

Psychometrische Analysen von Skalen zur Erfassung psychosozialer Aspekte der Arbeitstätigkeit und Arbeitsumgebung

**Dissertation zur Erlangung des akademischen Grades
eines Doktors der Sicherheitswissenschaften
(Dr. rer. sec.)**



**Vorgelegt dem Fachbereich D Abteilung Sicherheitstechnik
an der Bergischen Universität Wuppertal**

1. Gutachter: Univ.-Prof. Dr.-Ing. B. H. Müller
2. Gutachter: Univ.-Prof. Dr.-Phil. H. O. Häcker

**von
Dipl.-Psych.
Martin Willner
aus Berlin**

Tag der mündlichen Prüfung:
25. Juni 2013

D468

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20130627-111056-8

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20130627-111056-8>]

Selbständigkeitserklärung

Hiermit erkläre ich, Martin Willner, geboren am 16. Februar 1980 in Berlin, dass ich die vorliegende Arbeit selbständig verfasst und dabei keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Diejenigen Stellen, welche anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, wurden durch Quellenangaben im Text kenntlich gemacht.

Wuppertal, 29. Oktober 2012

Martin Willner

Danksagung

Zuallererst danke ich Univ.-Prof. Dr. B. H. Müller und Univ.-Prof. Dr. H. O. Häcker für das große Vertrauen in meine Person sowie die konstruktive und unkomplizierte Betreuung der vorliegenden Dissertation.

Für die kritischen Reflexionen zu den methodischen Aspekten sowie zur inhaltlichen Struktur danke ich ebenso Dr. D. P. Seitz wie Dipl.-Psych. M. A. Iskenius.

Ferner bedanke ich mich für die freundliche Bereitstellung der verschiedenen Datensätze, die eine wesentliche Grundlage für die Berechnung der Referenzdaten in der vorliegenden Arbeit bilden, bei M. Galatsch, MScN (NEXT-Studie), Dr. M. Nübling (COPSOQ-I-Studie – kein direkter Zugriff auf die Daten), Prof. Dr. R. Tielsch sowie Dr. B. Gauggel (IPSO-Studie) und S. Schmidt, MScN (3Q-Studie).

Dem Initiator der lidA-Studie, apl. Prof. Dr. H.-M. Hasselhorn, danke ich, da er durch seine Schaffenskraft den Grundstein dieser Dissertation gelegt hat. In diesem Zusammenhang danke ich auch den Mitgliedern des lidA-Projektconsortiums für die interdisziplinäre Teamarbeit, die mich fachlich in vielfältiger Weise inspiriert hat.

Für die besonders gewissenhafte Orthografie- und Grammatikprüfung sowie die Konsistenzprüfung der vorliegenden Arbeit bin ich Dipl.-Psych. M. Ebener sehr dankbar.

Ganz besonders herzlich danke ich Dipl.-Soz.-Wiss. S. Weisbrich für die emotionale Unterstützung während des gesamten Entstehungsprozesses dieser Dissertation und das bedingungslose Verständnis für die häufige Beschäftigung mit diesem Werk. Außerdem danke ich ihr für die Unterstützung bei der Konsistenzprüfung dieser Arbeit.

Zusammenfassung

Das Verbundprojekt „lidA (leben in der Arbeit) – Kohortenstudie zu Gesundheit und Älterwerden in der Arbeit (lidA-Studie)“ konstruierte ein studienspezifisches „Computer-assisted Personal Interview (CAPI)“, in welchem Skalen der zweiten Version des „Copenhagen Psychosocial Questionnaire (COPSOQ-II)“ eingesetzt werden. Eine Überprüfung der psychometrischen Qualität dieser COPSOQ-II-Skalen ist für die deutschsprachige Anwendung, den Erhebungsmodus des CAPI sowie den Itemkontext und die Itemabfolge des lidA-CAPI noch nicht erfolgt. Zudem wurde bisher keine faktorenanalytische Testung der postulierten Dimensionalität dieser Skalen auf der Itemebene publiziert.

Im Rahmen der vorliegenden Untersuchung wurde überprüft, ob die COPSOQ-II-Skalen, wie sie im CAPI der 1. Welle der lidA-Hauptstudie eingesetzt werden, eine für inhaltliche statistische Analysen akzeptable psychometrische Qualität aufweisen. Es handelt sich hierbei um sechs partiell modifizierte Skalen: „Quantitative Anforderungen“, „Einfluss bei der Arbeit“, „Entwicklungsmöglichkeiten“, „Führungsqualität“, „Hilfe und Unterstützung von Kollegen“ sowie „Arbeit-Familie/Privatleben-Konflikt“.

Im Vorfeld der psychometrischen Item- und Skalenanalyse dieser COPSOQ-II-Skalen wurden statistische Kriterien ausgewählt (Verteilung der Werte, Item- und Skalenschwierigkeit, Trennschärfe, Homogenität, Dimensionalität, nomologisches Netzwerk), deren theoretische Grenzwerte bestimmt und vier empirische Vergleichsstudien herangezogen. Auf dieser Grundlage wurden anschließend aggregierte Hypothesen formuliert, die zunächst anhand einer kleinen lidA-Stichprobe (n = 171) und anschließend an einer per Zufall geteilten (jeweils ca. 50 Prozent) großen lidA-Stichprobe (n = 5 666) überprüft wurden.

Auf der Grundlage der Analyseergebnisse können keine bedeutsamen Unterschiede zwischen den sechs COPSOQ-II-Skalen im lidA-CAPI der lidA-Hauptstudie: 1. Welle und den korrespondierenden COPSOQ-II-Skalen der vier empirischen Vergleichsstudien berichtet werden. Außerdem sind die Ergebnisse im Wesentlichen konsistent zu denen internationaler Studien. Die postulierte Dimensionalität der fokussierten Skalen konnte nachgewiesen werden. Daher wird die psychometrische Qualität aller sechs COPSOQ-II-Skalen als akzeptabel und für inhaltliche statistische Analysen geeignet bewertet.

Hinsichtlich eines Items der Skala „Entwicklungsmöglichkeiten (EM)“ ist die Itemschwierigkeit nicht uneingeschränkt als akzeptabel zu beurteilen und die Trennschärfe auffallend niedrig. Zudem wurden im Rahmen der Hauptkomponentenanalyse Nebenladungen auf der Komponente „Einfluss bei der Arbeit (EA)“ gefunden. Eine inhaltliche Analyse legt eine mögliche Interdependenz zu einem Item der EA-Skala nahe. Hier ist eine Eliminierung des Items EM-1 in Erwägung zu ziehen, die allerdings eine psychometrische Reanalyse der EM-Skala (ohne das Item EM-1) erforderlich macht.

Abstract

The cooperative project 'lidA (leben in der Arbeit) – cohort study on health and aging in work (lidA-study)' constructed a study-specific 'Computer-assisted Personal Interview (CAPI)', in which scales of the second version of the 'Copenhagen Psychosocial Questionnaire (COPSOQ-II)' are used. A testing of the psychometric quality of these COPSOQ-II-scales has not yet occurred for the German version, the acquisition mode of the CAPI, the context and sequence of the items according to the lidA-CAPI. Moreover, so far no factor analysis of the postulated dimensionality of these scales on the level of the items was published.

In the present study it was tested whether the COPSOQ-II-scales, as used in the CAPI of the first wave of lidA, ensure an acceptable psychometric quality for substantial statistical analysis. Six partially modified scales became analyzed: 'quantitative demands', 'influence at work', 'possibilities for development', 'quality of leadership', 'social support from colleagues' and 'work-family conflict'.

In preparation of the psychometric item and scale analysis of these COPSOQ-II-scales statistical criteria were selected (distribution of values, item and scale difficulty, item-correlations with total scale, homogeneity, dimensionality, nomological network), their theoretical limits were determined and four empirical comparative studies were selected. On this basis aggregated hypotheses has been established which was tested first on a small lidA-sample (n = 171) and then tested on a large lidA-sample (n = 5666) that was split by chance in two parts (each with approximately 50 percent).

The analyses did not show any significant differences between the six COPSOQ-II-scales in the lidA-CAPI of the first wave of the lidA-study and the corresponding COPSOQ-II-scales of the four empirical comparative studies. Moreover, the results are essentially consistent with those of international studies. The postulated dimensionality of the focused scales could be verified. Therefore, the psychometric quality of all six COPSOQ-II-scale is judged as acceptable and suitable for substantial statistical analysis.

The item difficulty of one item of the scale 'possibilities for development (EM)' was not acceptable and its item-correlation with the EM-scale was very low. Moreover, according to the principal component analysis additionally loads on the component 'influence at work (EA)' were found. An analysis of the item contents suggests interdependence to an item of the EA-scale. According to this an elimination of the item EM-1 should be considered, but in preparation of this a psychometric reanalysis of the EM-scale (without the item EM-1) is required.

Inhaltsverzeichnis

Selbständigkeitserklärung.....	I
Danksagung.....	II
Zusammenfassung.....	III
Abstract.....	IV
Abbildungsverzeichnis.....	VIII
Tabellenverzeichnis.....	IX
Abkürzungen.....	XI
1. Einleitung.....	1
2. lidA-Studie.....	6
2.1 lidA-CAPI.....	8
2.1.1 COPSQ.....	10
2.1.2 Weitere Instrumente.....	12
3. Referenzwerte.....	14
3.1 Theoretische Referenzwerte.....	15
3.1.1 Verteilung der Werte.....	16
3.1.2 Item- und Skalenschwierigkeit.....	18
3.1.3 Trennschärfe.....	20
3.1.4 Homogenität.....	22
3.1.5 Dimensionalität.....	24
3.1.6 Nomologisches Netzwerk.....	26
3.2 Empirische Referenzwerte.....	27
3.2.1 NEXT-Studie.....	28
3.2.2 COPSQ-I-Studie.....	30
3.2.3 IPSO-Studie.....	32
3.2.4 3Q-Studie.....	34
3.2.5 Zusammenfassung.....	36
4. Fragestellungen und Hypothesen.....	39
4.1 Verteilung der Werte.....	40
4.2 Item- und Skalenschwierigkeit.....	42
4.3 Trennschärfe.....	43
4.4 Homogenität.....	44
4.5 Dimensionalität.....	45
4.6 Nomologisches Netzwerk.....	46

5. Methode	47
5.1 Stichproben	48
5.1.1 lidA-Machbarkeitsstudie	49
5.1.2 lidA-Hauptstudie: 1. Welle/A+B	51
5.2 COPSOQ	54
5.2.1 Quantitative Anforderungen.....	55
5.2.2 Einfluss bei der Arbeit.....	56
5.2.3 Entwicklungsmöglichkeiten.....	57
5.2.4 Führungsqualität.....	58
5.2.5 Hilfe und Unterstützung von Kollegen	59
5.2.6 Arbeit-Familie/Privatleben-Konflikt	60
5.3 Weitere Instrumente	61
5.3.1 Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen	62
5.3.2 Gedanke an Berufsaufgabe.....	64
5.3.3 Gegenwärtiger Gesundheitszustand	65
5.4 Statistische Analysen	66
6. Ergebnisse	68
6.1 lidA-Machbarkeitsstudie	69
6.1.1 Verteilung der Werte.....	70
6.1.2 Item- und Skalenschwierigkeit.....	73
6.1.3 Trennschärfe.....	75
6.1.4 Homogenität	77
6.1.5 Dimensionalität	78
6.1.6 Nomologisches Netzwerk	80
6.2 lidA-Hauptstudie: 1. Welle/A+B	81
6.2.1 Verteilung der Werte.....	82
6.2.2 Item- und Skalenschwierigkeit.....	87
6.2.3 Trennschärfe.....	91
6.2.4 Homogenität	94
6.2.5 Dimensionalität	96
6.2.6 Nomologisches Netzwerk	100
7. Diskussion	102
7.1 Quantitative Anforderungen	103
7.2 Einfluss bei der Arbeit	105
7.3 Entwicklungsmöglichkeiten	107
7.4 Führungsqualität	109
7.5 Hilfe und Unterstützung von Kollegen	111

7.6 Arbeit-Familie/Privatleben-Konflikt	113
7.7 Allgemeines	115
7.8 Grenzen und Ausblick	116
Literaturverzeichnis	118
Anhang	130
lidA-Studie	130
NEXT-Studie	132
COPSOQ-I-Studie	136
IPSO-Studie	142
3Q-Studie	146

Abbildungsverzeichnis

Abbildung 1: Belastungs-Beanspruchungs-Konzept.	3
Abbildung 2: Forschungsmodelle der lidA-Studie.....	7
Abbildung 3: Dimensionalität (theoretische Skalenstruktur).	45

Tabellenverzeichnis

Tabelle 1: Item- und Skalenübersicht (empirische Referenzwerte)	38
Tabelle 2: Verteilung der Werte (empirische Referenzwerte).....	40
Tabelle 3: Item- und Skalenschwierigkeit (empirische Referenzwerte).	42
Tabelle 4: Trennschärfe (empirische Referenzwerte).....	43
Tabelle 5: Homogenität (empirische Referenzwerte)	44
Tabelle 6: Nomologisches Netzwerk (empirisches Referenznetzwerk).....	46
Tabelle 7: Soziodemografie (lidA-Machbarkeitsstudie)	50
Tabelle 8: Soziodemografie (lidA-Hauptstudie: 1. Welle/A).....	52
Tabelle 9: Soziodemografie (lidA-Hauptstudie: 1. Welle/B).....	53
Tabelle 10: Verteilung der Werte (lidA-Machbarkeitsstudie)	71
Tabelle 11: Tests auf Normalverteilung (lidA-Machbarkeitsstudie).....	72
Tabelle 12: Item- und Skalenschwierigkeit (lidA-Machbarkeitsstudie).....	74
Tabelle 13: Trennschärfe (lidA-Machbarkeitsstudie)	75
Tabelle 14: Homogenität (lidA-Machbarkeitsstudie).....	77
Tabelle 15: Komponentenladungsmatrix (lidA-Machbarkeitsstudie).....	78
Tabelle 16: Nomologisches Netzwerk (lidA-Machbarkeitsstudie).....	80
Tabelle 17: Verteilung der Werte (lidA-Hauptstudie: 1. Welle/A).....	82
Tabelle 18: Verteilung der Werte (lidA-Hauptstudie: 1. Welle/B).....	83
Tabelle 19: Tests auf Normalverteilung (lidA-Hauptstudie: 1. Welle/A).....	85
Tabelle 20: Tests auf Normalverteilung (lidA-Hauptstudie: 1. Welle/B).....	86
Tabelle 21: Item- und Skalenschwierigkeit (lidA-Hauptstudie: 1. Welle/A)	88
Tabelle 22: Item- und Skalenschwierigkeit (lidA-Hauptstudie: 1. Welle/B)	89
Tabelle 23: Trennschärfe (lidA-Hauptstudie: 1. Welle/A).....	92
Tabelle 24: Trennschärfe (lidA-Hauptstudie: 1. Welle/B).....	93
Tabelle 25: Homogenität (lidA-Hauptstudie: 1. Welle/A)	94
Tabelle 26: Homogenität (lidA-Hauptstudie: 1. Welle/B)	94
Tabelle 27: Komponentenladungsmatrix (lidA-Hauptstudie: 1. Welle/A).	97
Tabelle 28: Komponentenladungsmatrix (lidA-Hauptstudie: 1. Welle/B).	98
Tabelle 29: Nomologisches Netzwerk (lidA-Hauptstudie: 1. Welle/A)	100
Tabelle 30: Nomologisches Netzwerk (lidA-Hauptstudie: 1. Welle/B)	101
Tabelle 31: Konstrukte des lidA-CAPI (lidA-Hauptstudie: 1. Welle).....	130
Tabelle 32: Soziodemografie (lidA: ursprüngliche Nettostichprobe).....	131

Tabelle 33: Iteminhalt (NEXT-Studie).....	132
Tabelle 34: Soziodemografie (NEXT-Studie).....	133
Tabelle 35: Verteilung der Werte (NEXT-Studie).....	133
Tabelle 36: Tests auf Normalverteilung (NEXT-Studie).....	134
Tabelle 37: Item- und Skalenschwierigkeit (NEXT-Studie).....	134
Tabelle 38: Trennschärfe (NEXT-Studie)	134
Tabelle 39: Homogenität (NEXT-Studie)	135
Tabelle 40: Nomologisches Netzwerk (NEXT-Studie).....	135
Tabelle 41: Iteminhalt 1/2 (COPSOQ-I-Studie).....	136
Tabelle 42: Iteminhalt 2/2 (COPSOQ-I-Studie).....	137
Tabelle 43: Soziodemografie (COPSOQ-I-Studie)	138
Tabelle 44: Verteilung der Werte (COPSOQ-I-Studie)	139
Tabelle 45: Tests auf Normalverteilung (COPSOQ-I-Studie)	139
Tabelle 46: Item- und Skalenschwierigkeit (COPSOQ-I-Studie).....	140
Tabelle 47: Trennschärfe (COPSOQ-I-Studie).....	141
Tabelle 48: Homogenität (COPSOQ-I-Studie).....	141
Tabelle 49: Nomologisches Netzwerk (COPSOQ-I-Studie).....	141
Tabelle 50: Iteminhalt (IPSO-Studie).....	142
Tabelle 51: Soziodemografie (IPSO-Studie).....	143
Tabelle 52: Verteilung der Werte (IPSO-Studie).....	143
Tabelle 53: Tests auf Normalverteilung (IPSO-Studie).....	144
Tabelle 54: Item- und Skalenschwierigkeit (IPSO-Studie).....	144
Tabelle 55: Trennschärfe (IPSO-Studie)	145
Tabelle 56: Homogenität (IPSO-Studie)	145
Tabelle 57: Nomologisches Netzwerk (IPSO-Studie).....	145
Tabelle 58: Iteminhalt (3Q-Studie).....	146
Tabelle 59: Soziodemografie (3Q-Studie)	147
Tabelle 60: Verteilung der Werte (3Q-Studie)	148
Tabelle 61: Tests auf Normalverteilung (3Q-Studie)	148
Tabelle 62: Item- und Skalenschwierigkeit (3Q-Studie).....	149
Tabelle 63: Trennschärfe (3Q-Studie).....	149
Tabelle 64: Homogenität (3Q-Studie).....	150
Tabelle 65: Nomologisches Netzwerk (3Q-Studie).....	150

Abkürzungen

AF	Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen
AK	Arbeit-Familie-Privatleben-Konflikt (COPSOQ-Skala)
AM	Arithmetisches Mittel
AO	Antwortoptionen
BA	Bundesagentur für Arbeit
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
BMAS	Bundesministerium für Arbeit und Soziales
BMWI	Bundesministerium für Wirtschaft und Technologie
BUW	Bergische Universität Wuppertal
COPSOQ	Copenhagen Psychosocial Questionnaire
Destatis	Statistisches Bundesamt
EA	Einfluss bei der Arbeit (COPSOQ-Skala)
EM	Entwicklungsmöglichkeiten (COPSOQ-Skala)
F	Fragestellung
FFAS	Freiburger Forschungsstelle Arbeits- und Sozialmedizin
FQ	Führungsqualität (COPSOQ-Skala)
GB	Gedanke an Berufsaufgabe
GG	Gegenwärtiger Gesundheitszustand
H	Hypothese
HU	Hilfe und Unterstützung von Kollegen (COPSOQ-Skala)
et al.	et alii = und andere
I-I-K	Inter-Item-Korrelation
IAB	Institut für Arbeitsmarkt und Berufsforschung
IEB	Integrierte Erwerbsbiografie
infas	Institut für angewandte Sozialwissenschaft GmbH

IQR	Interquartil Range (Interquartilsabstand)
KMO	Kaiser-Meyer-Olkin-Kriterium
K-S-A	Kolmogorov-Smirnov-Anpassungstest mit Lilliefors-Schranken
lidA	leben in der Arbeit
NRCWE	National Research Centre for the Working Environment
PCA	Principal Component Analysis (Hauptkomponentenanalyse)
QA	Quantitative Anforderungen (COPSOQ-Skala)
S-W-T	Shapiro-Wilk-Test
SD	Standard Deviation (Standardabweichung)
SGB	Sozialgesetzbuch
TF	Teilfragestellung
TN	(Studien-)Teilnehmer (Singular/Plural – je nach Kontext)

1. Einleitung

Für den Großteil der Bevölkerung in Deutschland stellt die Erwerbsarbeit das Zentrum des Lebens dar (vgl. Schaper, 2011c). Die Erwerbsarbeit ist für den Menschen nicht nur während der Zeit ihrer Ausübung, sondern bereits davor (Schul- und Berufsausbildung) und danach (Ruhestand) ein konstituierendes Moment. Neben dem monetären Entgelt erfüllt die Erwerbsarbeit mannigfaltige weitere Funktionen (z. B. Blickle, 2011; Ulich, 2005; Warr, 1987; Jahoda, 1981), welche dem einzelnen Menschen mehr oder weniger bewusst zugänglich sind (Blickle, 2011). Jahoda (1981) differenziert hier fünf so genannte „latente Funktionen“: (1) zeitliche Strukturierung (Tag, Woche, Monat etc.), (2) Sicherstellung sozialer Kontakte außerhalb der Familie, (3) Teilhabe an den einzelnen Menschen transzendierenden Zielen, (4) Konstitution der Identität und des gesellschaftlichen Status sowie (5) physische und psychische Aktivierung. Hierdurch wird erkennbar, dass die Erwerbsarbeit die Integration in die Gesellschaft sowie die physische und psychische Gesundheit beeinflusst (z. B. Friedmann, 1959).

Der Arbeitsmarkt unterliegt ständigen, durch politische und wirtschaftliche Maßnahmen wesentlich beeinflussten Entwicklungen (Dietz & Walwei, 2008; Plath, 2000). Zusammenfassend können die Manifestationen jener Entwicklungen vornehmlich auf der Ebene der Organisation (zunehmende Modularisierung, Virtualisierung, Netzwerkbildung und Globalisierung) sowie der Ebene der Arbeitstätigkeit (zunehmende Wissensarbeit und Flexibilisierung) beobachtet und differenziert werden (vgl. Schaper, 2011d). Solche Entwicklungen sind oftmals mit einer Veränderung der Anforderungen an die Erwerbstätigen, namentlich deren Kompetenzen verbunden (vgl. Plath, 2000). Die Entwicklung des Arbeitsmarktes ist wechselseitig verknüpft mit gesamtgesellschaftlichen Entwicklungen.

Seit einigen Jahren sinkt die Geburtenzahl in Deutschland kontinuierlich (Schneider & Dorbritz, 2011). Gleichzeitig steigt die Lebenserwartung der Menschen hierzulande stetig (Destatis, 2011). Außerdem wächst der Wanderungssaldo, d. h. die Differenz zwischen Immigration und Emigration, in den kommenden Jahren voraussichtlich nur geringfügig (Destatis, 2009; Schimany, 2007). Diese drei gesamtgesellschaftlichen Entwicklungstrends begründen den so genannten „demografischen Wandel“ in Deutschland, der vor allem durch die Schrumpfung sowie die Alterung der Gesellschaft gekennzeichnet ist (z. B. Garloff, Pohl & Schanne, 2012; Börsch-Supan, 2011; Destatis, 2009). Hieraus resultiert zum Einen eine bereits beobachtbare und sich weiter fortsetzende Veränderung der Relation von Erwerbstätigen (sinkender Anteil) zu Ruheständlern (steigender Anteil) innerhalb der Gesamtbevölkerung (Börsch-Supan, 2011; Destatis, 2011; Schimany, 2007). Zum Anderen führen die drei genannten Entwicklungstrends zu einem zunehmend höheren relativen Anteil älterer Menschen an der Erwerbsbevölkerung (Garloff et al., 2012).

Zwar bedroht die Fortsetzung einer derartigen Entwicklung verschiedene gesellschaftliche Sicherungssysteme fundamental, gleichsam ermöglicht diese auch konstruktive System-

veränderungen, wenn mit adäquaten wirtschaftlichen und politischen Maßnahmen reagiert wird (vgl. Börsch-Supan, 2011; Bode & Richter, 2011).

