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000		
	N	6252	7571	7548	7553	7581	7662	
am11	r	0.1857*	0.1300*	0.1700*	0.1807*	0.2586*	0.5107*	10000
	rho	0.1952*	0.1327*	0.1718*	0.1867*	0.2529*	0.5130*	10000
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	N	6223	7527	7506	7508	7533	7557	7616
am12	r	0.2145*	0.1154*	0.1782*	0.1976*	0.2346*	0.5467*	0.6116*
	rho	0.2150*	0.1190*	0.1771*	0.2035*	0.2327*	0.5497*	0.6113*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	N	6202	7507	7483	7491	7512	7536	7511	7591					
sa1	r	0.3074*	0.2239*	0.2185*	0.2346*	0.2589*	0.3017*	0.2952*	0.3079*	10000				
	rho	0.3079*	0.2080*	0.2164*	0.2308*	0.2556*	0.2980*	0.3015*	0.3070*	10000				
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000					
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000					
	N	4218	4858	4839	4852	4848	4859	4850	4835	4897				
sa2	r	0.2849*	0.2929*	0.3172*	0.3249*	0.2245*	0.2670*	0.2594*	0.2708*	0.5737*	10000			
	rho	0.2841*	0.2798*	0.3080*	0.3166*	0.2195*	0.2642*	0.2635*	0.2749*	0.5615*	10000			
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
	N	4495	5239	5220	5229	5229	5240	5227	5213	4826	5282			
sa3	r	0.2792*	0.3207*	0.4230*	0.3604*	0.1939*	0.2300*	0.2237*	0.2521*	0.3878*	0.5720*	10000		
	rho	0.2846*	0.3051*	0.4153*	0.3555*	0.1840*	0.2308*	0.2298*	0.2603*	0.3851*	0.5558*	10000		
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
	N	5872	7072	7055	7053	7071	7079	7042	7029	4814	5198	7139		
sa4	r	0.2785*	0.2488*	0.3566*	0.2866*	0.2357*	0.2603*	0.2362*	0.2530*	0.3286*	0.4212*	0.5618*	10000	
	rho	0.2724*	0.2405*	0.3470*	0.2796*	0.2237*	0.2620*	0.2493*	0.2624*	0.3315*	0.4192*	0.5511*	10000	
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
	N	6160	7427	7409	7407	7431	7444	7405	7379	4836	5219	7079	7516	
sa5	r	0.1749*	0.1357*	0.2141*	0.2017*	0.2128*	0.1878*	0.1717*	0.1885*	0.2393*	0.2581*	0.3141*	0.3609*	10000
	rho	0.1767*	0.1397*	0.2141*	0.2017*	0.2179*	0.1959*	0.1852*	0.2008*	0.2487*	0.2571*	0.3165*	0.3671*	10000
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	N	6129	7411	7385	7384	7413	7431	7387	7366	4800	5182	7006	7396	7499

sa6	r	0.2710*	0.3131*	0.7133*	0.3974*	0.2250*	0.2330*	0.2145*	0.2284*	0.2725*	0.3711*	0.5055*	0.4315*	0.2867*
	rho	0.2647*	0.2949*	0.7025*	0.3930*	0.2182*	0.2346*	0.2207*	0.2349*	0.2728*	0.3700*	0.5014*	0.4206*	0.2972*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	6179	7471	7468	7448	7475	7493	7447	7429	4820	5202	7045	7403	7382
sa7	r	0.2833*	0.2697*	0.6226*	0.3531*	0.2454*	0.2484*	0.2150*	0.2271*	0.2674*	0.3307*	0.4083*	0.4467*	0.2660*
	rho	0.2777*	0.2534*	0.6111*	0.3437*	0.2410*	0.2486*	0.2232*	0.2374*	0.2702*	0.3307*	0.4032*	0.4387*	0.2756*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	6224	7529	7525	7507	7532	7550	7509	7485	4848	5230	7081	7457	7436
sa8	r	0.1994*	0.1798*	0.3713*	0.2816*	0.2023*	0.1854*	0.1603*	0.1825*	0.2081*	0.2427*	0.2707*	0.2750*	0.4505*
	rho	0.2094*	0.1835*	0.3785*	0.2805*	0.2108*	0.2000*	0.1769*	0.1975*	0.2178*	0.2528*	0.2794*	0.2822*	0.4508*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	6174	7477	7468	7451	7483	7502	7455	7432	4805	5183	7020	7400	7399
sa9	r	0.2574*	0.2749*	0.2395*	0.2715*	0.1734*	0.2054*	0.1887*	0.1988*	0.3213*	0.3989*	0.4692*	0.4180*	0.3084*
	rho	0.2636*	0.2787*	0.2341*	0.2568*	0.1572*	0.2047*	0.1998*	0.2060*	0.3144*	0.3957*	0.4653*	0.4178*	0.3239*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	4957	5620	5610	5627	5619	5633	5605	5597	3905	4175	5370	5600	5580
sa10	r	0.2484*	0.2226*	0.2097*	0.2429*	0.1633*	0.2114*	0.1991*	0.2016*	0.3211*	0.3722*	0.4154*	0.4681*	0.3184*
	rho	0.2514*	0.2272*	0.2085*	0.2332*	0.1507*	0.2084*	0.2121*	0.2108*	0.3189*	0.3697*	0.4132*	0.4657*	0.3253*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	4957	5621	5613	5630	5626	5635	5607	5605	3907	4177	5367	5608	5588