Eine wichtige Reaktion der Politik bzw. des Gesetzgebers auf den demografischen Wandel in der Vergangenheit war die gesetzlich geregelte Verlängerung der Lebensarbeitszeit in Deutschland (Destatis, 2011). Allerdings sinkt bereits seit mehreren Jahren die Erwerbsquote (Gesamtzahl der Erwerbspersonen, d. h. Erwerbstätige und Arbeitslose, dividiert durch die Gesamtzahl der Erwerbsbevölkerung, d. h. alle 15- bis 64-jährigen) mit steigendem Lebensalter, d. h. ab 45 Jahren (z. B. Garloff et al., 2012). Dieser Zusammenhang zwischen Erwerbsquote und Lebensalter wird schon seit Längerem nicht nur in Deutschland, sondern auch in anderen europäischen Staaten beobachtet. Demgemäß wurden in einem von fünf Zielen der Wachstumsstrategie Europas „Europe 2020“ die (Beschäftigungs- bzw.) Erwerbsquoten der EU-Mitgliedstaaten fokussiert und u. a. die Forderungen formuliert, dass die Erwerbsquote der 20- bis 64-jährigen in Deutschland bis 2020 mindestens 77.0 Prozent und diejenige der 55- bis 64-jährigen mindestens 60.0 Prozent betragen soll (BMAS, 2011; BMWI, 2011). Im Jahr 2010 betrug die Erwerbsquote unter den 15- bis 64-jährigen hierzulande lediglich 76.5 Prozent (Garloff et al., 2012).

„Zwischen 1991 und 2010 ist die Erwerbsbevölkerung von 55,1 (...) auf 54,0 Mio. gesunken (...). Der Anteil der jüngeren Bevölkerung (...) [15- bis 39-jährige] fiel (...) von 53,7 (...) auf 44,3 Prozent (...). (...) [Demgemäß] ist der Anteil der älteren erwerbsfähigen Bevölkerung (40- bis 64-jährige) von 46,3 (...) auf 55,7 Prozent (...) gestiegen. (Garloff et al., 2012, S. 1).

Vor diesem Hintergrund gewinnt der Erhalt und die Förderung der Fähigkeit, eine Beschäftigung zu finden und auszuüben (Beschäftigungsfähigkeit), für den (älteren) Arbeitnehmer wie auch den Arbeitgeber und die Gesellschaft zunehmend an Bedeutung (Garloff et al., 2012; Beermann, 2011). Die Beschäftigungsfähigkeit ist definiert als das Resultat der Interaktion von Arbeitsfähigkeit (diese wiederum setzt sich zusammen aus Leistungsfähigkeit und Leistungsbereitschaft eines Menschen) sowie Aspekten des Arbeitsmarktes (z. B. Beermann, 2011). Während einige für die Beschäftigungsfähigkeit relevante rechtliche Rahmenbedingungen bereits 1988 im „Gesetz zur Strukturreform im Gesundheitswesen (Gesundheits-Reformgesetz – GRG)“ und im selben Jahr im „Sozialgesetzbuch (SGB) Fünftes Buch (V)“ sowie 1996 im „Gesetz über die Durchführung von Maßnahmen des Arbeitsschutzes zur Verbesserung der Sicherheit und des Gesundheitsschutzes der Beschäftigten bei der Arbeit (Arbeitsschutzgesetz – ArbSchG)“ und 1997 im „Sozialgesetzbuch (SGB) Siebtes Buch (VII)“ schriftlich fixiert worden sind, wurden mit „Europe 2020“ konkrete Zielvorgaben verbindlich formuliert, mittels welchen eine konsequente und strukturierte Umsetzung sinnvoller Maßnahmen auf der politischen Ebene forciert werden soll.

Um die Erreichung dieser Zielvorgaben sicherzustellen, sind neben politischen Maßnahmen für eine zeitgemäße Fortentwicklung des Arbeitsmarktes vor allem organisationale verhältnis- und verhaltensorientierte Maßnahmen (vgl. Ulich, 2005; Greiner, 1998; Frieling & Sonntag, 1999; Mohr & Udris, 1997) zum Erhalt und zur Förderung der Arbeitsfähigkeit erforderlich (Beermann, 2011). Hierbei ist zu beachten, dass verhältnisorientierte Maßnahmen den verhaltensorientierten sinnvollerweise vorgeordnet sind (vgl. Oppolzer, 2010; Klotter, 1999) und stets eine menschengerechte Arbeitsgestaltung zum Ziel haben (vgl. Rothe, 2011; Müller, 1997; Dunckel & Volpert, 1997). Eine ebensolche Arbeitsgestaltung setzt eine fachlich fundierte Arbeitsanalyse sowie eine sich daran anschließende ebenso fundierte Bewertung (z. B. Ulich, 2005; Schaper, 2011a; Müller, 1997) der jeweils relevanten Aspekte des Menschen, der Technik sowie der Organisation voraus (Ulich, 2005). Im Kontext dieser doppelten Trias der Arbeitsgestaltung, d. h. der Analyse, Bewertung und Gestaltung von Mensch, Technologie und Organisation (MTO-Konzept), in deren Mittelpunkt immer die Arbeitsaufgabe steht, sind die Erkenntnisse und Methoden der verschiedenen jeweils relevanten Wissenschaftsdisziplinen heranzuziehen (Ulich, 2005).

Einen disziplinübergreifenden theoretischen Bezugsrahmen, der in diesem Zusammenhang von grundlegender Bedeutung ist, stellt das Belastungs-Beanspruchungs-Konzept dar, wie es in seiner allgemeinen Form in der Abbildung 1 zusammengefasst ist (z. B. Schaper, 2011c; Ulich, 2005; Richter & Hacker, 1998; Müller, 1997; Rohmert, 1984; Rohmert & Rutenfranz, 1975). Hierbei werden Belastungen (wertneutral) als diejenigen Aspekte der Arbeitstätigkeit und Arbeitsumgebung verstanden, welche auf die Person einwirken (z. B. Müller, 1997; Rohmert, 1984). In Abhängigkeit von individuellen Merkmalen (bspw. Anspruchsniveau, Motivation, Bewältigungsstrategien, Fähigkeiten, Erfahrung, Gesundheit, Alter und Ernährung – DIN EN ISO 10075-1) führen Belastungen zu Reaktionen in der Person, die hier (wertneutral) als Beanspruchungen bezeichnet werden (vgl. Müller, 1997; Rohmert, 1984).

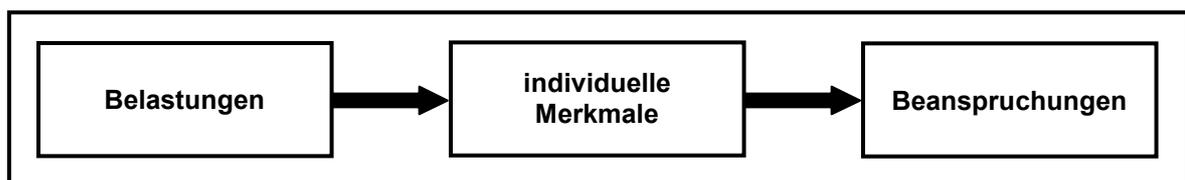


Abbildung 1: Belastungs-Beanspruchungs-Konzept.

Entsprechend den oben thematisierten Entwicklungen des Arbeitsmarktes und deren Manifestationen auf der Ebene der Organisation sowie der Ebene der Arbeitstätigkeit ist in Deutschland und weltweit eine Zunahme psychischer Belastungen im Erwerbsleben zu konstatieren (z. B. Schaper, 2011c; Oppolzer, 2010; Ulich, 2005; Richter & Hacker, 1998; Greif, 1991). Infolgedessen wurden im Jahr 2000 „Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung“ in einer deutschen Industrienorm festgeschrieben (DIN EN ISO 10075). Konsistent hierzu wird innerhalb der arbeitswissenschaftlichen Forschung be-

reits seit mehreren Jahren ein verstärkter Fokus auf die Untersuchung psychischer Belastungen gelegt (vgl. Buunk, de Jonge, Ybema & de Wolff, 1998).

Oppolzer (2010) differenziert fünf Kategorien psychischer Belastungen, die potentielle Risikofaktoren für die Genese psychischer Fehlbeanspruchungen darstellen: (1) Arbeitsaufgabe und -inhalt, (2) Arbeitsorganisation und -zeit, (3) Arbeitsumgebung und -platz, (4) soziale Beziehungen in vertikaler und horizontaler Hinsicht sowie (5) soziale Bedingungen des Arbeits- und Beschäftigungsverhältnisses. In Bezug auf die psychischen Fehlbeanspruchungen selbst unterscheidet Oppolzer (2010) drei Kategorien: (1) Stress, (2) psychische Ermüdung sowie (3) ermüdungsähnliche Zustände (Monotonie, psychische Sättigung und herabgesetzte Wachsamkeit). Diese psychischen Fehlbeanspruchungen gefährden vermittelt über physiologische Reaktionen des Organismus (bspw. Muskeltonus, Pulsfrequenz, Hirnströme und Stoffwechsel) das psychische Wohlbefinden sowie die Gesundheit (bspw. Herz-Kreislauf-Erkrankungen, psychische Störungen und Arbeits-Unfälle) eines Menschen (Oppolzer, 2010).

Obgleich bereits vielfältige Erkenntnisse über die verschiedenen Zusammenhänge zwischen psychischen Belastungen, individuellen Merkmalen, psychischen Beanspruchungen sowie psychischem Wohlbefinden und Gesundheit vorliegen, besteht diesbezüglich weiterer Forschungsbedarf, insbesondere in Bezug auf ältere Erwerbstätige (vgl. Rothe, 2011; Hasselhorn, 2011; Borgetto & Troschke, 2001; Plath, 2000).

„Die Frage ist, wie sich die Gesundheit einer stetig älter werdenden Erwerbsbevölkerung künftig entwickeln wird und welche Rolle die Arbeitsbedingungen hierbei spielen.“ (Hasselhorn, 2011, S. 4).

Als besondere Spezifikation menschengerechter Arbeitsgestaltung zum Erhalt und zur Förderung der Arbeitsfähigkeit tritt nun die altersgerechte Arbeitsgestaltung in den Fokus des arbeitswissenschaftlichen Interesses (vgl. Hasselhorn, 2011). Für konkrete und nachhaltig wirksame Arbeitsgestaltungsmaßnahmen sind ebenso konkrete (psychologische) Arbeitsanalysen („genaue Messungen“ – DIN EN ISO 10075-3) erforderlich (z. B. Dunckel, 1999). Vorbedingungen derartiger Analysen sind spezifische Annahmen bzw. Erkenntnisse hinsichtlich der Wirkungszusammenhänge zwischen den jeweils interessierenden Aspekten (vgl. Schaper, 2011b; Dunckel, 1999; Müller, 1997; Frei, 1981) von Arbeit, Alter und Gesundheit. Ferner ist zur Anwendung entsprechender Analysen die Investition wertvoller Ressourcen wie Zeit und Personal unabdingbar (z. B. Kreienbrock, Pigeot & Ahrens, 2012; Dunckel, 1999). Aufgrund des häufigen Mangels vorgenannter Voraussetzungen werden zumeist sehr kurze und eher allgemeine Verfahren („orientierende Messungen“ – DIN EN ISO 10075-3) zur (psychologischen) Arbeitsanalyse eingesetzt, die oftmals nicht auf umfangreichen theoretischen Vorannahmen basieren, nur wenig Vorbereitungs- und Durchführungszeit erfordern und lediglich verhältnismäßig geringfügige Kosten verursachen (vgl. Kreienbrock, et al., 2012; Zielhuis & Kiemeneij, 2001).

Die Resultate derartiger Studien liefern gleichsam kaum konkrete Gestaltungsansätze (vgl. Dunckel, 1999).

Die Anwendung solcher Kurzverfahren, welche insbesondere im Kontext sozial-epidemiologischer Forschung erfolgt, kann dennoch wertvolle Hinweise auf mögliche kausale Zusammenhänge relevanter Aspekte von Arbeit, Alter und Gesundheit generieren (vgl. Kreienbrock, et al., 2012; Siegrist, 2001) und ihrerseits den Ausgangspunkt genauer Messungen bilden (Dunckel, 1999). Entsprechend der oben dargestellten gesellschaftlichen Entwicklungen werden weltweit vermehrt sozialepidemiologische Längsschnittstudien zum Themenkomplex Arbeit, Alter und Gesundheit anhand großer Stichproben durchgeführt – exemplarisch sind hier acht jener Studien genannt: „Health and Retirement Study (HRS)“ in den Vereinigten Staaten von Amerika, „English Longitudinal Study of Ageing (ELSA)“ in England, „Survey of Health, Ageing and Retirement in Europe (SHARE)“ in 20 europäischen Staaten, „Korean Longitudinal Study of Ageing (KLoSA)“ in Korea, „China Health and Retirement Longitudinal Study (CHARLS)“ in China, „Longitudinal Aging Study Amsterdam (LASA)“ in den Niederlanden, „Swedish Longitudinal Occupational Survey of Health (SLOSH)“ in Schweden und „leben in der Arbeit (lidA-Studie)“ in Deutschland.

Standardmäßig werden in derartigen Studien Fragebögen (bspw. „Papier-Bleistift-Verfahren“, „Computer-assisted Telephone Interview – CATI“, „Computer-assisted Personal Interview – CAPI“) eingesetzt, mit deren Hilfe diverse Konstrukte bzw. Dimensionen in kurzer Zeit erfasst werden (vgl. Kreienbrock, et al., 2012; Borgetto & Troschke, 2001; Atteslander, 1999). Zudem werden hierbei nicht selten die Inhalte und/oder Antwortoptionen von Skalen oder Einzelitems studienspezifisch modifiziert, um die Durchführungszeit zu verringern, die Verständlichkeit des Fragebogens durch die Befragten zu erhöhen oder die Augenscheinvalidität zu optimieren (vgl. Atteslander, 1999). Darüber hinaus werden Skalen oder Einzelitems, deren Validität für einen spezifischen Erhebungsmodus (bspw. „Papier-Bleistift-Verfahren“) nachgewiesen wurde, mitunter ungeprüft in einem anderen Erhebungsmodus eingesetzt (bspw. CAPI), obwohl hier messbare Unterschiede im Antwortverhalten zu erwarten sind (vgl. Bortz & Döring, 2009). Grundsätzlich und insbesondere im Anwendungskontext studienspezifisch konstruierter Fragebögen sollte vor der Durchführung inhaltlicher Analysen daher die psychometrische Qualität der eingesetzten Verfahren überprüft werden, um die (Konstrukt-)Validität der jeweils relevanten Skalen oder Einzelitems kritisch einschätzen zu können (vgl. Kreienbrock, et al., 2012; Bühner, 2011; Bortz & Döring, 2009; Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998).

Zur Förderung der Übersichtlichkeit in der Darstellungsweise wichtiger Aspekte wurde im Einzelfall auf abwechslungsreiche Formulierungen verzichtet und stattdessen einer hohen Einheitlichkeit hinsichtlich der Sprachgestalt sowie der Absatzstruktur in ausgewählten Abschnitten der vorliegenden Arbeit der Vorzug gegeben.

2. lidA-Studie

Das nationale Verbundprojekt „lidA (leben in der Arbeit) – Kohortenstudie zu Gesundheit und Älterwerden in der Arbeit (lidA-Studie)“ wird vom Bundesministerium für Bildung und Forschung (BMBF) zunächst für den Zeitraum von April 2009 bis März 2015 gefördert (BMBF, 2009). Die Projektleitung und das Projektkonsortium der lidA-Studie setzen sich aus denselben fünf Institutionen zusammen: (1) Bergische Universität Wuppertal (BUW – Koordination), (2) Institut für Arbeitsmarkt- und Berufsforschung (IAB), (3) Otto-von-Guericke Universität Magdeburg, (4) Universität Ulm sowie (5) infas – Institut für angewandte Sozialforschung GmbH. Die Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) ist assoziierter Projektpartner der lidA-Studie. Durch das „International Review Board (IRB)“, bestehend aus einer interdisziplinären, internationalen Forschergruppe, erfolgen wissenschaftliche Begutachtungen der Studie. Wissenschaftliche Beratung erhält das Projektkonsortium zudem durch T. S. Kristensen sowie T. Theorell.

„Ziel dieser Studie ist die Untersuchung von Langzeiteffekten der Arbeit auf die Gesundheit in einer älter werdenden Erwerbsbevölkerung aus interdisziplinärer Perspektive. Die lidA-Studie ist die erste repräsentative Kohortenstudie in Deutschland, die Zusammenhänge von Arbeit, Alter und Gesundheit untersucht, und wird eine der umfassendsten Langzeitstudien im Bereich der Arbeitswelt sein.“ (lidA-Projektkonsortium, 2010)

Konkretisiert sind die zentralen inhaltlichen Forschungsfragen der lidA-Studie, welche dem zuvor benannten Forschungsziel entspringen, im Wesentlichen in drei zentralen Forschungsmodellen (siehe Abbildung 2), die verschiedene Variablen aus den sechs nachstehend angeführten Inhaltsbereichen umfassen (du Prel et al., 2010): (1) Arbeit, (2) sozio-ökonomischer Statuts, (3) private Verhältnisse, (4) personale Merkmale, (5) Geschlecht sowie (6) Gesundheit. Ein besonderer Schwerpunkt jedes dieser drei Forschungsmodelle liegt auf dem Bereich der Arbeit, welcher demgemäß am differenziertesten erhoben wird.

Zur Bearbeitung dieser zentralen wie einiger weiterer Forschungsmodelle werden im Rahmen der lidA-Studie die drei folgenden Datenquellen herangezogen:

- individuelle Befragungsdaten des studienspezifischen Computer-assisted Personal Interviews (CAPI) – eventuell Modifikationen für verschiedene (Erhebungs-)Wellen
- individuelle (Zuspielungsbereitschaft der Teilnehmer – TN – erforderlich), administrative Daten (d. h. „Integrierte Erwerbsbiografie – IEB“) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB)
- individuelle (Zuspielungsbereitschaft der TN erforderlich) Krankenkassendaten und aggregierte Krankenkassendaten kooperierender Krankenkassen

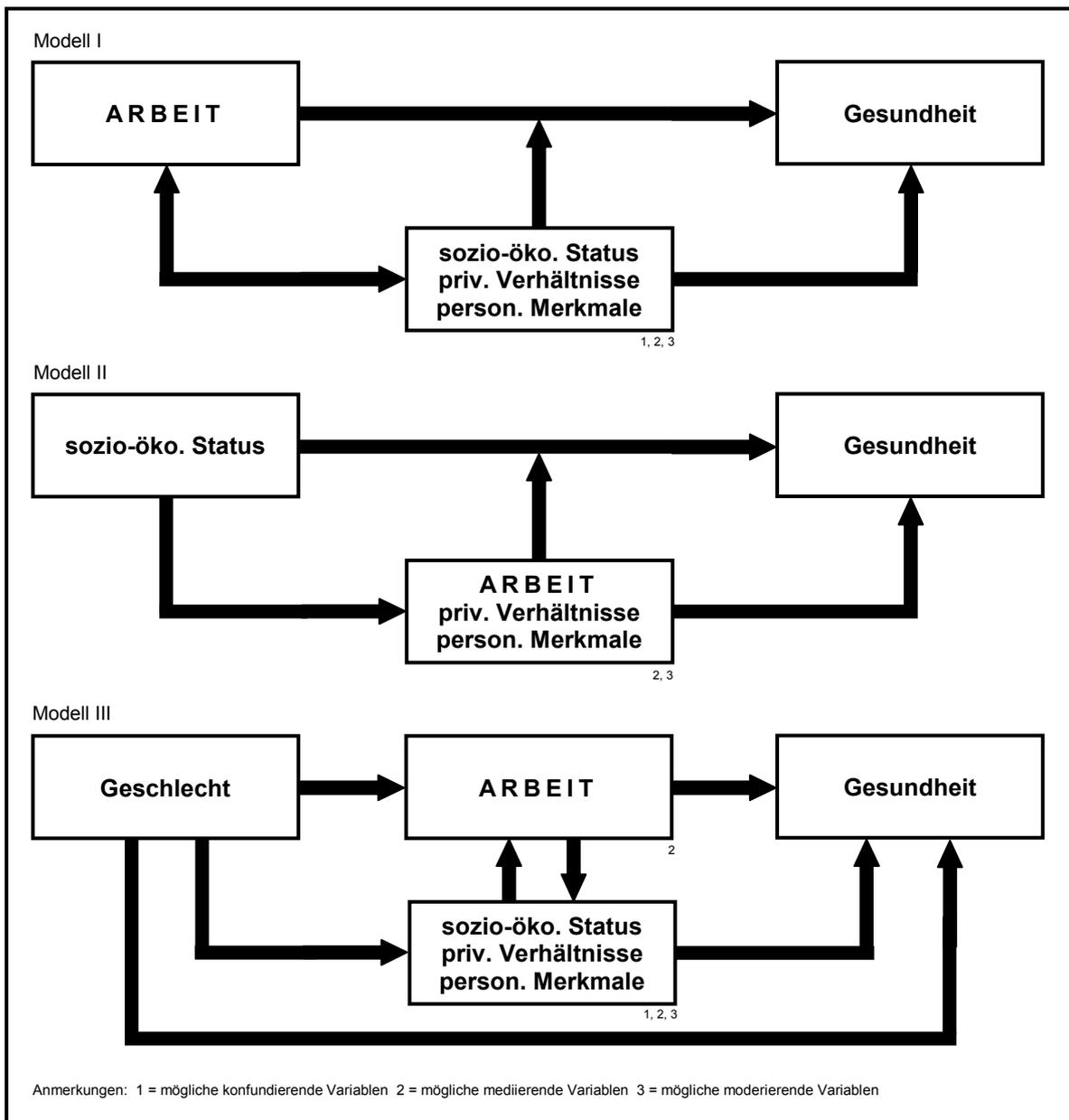


Abbildung 2: Forschungsmodelle der lidA-Studie.

Die individuellen Befragungsdaten der lidA-CAPI stellen eine zentrale Datenquelle der lidA-Studie dar. Befragt werden zunächst zwei Geburtsjahrgänge (1959 und 1965) in einem Turnus von drei Jahren (du Prel et al., 2010). Aktuell ist die finanzielle Förderung von zwei (Erhebungs-)Wellen sichergestellt: 1. Welle der lidA-Hauptstudie im Jahr 2011 sowie 2. Welle der lidA-Hauptstudie im Jahr 2014. Geplant ist eine Verlängerung um zwei weitere Wellen: 3. Welle der lidA-Hauptstudie im Jahr 2017 sowie 4. Welle der lidA-Hauptstudie im Jahr 2020. Im Falle einer Studienverlängerung wird noch ein dritter Geburtsjahrgang (= 1971) hinzugenommen (du Prel et al., 2010). Ein solches Studiendesign ermöglicht die Identifikation (1) themenspezifischer Differenzen zwischen Geburtsjahrgängen, (2) Modifikationen über die Zeit innerhalb der Geburtsjahrgänge sowie (3) Differenzen zwischen „Generationen“ (Hardt et al., 2010).

2.1 lidA-CAPI

Alle persönlichen Befragungen der lidA-Studie erfolgen mittels eines studienspezifischen CAPI, wobei dasjenige der 1. Welle der lidA-Hauptstudie eventuell für jede nächstfolgende Welle inhaltlich und formal leicht modifiziert wird, da einige Konstrukte lediglich einmal erhoben werden müssen. Eine Übersicht über die im lidA-CAPI der 1. Welle der lidA-Hauptstudie erhobenen Konstrukte befindet sich im Anhang (siehe Tabelle 31).

Die drei zentralen Forschungsmodelle der lidA-Studie (siehe Abbildung 2) verknüpfen, über die sechs oben benannten Inhaltsbereiche, eine verhältnismäßig große Anzahl verschiedener Konstrukte miteinander. Hieraus erwächst u. a. die Notwendigkeit, ein entsprechend umfangreiches lidA-CAPI zu konstruieren. Gleichzeitig setzt die zu erwartende Belastbarkeit sowie das als verhältnismäßig gering vorauszusetzende (fachliche) Verständnis der TN bezüglich der formalen und inhaltlichen Bedingungen der persönlichen Befragungen der CAPI-Konstruktion Grenzen. Hiermit ist die Akzeptanz des lidA-CAPI durch die TN angesprochen, welche für die Panelbereitschaft, d. h. die Bereitschaft für die nächstfolgende Welle erneut als TN zur Verfügung zu stehen, und damit für den Realisationserfolg der Datenerhebung einen kaum zu überschätzenden Einfluss hat. Zudem begrenzen die finanziellen Rahmenbedingungen das Quantum möglicher Items des lidA-CAPI, da die Bearbeitung jedes einzelnen Items durch die TN Interviewzeit und hierdurch Geld kostet. Vor diesem Hintergrund hat sich das lidA-Projektkonsortium dazu entschlossen, im Rahmen des CAPI der 1. Welle der lidA-Hauptstudie keine konvergenten, sondern ausschließlich diskriminante Konstrukte zu erheben und insgesamt 60 Minuten reine Interviewzeit anzuvisieren (Kersting, Schröder & Steinwede, 2012; Kersting, Schröder & Steinwede, 2011).

Neben wenigen neuen Items werden zur Konstruktion des CAPI der 1. Welle der lidA-Hauptstudie vor allem psychometrisch geprüfte und in der (sozialepidemiologischen) Forschung etablierte Skalen und Einzelitems herangezogen, von denen einige partiell modifiziert werden. Bestehen inhaltliche und/oder formale Divergenzen zwischen einem Originalitem und dem tatsächlich verwendeten Item, kann nicht ohne Weiteres davon ausgegangen werden, dass die nachgewiesene psychometrische Qualität des Originalitems auch für das modifizierte Item Gültigkeit hat (vgl. Bortz & Döring, 2009; Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998).

Außerdem ist die psychometrische Qualität einiger Originalitems ausschließlich im Rahmen anonymisierter, schriftlicher Befragungen („Papier-Bleistift-Verfahren“) überprüft worden. Die in der sozialwissenschaftlichen Literatur hinlänglich bekannten Differenzen zwischen schriftlichen und mündlichen (persönlichen) Befragungen lassen die Übertragbarkeit der Resultate psychometrischer Analysen zwischen diesen beiden Erhebungsmodi im Einzelfall als zweifelhaft erscheinen (z. B. Li, Ford, Zhao, Tsai & Balluz, 2012; Ingenkamp & Lissmann, 2008; Naplava & Oberwittler, 2002; Konrad, 2001). Nachfolgend

sind sieben solche Differenzen mit dem Fokus auf die persönliche Befragung aus der Perspektive der TN zusammenfassend aufgelistet:

- geringere wahrgenommene Anonymität (z. B. Bortz & Döring, 2009)
- geringere Bereitschaft zur Selbstoffenbarung (z. B. Chelune, 1979)
- höhere soziale Erwünschtheit (z. B. Kury, 1994), insb. Bemühen, dem Interviewer gefallen zu wollen (z. B. Roethlisberger & Dickson, 1964)
- vielfältige Reaktionen auf personale Merkmale des Interviewers (z. B. Schnell, Hill & Esser, 2008)
- weniger Bearbeitungszeit – hierdurch geringere Bearbeitungstiefe (z. B. Metzner & Mann, 1952)
- geringere Textverarbeitungskompetenz erforderlich – potentieller Einfluss auf die Teilnahmebereitschaft bestimmter Personengruppen (z. B. Bortz & Döring, 2009)
- geringerer Grad der Standardisierung infolge einer hohen Anzahl verschiedener Interviewer (z. B. Amelang & Schmidt-Atzert, 2006)

Schließlich sind auch der Itemkontext (Bortz & Döring, 2009) sowie die Itemabfolge (Schnell, Hill & Esser, 2008) mögliche Einflussgrößen in Bezug auf die psychometrische Qualität von Skalen und Einzelitems, wobei das Streben der TN nach inhaltlicher Konsistenz hier eine nicht unwesentliche Rolle spielt (Laux & Weber, 1993; Mummendey, 1990). Hinsichtlich des CAPI der 1. Welle der lidA-Hauptstudie befinden sich viele der eingesetzten Items in einem neuen Itemkontext und einige werden in einer von der ursprünglichen Itemabfolge abweichenden Abfolge angewandt.

Die zuvor dargelegten Überlegungen begründen, dass eine Überprüfung der psychometrischen Qualität einzelner Skalen und Einzelitems des lidA-CAPI für eine kritische Einschätzung der (Konstrukt-)Validität dringend angezeigt und zugleich in besonderem Maße diffizil ist. Ausgehend vom thematischen Schwerpunkt der lidA-Studie, namentlich dem Inhaltsbereich „Arbeit“, werden vor diesem Hintergrund in der vorliegenden Arbeit sechs arbeitsbezogene Skalen der zweiten Version des Copenhagen Psychosocial Questionnaire (COPSOQ-II) von Pejtersen, Kristensen, Borg & Bjorner (2010) einer psychometrischen Item- und Skalenanalyse unterzogen.

2.1.1 COPSOQ

Die erste Version des Copenhagen Psychosocial Questionnaire (COPSOQ-I) wurde 1997 in Dänemark am „National Research Centre for the Working Environment (NRCWE)“ entwickelt (Kristensen, 2000). Die Autoren geben an, bei der Konstruktion des COPSOQ vier Kriterien gefolgt zu sein (Kristensen, Hannerz, Hogh & Borg, 2005):

- wichtige Aspekte psychosozialer Belastungen, individueller Merkmale und Beanspruchungen im Arbeitskontext berücksichtigen
- ausschließlich theoriebasierte Skalen und Einzelitems auswählen bzw. konstruieren, ohne sich auf eine Theorie zu beschränken
- verschiedene Analyseebenen, von der organisationalen bis zur personalen Ebene, einbeziehen
- von Branchen und Berufsgruppen unabhängige, allgemeine Formulierungen sicherstellen

Die Mehrheit der Skalen und Einzelitems des COPSOQ-I wurde geprüften Instrumenten entnommen, welche zuvor bei verschiedenen, i. d. R. größeren (Bevölkerungs-)Studien eingesetzt worden sind (Nübling, Stößel, Hasselhorn, Michaelis & Hofmann, 2005). Neben den arbeitsbezogenen Items umfasst der COPSOQ-I Fragen zu grundlegenden soziodemografischen Merkmalen.

Der COPSOQ-I wurde bis zum Jahr 2010 in zwölf verschiedene Sprachen übersetzt: Chinesisch, Deutsch, Englisch, Flämisch, Japanisch, Malaysisch, Norwegisch, Persisch, Portugiesisch, Spanisch, Schwedisch und Türkisch (Pejtersen et al., 2010). Darüber hinaus existiert bereits seit mehreren Jahren ein internationales COPSOQ-Netzwerk (www.copsoq-network.org), das von der „Freiburger Forschungsstelle Arbeits- und Sozialmedizin (FFAS)“ organisiert wird. Dementsprechend wird der COPSOQ als ein weltweit verbreitetes Verfahren zur Erfassung psychosozialer Aspekte der Arbeitsbedingungen bezeichnet (vgl. Nübling et al., 2010).

Vielfältige Rückmeldungen von Wissenschaftlern und Praktikern sowie statistische Analysen der Daten des COPSOQ-I und Analysen der Daten einer anonymisierten, schriftlichen Befragung von dänischen Erwerbstätigen in den Jahren 2004 und 2005 mit einer Testversion des COPSOQ-II begründeten im Jahr 2007 die Fertigstellung der zweiten Version des COPSOQ (Pejtersen et al., 2010; NRCWE, 2007c). Beide COPSOQ-Versionen, d. h. COPSOQ-I und COPSOQ-II, liegen in der (dänischen und) englischen Sprache jeweils in drei unterschiedlichen Längen vor, die jeweils auseinander hervorgehen (vgl. Pejtersen et al., 2010; Kristensen, 2000): (1) Lang (COPSOQ-I: 141 Items. COPSOQ-II: 127 Items), (2) Mittel (COPSOQ-I: 95 Items. COPSOQ-II: 87 Items) und (3) Kurz (COPSOQ-I: 44 Items. COPSOQ-II: 40 Items).

Der COPSOQ ist ein Selbstausskunftsinstrument, das die subjektive Bewertung inhaltlicher Konstrukte mittels vier- bis fünfstufiger, unipolarer Antwortskalen mit verbalen Marken abbildet. Es handelt sich hierbei um ein „Papier-Bleistift-Verfahren“, dessen Anwendung i. d. R. als anonymisierte, schriftliche Befragung erfolgt. Die meisten COPSOQ-Skalen beinhalten höchstens vier Items (COPSOQ-I-Lang: 19 von 30 Skalen. COPSOQ-II-Lang: 32 von 33 Skalen). Neben inhaltlichen Skalen umfasst der COPSOQ inhaltliche Einzelitems (COPSOQ-I-Lang: 6 Einzelitems. COPSOQ-II-Lang: 8 Einzelitems). Der COPSOQ ist ein Instrument zur Datenerhebung im Kontext der Erwerbsarbeit, wobei die Autoren drei mögliche Anwendungsbereiche definieren, die mit den drei unterschiedlichen Längen korrespondieren (vgl. Pejtersen et al., 2010; Kristensen, 2000): (1) Wissenschaft (Lang), (2) Beratung (Mittel) und (3) Unternehmen (Kurz). Alle drei Längen beider COPSOQ-Versionen sowie Literatur und einige Ergebnisse statistischer Analysen sind kostenlos bei den Autoren erhältlich und können zudem auf der Internetpräsenz des NRCWE als PDF herunter geladen werden. Nübling et al. (2005) haben den COPSOQ-I auf der Grundlage einer deutschen Beschäftigtenstichprobe übersetzt, kulturell adaptiert und psychometrisch analysiert. Darüber hinaus haben die Autoren eine Kurzversion des COPSOQ-I erstellt, welche als so genannte „COPSOQ: deutsche Standard-Version“ (FFAS, 2007) in Deutschland Verbreitung gefunden hat. Nübling et al. (2005) bewerten den COPSOQ-I auf der Grundlage Ihrer Analyseergebnisse bezüglich der DIN EN ISO 10075-3 als ein (Screening-)Verfahren für Übersichtszwecke.

Im lidA-CAPI werden Items der nachfolgend angeführten sechs Skalen des COPSOQ-II verwendet: „Quantitative Anforderungen“, „Einfluss bei der Arbeit“, „Entwicklungsmöglichkeiten“, „Führungsqualität“, „Hilfe und Unterstützung von Kollegen“ sowie „Arbeit-Familie/Privatleben-Konflikt“. Die verwendeten Items jener Skalen sind im Abschnitt 5.2 detailliert wiedergegeben. Die Überprüfung der psychometrischen Skalen-Qualität ist für

- die deutschsprachige Anwendung,
- den Erhebungsmodus des CAPI sowie
- den Itemkontext und die Itemabfolge des lidA-CAPI bisher nicht erfolgt.

Außerdem ist dem Autor keine deutsch- oder englischsprachige Publikation bekannt (Stand: August 2012), in welcher eine faktorenanalytische Überprüfung der postulierten Dimensionalität der hier interessierenden COPSOQ-II-Skalen auf der Ebene der Items realisiert wurde.

Eine kritische Analyse der benannten Skalen sowie deren Items ist demgemäß dringend angezeigt und u. a. dadurch erschwert, dass ausschließlich diskriminante Konstrukte im lidA-CAPI erhoben werden (vgl. Abschnitt 2.1).

2.1.2 Weitere Instrumente

Im Kontext der psychometrischen Item- und Skalenanalyse der im vorherigen Abschnitt bezeichneten sechs COPSOQ-II-Skalen werden für die Entwicklung eines empirischen nomologischen Netzwerkes (siehe Abschnitt 3.1.6) drei zusätzliche Instrumente herangezogen, die im folgenden Abschnitt kurz dargestellt sind. Sämtliche Instrumente sind zum Einen Bestandteil der lidA-Studie und zum Anderen Bestandteil verschiedener Quer- und Längsschnittstudien in Deutschland, so dass hierfür empirische Vergleichswerte (siehe Abschnitt 3.2) herangezogen werden können.

Das erste zusätzliche Instrument ist die zweite von insgesamt sieben Dimensionen des „Work Ability Index (WAI)“ von Tuomi, Ilmarinen, Jahkola, Katajarinne & Tulkki, (1998) mit dem Namen „Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen“. Diese WAI-Dimension weist innerhalb verschiedener Berufsgruppen in Quer- und Längsschnittstudien hinsichtlich des WAI-Gesamtwertes eine hohe Prognosegüte auf (z. B. Ebener, Hardt, Galatsch & Hasselhorn, 2011). Mittels des WAI-Gesamtwertes wiederum konnte in mehreren Längsschnittstudien der frühzeitige Ausstieg aus dem Erwerbsleben prognostiziert werden (z. B. Ilmarinen, 2006). Die englische Bezeichnung „Work Ability Index“ wird im Deutschen als „Arbeitsfähigkeitsindex“ (BAuA, 2011, S. 9) bzw. „Arbeitsbewältigungsindex“ (BAuA, 2011, S. 9) übersetzt. In Anlehnung an Ilmarinen & Tuomi (2004) definieren Hasselhorn & Freude (2007, S. 9) Arbeitsfähigkeit als „die Fähigkeit eines Menschen, eine gegebene Arbeit zu einem bestimmten Zeitpunkt zu bewältigen“. Hiernach stellt die Arbeitsfähigkeit, wie weiter oben bereits herausgestellt (siehe Kapitel 1), einen wesentlichen Teil der Beschäftigungsfähigkeit dar (Beermann, 2011). Der WAI findet mittlerweile seit über einer Dekade in vielen europäischen Staaten in ganz unterschiedlichen Berufsfeldern Anwendung (vgl. BAuA, 2002). Es handelt sich bei dem WAI um ein modular aufgebautes Selbstauskunftsinstrument mit der Option einer Gesamtwertbildung, welches in der deutschen Sprache bspw. über das nationale WAI-Netzwerk (Gauggel, 2008) kostenlos bezogen werden kann. Eine detaillierte Darstellung der Operationalisierung der „Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen“, wie sie in dem lidA-CAPI erhoben wird, ist im Abschnitt 5.3.1 vorgenommen.

Das zweite zusätzliche Instrument wird uneinheitlich entweder als „Gedanke an Berufsaufgabe“ (NEXT-Studie, 2002) oder als „Gedanke an Berufsausstieg“ (Hasselhorn & Müller, 2004) bezeichnet. Die operationale Entwicklung in der Form eines Einzelitems wurde im Rahmen der NEXT-Studie (siehe Abschnitt 3.2.1) vollzogen. Dieses Einzelitem ist Bestandteil von Erhebungsinstrumenten, die bspw. in Krankenhäusern und (Alten-)Pflegeeinrichtungen (z. B. Gregersen, Dulon, Nienhaus & Nübling, 2010; Hasselhorn, Tackenberg, Müller & NEXT-Studiengruppe, 2005) im Rettungsdienst (z. B. Schumann, 2012) und vielen weiteren Berufsfeldern (Nübling et al., 2005) Anwendung finden. Je nach Studienkontext erfolgt zumeist eine berufsspezifische Anpassung der Itemformulierung,

bspw. „(...) daran gedacht, den Pflegeberuf zu verlassen (...)“ (Hasselhorn, Tackenberg, Büscher, Stelzig, Kümmerling & Müller, 2005, S. 138). Das Einzelitem wird als Selbstausskunftsinstrument eingesetzt und kann der einschlägigen wissenschaftlichen Literatur kostenlos entnommen werden. Ein empirischer Nachweis über den Zusammenhang zwischen dem benannten Einzelitem und dem tatsächlichen Berufsausstieg im Rahmen einer Längsschnittuntersuchung wurde im Rahmen einer deutsch- oder englischsprachigen Publikation, dem Kenntnisstand des Autors zufolge (Stand: August 2012), bisher nicht vorgenommen. Eine detaillierte Darstellung der Operationalisierung des Instrumentes „Gedanke an Berufsaufgabe“, wie es in dem lidA-CAPI verwendet wird, ist im Abschnitt 5.3.2 vollzogen.

Das dritte zusätzliche Instrument ist ein Einzelitem zur Erfassung der selbsteingeschätzten Gesundheit, welches weltweit in Quer- und Längsschnittstudien Anwendung findet (Robert Koch-Institut, 2011). Obgleich die Weltgesundheitsorganisation (World Health Organization – WHO) vor mehr als einer Dekade empfohlen hat, eine einheitliche Konstruktbezeichnung sowie eine ebenso einheitliche korrespondierende Operationalisierung (Itemformulierung und Antwortskala) zu verwenden (de Bruin, Picavet & Nossikoy, 1996), zeichnet sich die Forschungslandschaft diesbezüglich noch immer eher durch Heterogenität aus, bspw. „Allgemeine Gesundheit“ (Bellach, Knopf, & Thefeld, 1998), „Gesundheitszustand“ (Schulenburg, Claes, Greiner & Uber, 1998) oder „Gegenwärtiger Gesundheitszustand“ (Infratest Sozialforschung, 2012). Gemeinsam ist den verschiedenen Bezeichnungen und Formulierungen die Form des Einzelitems sowie eine hohe Augenscheinvalidität und ein leichtes Verständnis, d. h. eine hohe Akzeptanz der Frage nach der selbsteingeschätzten bzw. subjektiven Gesundheit. Hierbei wird nicht zwischen der psychischen und physischen Gesundheit differenziert, sondern vielmehr nach der „allgemeinen“ Gesundheit gefragt. Der selbsteingeschätzte Gesundheitszustand weist, unabhängig von den einzelnen Itemformulierungen und Antwortskalen, in Quer- und Längsschnittstudien Zusammenhänge auf, u. a. mit Major Depression (Badawi et al., in Druck), koronaren Herzerkrankungen (z. B. Moller, Kristensen & Hollnagel, 1996) sowie Mortalität (z. B. de Salvo, Bloser, Reynolds, He, & Muntner, 2006; Idler & Benyamini, 1997). Eine detaillierte Darstellung des Einzelitems „Gegenwärtigen Gesundheitszustandes“, wie sie in dem lidA-CAPI verwendet wird, wird im Abschnitt 5.3.3 vorgenommen.

3. Referenzwerte

Prämisse für die Durchführung einer hochwertigen psychometrischen Item- und Skalenanalyse der fokussierten sechs COPSOQ-II-Skalen ist zunächst eine fachlich fundierte Auswahl relevanter theoretischer Kennwerte. Ergänzend hierzu sind adäquate (bspw. hinsichtlich grundlegender soziodemografischer Merkmale äquivalente Stichproben der einzelnen Datenerhebungen) empirische Vergleichswerte zu ermitteln, um der Kontextspezifität der Konstrukte Rechnung zu tragen. Schließlich ist, unter Heranziehung der theoretischen wie empirischen Referenzwerte, die Festlegung eines als akzeptabel zu bewertenden Spektrums möglicher quantitativer Ausprägungen jener Kennwerte vonnöten, um zumindest in den Grenzbereichen (akzeptabel vs. nicht akzeptabel) eine eindeutige Qualitätsbewertung der sechs COSOQ-II-Skalen, wie sie im lidA-CAPI Anwendung finden, vornehmen zu können.

Vor diesem Hintergrund wird nachfolgend in einem ersten Schritt eine wohl begründete Auswahl relevanter Kennwerte vorgenommen (siehe Abschnitt 3.1). Die Selektion orientiert sich hierbei zum Einen an den etablierten wissenschaftlichen Standards zur Durchführung von Item- und Skalenanalysen psychometrischer Instrumente (z. B. Fisseni, 1997; Lienert & Raatz, 1998; Amelang & Schmidt-Atzert, 2006; Bortz & Döring, 2009; Bühner, 2011). Zum Anderen richtet sie sich an den gegebenen Bedingungen der lidA-Studie, insb. des lidA-CAPI der 1. Welle der lidA-Hauptstudie, aus und stellt damit eine pragmatische Durchführung der Analysen sicher.

In einem zweiten Schritt werden die Datensätze von vier Studien mit deutschen (älteren) Erwerbstätigen zur Ermittlung empirischer Referenzwerte herangezogen, in welchen die sechs COPSOQ-II-Skalen sowie die drei zusätzlich herangezogenen Instrumente Bestandteil der Datenerhebung sind (siehe Abschnitt 3.2). Obgleich aktuell eine deutsche COPSOQ-Datenbank aufgebaut wird und bereits teilweise nutzbar ist (Nübling, Vomstein, Nübling, Stößel, Hasselhorn & Hofmann, 2011), stellt diese für die vorliegende Arbeit keine relevanten empirischen Referenzwerte bereit, da sie ausschließlich auf verschiedenen Datenerhebungen mit dem COPSOQ-I bzw. der „COPSOQ: deutsche Standard-Version“ (FFAS, 2007) basiert und die Divergenzen zu den COPSOQ-II-Skalen, wie sie in der lidA-Studie verwendet werden, z. T. sehr stark sind.

3.1 Theoretische Referenzwerte

Die Auswahl relevanter Kennwerte sowie die Bestimmung ihrer kritischen Grenzwerte im Rahmen einer psychometrischen Item- und Skalenanalyse sind in der vorliegenden Arbeit an etablierten Standards wissenschaftlicher Personenbefragungen ausgerichtet. Im Wesentlichen wird daher, unter der Berücksichtigung der Rahmenbedingungen der lidA-Studie, den grundlegenden Empfehlungen von Fisseni (1997), Lienert & Raatz (1998), Amelang & Schmidt-Atzert (2006) sowie Bortz & Döring (2009) und Bühner (2011) zur Durchführung einer psychometrischen Item- und Skalenanalyse gefolgt. In den folgenden sechs Abschnitten sind die hiernach selektierten Gütekriterien dargestellt:

- Parameter der Verteilung der Werte von Items und Skalen (siehe Abschnitt 3.1.1)
- Item- und Skalenschwierigkeit (siehe Abschnitt 3.1.2)
- korrigierte und nicht-korrigierte Trennschärfe (siehe Abschnitt 3.1.3)
- Homogenität (siehe Abschnitt 3.1.4)
- Dimensionalität der gesamten Itemmenge (siehe Abschnitt 3.1.5)
- nomologisches Netzwerk aller Skalen und Einzelitems (siehe Abschnitt 3.1.6)

Jeder dieser sechs Abschnitte beginnt mit einer Minimaldefinition des entsprechenden Aspektes der psychometrischen Item- und Skalenanalyse. Fortgesetzt werden die Abschnitte jeweils mit der Darstellung ebenso etablierter wie praktikabler Quantifizierungen, d. h. statistischer Kennwerte. Sämtliche sechs Abschnitte schließen mit literaturbasierten Empfehlungen für konkrete Grenzwerte der zuvor benannten Kennwerte.

3.1.1 Verteilung der Werte

Die auf der Grundlage einer empirischen Stichprobe ausgezählten absoluten Häufigkeiten der einzelnen Antwortoptionen (AO) eines Items bzw. einer Skala (Rohwerte oder positiv-linear transformierte Werte) bilden eine Gestalt, die als Verteilung der Werte bezeichnet wird (vgl. Lienert & Raatz, 1998; Fisseni, 1997). Jede Item- und Skalenanalyse sollte mit einer Betrachtung dieser Verteilung beginnen, da die Auswahl der statistischen Verfahren für die nächstfolgenden Analyseschritte in vielen Fällen maßgeblich von der Verteilungsform insb. der Normalverteilung abhängen (Bortz & Döring, 2009; Lienert & Raatz, 1998).

Zur Beurteilung der Qualität eines Items bzw. einer Skala im Kontext der Betrachtung einer empirischen Verteilung ist grundsätzlich die Berücksichtigung der drei nachstehend erläuterten Sachverhalte angezeigt (Bühner, 2011; Bortz & Döring, 2009; Bortz, Lienert, & Boehnke, 2008; Lienert & Raatz, 1998; Fisseni, 1997).

Erstens ist zu prüfen, ob die gesamte mögliche Spannweite der AO des Items bzw. der Skala von den TN tatsächlich ausgeschöpft wurde (Lienert & Raatz, 1998). D. h., die empirische Spannweite sollte grundsätzlich der theoretischen Spannweite entsprechen, um in ausreichendem Maße Streuung und hierdurch Differenzierungspotential der Items und/oder Skala zu erhalten. Ein hohes Differenzierungspotential wiederum stellt eine wichtige Grundlage zur Identifikation von Korrelationen mit anderen Konstrukten dar.

Zweitens sollte weder innerhalb der ersten (Bodeneffekt) noch der letzten AO (Deckeneffekt) eines Items die relative Mehrheit der Häufigkeiten einer gegebenen, empirischen Stichprobe lokalisiert sein (vgl. Lienert & Raatz, 1998). Auch hierbei geht es, wie bei der zuvor erwähnten empirischen Spannweite, um das Differenzierungspotential des Items. Da es keinen allgemein verbindlichen Standard bezüglich der quantitativen Definition eines Boden- bzw. Deckeneffektes gibt, folgt der Autor hierbei dem Vorgehen von Nübling et al. (2005). Dementsprechend liegt ein Boden- bzw. Deckeneffekt dann vor, wenn die erste bzw. letzte AO eines Items von mindestens 50 Prozent sämtlicher TN der empirischen Stichprobe, deren Antworten als gültig bewertet wurden, gewählt wird.