sa11	r	0.1615*	0.1309*	0.1315*	0.1422*	0.1793*	0.1596*	0.1428*	0.1392*	0.2202*	0.2219*	0.2350*	0.2539*	0.5404*
	rho	0.1690*	0.1459*	0.1348*	0.1475*	0.1837*	0.1738*	0.1590*	0.1488*	0.2320*	0.2322*	0.2448*	0.2604*	0.5361*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	4989	5670	5657	5670	5677	5683	5656	5653	3921	4199	5392	5650	5637
sa12	r	0.2070*	0.2409*	0.3350*	0.3035*	0.1513*	0.1633*	0.1406*	0.1613*	0.2113*	0.2924*	0.3851*	0.3769*	0.3153*
	rho	0.2039*	0.2307*	0.3241*	0.2903*	0.1377*	0.1563*	0.1369*	0.1629*	0.2047*	0.2814*	0.3676*	0.3610*	0.3139*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	6021	7235	7226	7217	7238	7256	7209	7191	4694	5059	6814	7168	7161
um1	r	0.1946*	0.0840*	0.1714*	0.1777*	0.2822*	0.2752*	0.3262*	0.3742*	0.2875*	0.2493*	0.2157*	0.2375*	0.2073*
	rho	0.2004*	0.0778*	0.1702*	0.1765*	0.2811*	0.2768*	0.3296*	0.3752*	0.2851*	0.2477*	0.2166*	0.2420*	0.2142*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	6235	7555	7532	7528	7557	7579	7540	7520	4857	5240	7080	7452	7433
um2	r	0.1847*	0.1032*	0.1737*	0.1740*	0.2301*	0.3419*	0.3804*	0.4742*	0.2477*	0.2253*	0.2129*	0.2277*	0.1831*
	rho	0.1866*	0.1007*	0.1727*	0.1749*	0.2301*	0.3419*	0.3781*	0.4715*	0.2468*	0.2251*	0.2165*	0.2311*	0.1916*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	6206	7517	7494	7490	7523	7539	7505	7489	4835	5219	7048	7416	7403
um3	r	0.2697*	0.1759*	0.1746*	0.2071*	0.1432*	0.2637*	0.2513*	0.2962*	0.3105*	0.2652*	0.2446*	0.2515*	0.1750*
	rho	0.2903*	0.1748*	0.1790*	0.2109*	0.1437*	0.2757*	0.2698*	0.3172*	0.3105*	0.2689*	0.2516*	0.2625*	0.1855*
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	6209	7471	7442	7443	7477	7497	7450	7430	4813	5184	7009	7377	7352
um4	r	0.1987*	0.1026*	0.1524*	0.1661*	0.2346*	0.3969*	0.4421*	0.6177*	0.2569*	0.2350*	0.2198*	0.2310*	0.1789*

rho	0.2086*	0.1001*	0.1551*	0.1724*	0.2313*	0.3974*	0.4432*	0.6124*	0.2607*	0.2383*	0.2299*	0.2406*	0.1930*
p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	6124	7422	7401	7402	7431	7440	7419	7415	4824	5202	6982	7326	7307