Drittens sollte überprüft werden, ob die empirischen Werte normalverteilt sind, weil die Normalverteilung neben der Skalendignität (nicht-kardinales vs. kardinales Skalenniveau), der Varianzhomogenität und verschiedenen Stichprobencharakteristika (z. B. Repräsentativität), ein zentrales Entscheidungskriterium für oder gegen parametrische (parametrische vs. verteilungsfreie) Verfahren ist (Bühner, 2011; Bortz & Döring, 2009; Bortz, Lienert, & Boehnke, 2008; Lienert & Raatz, 1998). Zur Überprüfung der Normalverteilungsannahme finden, in Ergänzung zur Augenscheinkontrolle der Rohwerte bzw. der positiv-linear transformierten Werte (Bortz et al., 2008), zwei statistische Testverfahren besonders häufig Anwendung (Seier, 2002): Kolmogorov-Smirnov-Anpassungstest mit Lilliefors-Schranken (Lilliefors, 1967) sowie Shapiro-Wilk-Test (Rahman & Govindarajulu, 1997).

Bei dem Kolmogorov-Smirnov-Anpassungstest mit Lilliefors-Schranken (K-S-A) lautet die Nullhypothese, dass die Grundgesamtheit der empirischen Stichprobe normalverteilt sei (Bortz, et al., 2008). Der K-S-A basiert auf Berechnungen der Schiefe sowie der Kurtosis (Seier, 2002). Das Ergebnis des K-S-A ist eine D-Statistik sowie ein korrespondierender p-Wert. Wenn die D-Statistik größer ist als der kritische Wert (D-Kritisch), wird die Nullhypothese verworfen und stattdessen die Alternativhypothese, die Grundgesamtheit der Stichprobe sei nicht normalverteilt, angenommen (Bortz et al., 2008). Um den Wert D-Kritisch explizit zu ermitteln, können entweder in der Literatur vorliegende Tabellen (z. B. Lilliefors, 1967) oder eine exakte Berechnungsformel (z. B. Conover, 1971) herangezogen werden, wobei hier jeweils das a priori gewählte Signifikanzniveau und die Stichprobengröße berücksichtigt werden (Bortz, et al., 2008). Der K-S-A weist zwar keine Anwendungslimitierung bezüglich der Stichprobengröße auf, jedoch reagiert er mit zunehmendem Stichprobenumfang stärker auch auf kleine Abweichungen von der Normalverteilung (Lilliefors, 1967).

Die Nullhypothese des Shapiro-Wilk-Testes (S-W-T) besagt ebenfalls, dass die Grundgesamtheit der empirischen Stichprobe normalverteilt sei (Duller, 2008; Sachs & Hedderich, 2006; Precht, Kraft & Bachmaier, 2005). Der S-W-T basiert auf regressionsanalytischen und korrelativen Berechnungen (Seier, 2002). Das Ergebnis des S-W-T ist eine W-Statistik sowie ein korrespondierender p-Wert. Wenn die W-Statistik kleiner ist als der kritische Wert (W-Kritisch), wird die Nullhypothese abgelehnt und die Alternativhypothese, die Grundgesamtheit der Stichprobe sei nicht normalverteilt, angenommen (Precht et al., 2005). Zur Bestimmung des Wertes W-Kritisch können entweder spezielle Tabellen (z. B. Shapiro & Wilk, 1965) oder eine präzise Berechnungsformel (z. B. Precht et al., 2005) verwendet werden, wobei jeweils die Stichprobengröße und das a priori gewählte Signifikanzniveau in Rechnung gestellt werden (Precht et al., 2005). Der S-W-T zeichnet sich gegenüber vergleichbaren Verfahren durch eine besonders hohe Teststärke aus (Coin, 2008; Sachs & Hedderich, 2006) und wird für Stichproben von $n = 3$ bis $n = 5\,000$ herangezogen (Rahman & Govindarajulu, 1997).

3.1.2 Item- und Skalenschwierigkeit

Das Konzept der Schwierigkeit im Kontext einer psychometrischen Item- und Skalenanalyse findet vornehmlich in der psychologischen Intelligenz- und Leistungsdiagnostik Anwendung (Lienert & Raatz, 1998). Die Bezeichnung dieses Konzeptes als Schwierigkeit ist in der Tatsache begründet, dass in dem benannten Forschungs- und Anwendungsfeld die Bearbeitungsergebnisse bzw. Antworten eines Items entweder als falsch oder als richtig bezeichnet werden können. Demgemäß können die Items für die TN hier unterschiedlich schwierig bzw. leicht sein. Gleichsam ist die Berücksichtigung dieses Konzeptes weit über den Bereich der Intelligenz- und Leistungsdiagnostik hinaus (Schwierigkeit als „Lösungsrate“ – Bortz, & Döring, 2009, S. 218) bei psychometrischen Datenerhebungen allgemein (Schwierigkeit als „Zustimmungsrate“ – Bortz, & Döring, 2009, S. 218) üblich und sinnvoll (z. B. Bortz, & Döring, 2009; Lienert & Raatz, 1998; Fisseni, 1997). Der Schwierigkeitsbegriff bezeichnet zunächst und allgemein ein Gütekriterium (z. B. Fisseni, 1997). Zusätzlich bezeichnet der Begriff die maximale Ausprägung eben dieses Gütekriteriums, während die minimale Ausprägung als Leichtigkeit bezeichnet wird. Die Schwierigkeit (bzw. Leichtigkeit) wird definiert als derjenige Anteil von TN einer Stichprobe, welcher das jeweilige Item resp. die jeweilige Skala nicht im Sinne des Konstruktes (bzw. im Sinne des Konstruktes) bearbeitet hat (vgl. Bortz, & Döring, 2009; Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998; Fisseni, 1997).

Die Quantifizierung der Schwierigkeit (Schwierigkeitsindex) bei mehrstufigen Antwortskalen erfolgt i. d. R. durch die Berechnung des Medians (Md) mit dem Interquartilsabstand (IQR – Differenz zwischen dem 1. und 3. Quartil) als Streuungsmaß (Bühner, 2011) und/oder durch die Berechnung des Arithmetischen Mittels (AM) mit der Standardabweichung (SD) als Streuungsmaß (Bortz, & Döring, 2009; Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998; Fisseni, 1997). Während der Md die Daten in zwei „gleichgroße“ Hälften teilt, teilt das AM die Daten gewissermaßen in zwei „gleichstarke“ Hälften (vgl. Bortz, 1999). Der Md reagiert im Vergleich zum AM nicht auf Ausreißer (d. h. besonders große oder kleine Werte), so dass eine allzu starke Differenz zwischen AM und Md ein milder Hinweis auf das Vorliegen von Ausreißern sein kann. Außerdem gibt der Md stets eine beobachtete Ausprägung an, während das AM oftmals einen artifiziellen Wert annimmt, d. h. ein Wert, der nicht im Spektrum der für den TN verfügbaren AO existiert. In vielen Analysen wird häufig das AM als Item- und/oder Skalenschwierigkeit angegeben (vgl. Bortz, 1999), so dass jener Parameter für eine Vergleichbarkeit hinsichtlich anderer Studien grundsätzlich berechnet werden sollte. Ein quantitativ hoher (bzw. niedriger) Schwierigkeitsindex bedeutet, dass viele (bzw. wenige) TN das Item lösen resp. dem Item zustimmen (bzw. nicht lösen resp. nicht zustimmen), d. h. dieses Item ist leicht (bzw. schwer) (Lienert & Raatz, 1998). Wird der Schwierigkeitsindex durch einen der zuvor genannten statistischen Kennwerte berechnet, entspricht die theoretische Spannweite des

Schwierigkeitsindex der theoretischen Spannweite der Antwortskala des jeweiligen Items (bspw. 0 bis 100). Das zuvor Dargestellte ist nicht nur für Items, sondern äquivalent auch für Skalen anwendbar.

Die quantitative Ausprägung des Schwierigkeitsindex hängt in unterschiedlicher Stärke mit anderen Kennwerten (insb. Trennschärfe und Homogenität) der Item- und Skalenanalyse zusammen – einen Überblick hierzu bieten Lienert und Raatz (1998) sowie Fisseni (1997). Die wichtigsten Korrelationen zwischen der Ausprägung der Schwierigkeit und der Trennschärfe (siehe Abschnitt: 3.1.3) sowie der Homogenität (siehe Abschnitt: 3.1.4) sind in den nachfolgenden Abschnitten skizziert.

Die a priori Bestimmung der optimalen Ausprägung des Schwierigkeitsindex wird durch verschiedene Rahmenbedingungen der jeweiligen Datenerhebung und der Instrumentenkonstruktion beeinflusst. Dazu zählen vor allem die Zielsetzung der Datenerhebung, der inhaltliche Einsatzbereich des jeweils herangezogenen Instrumentes selbst (Amelang & Schmidt-Atzert, 2006) sowie bereichsspezifische Merkmale der Population (Lienert & Raatz, 1998). Außerdem sind die Priorisierungen anderer relevanter Gütekriterien der Item- und Skalenanalyse (insb. Trennschärfe und Homogenität) vor dem Hintergrund ihrer Interdependenz zu beachten, so dass im Einzelfall ein niedriger Schwierigkeitsindex zum Zwecke eines höheren anderen Kennwertes (und vice versa) akzeptiert wird (Lienert & Raatz, 1998; Fisseni, 1997). Um a priori realistische Erwartungen festzulegen, sollte ein theoretisch bestimmter Schwierigkeitsindex schließlich auch an bereits vorliegenden, empirischen Ergebnissen (empirische Referenzwerte) vergleichbarer Analysen ausgerichtet werden. Vor dem Hintergrund dieser Überlegungen kann die vorsichtige Schlussfolgerung abgeleitet werden, dass Skalen „zugunsten der Differenzierungsvielfalt“ (Fisseni, 1997, S. 47) im Allgemeinen vornehmlich mittelschwere (auf einer Antwortskala von 0 bis 100 bspw. etwa 50), ergänzend aber auch leichte (auf einer Antwortskala von 0 bis 100 bspw. größer als 50) und schwere (auf einer Antwortskala von 0 bis 100 bspw. kleiner als 50) Items beinhalten sollten (vgl. Lienert & Raatz, 1998; Fisseni, 1997). Auf der Grundlage der vorgenannten Ausführungen sollte im Rahmen einer psychometrischen Item- und Skalenanalyse die quantitative Ausprägung der Schwierigkeit im Allgemeinen innerhalb eines Wertebereiches von 30 bis 70 Prozent des theoretischen Spektrums der Antwortskala (auf einer Antwortskala von 0 bis 100 bspw. 30 bis 70) lokalisiert sein.

3.1.3 Trennschärfe

Die („Eigen-“)Trennschärfe (vgl. Bühner, 2011, S. 171) bezeichnet den Grad der Genauigkeit (bzw. Schärfe), mit welcher ein Itemwert zwischen TN, die einen hohen, und solchen, die einen niedrigen Skalenwert aufweisen, unterscheidet (bzw. trennt) (Lienert & Raatz, 1998) – es geht hierbei folglich um die „Repräsentativität“ (vgl. Bortz & Döring, 2009, S. 219) eines Items in Bezug auf die Skala, zu der es gehört, bzw. um dessen „Itemvalidität“ (Bühner, 2011, S. 171).

Zur zahlenmäßigen Bestimmung der Trennschärfe wird zumeist der so genannte Trennschärfekoeffizient als Koeffizient einer Produkt-Moment-Korrelation (nach Pearson) oder als Koeffizient einer Rangkorrelation (nach Spearman) berechnet (z. B. Bortz & Döring, 2009; Lienert & Raatz, 1998). Dieser ist i. d. R. der Koeffizient einer Korrelation zwischen dem Wert des interessierenden Items und dessen Skalenwert (Lienert & Raatz, 1998; Fisseni, 1997). Wenn der Skalenwert den Wert des interessierenden Items enthält, entspricht der resultierende Koeffizient der so genannten nicht-korrigierten Trennschärfe (z. B. Bühner, 2011; Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998). Der Koeffizient der korrigierten Trennschärfe hingegen ist um diesen Itemwert bzw. diese Eigenwertkorrelation bereinigt und dadurch in den allermeisten Fällen niedriger als derjenige der nicht-korrigierten Trennschärfe (z. B. Bühner, 2011; Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998). Im Rahmen der psychometrischen Qualitätsprüfung eines Items sollte in erster Linie die korrigierte Trennschärfe Verwendung finden (Nunnally & Bernstein, 1994). Gleichsam kann zusätzlich für eine tiefergehende Analyse die Differenz zwischen nicht-korrigierter und korrigierter Trennschärfe herangezogen werden (vgl. Lienert & Raatz, 1998). Hierbei ist zu beachten, dass die Größe des Einflusses der Eigenwertkorrelation auf die Höhe der zuvor benannten Differenz positiv mit der Skalenlänge, d. h. mit der Anzahl der Items der Skala zusammenhängt (vgl. Lienert & Raatz, 1998). Die mögliche Spannweite der zwei oben skizzierten Trennschärfekoeffizienten erstreckt sich stets über einen Wertebereich von -1.00 bis +1.00 (z. B. Bortz & Döring, 2009).

Der Zusammenhang zwischen der Ausprägung des Trennschärfekoeffizienten und anderer Gütekriterien der psychometrischen Item- und Skalenanalyse (insb. Schwierigkeit und Homogenität) ist sehr übersichtlich dargestellt von Lienert und Raatz (1998) sowie von Fisseni (1997). Grundsätzlich kann festgestellt werden, dass eine niedrige und eine hohe Schwierigkeit eher eine mittlere Trennschärfe begünstigen, während eine mittlere Schwierigkeit eher eine hohe Trennschärfe begünstigt (z. B. Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998; Fisseni, 1997). Ferner besteht tendenziell eine positive Korrelation zwischen dem Ausmaß der Homogenität einer Skala und der Höhe korrespondierender Trennschärfen (z. B. Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998; Fisseni, 1997).

Die Festlegung der optimalen Ausprägung des Trennschärfekoeffizienten im Vorfeld einer psychometrischen Item- und Skalenanalyse wird u. a. durch die Zielsetzung der Datenerhebung sowie der Instrumentenkonstruktion beeinflusst. Zusätzlich sind, infolge ihrer partiellen Interdependenzen, die Priorisierungen anderer relevanter Gütekriterien (insb. Schwierigkeit und Homogenität) zu beachten, so dass im Zweifelsfall eine niedrige Trennschärfe für ein höheres anderes Gütekriterium (und anders herum) angenommen wird (Lienert & Raatz, 1998; Fisseni, 1997). Um vor einer entsprechenden Analyse realistische, quantitative Ausprägungen der Trennschärfen zu bestimmen, sollten stets bereits vorliegende empirische Ergebnisse (empirische Referenzwerte) vergleichbarer Analysen in angemessener Weise berücksichtigt werden. Grundsätzlich sollten Trennschärfekoeffizienten keine negativen Werte aufweisen, da die jeweiligen Items die TN bezüglich des korrespondierenden Skaleninhaltes invers differenzieren (Lienert & Raatz, 1998). Im Einzelfall kann dieses Problem durch eine einfache Umpolung der (Roh-)Werte gelöst werden (z. B. Lienert & Raatz, 1998). Weise (1975) nimmt eine Analyse der Höhe quantitativer Ausprägungen der Trennschärfe vor und empfiehlt drei Bewertungskategorien für Trennschärfekoeffizienten:

- Koeffizienten mit einem Wert niedriger als .30 sind als „klein“ zu bezeichnen
- Koeffizienten mit einem Wert von .30 bis .50 sind als „mittel“ zu bezeichnen
- Koeffizienten mit einem Wert höher als .50 sind als „groß“ zu bezeichnen

Kerlinger (1986) fordert, konsistent zu den Werten von Weise (1975), eine generelle Mindesthöhe des Trennschärfekoeffizienten von .30 für jedes Item einer Skala. Bei den von Weise (1975) und Kerlinger (1986) vorgenommenen Bewertungen der Trennschärfe werden ausschließlich nicht-korrigierte Trennschärfekoeffizienten, berechnet als Produkt-Moment-Korrelationen (nach Pearson), betrachtet.

3.1.4 Homogenität

Der Begriff Homogenität bezeichnet ein Gütekriterium im Kontext der Item- und Skalenanalyse (z. B. Fisseni, 1997). Darüber hinaus markiert jener Begriff gleichzeitig das Ausprägungsmaximum dieses Gütekriteriums, während das Ausprägungsminimum als Heterogenität bezeichnet wird (Lienert & Raatz, 1998). Die Homogenität (bzw. Heterogenität) einer Skala ist definiert als das Ausmaß der inhaltlichen Konvergenz (bzw. Divergenz) sämtlicher Items, die sie konstituiert (vgl. Bühner, 2011; Bortz, & Döring, 2009; Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998; Fisseni, 1997).

Zur Quantifizierung der Homogenität einer Skala werden i. d. R. zunächst alle Items jener Skala, zumeist mittels der Produkt-Moment-Korrelation (nach Pearson) oder gemäß der Rangkorrelation (nach Spearman), paarweise miteinander korreliert (Inter-Item-Korrelation = I-I-K), und anschließend wird mittels der Fisher-z-Transformation das AM aus sämtlichen I-I-K-Koeffizienten gebildet (vgl. Bortz, & Döring, 2009; Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998). Die theoretische Spannweite des AM der I-I-K-Koeffizienten weist ein Minimum von -1.00 und ein Maximum von +1.00 auf (z. B. Bortz, & Döring, 2009). Zusätzlich kann die „itemspezifische Homogenität“ (Bortz, & Döring, 2009, S. 220) ermittelt werden, indem ausschließlich alle möglichen bivariaten Korrelationen eines Items mit allen übrigen Items einer Skala berechnet und anschließend das AM dieser I-I-K-Koeffizienten gebildet wird (Bortz & Döring, 2009). Nicht selten wird im Rahmen psychometrischer Item- und Skalenanalysen Cronbach's Alpha (Cronbach, 1951) zur Ermittlung der Skalen-Homogenität herangezogen (Bortz & Döring, 2009; Bortz, 2005; Schmitt, 1996). Dieser Kennwert basiert im Wesentlichen auf der Berechnung von Produkt-Moment-Korrelationen (nach Pearson) und setzt daher nicht nur ein kardinales Messniveau, sondern zudem normalverteilte Werte voraus. Überdies korreliert der Alphakoeffizient von Cronbach positiv mit der Skalenlänge, d. h. der Itemanzahl einer Skala (Bortz & Döring, 2009; Bortz, 2005).

Die Kohärenz zwischen der quantitativen Ausprägung der Homogenität und anderen Gütekriterien der Item- und Skalenanalyse (insb. Schwierigkeit und Trennschärfe) ist von Lienert und Raatz (1998) sowie von Fisseni (1997) zusammenfassend dargestellt. Generell ist eine Skala umso weniger homogen, je stärker die Schwierigkeiten der einzelnen Items variieren (vgl. Fisseni, 1997). Gleichzeitig ist eine geringe Variation der Item-Schwierigkeiten lediglich eine notwendige, aber keine hinreichende Bedingung für eine hohe Ausprägung der Homogenität (vgl. Fisseni, 1997). Generell sind die einzelnen Ausprägungen der I-I-K bei homogenen Skalen einander wesentlich ähnlicher als diejenigen heterogener Skalen (Lienert & Raatz, 1998). Die verschiedenen Zusammenhänge zwischen der Ausprägung der Homogenität und der Trennschärfe sind im vorherigen Abschnitt grob dargestellt (siehe Abschnitt: 3.1.3).

Zur a priori Bestimmung der optimalen quantitativen Ausprägung der Homogenität im Kontext einer psychometrischen Item- und Skalenanalyse sollten vornehmlich die Zielsetzung der Datenerhebung und die theoretischen Grundlagen der eingesetzten Skalen berücksichtigt werden. Gemäß jener theoretischen Grundlagen ist zunächst primär zu entscheiden, ob die interessierende Skala lediglich ein einziges Merkmal und deren Items verschiedene (Merkmals-)Facetten erfassen (homogene Skala) oder ob die Items der interessierenden Skala mehrere, unterschiedliche Merkmale erfassen sollen (heterogene Skala) (vgl. Bortz, & Döring, 2009; Fisseni, 1997). Außerdem sind aufgrund ihrer partiell wechselseitigen Abhängigkeiten die Priorisierungen anderer Gütekriterien (insb. Schwierigkeit und Trennschärfe) zu beachten, so dass im Einzelfall eine niedrige Homogenität für einen höhere, anderen Kennwert (und vice versa) von Vorteil ist (Lienert & Raatz, 1998; Fisseni, 1997). Um vor einer psychometrischen Item- und Skalenanalyse realistische Ausprägungen des AM der I-I-K zu bestimmen, werden zudem idealerweise bereits vorliegende, empirische Ergebnisse (empirische Referenzwerte) äquivalenter Analysen berücksichtigt. Die quantitative Ausprägung der Homogenität einer theoretisch homogenen Skala sollte einerseits nicht zu niedrig sein, da die Inhaltsbereiche der Items dann zu stark voneinander abweichen, d. h. zu heterogen sind (Bortz & Döring, 2009; Amelang & Schmidt-Atzert, 2006). Andererseits sollte die quantitative Ausprägung der Homogenität einer theoretisch homogenen Skala auch nicht zu hohe Werte annehmen, da die Inhaltsbereiche der Items dann zu wenig facettenreich sind, d. h. zu homogen sind (Bortz & Döring, 2009; Amelang & Schmidt-Atzert, 2006). Bei einer homogenen Skala sind die verschiedenen Items Operationalisierungen verschiedener Facetten desselben Konstruktes, welches die Skala misst (z. B. Bortz & Döring, 2009). Vor diesem Hintergrund empfehlen Briggs & Cheek (1986) ein als akzeptabel aufzufassendes Spektrum für das AM der I-I-K mit einem Minimum von .20 und einem Maximum von .40. Kerlinger (1986) hingegen definiert für das AM der I-I-K einen anzustrebenden Wertebereich, der sich von .30 bis .70 erstreckt und eine verhältnismäßig große Verbreitung innerhalb der Literatur zur psychometrischen Item- und Skalenanalyse erfahren hat (Knapp & Brown, 1995). Dieser zuletzt genannte Wertebereich basiert auf einem AM der I-I-K, welche als Produkt-Moment-Korrelationen (nach Pearson) berechnet werden.

3.1.5 Dimensionalität

Der Begriff Dimensionalität bezeichnet die Anzahl latenter, d. h. der unmittelbaren Beobachtung nicht zugänglicher Konstrukte (bzw. Dimensionen), für welche eine Menge wohl definierter Items Indikatoren darstellen (Bortz & Döring, 2009). Die Dimensionalitätsprüfung stellt einen ebenso praktikablen wie wichtigen Aspekt der Konstruktvalidität (faktorielle Konstruktvalidität) einer psychometrischen Skala bzw. eines Instrumentes, welches mehrere Skalen in sich vereinigt, dar (z. B. Amelang & Schmidt-Atzert, 2006; Lienert & Raatz, 1998).

Zur empirischen Überprüfung der Dimensionalität einer Itemmenge wird zumeist mittels einer so genannten Faktorenanalyse die faktorielle Struktur der Items berechnet, wobei die Faktoren den latenten Konstrukten bzw. Dimensionen entsprechen (z. B. Bühner, 2011; Wolff & Bacher, 2010; Lienert & Raatz, 1998). Hierbei sind grundsätzlich zwei Verfahrensklassen voneinander zu differenzieren: explorative Faktorenanalysen und konfirmatorische Faktorenanalysen (z. B. Bühner, 2011; Wolff & Bacher, 2010). Während die zuerst genannte Klasse quasi erwartungsfrei die empirisch beobachtete faktorielle Struktur der Itemmenge abbildet, wird mit Hilfe der zweiten Verfahrensklasse die graduelle Abweichung zwischen theoretisch erwarteter und empirisch beobachteter faktorieller Struktur statistisch ermittelt (Wolff & Bacher, 2010; Bortz, 1999). Im Rahmen von Dimensionalitätsprüfungen, insb. bei empirischen Daten aus Personenbefragungen, wird sehr häufig die Hauptkomponentenanalyse (Principal Component Analysis – PCA) als ein spezielles Verfahren explorativer Faktorenanalysen verwendet (Fromm, 2012; Handl, 2010; Wolff & Bacher, 2010; Diehl & Staufenbiel, 2002; Bortz, 1999). Die PCA gilt als ein besonders stabiles Verfahren und ermöglicht auch bei nicht-normalverteilten sowie nicht kardinalen Daten eine inhaltlich sinnvolle Interpretation der resultierenden Faktoren bzw. Komponenten (Bühner, 2006). Die Ergebnisse faktorenanalytischer Berechnungen sind grundsätzlich nicht selbsterklärend, vielmehr bedarf es hierzu eines fachlich fundierten Vorwissens sowie fortgeschrittener Methodenkenntnisse (vgl. Bortz & Döring, 2009; Bühner, 2006).

Um die Adäquanz der Anwendung einer Faktorenanalyse im Vorfeld zu bewerten, sollten zumindest, neben den oben erläuterten Aspekten einer psychometrischen Item- und Skalenanalyse (bspw. Homogenität – Bortz & Döring, 2009), die Korrelationen der interessierenden Items mittels des Kaiser-Meyer-Olkin-Kriteriums (KMO) untersucht werden (Backhaus, Erichson, Plinke & Weiber, 2011; Wolff & Bacher, 2010). Das KMO basiert im Wesentlichen auf den partiellen Korrelationen der interessierenden Itemmenge und kann einen Wert zwischen 0.00 und 1.00 annehmen. Die Mindesthöhe des KMO als Vorbedingung zur inhaltlich sinnvollen Anwendung einer Faktorenanalyse sollte .50 (vgl. Backhaus et al., 2011) bzw. .60 (vgl. Dziuban & Shirkey, 1974; Kaiser, 1974) betragen. Idealerweise weist das KMO einen Wert von .80 oder mehr auf (Backhaus et al., 2011).

Die Faktoren werden bei einer PCA Komponenten genannt und erklären in aufsteigender Reihenfolge die höchstmögliche Varianz (z. B. Wolff & Bacher, 2010; Bortz, 1999). Damit die Anzahl der extrahierten Komponenten einer PCA nicht größer ist als die Anzahl der verwendeten Items, wird zumeist gemäß dem Kaiser-Guttman-Kriterium ein Eigenwert der Komponenten von mindestens 1.00 gefordert (Wolff & Bacher, 2010; Bortz, 1999; Zeller & Carmines, 1980; Kaiser, 1958). Für eine verhältnismäßig einfache, inhaltliche Interpretation der resultierenden Komponenten wird darüber hinaus oftmals eine Rotation der Komponenten durchgeführt, wobei hier besonders häufig die so genannte orthogonale Varimaxrotation mit Kaiser-Normalisierung die Methode der Wahl ist, um statistisch voneinander unabhängige Komponenten zu erhalten (Fromm, 2012; Wolff & Bacher, 2010). Die Korrelation zwischen einer Komponente und einem Item wird (Komponenten-)Ladung genannt und kann einen Wert zwischen -1.00 und +1.00 annehmen (z. B. Fromm, 2012; Handl, 2010; Wolff & Bacher, 2010; Diehl & Staufenbiel, 2002; Bortz, 1999).

Da Items i. d. R. nicht nur auf einer (Haupt-)Komponente, sondern zusätzlich auf weiteren (Neben-)Komponenten laden, ist die Bestimmung eines quantitativen Grenzwertes zur Differenzierung von Haupt- und Nebenladungen angezeigt. Hinsichtlich der Festlegung eines solchen Grenzwertes gibt es verschiedene, auch fachspezifisch variierende Empfehlungen (z. B. Wolff & Bacher, 2010; Bortz, 1999). Als Nebenladungen gelten oftmals Ladungen ab einem Wert von .30 (vgl. Grice, 2001; Gorsuch, 1983; Nunnally, 1978). Ladungen mit einer Höhe ab .50 werden häufig als Hauptladungen definiert (vgl. Backhaus, Erichson, Plinke & Weiber, 2008). Ausschließlich Hauptladungen ($\geq .50$) und Nebenladungen ($\geq .30$ und $< .50$), d. h. insgesamt Ladungen mit absoluten Werten ab .30, werden i. d. R. als inhaltlich interpretierbar bewertet (Backhaus et al., 2008; Grice, 2001; Gorsuch, 1983; Nunnally, 1978).

3.1.6 Nomologisches Netzwerk

Das nomologische Netzwerk bezeichnet einen Komplex theoretisch begründeter und/oder empirisch nachgewiesener Korrelationen zwischen konvergenten und/oder diskriminanten Konstrukten, die als Einzelitems und/oder Skalen operationalisiert sind (vgl. Bühner, 2011; Asendorpf, 2007; Amelang & Schmidt-Atzert, 2006; Cronbach & Meehl, 1955). Die Erstellung eines nomologischen Netzwerkes wird in Abhängigkeit von verfügbaren Vergleichsdaten als ein wesentlicher Teil der Konstruktvalidierung (insb. im Kontext der Analyse von Daten aus Querschnittsuntersuchungen) einer psychometrischen Skala bzw. eines Verfahrens, das verschiedene Skalen beinhaltet, eingesetzt (z. B. Bühner, 2011; Amelang & Schmidt-Atzert, 2006; Cronbach & Meehl, 1955).

Zur empirischen Quantifizierung eines nomologischen Netzwerkes werden alle interessierenden Einzelitems und/oder Skalen entweder gemäß der Produkt-Moment-Korrelation (nach Pearson) oder mittels der Rangkorrelation (nach Spearman) jeweils bivariat miteinander korreliert (Asendorpf, 2007; Amelang & Schmidt-Atzert, 2006). Anschließend wird das hieraus resultierende Netzwerk mit einem theoretisch begründeten und/oder empirisch nachgewiesenen nomologischen Netzwerk per Augenmaß oder unter Verwendung statistischer Tests miteinander verglichen (z. B. Bühner, 2011; Asendorpf, 2007; Amelang & Schmidt-Atzert, 2006; Cronbach & Meehl, 1955).

Als Kriterien zur Erstellung eines empirischen nomologischen Netzwerkes, welches als Referenz zur Beurteilung eines anderen nomologischen Netzwerkes herangezogen wird, haben die drei nachstehend umrissenen Aspekte praktische Relevanz.

Erstens sollten die einzelnen bivariaten Korrelationen, welche die Grundlage des nomologischen Referenznetzwerkes bilden, in mehreren, voneinander unabhängigen Datenerhebungen nachgewiesen worden sein. Hierbei ist zu beachten, dass die wesentlichen Merkmale der herangezogenen Datenerhebungen (Soziodemografie der Stichprobe, Erhebungsmodus der Skalen und Einzelitems etc.) im günstigsten Fall denen derjenigen Datenerhebung, welche die Grundlage für das zu überprüfende nomologische Netzwerk darstellen, entsprechen.

Zweitens sollten die Valenzen der Koeffizienten spezifischer Korrelationen, welche die Grundlage des nomologischen Referenznetzwerkes bilden, idealerweise in allen Datenerhebungen konsistent sein, d. h. positiv oder negativ oder unkorreliert (Null-Korrelation).

Und drittens sollte die quantitative Ausprägung der Koeffizienten spezifischer Korrelationen, welche die Grundlage des nomologischen Referenznetzwerkes bilden, gemäß etablierter Beurteilungskriterien bewertet werden. Hierfür kann eine Einteilung von Cohen (1988) herangezogen werden, wonach Koeffizienten von .10 bis .29 als „klein“, solche von .30 bis .49 als „mittel“ und Koeffizienten von .50 bis 1.00 als „groß“ bezeichnet werden. Diese Wertebereiche beziehen sich auf Produkt-Moment-Korrelationen (nach Pearson).

3.2 Empirische Referenzwerte

Zur Ermittlung empirischer Referenzwerte für die vorliegende Untersuchung werden die Datensätze von vier empirischen Vergleichsstudien mit Erwerbstätigen (in der überwiegenden Mehrheit sozialversicherungspflichtig Beschäftigte) in Deutschland herangezogen. Es sind dies namentlich die nachstehend benannten Studien:

- NEXT-Studie (siehe Abschnitt 3.2.1)
- COPSOQ-I-Studie (siehe Abschnitt 3.2.2)
- IPSO-Studie (siehe Abschnitt 3.2.3)
- 3Q-Studie (siehe Abschnitt 3.2.4)

Im Folgenden sind jeweils die formalen Rahmenbedingungen (Leitung und Koordination, Kooperationen sowie Finanzierung), die für die vorliegende Untersuchung relevanten Bestandteile der eingesetzten Instrumente zur Datenerhebung (inkl. der jeweiligen Konstrukte), eine Auswahl wichtiger soziodemografischer Merkmale der Stichproben sowie die ermittelten statistischen Parameter jeder einzelnen Vergleichsstudie grob dargestellt. In jedem der vier Abschnitte wird ferner auf Tabellen im Anhang dieser Arbeit verwiesen, denen umfängliche Informationen zu den Datenerhebungsinstrumenten, den Stichproben und den statistischen Kennwerten entnommen werden können. Die Auswahl jener statistischer Kennwerte erfolgte auf der Grundlage der im Abschnitt „Theoretische Referenzwerte“ (siehe Abschnitt 3.1) dargestellten Überlegungen. Gemäß dem klassischen Vorgehen wurden in allen Vergleichsstudien einheitlich im Vorfeld der statistischen Datenanalyse fehlende Werte listenweise ausgeschlossen (vgl. Schendera, 2007), um eine zahlenmäßig homogene Stichprobenbasis für die Interpretationen der zahlreichen statistischen Kennwerte zu haben und darüber hinaus die Risiken der verschiedenen Verzerrungen als Folge der Verwendung von Imputationsverfahren (z. B. Janssen & Laatz, 2007; Höfler, 2004) zu vermeiden.

In dem letzten Abschnitt befindet sich eine Zusammenfassung der mitunter voneinander abweichenden Bezeichnungen der eingesetzten Skalen und Einzelitems sowie der unterschiedlichen Skalenstrukturen, d. h. Itemkonstellationen, und der partiell divergierenden Iteminhalte innerhalb der vier empirischen Vergleichsstudien (siehe Abschnitt 3.2.5). In diesem Abschnitt wird ebenfalls explizit auf etwaige Konvergenzen und Divergenzen in Bezug auf die in der lidA-Studie verwendeten Skalen und Einzelitems eingegangen.

3.2.1 NEXT-Studie

Die Bergische Universität Wuppertal (BUW) koordinierte von 2002 bis 2005 die europäische Studie „zum vorzeitigen Ausstieg aus dem Pflegeberuf“ mit dem englischen Titel „Nurses Early Exit Study (NEXT)“ (Simon, Tackenberg, Hasselhorn, Kümmerling, Büscher, & Müller, 2005). H.-M. Hasselhorn (wissenschaftliche Koordination) und B. H. Müller (administrative und finanzielle Koordination) leiteten die NEXT-Studie, die von der Europäischen Union (EU) finanziert wurde (Simon et al., 2005). Das Untersuchungsfeld der Studie erstreckte sich über zehn europäische Staaten: Belgien, Deutschland, Frankreich, Finnland, England, Italien, Niederlande, Norwegen, Polen und Slowakei (Hasselhorn et al., 2005). Die private Universität Witten/Herdecke zeichnete für die Realisation der Datenerhebung in Deutschland verantwortlich (Simon et al., 2005).

Im Rahmen der Studiendurchführung erfolgte im Jahr 2002 in 75 deutschen Einrichtungen (Krankenhäuser, Alten-/Pflegeheime und ambulante Pflegeheime) eine anonymisierte, schriftliche Mitarbeiterbefragung (Simon et al., 2005). Wenn im Rahmen der vorliegenden Arbeit die NEXT-Studie erwähnt wird, ist ausschließlich auf diese Teilbefragung rekurriert. Gegenstand jener Mitarbeiterbefragung waren verschiedene physische Expositionen und Anforderungen, die soziale Umgebung und die Organisation des Arbeitsplatzes sowie weitere Konstrukte, die potentiell durch die vorgenannten Aspekte beeinflusst werden (Simon et al., 2005). Das Instrument zur Datenerhebung in Deutschland beinhaltet mehr als 20 inhaltliche Skalen und Einzelitems sowie mehrere soziodemografische Items (Kümmerling, Hasselhorn & Tackenberg, 2005). Sechs der eingesetzten Skalen und Einzelitems sind für die vorliegende Untersuchung von Interesse:

- **Skala** Entwicklungsmöglichkeiten (Kristensen, 2000)
- **Skala** Führungsqualität (Kristensen, 2000)
- **Skala** Arbeit-Familie-Konflikt (Netemeyer, Boles & McMurrian, 1996)
- **Skala** Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen (Hasselhorn & Freude, 2007)
- **Einzelitem** Gedanke an Berufsausstieg (NEXT-Studie, 2002)
- **Einzelitem** Allgemeine Gesundheit (Bellach, Knopf, & Thefeld, 1998)

Die genauen Iteminhalte sowie die korrespondierenden AO der benannten Skalen und Einzelitems, welche als Grundlage der Berechnung empirischer Referenzwerte für die vorliegende Untersuchung herangezogen werden, befinden sich im Anhang (siehe Tabelle 33).

Die unbereinigte Stichprobengröße der NEXT-Studie betrug $n = 3\,565$. Vor der Ermittlung der empirischen Referenzwerte wurden fehlende Werte bezüglich der oben angeführten vier Skalen und zwei Einzelitems listenweise ausgeschlossen. Hinsichtlich der Skala

„Führungsqualität“ wurden fehlende Werte allerdings nur unter der Bedingung ausgeschlossen, dass nicht alle Items unbeantwortet blieben, weil die TN im Einzelfall keinen unmittelbar Vorgesetzten hatten und sämtliche Items der Skala in diesem Fall nicht beantworten konnten. Infolge des listenweisen Ausschlusses fehlender Werte wurden insgesamt 267 Fälle (7.5 Prozent) gelöscht. Daraus resultierte eine finale Stichprobengröße von $n = 3\,298$. Hiervon waren 49.0 Prozent mindestens 40 Jahre alt und 83.1 Prozent der TN waren weiblich. Alle TN befanden sich zum Zeitpunkt der Datenerhebung in einem sozialversicherungspflichtigen Beschäftigungsverhältnis und 69.4 Prozent arbeiteten 30 Stunden pro Woche und mehr. Detaillierte soziodemografische Daten dieser Stichprobe sind dem Anhang beigelegt (siehe Tabelle 34).

Die statistischen Analyseergebnisse der zwei Skalen „Entwicklungsmöglichkeiten“ und „Führungsqualität“ stellen einen wichtigen Teil der empirischen Referenzwerte dar, weil diese, neben vier weiteren Skalen, im Fokus der vorliegenden Arbeit stehen. Jene Referenzwerte setzen sich zusammen aus den nachfolgenden statistischen Kennwerten:

- absolute und relative Häufigkeiten der Itemwerte (siehe Tabelle 35)
- Ergebnisse des K-S-A (D-Statistik, p-Wert, D-Kritisch, α -Niveau) sowie des S-W-T (W-Statistik, p-Wert, W-Kritisch, α -Niveau), jeweils für sämtliche Items und Skalen (siehe Tabelle 36)
- erstes, zweites und drittes Quartil sowie AM und SD sämtlicher Items und Skalen (siehe Tabelle 37)
- korrigierte und nicht-korrigierte verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen Item und dem jeweils korrespondierenden Skalenwert (siehe Tabelle 38)
- (gemittelte) verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen den einzelnen Items innerhalb der jeweiligen Skala (siehe Tabelle 39)
- verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen allen sechs oben benannten Skalen bzw. Einzelitems (siehe Tabelle 40)

3.2.2 COPSOQ-I-Studie

Im Auftrag der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) führte ein Projektverbund in den Jahren 2003 und 2004 eine Studie zur kulturellen Adaptation und Reanalyse der ersten Version des „Copenhagen Psychosocial Questionnaire (COPSOQ-I)“ (Kristensen, 2000) mit dem Titel „Methoden zur Erfassung psychischer Belastungen – Erprobung eines Messinstrumentes (COPSOQ)“ durch (Nübling et al., 2005). Jener Projektverbund setzte sich zusammen aus der „Freiburger Forschungsstelle Arbeits- und Sozialmedizin (FFAS)“ (Projektleitung), der „Abteilung für Medizinische Soziologie der Universität Freiburg“ sowie dem „Lehrstuhl für Arbeitsphysiologie, Arbeitsmedizin und Infektionsschutz“ an der BUW (Nübling, Stössel & Michaelis, 2010). Außerdem fungierten das „Institut für Psychologie der Universität Freiburg“ und T. S. Kristensen als externe Kooperationspartner der COPSOQ-I-Studie (Nübling et al., 2005).

Nach der kulturellen Adaptation des COPSOQ-I wurde im Jahr 2003 eine Pilotstudie mit einer Stichprobengröße von $n = 352$ realisiert (Nübling et al., 2005). Die Ergebnisse jener Pilotstudie führten zu geringfügigen Modifikationen der kulturell adaptierten deutschen Version des COPSOQ-I, nach denen im Jahr 2004 die Hauptstudie mit einer Stichprobengröße von $n = 2\,209$ durchgeführt wurde (Nübling et al., 2005). Beide Datenerhebungen erfolgten als anonymisierte, schriftliche Befragungen in einer Stadtverwaltung (Pilotstudie) bzw. in mehreren deutschen Betrieben (Hauptstudie) (Nübling et al., 2005). Für die statistische Datenverarbeitung wurden die Stichproben beider Erhebungen zusammengefügt, wodurch eine Gesamtstichprobe mit einer Größe von $n = 2\,561$ entstand. Wenn im Rahmen der vorliegenden Arbeit die COPSOQ-I-Studie erwähnt wird, ist stets diese Gesamtstichprobe gemeint. Neun der in der COPSOQ-I-Studie eingesetzten Skalen und Einzelitems sind für die vorliegende Arbeit von Interesse:

- **Skala** Quantitative Anforderungen (Kristensen, 2000)
- **Skala** Einfluss bei der Arbeit (Kristensen, 2000)
- **Skala** Entwicklungsmöglichkeiten (Kristensen, 2000)
- **Skala** Führungsqualität (Kristensen, 2000)
- **Skala** Soziale Unterstützung (Kristensen, 2000)
- **Skala** Arbeit-Familie-Konflikt (Netemeyer et al., 1996)
- **Skala** Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen (Hasselhorn & Freude, 2007)
- **Einzelitem** Gedanke an Berufsaufgabe (NEXT-Studie, 2002 – leicht modifiziert)
- **Einzelitem** Allgemeiner Gesundheitszustand (Schulenburg et al., 1998 – Konstruktbezeichnung modifiziert)

Alle konkreten Iteminhalte sowie die jeweiligen AO dieser sieben Skalen und zwei Einzelitems sind im Anhang (siehe Tabelle 41 und Tabelle 42) ausführlich dargestellt.

Die unbereinigte Stichprobengröße der COPSOQ-I-Studie betrug $n = 2\,561$. Zur Ermittlung empirischer Referenzwerte für die vorliegende Arbeit konnten fehlende Werte nicht konsistent listenweise ausgeschlossen werden, weil dem Autor aus Gründen des Datenschutzes kein direkter Zugriff auf die Datenverarbeitung vorlag. Bei welchen Berechnungen fehlende Werte listenweise und bei welchen Berechnungen fallweise ausgeschlossen wurden, kann den jeweiligen Tabellen im Anhang entnommen werden (siehe COPSOQ-I-Studie im Anhang). Das durchschnittliche Alter der TN betrug 42.6 Jahre ($SD = 10.6$), und 56.8 Prozent der TN waren Frauen. Alle TN befanden sich zum Zeitpunkt der Datenerhebung in einem Beschäftigungsverhältnis und 70.2 Prozent der TN arbeiteten im Rahmen dieses Verhältnisses 35 Stunden pro Woche und mehr. Detaillierte soziodemografische Daten der Stichprobe sind dem Anhang beigefügt (siehe Tabelle 43).

Vor allem die statistischen Analyseergebnisse der fünf Skalen „Quantitative Anforderungen“, „Einfluss bei der Arbeit“, „Entwicklungsmöglichkeiten“, „Führungsqualität“ und „Soziale Unterstützung“ stellen einen zentralen Bestandteil der empirischen Referenzwerte dar, weil diese, neben einer weiteren COPSOQ-Skala, im Fokus der vorliegenden Arbeit stehen. Diese Referenzwerte setzen sich im Einzelnen zusammen aus den folgenden statistischen Kennwerten:

- absolute und relative Häufigkeiten der Itemwerte (siehe Tabelle 44)
- Ergebnisse des K-S-A (D-Statistik, p-Wert, D-Kritisch, α -Niveau) sowie des S-W-T (W-Statistik, p-Wert, W-Kritisch, α -Niveau), jeweils für sämtliche Items und Skalen (siehe Tabelle 45)
- erstes, zweites und drittes Quartil sowie AM und SD sämtlicher Items und Skalen (siehe Tabelle 46)
- korrigierte und nicht-korrigierte verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen Item und dem jeweils korrespondierenden Skalenwert (siehe Tabelle 47)
- (gemittelte) verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen den einzelnen Items innerhalb der jeweiligen Skala (siehe Tabelle 48)
- verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen allen neun oben benannten Skalen bzw. Einzelitems (siehe Tabelle 49)

3.2.3 IPSO-Studie

Die BUW leitete in den Jahren 2007 bis 2011 die Studie zur „Integration psychosozialer Faktoren in den Arbeits- und Gesundheitsschutz und die Organisationsentwicklung in Gerichten und Staatsanwaltschaften in Nordrhein-Westfalen (IPSO)“ (Gauggel & Tielsch, 2008). Finanziert wurde die IPSO-Studie durch die Unfallkasse Nordrhein-Westfalen (UK NRW), welche zugleich als Kooperationspartner der BUW fungierte (Gauggel, Tielsch & Portuné, 2011). An dieser Studie nahmen insgesamt 35 Behörden aus drei nordrhein-westfälischen Oberlandesgerichtsbezirken (OLG-Bezirke) teil (Gauggel & Tielsch, 2008).

Im Zuge der Studiendurchführung erfolgte im Jahr 2008 in zwölf Gerichten und Staatsanwaltschaften eine anonymisierte schriftliche Mitarbeiterbefragung (Gauggel, Portuné, Tielsch, Müller & Hasselhorn, 2009). Wenn im Rahmen der vorliegenden Arbeit die IPSO-Studie Erwähnung findet, ist ausschließlich auf diese Teilbefragung rekurriert. Gegenstand jener Mitarbeiterbefragung waren vornehmlich Aspekte des Arbeitsinhaltes, der Arbeitsorganisation und des sozialen Arbeitsumfeldes sowie Belastungsfolgen als mögliche Konsequenzen der Ausprägung zuvor angeführter Aspekte (Gauggel et al., 2011). Das eingesetzte Fragebogeninstrument basierte im Wesentlichen auf der gekürzten Version des für den deutschsprachigen Raum adaptierten COPSOQ-I (Nübling et al., 2005), welche auch als „COPSOQ: deutsche Standard-Version“ (FFAS, 2007) bezeichnet wird. Im Einzelnen sind sieben Skalen und Einzelitems dieses Instrumentes für die vorliegende Arbeit von Interesse:

- **Skala** Quantitative Anforderungen (Nübling et al., 2005)
- **Skala** Einfluss bei der Arbeit (Nübling et al., 2005)
- **Skala** Entwicklungsmöglichkeiten (Nübling et al., 2005)
- **Skala** Soziale Unterstützung von Kollegen (Nübling et al., 2005)
- **Skala** Arbeit-Familie-Konflikt (Nübling et al., 2005)
- **Einzelitem** Gedanke an Berufsaufgabe (NEXT-Studie, 2002 – leicht modifiziert)
- **Einzelitem** Generelle Gesundheit (Schulenburg et al., 1998 – Konstruktbezeichnung modifiziert)

Die konkreten Iteminhalte wie auch die entsprechenden AO jener fünf Skalen und zwei Einzelitems, die als Basis der Ermittlung empirischer Referenzwerte für die vorliegende Untersuchung herangezogen werden, sind im Anhang ausführlich dargestellt (siehe Tabelle 50).

Die unbereinigte Stichprobengröße der IPSO-Studie betrug $n = 1\,408$. Vor der Ermittlung empirischer Referenzwerte wurden fehlende Werte in den oben angeführten, sieben Skalen bzw. Einzelitems listenweise ausgeschlossen. In Bezug auf die Skala „Soziale Unter-

stützung“ wurden fehlende Werte gleichsam nur ausgeschlossen, wenn nicht alle Items unbeantwortet blieben, da die TN im Einzelfall möglicherweise keine Kollegen hatten und die entsprechenden Items demgemäß nicht beantworten konnten. Infolge dieser Vorgehensweise zum listenweisen Ausschließen fehlender Werte wurden 170 Fälle (12.1 Prozent) gelöscht. Hieraus resultierte eine finale Stichprobengröße von $n = 1\,238$. Mindestens 45 Jahre und älter waren 49.6 Prozent, und 64.0 Prozent der TN waren Frauen. Alle TN befanden sich zum Zeitpunkt der Datenerhebung in einem sozialversicherungspflichtigen Beschäftigungsverhältnis, wobei 71.3 Prozent mehr als 35 Stunden pro Woche arbeiteten. Weiterführende soziodemografische Angaben zur Stichprobe der IPSO-Studie sind in der Tabelle 51 im Anhang dargestellt.