Item		sa6	sa7	sa8	sa9	sa10	sa11	sa12	um1	um2	um3
sa7	r	0.7493*	10000								
	rho	0.7335*	10000								
	p (r)	0.0000									
	p (rho)	0.0000									
	N	7523	7622								
sa8	r	0.4631*	0.5043*	10000							
	rho	0.4725*	0.5193*	10000							
	p (r)	0.0000	0.0000								
	p (rho)	0.0000	0.0000								
	N	7457	7524	7569							
sa9	r	0.3303*	0.2844*	0.2386*	10000						
	rho	0.3341*	0.2915*	0.2536*	10000						
	p (r)	0.0000	0.0000	0.0000							
	p (rho)	0.0000	0.0000	0.0000							
	N	5610	5639	5597	5675						
sa10	r	0.2929*	0.2886*	0.2292*	0.7028*	10000					
	rho	0.2979*	0.3001*	0.2426*	0.6942*	10000					
	p (r)	0.0000	0.0000	0.0000	0.0000						
	p (rho)	0.0000	0.0000	0.0000	0.0000						
	N	5610	5643	5604	5535	5677					

sa11	r	0.1907*	0.1836*	0.4586*	0.4419*	0.4829*	10000				
	rho	0.1998*	0.1992*	0.4476*	0.4658*	0.4958*	10000				
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000					
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000					
	N	5658	5690	5661	5507	5594	5725				
sa12	r	0.3808*	0.3671*	0.2912*	0.3813*	0.3837*	0.2669*	10000			
	rho	0.3655*	0.3574*	0.2891*	0.3741*	0.3761*	0.2780*	10000			
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
	N	7208	7264	7218	5522	5539	5586	7318			
um1	r	0.2153*	0.2147*	0.1830*	0.2002*	0.1934*	0.1621*	0.1800*	10000		
	rho	0.2195*	0.2220*	0.1945*	0.2042*	0.1987*	0.1724*	0.1773*	10000		
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
	N	7490	7556	7503	5634	5638	5688	7257	7653		
um2	r	0.2118*	0.2066*	0.1656*	0.1961*	0.1791*	0.1415*	0.1642*	0.5751*	10000	
	rho	0.2148*	0.2105*	0.1796*	0.1974*	0.1824*	0.1475*	0.1618*	0.5717*	10000	
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
	N	7454	7518	7468	5611	5617	5669	7222	7585	7616	
um3	r	0.2224*	0.2377*	0.1956*	0.2516*	0.2412*	0.1558*	0.1776*	0.3377*	0.3452*	10000
	rho	0.2282*	0.2459*	0.2159*	0.2643*	0.2532*	0.1758*	0.1796*	0.3521*	0.3579*	10000
	p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
	N	7413	7470	7421	5589	5591	5638	7188	7519	7481	7566
um4	r	0.2048*	0.2152*	0.1782*	0.1975*	0.1833*	0.1565*	0.1493*	0.4537*	0.5358*	0.3741*

rho	0.2122*	0.2252*	0.1958*	0.2063*	0.1933*	0.1686*	0.1545*	0.4528*	0.5341*	0.3947*
p (r)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
p (rho)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	7367	7424	7368	5576	5581	5630	7137	7485	7472	7382

r = Korrelationskoeffizient (Pearson), rho = Rangkorrelationskoeffizient (Spearman), p = Irrtumswahrscheinlichkeit, * = $p \leq 0,05$

Anlage 7: Ergebnisse der explorativen Faktorenanalyse IV: Faktorladungen bei 4 extrahierten Faktoren nach Varimax-Rotation (n = 8.134)