Einen wichtigen Teil empirischer Referenzwerte stellen die statistischen Analyseergebnisse der vier Skalen „Quantitative Anforderungen“, „Einfluss bei der Arbeit“, „Entwicklungsmöglichkeiten“ und „Soziale Unterstützung“ dar, weil sich diese, neben zwei weiteren Skalen, im Fokus der vorliegenden Arbeit befinden. Eben jene Referenzwerte setzen sich im Einzelnen zusammen aus den folgenden statistischen Kennwerten:

- absolute und relative Häufigkeiten der Itemwerte (siehe Tabelle 52)
- Ergebnisse des K-S-A (D-Statistik, p-Wert, D-Kritisch, α -Niveau) sowie des S-W-T (W-Statistik, p-Wert, W-Kritisch, α -Niveau), jeweils für sämtliche Items und Skalen (siehe Tabelle 53)
- erstes, zweites und drittes Quartil sowie AM und SD sämtlicher Items und Skalen (siehe Tabelle 54)
- korrigierte und nicht-korrigierte verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen Item und dem jeweils korrespondierenden Skalenwert (siehe Tabelle 55)
- (gemittelte) verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen den einzelnen Items innerhalb der jeweiligen Skala (siehe Tabelle 56)
- verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen allen sieben oben benannten Skalen bzw. Einzelitems (siehe Tabelle 57)

3.2.4 3Q-Studie

Von 2007 bis 2011 leitete die BUW eine Längsschnittstudie zur Bestimmung und Verbesserung der „Qualität in Altenpflegeeinrichtungen – Qualität der Pflege, Qualität der Arbeit und Qualität der Organisation (3Q)“ (Dichter, Schmidt, Müller & Hasselhorn, 2007). Die 3Q-Studie ist ein Folgeprojekt der oben dargestellten NEXT-Studie (siehe Abschnitt 3.2.1) und wurde von der „Johanniter GmbH Deutschland“ sowie der „Initiative Neue Qualität der Arbeit (INQA)“ finanziert (Schmidt & Hasselhorn, 2007). Die BUW und die „Johanniter GmbH Deutschland“ realisierten diese Studie, an welcher über 50 deutsche Altenpflegeeinrichtungen teilnahmen, gemeinsam (Dichter et al., 2007).

Im Kontext der Studiendurchführung fand im Jahr 2011 in mehreren Einrichtungen der Altenpflege eine anonymisierte schriftliche Mitarbeiterbefragung statt. Wenn im Rahmen dieser Arbeit die 3Q-Studie erwähnt wird, ist stets auf diese Teilbefragung rekurriert. Gegenstand jener Mitarbeiterbefragung waren Aspekte der Arbeitsanforderung, Arbeitsorganisation, soziale Aspekte, Arbeitszufriedenheit, potentielle Wirkungen der Ausgestaltung verschiedener Arbeitsbedingungen sowie berufliche Ressourcen und einige grundlegende soziodemografische Merkmale (Schmidt, Hasselhorn & Müller, 2007). Acht der angewandten Skalen und Einzelitems sind für die vorliegende Untersuchung von Interesse:

- **Skala** Quantitative Anforderungen (Nübling et al., 2005)
- **Skala** Entwicklungsmöglichkeiten (Nübling et al., 2005)
- **Skala** Führungsqualität (Nübling et al., 2005)
- **Skala** Soziale Unterstützung (Nübling et al., 2005)
- **Skala** Arbeit-Familie-Konflikt (Netemeyer et al., 1996)
- **Skala** Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen (Hasselhorn & Freude, 2007)
- **Einzelitem** Gedanke an Berufsausstieg (NEXT-Studie, 2002 – leicht modifiziert)
- **Einzelitem** Generelle Gesundheit (Schulenburg et al., 1998 – Konstruktbezeichnung modifiziert)

Die exakten Iteminhalte sowie die jeweiligen AO der benannten Skalen bzw. Einzelitems, welche als Grundlage der Ermittlung empirischer Referenzwerte für die vorliegende Untersuchung herangezogen werden, befinden sich im Anhang (siehe Tabelle 58).

Die unbereinigte Stichprobengröße der 3Q-Studie betrug $n = 2\,294$. Vor der Ermittlung empirischer Referenzwerte wurden fehlende Werte bezüglich der oben angeführten, vier Skalen und zwei Einzelitems listenweise ausgeschlossen. Hinsichtlich der Skalen „Führungsqualität“ sowie „Soziale Unterstützung“ wurden fehlende Werte allerdings nur unter der Bedingung ausgeschlossen, dass nicht alle Items unbeantwortet bleiben, weil die TN

im Einzelfall keinen unmittelbar Vorgesetzten resp. keine Kollegen hatten und demgemäß diejenigen Items der jeweiligen Skala nicht beantworten konnten. Infolge dieser Vorgehensweise zum listenweisen Ausschließen fehlender Werte wurden insgesamt 467 Fälle (20.4 Prozent) gelöscht. Hieraus resultierte eine finale Stichprobengröße von $n = 1\,827$. Mindestens 40 Jahre alt waren 63.2 Prozent der TN, und 81.3 Prozent der TN waren weiblich. Alle TN befanden sich zum Zeitpunkt der Datenerhebung in einem sozialversicherungspflichtigen Beschäftigungsverhältnis und 66.3 Prozent arbeiteten 30 Stunden pro Woche und mehr. Detaillierte soziodemografische Daten der Stichprobe sind dem Anhang beigefügt (siehe Tabelle 59).

Insbesondere die statistischen Analyseergebnisse der vier Skalen „Quantitative Anforderungen“, „Entwicklungsmöglichkeiten“, „Führungsqualität“ sowie „Soziale Unterstützung“ stellen einen wesentlichen Teil der empirischen Referenzwerte dar, weil diese, neben zwei weiteren Skalen, im Mittelpunkt der vorliegenden Arbeit stehen. Diese Referenzwerte setzen sich im Einzelnen zusammen aus den folgenden statistischen Kennwerten:

- absolute und relative Häufigkeiten der Itemwerte (siehe Tabelle 60)
- Ergebnisse des K-S-A (D-Statistik, p-Wert, D-Kritisch, α -Niveau) sowie des S-W-T (W-Statistik, p-Wert, W-Kritisch, α -Niveau), jeweils für sämtliche Items und Skalen (siehe Tabelle 61)
- erstes, zweites und drittes Quartil sowie AM und SD sämtlicher Items und Skalen (siehe Tabelle 62)
- korrigierte und nicht-korrigierte verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen Item und dem jeweils korrespondierenden Skalenwert (siehe Tabelle 63)
- (gemittelte) verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen den einzelnen Items innerhalb der jeweiligen Skala (siehe Tabelle 64)
- verteilungsfreie Korrelationen (Rangkorrelationen nach Spearman) zwischen allen acht oben benannten Skalen bzw. Einzelitems (siehe Tabelle 65)

3.2.5 Zusammenfassung

Die Stichproben der skizzierten empirischen Vergleichsstudien setzen sich ausschließlich aus Erwerbstätigen in Deutschland zusammen und sind jeweils distinkt. Die Erhebung der interessierenden Konstrukte erfolgte in allen Vergleichsstudien mittels einer anonymisierten, schriftlichen Befragung. Allerdings wurden mitunter voneinander abweichende Bezeichnungen der Skalen und Einzelitems sowie unterschiedliche Skalenstrukturen, d. h. Itemkonstellationen, und partiell divergierende Iteminhalte und AO verwendet. Nachfolgend werden diesbezüglich bestehende Konvergenzen und Divergenzen, unter Berücksichtigung der im lidA-CAPI eingesetzten Skalen sowie Einzelitems (siehe Abschnitt 5.2 und 5.3), konstatiert und kommentiert. Zusammenfassend sind die in den vier empirischen Vergleichsstudien eingesetzten Skalen und Einzelitems in der Tabelle 1 dargestellt.

Die Bezeichnungen der drei Skalen „Quantitative Anforderungen“, „Einfluss bei der Arbeit“ und „Führungsqualität“ wurden in allen empirischen Vergleichsstudien konsistent und für dieselben Konstrukte verwendet und entsprechen zudem den Bezeichnungen der korrespondierenden Skalen in der lidA-Studie. Außerdem sind die Übereinstimmungen zwischen den empirischen Vergleichsstudien und der lidA-Studie hinsichtlich der skalenspezifischen Itemkonstellationen sowie den Iteminhalten und AO perfekt. Die Skalen und die dazugehörigen Items werden in dieser Arbeit einheitlich mit „QA“ (Quantitative Anforderungen), „EA“ (Einfluss bei der Arbeit) und „FQ“ (Führungsqualität) abgekürzt. Weiter unten sind die QA-Skala (siehe Abschnitt 5.2.1), die EA-Skala (siehe Abschnitt 5.2.2) und die FQ-Skala (siehe Abschnitt 5.2.4), wie sie in der lidA-Studie verwendet werden, im Detail dargestellt.

Die Bezeichnung der Skala „Entwicklungsmöglichkeiten“ wurde in allen empirischen Vergleichsstudien konsistent und für dasselbe Konstrukt verwendet und entspricht überdies der Bezeichnung der korrespondierenden Skala in der lidA-Studie. Allerdings kann die Skala lediglich in zwei Vergleichsstudien (COPSOQ-I-Studie und IPSO-Studie) mit allen vier erforderlichen Items und in zwei weiteren Vergleichsstudien (NEXT-Studie und 3Q-Studie) lediglich mit drei Items gebildet werden. Jedoch stimmen die Iteminhalte und die AO in den empirischen Vergleichsstudien mit denen der lidA-Studie vollkommen überein. Unabhängig von der Vollständigkeit der Itemanzahl werden die Skalen sowie die korrespondierenden Items in der vorliegenden Arbeit einheitlich mit „EM“ (Entwicklungsmöglichkeiten) abgekürzt. Unten ist die EM-Skala (siehe Abschnitt 5.2.3), wie sie in der lidA-Studie angewendet wird, detailliert dargestellt.

Die unterschiedlichen Skalenbezeichnungen „Soziale Unterstützung“ und „Soziale Unterstützung von Kollegen“ wurden in den empirischen Vergleichsstudien für dasselbe Konstrukt verwendet. Die jeweils bezeichneten Skalen stimmen hinsichtlich der Itemkonstellation, der Iteminhalte sowie der AO innerhalb der empirischen Vergleichsstudien und auch in Bezug auf die korrespondierende Skala in der lidA-Studie, namentlich

„Hilfe und Unterstützung von Kollegen“, vollkommen überein. Vor dem Hintergrund dieser Konvergenzen werden alle Skalen sowie deren Items, entsprechend der Bezeichnung in der lidA-Studie, in der vorliegenden Arbeit einheitlich mit „HU“ (Hilfe und Unterstützung von Kollegen) abgekürzt. Im Abschnitt 5.2.5 ist die HU-Skala, wie sie in der lidA-Studie Verwendung findet, im Detail wiedergegeben.

Die Skalenbezeichnung, die Itemkonstellation und die Inhalte sowie die AO sämtlicher Items der Skala „Arbeit-Familie-Konflikt“ wurden in allen Vergleichsstudien zwar konsistent und für dasselbe Konstrukt verwendet. Allerdings ist in der lidA-Studie eine andere Skala, namentlich „Arbeit-Familie/Privatleben-Konflikt“, eingesetzt. Da beide Skalen aber dasselbe Konstrukt erfassen, werden alle Skalen sowie die korrespondierenden Items in der vorliegenden Arbeit einheitlich mit „AK“ (Arbeit-Familie/Privatleben-Konflikt) abgekürzt. Zur leichteren Differenzierung der beiden Skalen in den Tabellen, erfolgt die Kennzeichnung der fünf Items der empirischen Vergleichsstudien mit Kleinbuchstaben (AK-a, AK-b, AK-c, AK-d und AK-e) und die Kennzeichnung der zwei Items in der lidA-Studie anhand von Ziffern (AK-1 und AK-2). Im Abschnitt 5.2.6 ist die AK-Skala, wie sie in der lidA-Studie eingesetzt ist, detailliert dargestellt.

Die Bezeichnung der Skala „Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen“ wurde in den jeweiligen empirischen Vergleichsstudien konsistent und für dasselbe Konstrukt verwendet und entspricht der Bezeichnung der korrespondierenden Skala in der lidA-Studie. Außerdem sind die Übereinstimmungen zwischen den Vergleichsstudien und der lidA-Studie hinsichtlich der skalenspezifischen Itemkonstellation, den Iteminhalten und den AO nahezu perfekt. Während in einer empirischen Vergleichsstudie (COPSOQ-I-Studie) die TN alle drei Items dieser Skala direkt beantworten, wird das erste Item in zwei anderen Studien (NEXT-Studie und 3Q-Studie) jeweils indirekt über die Abfrage der beruflichen Tätigkeit beantwortet. Die Skala sowie die korrespondierenden Items werden in der vorliegenden Arbeit einheitlich mit „AF“ (Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen) abgekürzt. Unten ist die AF-Skala (siehe Abschnitt 5.3.1), wie sie in der lidA-Studie verwendet wird, im Detail wiedergegeben.

Die unterschiedlichen Bezeichnungen „Gedanke an Berufsausstieg“ sowie „Gedanke an Berufsaufgabe“ wurden in den Vergleichsstudien für dasselbe Konstrukt verwendet, welches in allen empirischen Vergleichsstudien mit einem Einzelitem erfasst wird. Die Formulierungen der Iteminhalte divergierten geringfügig, wobei die AO jeweils identisch waren, so dass insgesamt eine sehr hohe Übereinstimmung zwischen den empirischen Vergleichsstudien sowie mit dem in der lidA-Studie eingesetzten Einzelitem „Gedanke an Berufsaufgabe“ besteht. Das Einzelitem wird in der vorliegenden Arbeit daher einheitlich mit „GB“ (Gedanke an Berufsaufgabe) abgekürzt. Unten ist das Einzelitem GB (siehe Abschnitt 5.3.2), wie es in der lidA-Studie eingesetzt ist, dargestellt.

Die divergierenden Bezeichnungen „Allgemeine Gesundheit“, „Allgemeiner Gesundheitszustand“ sowie „Generelle Gesundheit“ wurden in den empirischen Vergleichsstudien für dasselbe Konstrukt verwendet, welches überdies in allen Studien mit einem Einzelitem erfasst wurde. Die Iteminhalte unterscheiden sich nur geringfügig und die Anzahl der AO (fünf vs. elf) variierten teilweise. Trotzdem wird insgesamt eine ausreichend hohe Übereinstimmung innerhalb der Vergleichsstudien sowie mit dem in der lidA-Studie eingesetzten Einzelitem „Gegenwärtiger Gesundheitszustand“ konstatiert. Vor diesem Hintergrund wird das Einzelitem in der vorliegenden Arbeit einheitlich mit „GG“ (Gegenwärtiger Gesundheitszustand) abgekürzt. Weiter unten ist das Einzelitem GG (siehe Abschnitt 5.3.3), wie es in der lidA-Studie verwendet wird, im Detail beschrieben.

Tabelle 1: Item- und Skalenübersicht (empirische Referenzwerte)

	NEXT- Studie n = 3 298	COPSOQ-I- Studie n = 2 561	IPSO- Studie n = 1 238	3Q- Studie n = 1 827	Häufigkeit
QA-1	-	X	X	X	3
QA-2	-	X	X	X	3
QA-3	-	X	X	X	3
QA-4	-	X	X	X	3
EA-1	-	X	X	-	2
EA-2	-	X	X	-	2
EA-3	-	X	X	-	2
EM-1	X	X	X	X	4
EM-2	X	X	X	X	4
EM-3	X	X	X	X	4
EM-4	-	X	X	-	2
FQ-1	X	X	-	X	3
FQ-2	X	X	-	X	3
FQ-3	X	X	-	X	3
FQ-4	X	X	-	X	3
HU-1	-	X	X	X	3
HU-2	-	X	X	X	3
HU-3	-	X	X	X	3
AK-a^N	X	X	X	X	4
AK-b^N	X	X	X	X	4
AK-c^N	X	X	X	X	4
AK-d^N	X	X	X	X	4
AK-e^N	X	X	X	X	4
AK-1	-	-	-	-	0
AK-2	-	-	-	-	0
AF-1	(x)	X	-	(x)	1 + (2)
AF-2	X	X	-	X	3
AF-3	X	X	-	X	3
GB	X	X	X	X	4
GG	X	X	X	X	4

^N. Dieses Item wird nicht im lidA-CAPI verwendet. (x) = Dieses Item wird in den empirischen Vergleichsstudien nur indirekt über die Abfrage der beruflichen Tätigkeit erhoben.

4. Fragestellungen und Hypothesen

Wie bereits im Abschnitt 2.1 dargelegt, stehen im Zentrum der vorliegenden Arbeit sechs partiell modifizierte Skalen der zweiten Version des COPSOQ (Pejtersen et al., 2010): QA-Skala (siehe Abschnitt 5.2.1), EA-Skala (siehe Abschnitt 5.2.2), EM-Skala (siehe Abschnitt 5.2.3), FQ-Skala (siehe Abschnitt 5.2.4), HU-Skala (siehe Abschnitt 5.2.5) sowie AK-Skala (siehe Abschnitt 5.2.6). Die handlungsleitende Fragestellung (F) zur Überprüfung der psychometrischen Qualität dieser COPSOQ-II-Items und -Skalen lautet wie folgt:

- **F** Weisen die sechs COPSOQ-II-Skalen (QA-Skala, EA-Skala, EM-Skala, FQ-Skala, HU-Skala und AK-Skala) im CAPI der 1. Welle der lidA-Hauptstudie eine für inhaltliche statistische Analysen ausreichend hohe psychometrische Qualität auf?

Zum Zwecke einer empirischen Überprüfung dieser Fragestellung (F) werden zunächst unter Berücksichtigung der im Abschnitt 3.1 (Theoretische Referenzwerte) dargestellten Auswahl von sechs essentiellen Aspekten der psychometrischen Qualität von Items und Skalen jeweils eine korrespondierende Teilfragestellung (TF) formuliert:

- **TF.1** Ermöglicht die empirische Verteilung der positiv-linear transformierten Werte der COPSOQ-II-Items und -Skalen ein akzeptables Differenzierungspotential?
- **TF.2** Weisen die COPSOQ-II-Items und -Skalen eine akzeptable Schwierigkeit auf?
- **TF.3** Weisen die COPSOQ-II-Items eine akzeptable Trennschärfe auf?
- **TF.4** Weisen die COPSOQ-II-Skalen eine akzeptable Homogenität auf?
- **TF.5** Bilden die COPSOQ-II-Items die postulierte Skalenstruktur ab?
- **TF.6** Ist ein empirisches nomologisches Referenznetzwerk adäquat replizierbar?

In den nachfolgenden sechs Abschnitten werden für jede dieser sechs Teilfragestellungen (TF) statistische, gerichtete, aggregierte Hypothesen (H) ausgearbeitet (vgl. Bortz & Döring, 2009): drei Hypothesen zu TF.1 in Abschnitt 4.1, vier Hypothesen zu TF.2 in Abschnitt 4.2, zwei Hypothesen zu TF.3 in Abschnitt 4.3, eine Hypothese zu TF.4 in Abschnitt 4.4, zwei Hypothesen zu TF.5 in Abschnitt 4.5 und eine Hypothese zu TF.6 in Abschnitt 4.6. Hierzu werden ebenso die aus der Literatur entnommenen theoretischen Grenzwerte (siehe Abschnitt 3.1) wie auch die empirischen Referenzwerte der fokussierten statistischen Kennwerte aus vier verschiedenen empirischen Vergleichsstudien mit Erwerbstätigen in Deutschland (siehe Abschnitt 3.2) berücksichtigt und zusammengeführt.

4.1 Verteilung der Werte

Zur Formulierung operationaler Hypothesen im Rahmen der TF.1 werden zunächst zwei in dem Abschnitt 3.1.1 dargestellte statistische Kennwerte herangezogen: (1) Spannweite und (2) relative Häufigkeit der niedrigsten und höchsten zentralen AO. Ergänzend zu deren theoretischen Grenzwerten werden die korrespondierenden Analyseergebnisse der vier oben umrissenen empirischen Vergleichsstudien mit Erwerbstätigen in Deutschland berücksichtigt (siehe Abschnitt 3.2). Die zuletzt genannten empirischen Analyseergebnisse sind in der Tabelle 2 komprimiert dargestellt.

Tabelle 2: Verteilung der Werte (empirische Referenzwerte).

	Spannweite		Antwortoptionen (AO)		Tests auf Normalverteilung	
	Min. - Max.	niedrigste (%)	höchste (%)	D-Stat.	W-Stat.	
QA-1	0.00 - 100.00	6.1 bis 13.3	3.0 bis 5.4	.21 bis .24	.89 bis .90	
QA-2	0.00 - 100.00	3.7 bis 5.1	5.8 bis 10.8	.20 bis .23	.89 bis .90	
QA-3	0.00 - 100.00	11.1 bis 17.6	2.6 bis 5.1	.20 bis .21	.90 bis .91	
QA-4	0.00 - 100.00	5.9 bis 11.1	5.1 bis 6.2	.18 bis .19	.90 bis .91	
QA-Skala	0.00 - 100.00	---	---	.06 bis .07	.99	
EA-1	0.00 - 100.00	24.3 bis 31.6	1.6 bis 2.6	.21 bis .34	.72 bis .87	
EA-2	0.00 - 100.00	21.2 bis 31.0	0.8 bis 1.4	.22 bis .39	.65 bis .87	
EA-3	0.00 - 100.00	6.1 bis 20.4	15.1 bis 17.9	.17 bis .25	.89	
EA-Skala	0.00 - 100.00	---	---	.09 bis .12	.93 bis .98	
EM-1	0.00 - 100.00	1.5 bis 7.6	18.8 bis 49.5	.17 bis .28	.73 bis .91	
EM-2	0.00 - 100.00	2.6 bis 12.5	8.0 bis 30.0	.22 bis .30	.83 bis .91	
EM-3	0.00 - 100.00	1.4 bis 4.6	24.0 bis 39.8	.23 bis .29	.77 bis .87	
EM-4	0.00 - 100.00	4.3 bis 14.1	6.5 bis 9.8	.21 bis .23	.89 bis .91	
EM-Skala	0.00 - 100.00	---	---	.08 bis .20	.85 bis .98	
FQ-1	0.00 - 100.00	5.7 bis 10.5	7.2 bis 25.4	.21 bis .27	.86 bis .91	
FQ-2	0.00 - 100.00	6.6 bis 9.0	10.7 bis 32.2	.19 bis .27	.85 bis .91	
FQ-3	0.00 - 100.00	6.2 bis 8.1	7.3 bis 27.4	.20 bis .28	.86 bis .91	
FQ-4	0.00 - 100.00	7.7 bis 14.3	7.9 bis 26.5	.19 bis .26	.86 bis .91	
FQ-Skala	0.00 - 100.00	---	---	.08 bis .16	.92 bis .98	
HU-1	0.00 - 100.00	1.6 bis 8.2	14.9 bis 29.2	.17 bis .26	.86 bis .91	
HU-2	0.00 - 100.00	1.4 bis 6.0	22.0 bis 28.0	.21 bis .27	.86 bis .89	
HU-3	0.00 - 100.00	3.6 bis 20.4	2.9 bis 10.4	.19 bis .26	.88 bis .91	
HU-Skala	0.00 - 100.00	---	---	.10 bis .14	.96 bis .98	

D-Stat. = D-Statistik des K-S-A (Kolmogorov-Smirnov-Anpassungstest). W-Stat. = W-Statistik des S-W-T (Shapiro-Wilk-Test). Anmerkungen: Zusammenfassung der Ergebnisse vier empirischer Vergleichsstudien (siehe Abschnitt 3.2).

Basierend auf den theoretischen und empirischen Referenzwerten lautet die statistische Hypothese für alle COPSQ-II-Items und -Skalen (inkl. der AK-Skala – ohne empirische Vergleichsdaten) aggregiert folgendermaßen:

- **TF.1-H.1** Bei sämtlichen COPSQ-II-Items und COPSQ-II-Skalen entspricht die theoretische Spannweite (von 0 bis 100) der empirischen Spannweite der Werte (von 0 bis 100).
- **TF.1-H.2** Die niedrigste und die höchste zentrale AO sämtlicher COPSQ-II-Items weist jeweils weniger als 50 Prozent der Gesamtzahl gültiger AO auf.

Ergänzend zu den zwei zuvor thematisierten Aspekten der Verteilung der Werte (Spannweite und relative Häufigkeiten der AO) wird die Normalverteilungsannahme sämtlicher COPSOQ-II-Items und -Skalen überprüft, weil diese für die Auswahl der weiteren statistischen Analysen ein zentrales Kriterium hinsichtlich der Entscheidung für oder gegen parametrische Verfahren darstellt. Hierbei wird auf der Grundlage der empirischen Referenzwerte (siehe Abschnitt 3.2) die nachfolgende statistische Hypothese für alle COPSOQ-II-Items und Skalen (inkl. der AK-Skala – ohne empirische Vergleichsdaten) aggregiert formuliert:

- **TF.1-H.3:** Die Werte sämtlicher COPSOQ-II-Items sowie aller COPSOQ-II-Skalen sind nicht normalverteilt.

4.2 Item- und Skalenschwierigkeit

Vier Kennwerte (siehe Abschnitt 3.1.2) werden für die Bildung operationaler Hypothesen im Rahmen der TF.2 verwendet: (1) Md, (2) IQR, (3) AM und (4) SD. Zusätzlich zu deren theoretischen Grenzwerten werden die Analyseergebnisse mehrerer empirischer Vergleichsstudien berücksichtigt (siehe Abschnitt 3.2). Diese Ergebnisse sind in der Tabelle 3 zusammengefasst.

Tabelle 3: Item- und Skalenschwierigkeit (empirische Referenzwerte).

	1. Quartil Min. - Max.	Md Min. - Max.	3. Quartil Min. - Max.	AM Min. - Max.	SD Min. - Max.
QA-1	25.00 - 50.00	50.00	50.00 - 75.00	41.16 - 53.72	24.04 - 24.75
QA-2	25.00 - 50.00	50.00 - 75.00	75.00	53.11 - 60.28	24.10 - 24.71
QA-3	25.00	25.00 - 50.00	50.00	37.03 - 44.47	24.75 - 25.42
QA-4	25.00	50.00	75.00	47.85 - 52.22	24.71 - 27.88
QA-Skala	31.25 - 37.50	43.75 - 50.00	62.50 - 68.75	46.58 - 50.60	20.05 - 21.81
EA-1	0.00	0.00 - 25.00	25.00 - 50.00	17.49 - 29.72	24.67 - 26.61
EA-2	0.00	0.00 - 25.00	25.00 - 50.00	12.42 - 28.16	20.72 - 24.58
EA-3	25.00 - 50.00	50.00 - 75.00	75.00	47.94 - 61.86	27.41 - 34.22
EA-Skala	8.33	25.00 - 41.67	33.33	25.95 - 39.91	19.26 - 19.97
EM-1	50.00 - 75.00	50.00 - 75.00	75.00 - 100.00	58.28 - 82.69	21.48 - 29.26
EM-2	25.00 - 50.00	50.00 - 75.00	75.00 - 100.00	47.64 - 71.28	22.26 - 27.36
EM-3	50.00 - 75.00	75.00	75.00 - 100.00	67.87 - 78.77	20.84 - 26.33
EM-4	25.00 - 50.00	50.00	50.00 - 75.00	45.54 - 57.66	23.53 - 27.29
EM-Skala	66.67	56.25 - 83.33	91.67	54.83 - 77.23	17.35 - 23.10
FQ-1	25.00 - 50.00	50.00 - 75.00	75.00 - 100.00	50.02 - 67.00	26.62 - 29.76
FQ-2	50.00	50.00 - 75.00	75.00 - 100.00	54.70 - 68.48	27.53 - 30.27
FQ-3	25.00 - 50.00	50.00 - 75.00	75.00 - 100.00	52.84 - 67.57	26.08 - 29.06
FQ-4	25.00 - 50.00	50.00 - 75.00	75.00 - 100.00	48.23 - 65.99	28.71 - 30.95
FQ-Skala	43.75 - 50.00	50.00 - 75.00	81.25 87.50	51.45 - 67.26	23.72 - 27.11
HU-1	50.00	50.00 - 75.00	75.00 - 100.00	56.33 - 71.72	22.07 - 28.47
HU-2	50.00	75.00	75.00 - 100.00	64.02 - 70.72	22.13 - 28.93
HU-3	25.00 - 50.00	25.00 - 75.00	50.00 - 75.00	38.30 - 60.95	24.23 - 27.38
HU-Skala	33.33 - 58.33	58.33 - 66.67	66.67 - 83.33	52.88 - 67.73	17.83 - 22.73

Anmerkungen: Interquartilsabstand (IQR) = 2. - 1. Quartil. Md = Median. AM = Arithmetisches Mittel. SD = Standard Deviation (Standardabweichung). Zusammenfassung der Ergebnisse vier empirischer Vergleichsstudien (siehe Abschnitt 3.2).

Nachstehend sind auf der Grundlage ebenso theoretischer wie empirischer Referenzwerte vier statistische Hypothese für alle COPSOQ-II-Items und -Skalen (inkl. der AK-Skala – ohne empirische Vergleichsdaten) aggregiert formuliert:

- **TF.2-H.1** Der Md aller COPSOQ-II-Items und -Skalen ist in einem Wertebereich von 25.00 bis 75.00 lokalisiert.
- **TF.2-H.2** Der IQR aller COPSOQ-II-Items und -Skalen beträgt mindestens 25.00.
- **TF.2-H.3** Das AM aller COPSOQ-II-Items und -Skalen ist in einem Wertebereich von 25.00 bis 75.00 lokalisiert.
- **TF.2-H.4** Die SD aller COPSOQ-II-Items und -Skalen beträgt mindestens 20.00.

4.3 Trennschärfe

Unter Heranziehung der im Abschnitt 3.1.3 dargestellten Inhalte werden zur Formulierung operationaler Hypothesen der TF.3 zwei statistische Kennwerte verwendet: (1) nicht-korrigierter und (2) korrigierter Trennschärfekoeffizient. Ergänzend zu deren theoretischen Grenzwerten werden die Analyseergebnisse von vier verschiedenen Vergleichsstudien mit Erwerbstätigen in Deutschland berücksichtigt (siehe Abschnitt 3.2), welche in der Tabelle 4 komprimiert dargestellt sind.

Tabelle 4: Trennschärfe (empirische Referenzwerte)

	nicht-korrigierte Trennschärfe		korrigierte Trennschärfe	
	Minimum	Maximum	Minimum	Maximum
QA-1	.78 ***	.79 ***	.62 ***	.65 ***
QA-2	.81 ***	.90 ***	.66 ***	.82 ***
QA-3	.83 ***	.85 ***	.69 ***	.74 ***
QA-4	.77 ***	.86 ***	.59 ***	.73 ***
EA-1	.67 ***	.77 ***	.40 ***	.42 ***
EA-2	.62 ***	.73 ***	.40 ***	.41 ***
EA-3	.67 ***	.84 ***	.27 ***	.34 ***
EM-1	.68 ***	.77 ***	.42 ***	.57 ***
EM-2	.78 ***	.83 ***	.49 ***	.68 ***
EM-3	.72 ***	.83 ***	.49 ***	.62 ***
EM-4	.82 ***	.84 ***	.63 ***	.71 ***
FQ-1	.85 ***	.88 ***	.74 ***	.79 ***
FQ-2	.88 ***	.93 ***	.78 ***	.87 ***
FQ-3	.84 ***	.91 ***	.72 ***	.86 ***
FQ-4	.88 ***	.92 ***	.77 ***	.86 ***
HU-1	.81 ***	.84 ***	.57 ***	.62 ***
HU-2	.79 ***	.84 ***	.53 ***	.62 ***
HU-3	.70 ***	.77 ***	.34 ***	.51 ***

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: Zusammenfassung der Analyseergebnisse vier empirischer Vergleichsstudien (siehe Abschnitt 3.2).

Ausgehend von den theoretischen und empirischen Referenzwerten sind zwei statistische Hypothesen für sämtliche COPSOQ-II-Items (mit Ausnahme der AK-Skala, da diese aus zwei Items besteht) aggregiert formuliert:

- **TF.3-H.1** Der nicht-korrigierte Trennschärfekoeffizient sämtlicher COPSOQ-II-Items beträgt mindestens .50.
- **TF.3-H.2** Der korrigierte Trennschärfekoeffizient sämtlicher COPSOQ-II-Items beträgt mindestens .30.

4.4 Homogenität

Für die Bildung einer operationalen Hypothese im Rahmen der TF.4 wird ein in Abschnitt 3.1.4 dargestellter statistischer Kennwert herangezogen: AM der I-I-K. Über die theoretischen Grenzwerte dieses Kennwertes hinaus werden einzelne Analyseergebnisse mehrerer empirischer Vergleichsstudien berücksichtigt (siehe Abschnitt 3.2). Diese Analyseergebnisse sind in der Tabelle 5 komprimiert wiedergegeben.

Tabelle 5: Homogenität (empirische Referenzwerte)

	Inter-Item-Korrelation (I-I-K)		
	Minimum	AM	Maximum
QA-Items	.43 *** bis .55 ***	.54 bis .64	.64 *** bis .74 ***
EA-Items	.22 *** bis .28 ***	.30 bis .33	.42 *** bis .43 ***
EM-Items	.33 *** bis .47 ***	.39 bis .54	.45 *** bis .69 ***
FQ-Items	.61 *** bis .73 ***	.67 bis .79	.71 *** bis .84 ***
HU-Items	.29 *** bis .45 ***	.43 bis .51	.59 *** bis .66 ***

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: Zusammenfassung der Analyseergebnisse vier empirischer Vergleichsstudien (siehe Abschnitt 3.2).

Auf der Grundlage zum Einen der theoretischen und zum Anderen der empirischen Referenzwerte ist nachfolgend eine statistische Hypothese für sämtliche COPSOQ-II-Skalen (inkl. der AK-Skala – ohne empirische Vergleichsdaten) aggregiert formuliert:

- **TF.4-H.1** Das AM der I-I-K sämtlicher Items einer Skala, ist bei allen COPSOQ-II-Skalen innerhalb eines Spektrums von .30 bis .80 lokalisiert.

4.5 Dimensionalität

Vor dem Hintergrund der Verwendung einer Hauptkomponentenanalyse werden zur Formulierung operationaler Hypothesen hinsichtlich der TF.5 die zwei folgenden statistischen Kennwerte herangezogen: (1) Anzahl der resultierenden Komponenten sowie (2) absolute Höhe der (Haupt- und Neben-)Ladungen dieser Komponenten. Da in keiner der oben dargestellten, empirischen Vergleichsstudien (siehe Abschnitt 3.2) sämtliche 20 Items der für die vorliegende Arbeit relevanten sechs COPSOQ-II-Skalen erhoben wurde, liegen keine adäquaten empirischen Referenzwerte hinsichtlich der Dimensionalität der fokussierten COPSOQ-II-Items vor. Auf dieser Grundlage wird die Ladung auf einer Komponente ganz allgemein dann als Hauptladung definiert, wenn sie mindestens einen absoluten Wert von .50 erreicht (vgl. Backhaus et al., 2008). Eine relevante und damit berichtenswerte Nebenladung ist durch einen Wert definiert, welcher innerhalb eines Wertespektrums von .30 bis unter .50 liegt (vgl. Grice, 2001; Gorsuch, 1983). Die theoretische Skalenstruktur der 20 fokussierten COPSOQ-II-Items ist in der Abbildung 3 dargestellt und wird ausführlich in dem Abschnitt 5.2 erläutert.

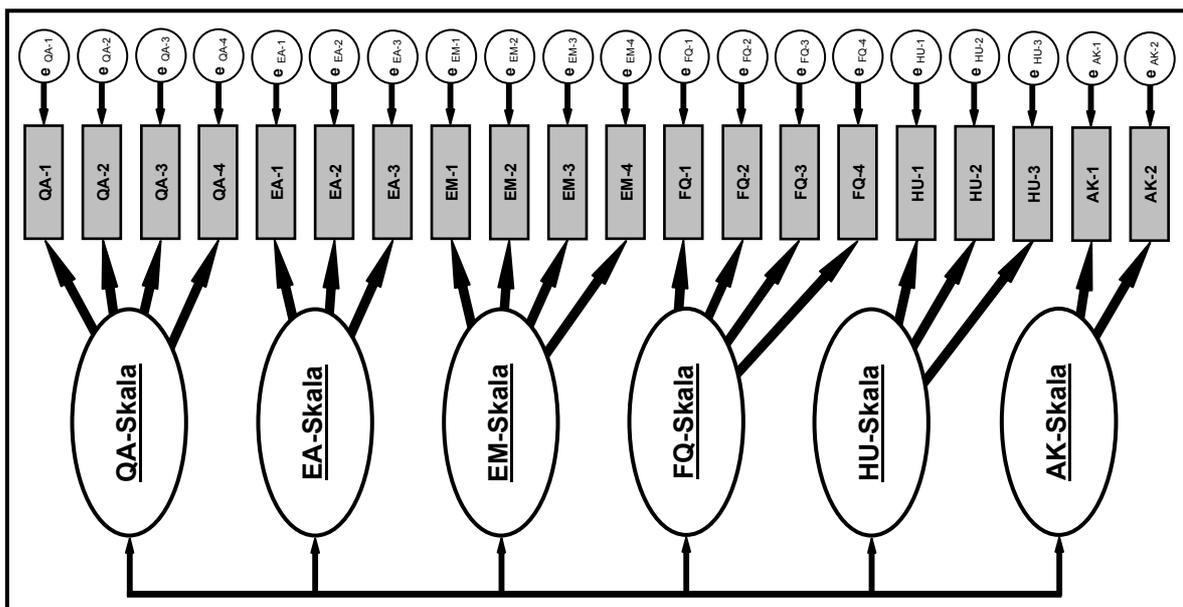


Abbildung 3: Dimensionalität (theoretische Skalenstruktur).

Ausgehend von dieser postulierten Skalenstruktur lauten die statistischen Hypothesen für die 20 COPSOQ-II-Items aggregiert folgendermaßen:

- **TF.5-H.1** Die Anwendung einer Hauptkomponentenanalyse aller 20 COPSOQ-II-Items resultiert in sechs Komponenten, die den sechs theoretisch erwarteten COPSOQ-II-Skalen (QA, EA, EM, FQ, HU, AK) entsprechen.
- **TF.5-H.2** Jedes der 20 COPSOQ-II-Items weist ausschließlich auf einer Komponente eine Hauptladung ($\geq .50$) und darüber hinaus auf keiner weiteren Komponente eine Nebenladung ($\geq .30$ und $< .50$) auf.

4.6 Nomologisches Netzwerk

Bezüglich der TF.6 wird für die Bildung operationaler Hypothesen ein in Abschnitt 3.1.6 benannter statistischer Kennwert herangezogen: Rangkorrelationskoeffizient (nach Spearman). Das empirisch fundierte nomologische Netzwerk, welches als Referenz herangezogen wird, basiert auf den Analyseergebnissen der oben vorgestellten empirischen Vergleichsstudien (siehe Abschnitt 3.2). Zur Bestimmung des Netzwerkes werden die vier nachfolgenden Entscheidungsregeln verwendet:

- die jeweilige bivariate Korrelation wird in mindestens zwei der vier empirischen Vergleichsstudien untersucht,
- sämtliche Koeffizienten der jeweils relevanten bivariaten Korrelation weisen konsistent äquivalente Vorzeichen auf, d. h. positiv, negativ oder Null-Korrelation,
- alle Koeffizienten erreichen mindestens einen absoluten Wert von .10 und
- sämtliche Koeffizienten sind auf einem Niveau von $\alpha = .001$ hoch signifikant.

In der Tabelle 6 ist das nach diesen vier Entscheidungsregeln erstellte finale nomologische Netzwerk, das als empirisches Referenznetzwerk definiert wird, dargestellt.

Tabelle 6: Nomologisches Netzwerk (empirisches Referenznetzwerk)

	QA	EA	EM	FQ	HU	AK	AF	GB	GG
QA-Skala		---	---	neg.	---	pos.	neg.	pos.	neg.
EA-Skala			pos.	---	pos.	---	---	neg.	pos.
EM-Skala				pos.	pos.	---	pos.	neg.	pos.
FQ-Skala					pos.	neg.	pos.	neg.	pos.
HU-Skala						neg.	pos.	neg.	pos.
AK-Skala							neg.	pos.	neg.
AF-Skala								---	---
GB									---
GG									

Anmerkungen: neg. = negatives Vorzeichen der bivariaten Korrelation. pos. = positives Vorzeichen der bivariaten Korrelation. Zusammenfassung der Analyseergebnisse vier empirischer Vergleichsstudien (siehe Abschnitt 3.2).

Das empirische Referenznetzwerk beinhaltet 26 für die sechs COPSOQ-II-Skalen relevante Korrelationen, wobei jede Skala jeweils in vier bis sieben bivariate Korrelationen eingebunden ist. Die statistische Hypothese lautet für alle COPSOQ-II-Skalen aggregiert (inkl. der AK-Skala – ohne exakte empirische Vergleichsdaten) wie folgt:

- **TF.6-H.1** Die Vorzeichen (positiv vs. negativ) der Koeffizienten sämtlicher 26 bivariaten Korrelationen hinsichtlich der sechs COPSOQ-II-Skalen entsprechen denjenigen des empirischen Referenznetzwerkes.

5. Methode

In diesem Kapitel sind diejenigen drei Stichproben, auf deren Grundlage die oben formulierte Fragestellung (F) bzw. deren Teilfragestellungen (TF) über die korrespondierenden Hypothesen (H) beantwortet wird (siehe Abschnitt 4), detailliert beschrieben (siehe Abschnitt 5.1).

Ferner sind die sechs in der vorliegenden Arbeit fokussierten COPSOQ-II-Skalen, wie sie im lidA-CAPI Anwendung finden, im Detail wiedergegeben (siehe Abschnitt 5.2).

Darüber hinaus sind diejenigen Instrumente, die ebenso in den empirischen Vergleichsstudien wie auch im lidA-CAPI eingesetzt sind und eine wichtige Rolle im Kontext des empirischen nomologischen Referenznetzwerkes spielen (siehe Abschnitt 4.6), ausführlich vorgestellt (siehe Abschnitt 5.3).

Und abschließend sind im letzten Abschnitt dieses Kapitels der genaue Ablaufplan sowie die konkreten Inhalte der statistischen Analysen für die Überprüfung der oben erarbeiteten Hypothesen dargestellt und erläutert (siehe Abschnitt 5.4).

5.1 Stichproben

In den zwei nachfolgenden Abschnitten werden die drei für die vorliegende Arbeit zentralen Stichproben ausführlich beschrieben. Die Stichproben entstammen im Einzelnen der lidA-Machbarkeitsstudie (siehe Abschnitt 5.1.1) sowie der 1. Welle/A+B der lidA-Hauptstudie (siehe Abschnitt 5.1.2). Im Wesentlichen gliedern sich jene beiden Abschnitte in die Darstellung der jeweiligen Brutto- und Nettostichproben (inkl. stichprobenneutraler und systematischer Ausfälle), den Umgang mit fehlenden Werten sowie eine ausführliche Soziodemografie der drei Stichproben. Gemäß dem klassischen Vorgehen wurden in allen drei benannten Stichproben einheitlich im Vorfeld der statistischen Datenanalyse fehlende Werte listenweise ausgeschlossen (vgl. Schendera, 2007), um eine homogene Stichprobenbasis für die Interpretationen der zahlreichen statistischen Kennwerte zu haben und die Risiken der verschiedenen Verzerrungen als Folge der Verwendung von Imputationsverfahren (z. B. Janssen & Laatz, 2007; Höfler, 2004) zu vermeiden.

Die Population, welcher die herangezogenen lidA-Stichproben entstammen, ist konstituiert durch diejenigen Personen der Geburtsjahrgänge 1958 und 1964 (lidA-Machbarkeitsstudie – hier liegen in drei Fällen Abweichungen vor) bzw. 1959 und 1965 (lidA-Hauptstudie: 1. Welle/A+B), die am 31. Dezember 2009 bei der BA als sozialversicherungspflichtig Beschäftigte in Deutschland gemeldet sind (vgl. Kersting et al., 2012; Kersting et al., 2011; Schröder, Gilberg, Tisch, Peter, du Prel & Trappmann, 2011; Gilberg, Schröder & Trappmann, 2010). Aus diesem Datensatz werden die lidA-Stichproben vom IAB gezogen und, auf der Grundlage der datenschutzrechtlichen Genehmigung durch das Bundesministerium für Arbeit und Soziales (BMAS) gemäß § 75 SGB X (zehntes Buch Sozialgesetzbuch), der infas GmbH zur Verfügung gestellt (Kersting et al., 2012; Kersting et al., 2011).

Die für die vorliegende Arbeit relevante Datenerhebung wird durch mehrere professionelle Interviewer der infas GmbH realisiert und erfolgt sowohl bei der lidA-Machbarkeitsstudie als auch bei der 1. Welle/A+B der lidA-Hauptstudie mittels eines ca. 60-minütigen CAPI bei den TN zu Hause (Kersting et al., 2012).

5.1.1 lidA-Machbarkeitsstudie

Im Vorfeld der lidA-Hauptstudie wurde ein Pretest, im weiteren Verlauf dieser Arbeit „lidA-Machbarkeitsstudie“ genannt, durchgeführt. Diese Studie erfolgte im Wesentlichen zum Zwecke der Überprüfung (1) der Stichprobenziehung und -übermittlung, (2) der Adressqualität, (3) der Dauer des lidA-CAPI, (4) der Verständlichkeit und Akzeptanz sowie (5) der Adäquanz der Interviewerschulung (vgl. Kersting et al., 2011). Aufgrund der geringen Stichprobengröße der lidA-Machbarkeitsstudie und der damit verbundenen geringen Aussagekraft entsprechender Analyseergebnisse, stellt die Überprüfung der (6) psychometrischen Qualität der eingesetzten Skalen und Einzelitems einen nachrangigen Zweck dar (vgl. Kersting et al., 2011). Als Konsequenz dieser psychometrischen Item- und Skalenanalysen wurden geringfügige Modifikationen, vornehmlich die Löschung einzelner Items, vorgenommen, welche für die in der vorliegenden Arbeit relevanten Skalen und Einzelitems bedeutungslos sind.

Die Bruttostichprobe, d. h. die Anzahl vorliegender Kontaktdaten, betrug $n = 970$ (= 100.0 Prozent – Kersting et al., 2011). Alle potentiellen TN wurden zunächst postalisch über die lidA-Machbarkeitsstudie informiert. Hiervon wurden 56 TN (= 5.8 Prozent) als stichprobenneutrale Ausfälle (z. B. verzogen, verstorben) und 714 TN (= 73.6 Prozent) als systematische Ausfälle (z. B. Verweigerung der Teilnahme, sprachbedingter Ausfall, nicht erreichbar) verzeichnet (Kersting et al., 2011). Schließlich realisierten insgesamt 16 Interviewer der infas GmbH mit 200 TN (= 20.6 Prozent) das lidA-CAPI (Kersting et al., 2011). Die Realisation dieser 200 lidA-CAPI erfolgte in dem Zeitraum von Juli 2010 bis Oktober 2010 (Kersting et al., 2011).

Vor der Analyse dieser 200 lidA-CAPI (= 100 Prozent) wurden fehlende Werte in den sechs COPSOQ-II-Skalen (siehe Abschnitt 5.2) sowie in den drei weiteren Instrumenten (siehe Abschnitt 5.3) vom Autor listenweise ausgeschlossen (vgl. Dupret, Bocéréan, Teherani, Feltrin & Pejtersen, 2012). Hinsichtlich der FQ-Skala (siehe Abschnitt 5.2.4) und der HU-Skala (siehe Abschnitt 5.2.5) wurden fehlende Werte nur dann ausgeschlossen, wenn nicht alle Items unbeantwortet blieben, weil die TN im Einzelfall keinen unmittelbaren Vorgesetzten (FQ-Skala) resp. keine Kollegen (HU-Skala) hatten und demgemäß die entsprechenden Items nicht beantworten konnten. Infolge dieser Vorgehensweise zum Ausschließen fehlender Werte wurden insgesamt 29 Fälle (= 14.5 Prozent) gelöscht. Hieraus resultierte die für die vorliegende Untersuchung finale Nettostichprobe der lidA-Machbarkeitsstudie mit einem Umfang von $n = 171$. Nachstehend ist die Soziodemografie dieser Stichprobe ausführlich in zehn Variablen dargestellt (siehe Tabelle 7).