	Faktor 1	Faktor 3	Faktor 3	Faktor 4
aa1	0,11	0,66	0,10	0,10
aa2	0,13	0,58	0,12	0,12
aa3	0,12	0,67	0,11	0,2
aa4	0,10	0,64	0,09	0,16
aa5	0,11	0,64	0,10	0,14
aa6	0,12	0,55	0,12	0,24
am1	0,25	0,28	0,2	0,19
am2	0,20	0,26	0,21	0,20
am3	0,29	0,16	0,23	0,16
am4	0,35	0,23	0,20	0,12
am5	0,30	0,26	0,19	0,16
am6	0,27	0,17	0,30	0,07
am7	0,18	0,12	0,77	0,08
am8	0,28	0,15	0,35	0,16
am9	0,13	0,45	0,11	0,19
am10	0,14	0,18	0,12	0,58
am11	0,11	0,17	0,09	0,65
am12	0,12	0,12	0,10	0,80
sa1	0,37	0,28	0,15	0,26
sa2	0,48	0,18	0,23	0,21
sa3	0,54	0,13	0,39	0,15
sa4	0,53	0,19	0,32	0,15
sa5	0,43	0,14	0,17	0,12
sa6	0,23	0,14	0,85	0,12
sa7	0,22	0,19	0,77	0,12
sa8	0,27	0,14	0,44	0,10
sa9	0,77	0,07	0,12	0,09
sa10	0,78	0,08	0,09	0,10
sa11	0,57	0,09	0,06	0,06
sa12	0,44	0,05	0,29	0,09
um1	0,14	0,22	0,08	0,46
um2	0,10	0,19	0,09	0,56
um3	0,23	0,24	0,10	0,34
um4	0,11	0,15	0,07	0,69

Anlage 8: Ergebnisse der explorativen Faktorenanalyse V: Faktorladungen bei 6 extrahierten Faktoren nach Varimax-Rotation (n = 8.134)

Item	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6
aa1	0,08	0,10	0,04	0,65	0,16	0,08
aa2	0,09	0,11	0,07	0,56	0,18	0,08
aa3	0,09	0,19	0,03	0,65	0,16	0,09
aa4	0,07	0,16	0,00	0,63	0,11	0,11
aa5	0,08	0,15	0,05	0,64	0,09	0,09
aa6	0,11	0,24	0,12	0,58	0,03	0,08
am1	0,14	0,18	0,03	0,23	0,34	0,19
am2	0,16	0,19	0,04	0,21	0,31	0,14
am3	0,16	0,13	0,01	0,09	0,44	0,22
am4	0,09	0,07	0,09	0,11	0,76	0,16
am5	0,09	0,12	0,08	0,16	0,68	0,12
am6	0,25	0,06	0,02	0,12	0,32	0,24
am7	0,73	0,08	0,05	0,12	0,18	0,20
am8	0,30	0,15	0,07	0,10	0,31	0,23
am9	0,10	0,20	0,12	0,47	0,01	0,10
am10	0,10	0,58	0,06	0,18	0,09	0,12
am11	0,07	0,65	0,05	0,17	0,07	0,09
am12	0,07	0,80	0,03	0,11	0,10	0,10
sa1	0,09	0,25	0,05	0,26	0,21	0,40
sa2	0,17	0,20	0,03	0,17	0,20	0,56
sa3	0,34	0,14	0,05	0,12	0,20	0,63
sa4	0,28	0,15	0,14	0,19	0,17	0,55
sa5	0,15	0,13	0,55	0,15	0,06	0,23
sa6	0,83	0,12	0,10	0,14	0,16	0,24
sa7	0,75	0,13	0,14	0,19	0,16	0,19
sa8	0,45	0,11	0,50	0,14	0,10	0,05
sa9	0,08	0,09	0,33	0,07	0,18	0,65
sa10	0,05	0,10	0,38	0,09	0,16	0,63
sa11	0,03	0,07	0,79	0,09	0,06	0,27
sa12	0,26	0,09	0,23	0,04	0,17	0,36
um1	0,07	0,46	0,07	0,22	0,09	0,11
um2	0,07	0,56	0,03	0,18	0,10	0,08
um3	0,06	0,33	0,08	0,21	0,23	0,16
um4	0,05	0,69	0,06	0,15	0,10	0,08

Anlage 9: Ergebnisse der explorativen Faktorenanalyse VI: Faktorladungen bei 8 extrahierten Faktoren nach Oblimin-Rotation (n = 8.134)

Item	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8
aa1	0,00	-0,01	-0,01	0,69	0,06	0,01	0,00	-0,06
aa2	0,01	0,01	0,00	0,58	0,09	0,00	0,02	-0,05
aa3	0,01	0,04	0,02	0,68	0,04	-0,01	-0,04	0,01
aa4	0,01	-0,02	0,05	0,66	0,00	-0,01	-0,07	0,06
aa5	0,01	-0,02	0,03	0,68	-0,02	-0,01	0,00	0,02
aa6	0,03	0,03	-0,02	0,57	-0,09	0,03	0,10	0,11
am1	0,07	0,08	0,02	0,15	0,30	0,08	-0,02	0,01
am2	0,11	0,07	0,00	0,13	0,27	0,02	0,00	0,06
am3	0,11	-0,01	0,05	-0,03	0,42	0,09	-0,05	0,09
am4	-0,02	-0,03	0,02	-0,02	0,80	0,00	0,01	0,02
am5	-0,01	0,05	-0,03	0,05	0,71	-0,01	0,03	-0,02
am6	0,21	-0,01	0,06	0,04	0,26	0,12	-0,06	-0,04
am7	0,79	-0,01	0,01	0,00	0,05	0,01	-0,05	0,01
am8	0,26	0,05	0,03	-0,02	0,25	0,11	0,01	0,04
am9	0,02	0,04	-0,03	0,46	-0,10	0,08	0,11	0,06
am10	0,03	0,62	0,00	0,09	0,00	0,06	0,03	-0,07
am11	0,00	0,65	0,00	0,06	-0,02	0,03	0,02	0,02
am12	0,01	0,88	0,02	-0,04	0,01	-0,01	-0,02	0,01
sa1	-0,08	0,09	-0,05	0,12	0,06	0,62	0,05	0,00
sa2	-0,01	0,00	0,01	-0,02	0,00	0,84	0,00	0,02
sa3	0,27	0,00	0,22	-0,04	0,04	0,45	-0,03	0,02
sa4	0,22	0,02	0,28	0,08	0,02	0,26	0,03	0,01
sa5	0,02	0,02	0,01	0,03	0,02	0,12	0,62	0,03
sa6	0,89	0,02	0,04	0,01	-0,01	0,01	0,00	0,02
sa7	0,78	0,03	0,01	0,07	0,00	-0,01	0,06	0,00
sa8	0,36	0,02	-0,15	0,01	0,05	0,01	0,56	0,02
sa9	0,05	0,00	0,75	0,01	0,03	0,05	0,02	0,02
sa10	0,01	0,03	0,81	0,04	0,01	-0,01	0,05	0,00
sa11	-0,10	0,01	0,26	0,01	0,02	-0,02	0,70	0,00
sa12	0,23	-0,01	0,25	-0,07	0,09	0,07	0,12	0,06
um1	0,00	-0,05	0,00	0,05	0,00	0,04	0,03	0,71
um2	0,02	0,12	0,01	0,00	0,02	-0,02	-0,03	0,66
um3	-0,02	0,05	0,05	0,09	0,18	0,07	0,02	0,30
um4	-0,01	0,49	0,02	0,00	0,02	-0,03	0,01	0,32

Danksagung

Mein Dank gilt meinem Betreuer, Prof. Dr. med. Dirk-Matthias Rose, für seine Unterstützung und grenzenlosen Optimismus auch in scheinbar ausweglosen Lagen.

Ich bin zudem sehr dankbar, die Arbeit in einem Team kluger und aufmerksamer Kolleginnen begonnen zu haben. Auf KB, MRW, MS, DW und NW konnte ich mich zu jederzeit als kritische und konstruktive Gesprächspartner und Korrekturleser verlassen und ohne sie hätte ich nicht halb so viel Freude an der Arbeit entwickeln können. Außerdem danke ich Dr. Jan Becker, Coach für Sinnkrisen, statistische Rätsel, angemessene Blödeleien und konstruktive Lösungsfindung.

Ohne die Unterstützung von Wolfram Schmid, der zwischenzeitlich ein größeres Interesse an der Fertigstellung dieser Arbeit gezeigt hat als ich, hätte mir in der letzten Phase der Antrieb und der nötige Freiraum gefehlt.

Mein lieber Ehemann, Gökhan, hat mehr daran geglaubt, dass ich diese Dissertation zum Abschluss bringen kann, als ich selbst. Ohne seinen Zuspruch und seine Motivation wäre ich nicht so weit gekommen. Ich schätze mich sehr glücklich, einen Partner zu haben, der klaglos alle Einschränkungen und Launen der letzten Monate in Kauf genommen hat.

Mein größter Dank gilt meinen Eltern, Katrin und Rigo Nicolai, für unendliche Unterstützung und Rückhalt bis hin zur Bereitschaft „notfalls den Kitt aus den Fenstern zu fressen“.

Lebenslauf