Tabelle 7: Soziodemografie (IidA-Machbarkeitsstudie)

		n	%
Gesamtzahl		171	100.0
Geburtsjahr	1958	86	50.3
	1964	82	48.0
	sonstige (1956, 1967, 1978)	3	1.7
Geschlecht	männlich	72	42.1
	weiblich	99	57.9
Familienstand	verheiratet, eingetragene Lebenspartnerschaft	113	66.1
	ledig	31	18.1
	geschieden	25	14.6
	verwitwet	2	1.2
Haushaltsgröße	1 Person	34	19.9
	2 Personen	64	37.4
	3 Personen	33	19.3
	4 Personen und mehr	40	23.4
Kinder im Haushalt unter 14 Jahren	keine Kinder	105	61.4
	1 Kind	16	9.4
	2 Kinder und mehr	16	9.4
	keine Angabe	34	19.9
Schulabschluss	ohne Schulabschluss	0	0.0
	Volks-, Hauptschulabschluss	29	17.0
	Realschulabschluss, polytechnische Oberschule	73	42.7
	Fach-/Hochschulreife, erweiterte Oberschule	67	39.2
	anderer Schulabschluss	2	1.2
Ausbildungsabschluss	ohne Ausbildungsabschluss	3	1.8
	beruflich-betrieblicher Ausbildungsabschluss	65	38.0
	beruflich-schulischer Ausbildungsabschluss	14	8.2
	Fach-, Meister-, Technikerschule etc.	32	18.7
	Fach-/Hochschulabschluss	56	32.7
	anderer Ausbildungsabschluss	1	0.6
Stellung im Beruf	un-, angelernte Arbeiter	14	8.2
	Facharbeiter	12	7.0
	Vorarbeiter, Meister etc.	2	1.2
	Angestellte ohne Personalverantwortung	63	36.8
	Angestellte mit Personalverantwortung	75	43.9
	keine Angabe	5	2.9
Wochenarbeitszeit inkl. Überstunden	unter 20 Stunden	12	7.0
	20 bis unter 30 Stunden	11	6.4
	30 bis unter 40 Stunden	44	25.7
	40 Stunden und mehr	101	59.1
	schwankt beträchtlich	3	1.8
Nettoeinkommen persönlich	unter 1.000 EUR	34	19.9
	1.000 bis unter 2.000 EUR	73	42.7
	2.000 bis unter 3.000 EUR	38	22.2
	3.000 bis unter 4.000 EUR	10	5.8
	4.000 EUR und mehr	9	5.3
	keine Angabe	7	4.1

Anmerkungen: Alle TN haben zum Zeitpunkt der Datenerhebung eine sozialversicherungspflichtige Beschäftigung.

5.1.2 lidA-Hauptstudie: 1. Welle/A+B

Die Bruttostichprobe, d. h. die Anzahl vorliegender Kontaktdaten potentieller TN, der 1. Welle der lidA-Hauptstudie betrug $n = 26\,697$ (Kersting et al., 2012). Alle potentiellen TN wurden im Vorfeld postalisch über die lidA-Studie durch die infas GmbH informiert. Hiervon wurden $1\,976$ TN (7.4 Prozent) als stichprobenneutrale Ausfälle (z. B. verzogen, verstorben) und $17\,685$ TN (66.4 Prozent) als systematische Ausfälle (z. B. Verweigerung der Teilnahme, sprachbedingter Ausfall, nicht erreichbar) verzeichnet (vgl. Kersting et al., 2012). Insgesamt wurden 399 Kontaktdaten (1.5 Prozent) nicht abschließend bearbeitet (Kersting et al., 2012). Schließlich führten insgesamt 347 Interviewer mit $6\,637$ TN (24.9 Prozent) das lidA-CAPI der 1. Welle der lidA-Hauptstudie durch (Kersting et al., 2012). Hiervon wurden 52 Datensätze (0.2 Prozent) durch die infas GmbH als nicht auswertbar eingestuft (Kersting et al., 2012). Die Realisation der lidA-CAPI erfolgte von März 2011 bis November 2011 (Kersting et al., 2012). Diese Stichprobe ist repräsentativ für die beiden Geburtsjahrgänge 1959 und 1965 der sozialversicherungspflichtig Beschäftigten gemäß der Beschäftigten-Historie des IAB zum Zeitpunkt des 31. Dezember 2009 (vgl. Kersting et al., 2012).

Im Vorfeld der statistischen Analyse dieser Nettostichprobe ($n = 6\,585 = 100.0$ Prozent) wurden fehlende Werte in den sechs interessierenden COPSOQ-II-Skalen (siehe Abschnitt 5.2) sowie in den drei weiteren Instrumenten (siehe Abschnitt 5.3) vom Autor listenweise ausgeschlossen (vgl. Dupret et al., 2012). Bezüglich der FQ-Skala (siehe Abschnitt 5.2.4) und der HU-Skala (siehe Abschnitt 5.2.5) wurden fehlende Werte nur dann ausgeschlossen, wenn nicht alle Items der jeweiligen Skala unbeantwortet blieben, da die TN im Einzelfall keinen unmittelbaren Vorgesetzten (FQ-Skala) bzw. keinen Kollegen (HU-Skala) hatten und hierdurch die entsprechenden Items nicht beantworten konnten (vgl. Abschnitt 3.2). Aufgrund dieses Vorgehens zum Ausschließen fehlender Werte wurden insgesamt 919 Fälle (= 14.0 Prozent) gelöscht, wodurch die Größe der Nettostichprobe der 1. Welle der lidA-Hauptstudie für die vorliegende Arbeit $n = 5\,666$ beträgt.

Diese Stichprobe wird mit Hilfe der Statistiksoftware „IBM SPSS Statistics 20“ per Zufall in zwei etwa gleichgroße, distinkte Stichproben (A+B) geteilt (IBM Corporation, 2011). Die eine Stichprobe (lidA-Hauptstudie: 1. Welle/A) umfasst $n = 2\,830$ Fälle, die andere Stichprobe (lidA-Hauptstudie: 1. Welle/B) umfasst $n = 2\,836$ Fälle. Auf den nachfolgenden zwei Seiten wird die Soziodemografie dieser zwei finalen Stichproben der 1. Welle der lidA-Hauptstudie in zehn Variablen ausführlich beschrieben (siehe Tabelle 8 und Tabelle 9). Im Einzelvergleich dieser zwei Stichproben (lidA-Hauptstudie: 1. Welle/A+B) mit der ursprünglichen Nettostichprobe ($n = 6\,585$ – siehe Tabelle 32) hinsichtlich der prozentualen Ausprägungen der erwähnten zehn soziodemografischen Variablen beträgt die überwiegende Mehrheit der Differenzen nicht mehr als einen Prozentpunkt.

Tabelle 8: Soziodemografie (IidA-Hauptstudie: 1. Welle/A)

		n	%
Gesamtzahl		2 830	100.0
Geburtsjahr	1959	1 255	44.3
	1965	1 575	55.7
Geschlecht	männlich	1 345	47.5
	weiblich	1 485	52.5
Familienstand	verheiratet, eingetragene Lebenspartnerschaft	2 068	73.1
	ledig	347	12.3
	geschieden	364	12.8
	verwitwet	48	1.7
	keine Angabe	3	0.1
Haushaltsgröße	1 Person	326	11.5
	2 Personen	834	29.5
	3 Personen	720	25.4
	4 Personen und mehr	947	33.5
	keine Angabe	3	0.1
Kinder im Haushalt unter 14 Jahren	keine Kinder	1 776	62.8
	1 Kind	469	16.6
	2 Kinder und mehr	256	9.0
	keine Angabe	329	11.6
Schulabschluss	ohne Schulabschluss	39	1.4
	Volks-, Hauptschulabschluss	720	25.4
	Realschulabschluss, polytechnische Oberschule	1 185	41.9
	Fach-/Hochschulreife, erweiterte Oberschule	870	30.7
	anderer Schulabschluss	16	0.6
Ausbildungsabschluss	ohne Ausbildungsabschluss	147	5.2
	beruflich-betrieblicher Ausbildungsabschluss	1 376	48.6
	beruflich-schulischer Ausbildungsabschluss	281	9.9
	Fach-, Meister-, Technikerschule etc.	431	15.2
	Fach-/Hochschulabschluss	569	20.1
	anderer Ausbildungsabschluss	24	0.8
	keine Angabe	2	0.1
Stellung im Beruf	un-, angelernte Arbeiter	304	10.7
	Facharbeiter	286	10.1
	Vorarbeiter, Meister etc.	71	2.5
	Angestellte ohne Personalverantwortung	1 019	36.0
	Angestellte mit Personalverantwortung	1 072	37.9
	keine Angabe	78	2.7
Wochenarbeitszeit inkl. Überstunden	unter 20 Stunden	258	9.1
	20 bis unter 30 Stunden	374	13.2
	30 bis unter 40 Stunden	730	25.8
	40 Stunden und mehr	1 418	50.1
	schwankt beträchtlich	49	1.7
	keine Angabe	1	0.0
Nettoeinkommen persönlich	unter 1.000 EUR	634	22.4
	1.000 bis unter 2.000 EUR	1 145	40.5
	2.000 bis unter 3.000 EUR	590	20.8
	3.000 bis unter 4.000 EUR	167	5.9
	4.000 EUR und mehr	128	4.5
	keine Angabe	166	5.9

Anmerkungen: Alle TN haben zum Zeitpunkt der Datenerhebung eine sozialversicherungspflichtige Beschäftigung.

Tabelle 9: Soziodemografie (IidA-Hauptstudie: 1. Welle/B)

		n	%
Gesamtzahl		2 836	100.0
Geburtsjahr	1959	1 215	42.8
	1965	1 621	57.2
Geschlecht	männlich	1 340	47.2
	weiblich	1 496	52.8
Familienstand	verheiratet, eingetragene Lebenspartnerschaft	2 110	74.4
	ledig	329	11.6
	geschieden	350	12.3
	verwitwet	40	1.4
	keine Angabe	7	0.3
Haushaltsgröße	1 Person	330	11.6
	2 Personen	805	28.4
	3 Personen	747	26.3
	4 Personen und mehr	953	33.6
	keine Angabe	1	0.0
Kinder im Haushalt unter 14 Jahren	keine Kinder	1 744	61.5
	1 Kind	500	17.6
	2 Kinder und mehr	261	9.2
	keine Angabe	331	11.6
Schulabschluss	ohne Schulabschluss	40	1.4
	Volks-, Hauptschulabschluss	704	24.8
	Realschulabschluss, polytechnische Oberschule	1 167	41.1
	Fach-/Hochschulreife, erweiterte Oberschule	907	32.0
	anderer Schulabschluss	18	0.6
Ausbildungsabschluss	ohne Ausbildungsabschluss	140	4.9
	beruflich-betrieblicher Ausbildungsabschluss	1 395	49.2
	beruflich-schulischer Ausbildungsabschluss	257	9.1
	Fach-, Meister-, Technikerschule etc.	443	15.6
	Fach-/Hochschulabschluss	581	20.5
	anderer Ausbildungsabschluss	19	0.7
	keine Angabe	1	0.0
Stellung im Beruf	un-, angelernte Arbeiter	317	11.2
	Facharbeiter	281	9.9
	Vorarbeiter, Meister etc.	78	2.8
	Angestellte ohne Personalverantwortung	1 016	35.8
	Angestellte mit Personalverantwortung	1 050	37.0
	keine Angabe	94	3.4
Wochenarbeitszeit inkl. Überstunden	unter 20 Stunden	277	9.8
	20 bis unter 30 Stunden	367	12.9
	30 bis unter 40 Stunden	712	25.1
	40 Stunden und mehr	1 426	50.3
	schwankt beträchtlich	54	1.9
	keine Angabe	0	0.0
Nettoeinkommen persönlich	unter 1.000 EUR	613	21.6
	1.000 bis unter 2.000 EUR	1 196	42.2
	2.000 bis unter 3.000 EUR	584	20.6
	3.000 bis unter 4.000 EUR	164	5.8
	4.000 EUR und mehr	139	4.9
	keine Angabe	140	4.9

Anmerkungen: Alle TN haben zum Zeitpunkt der Datenerhebung eine sozialversicherungspflichtige Beschäftigung.

5.2 COPSOQ

In den nachfolgenden sechs Abschnitten sind die für die vorliegende Arbeit relevanten sechs COPSOQ-II-Skalen detailliert dargestellt. Diese sind im Einzelnen:

- Quantitative Anforderungen (QA – siehe Abschnitt 5.2.1)
- Einfluss bei der Arbeit (EA – siehe Abschnitt 5.2.2)
- Entwicklungsmöglichkeiten (EM – siehe Abschnitt 5.2.3)
- Führungsqualität (FQ – siehe Abschnitt 5.2.4)
- Hilfe und Unterstützung von Kollegen (HU – siehe Abschnitt 5.2.5)
- Arbeit-Familie/Privatleben-Konflikt (AK – siehe Abschnitt 5.2.6)

Zu Beginn jedes Abschnittes sind zunächst die Literaturquelle für die Struktur der jeweiligen COPSOQ-II-Skala sowie diejenige der Übersetzung der Skalenbezeichnung und der korrespondierenden Items benannt. Nachfolgend sind die Modifikationen der Skalenstruktur und/oder der entsprechenden Items nachgezeichnet, die ausschließlich durch das li-dA-Projektconsortium erarbeitet wurden. Daran anknüpfend sind die operationalen Inhalte aller Items sowie ihre für die vorliegende Arbeit maßgebliche Kennzeichnung angeführt. Die Darstellung der inhaltlichen bzw. zentralen und auch der zusätzlichen Antwortoptionen (AO) sowie die Erläuterung der jeweiligen Transformationsregel der Rohwerte der zentralen AO bilden den Abschluss jedes Abschnittes.

Die Antwortskalen, bestehend aus den zentralen AO, sind grundsätzlich als mehrfach abgestufte, unipolare Ratingskalen (Likert-Skalen) konzipiert. Eine solche Ratingskala gilt in der sozialwissenschaftlichen Forschung sowie den wissenschaftlichen Personenbefragungen allgemein als intervallskalierte Antwortskala (Bortz & Döring, 2009). Die Polung der Skalen erfolgt stets im Sinne der Skalenbezeichnung, welche zugleich das latente Konstrukt bezeichnet. Zur Ermittlung des Skalenwertes wird bei allen sechs COPSOQ-II-Skalen gemäß den Empfehlungen von Pejtersen, Kristensen, Borg & Bjorner (2010) das AM aus den zuvor positiv-linear transformierten Einzelwerten sämtlicher Items der jeweiligen Skala berechnet. Hieraus resultiert eine für die Itemebene und Skalenebene äquivalente theoretische Spannweite der Werte.

5.2.1 Quantitative Anforderungen

Für die Erfassung der subjektiv bewerteten quantitativen Anforderungen bei der Arbeit wurde die Struktur der Skala „quantitative demands“ des „LONG COPSOQ II questionnaire“ (NRCWE – National Research Centre for the Working Environment, 2007a) herangezogen. Diese Skala besteht aus vier englischsprachigen Items, die in dem deutschen „COPSOQ I Lange Version“ (Nübling et al., 2005) übersetzt sind. Zur Förderung eines schnellen Verständnisses wurden drei (QA-1, QA-3, QA-4) der vier Items für das lidA-CAPI formal überarbeitet. Hierbei wurden die Wörter „Wie oft“ an den Satzanfang der Items hinzugefügt, um die sprachliche Passung zu den fünf zentralen AO zu optimieren. Dieser Satzanfang war bei dem unveränderten Item (QA-2) im englischsprachigen Original (NRCWE, 2007a) und in der deutschen Übersetzung (Nübling et al., 2005) bereits vorhanden. Die vier finalen Items, welche im lidA-CAPI die Skala „Quantitative Anforderungen (QA)“ bilden, sind nachstehend angeführt:

- **QA-1** Wie oft ist Ihre Arbeit ungleich verteilt, so dass sie sich aufhäuft?
- **QA-2** Wie oft kommt es vor, dass Sie nicht genügend Zeit haben, alle Ihre Aufgaben zu erledigen?
- **QA-3** Wie oft kommen Sie mit Ihrer Arbeit in Rückstand?
- **QA-4** Wie oft haben Sie genug Zeit für Ihre Arbeitsaufgaben? (AO invertiert)

Zur Bearbeitung dieser vier QA-Items steht den TN eine mehrfach abgestufte, unipolare Likert-Skala mit fünf verbalen Marken (fünf zentrale AO) zur Verfügung:

- **1** immer
- **2** oft
- **3** manchmal
- **4** selten
- **5** nie oder fast nie

Darüber hinaus können zwei zusätzliche AO („verweigert“ und „weiß nicht“) verwendet werden, die nicht sofort vom Interviewer vorgelesen, sondern erst angeboten werden, wenn die TN entsprechenden Bedarf anzeigen. Gemäß den Anwendungsempfehlungen von Pejtersen et al. (2010) werden die fünf zentralen AO auf einen Wertebereich von 0 (geringe Ausprägung) bis 100 (hohe Ausprägung) positiv-linear transformiert. Daraus resultieren die folgenden fünf möglichen Werte: (1 =) 100, (2 =) 75, (3 =) 50, (4 =) 25 und (5 =) 0. Hierbei ist zu beachten, dass die zentralen AO des Items QA-4 vor dessen Transformation umgepolt werden müssen (Pejtersen et al., 2010), woraus die inverse Rangreihenfolge der Werte resultiert: (1 =) 0, (2 =) 25, (3 =) 50, (4 =) 75 und (5 =) 100.

5.2.2 Einfluss bei der Arbeit

Zur Erhebung des subjektiv bewerteten Einflusses bei der Arbeit wurde auf die Struktur der Skala „influence at work“ des „LONG COPSOQ II questionnaire“ (NRCWE, 2007a) zurückgegriffen. Jene Skala setzt sich aus vier englischsprachigen Items zusammen, die in dem deutschen „COPSOQ I Lange Version“ (Nübling et al., 2005) übersetzt sind. Im Zuge einer ökonomischen Gestaltung der Datenerhebung wurde eines der insgesamt vier Items dieser Skala hierbei nicht verwendet. Diejenigen drei Items, welche im IidA-CAPI die Skala „Einfluss bei der Arbeit (EA)“ konstituieren, sind im Folgenden dargestellt:

- **EA-1** Haben Sie Einfluss darauf, mit wem Sie arbeiten?
- **EA-2** Haben Sie Einfluss auf die Menge der Arbeit, die Ihnen übertragen wird?
- **EA-3** Haben Sie Einfluss darauf, was Sie bei Ihrer Arbeit tun?

Für die Beantwortung dieser drei Fragen durch die TN ist eine mehrfach abgestufte, unipolare Likert-Skala mit fünf sprachlichen Marken (fünf zentrale AO) vorgesehen:

- **1** immer
- **2** oft
- **3** manchmal
- **4** selten
- **5** nie oder fast nie

Des Weiteren stehen zwei zusätzliche AO („verweigert“ und „weiß nicht“) zur Verfügung, die im Gegensatz zu den fünf zentralen AO nicht ungefragt vom Interviewer angeboten, sondern erst vorgelesen werden, wenn die TN entsprechenden Bedarf anzeigen. Den Anwendungsempfehlungen von Pejtersen et al. (2010) entsprechend werden die fünf zentralen AO auf einen Wertebereich von 0 (geringe Ausprägung) bis 100 (hohe Ausprägung) positiv-linear transformiert. Hieraus ergeben sich die fünf nachfolgenden Werte: (1 =) 100, (2 =) 75, (3 =) 50, (4 =) 25 und (5 =) 0.

5.2.3 Entwicklungsmöglichkeiten

Für die Erfassung der subjektiv bewerteten Entwicklungsmöglichkeiten bei der Arbeit wurde die Struktur der Skala „possibilities for development“ des „LONG COPSOQ II questionnaire“ (NRCWE, 2007a) herangezogen. Diese Skala besteht aus vier englischsprachigen Items, die in dem deutschen „COPSOQ I Lange Version“ (Nübling et al., 2005) übersetzt sind. Zur Förderung eines schnellen Verständnisses wurde eines (EM-2) der vier Items für das lidA-CAPI formal überarbeitet. Hierbei wurde das Wort „erlernen“ durch das Wort „lernen“ ausgetauscht, um ein möglichst einfaches Sprachniveau für die TN sicher zu stellen. Die vier finalen Items, welche im lidA-CAPI die Skala „Entwicklungsmöglichkeiten (EM)“ bilden, sind nachstehend angeführt:

- **EM-1** Verlangt es Ihre Arbeit, dass Sie die Initiative ergreifen?
- **EM-2** Haben Sie die Möglichkeit, durch Ihre Arbeit neue Dinge zu lernen?
- **EM-3** Können Sie Ihre Fertigkeiten oder Ihr Fachwissen bei Ihrer Arbeit anwenden?
- **EM-4** Ermöglicht es Ihre Arbeit, Ihre Fertigkeiten weiterzuentwickeln?

Zur Bearbeitung der vier EM-Items steht den TN eine mehrfach abgestufte, unipolare Likert-Skala mit fünf verbalen Marken (fünf zentrale AO) zur Verfügung:

- **1** in sehr hohem Maße
- **2** in hohem Maße
- **3** zum Teil
- **4** in geringem Maße
- **5** in sehr geringem Maße

Über die fünf zentralen AO hinaus können zwei zusätzliche AO („verweigert“ und „weiß nicht“) Verwendung finden, die vom Interviewer erst angeboten werden, wenn die TN entsprechenden Bedarf anzeigen. Gemäß den Anwendungsempfehlungen von Pejtersen et al. (2010) werden die fünf zentralen AO auf einen Wertebereich von 0 (geringe Ausprägung) bis 100 (hohe Ausprägung) positiv-linear transformiert. Daraus resultieren die folgenden Werte: (1 =) 100, (2 =) 75, (3 =) 50, (4 =) 25 und (5 =) 0.

5.2.4 Führungsqualität

Zur Erhebung der subjektiv bewerteten Führungsqualität des unmittelbaren Vorgesetzten bei der Arbeit wurde auf die Struktur der Skala „quality of leadership“ des „LONG COPSOQ II questionnaire“ (NRCWE, 2007a) zurückgegriffen. Jene Skala setzt sich aus vier englischsprachigen Items zusammen, die in dem deutschen „COPSOQ I Lange Version“ (Nübling et al., 2005) übersetzt sind. Zur Förderung eines schnellen Verständnisses wurden sämtliche Items für das lidA-CAPI formal überarbeitet. Hierbei wurde der Satzanfang „Bitte schätzen Sie ein“ bei allen Items entfernt. Weiter wurde eine geschlechtsspezifische Formulierung in Bezug auf die unmittelbar vorgesetzte Person („Ihr unmittelbarer Vorgesetzter“ bzw. „Ihre unmittelbare Vorgesetzte“) durch eine Geschlechtsabfrage hinsichtlich der unmittelbar vorgesetzten Person direkt vor der Skalenanwendung hinzugefügt. Außerdem wurde bei sämtlichen Items das einteilige Prädikat vom Satzende nach vorne hinter das Wort „Maße“ verlegt. Diejenigen vier finalen FQ-Items, welche im lidA-CAPI die Skala „Führungsqualität (FQ)“ bilden, sind unten aufgelistet:

- **FQ-1** In welchem Maße sorgt Ihr/e unmittelbare/r Vorgesetzte/r für gute Entwicklungsmöglichkeiten der einzelnen Mitarbeiter?
- **FQ-2** In welchem Maße misst Ihr/e unmittelbare/r Vorgesetzte/r der Arbeitszufriedenheit einen hohen Stellenwert bei?
- **FQ-3** In welchem Maße plant Ihr/e unmittelbare/r Vorgesetzte/r die Arbeit gut?
- **FQ-4** In welchem Maße löst Ihr/e unmittelbare/r Vorgesetzte/r Konflikte gut?

Für die Beantwortung dieser vier Fragen durch die TN ist eine mehrfach abgestufte, unipolare Likert-Skala mit fünf verbalen Marken (fünf zentrale AO) vorgesehen:

- **1** in sehr hohem Maße,
- **2** in hohem Maße
- **3** zum Teil
- **4** in geringem Maße
- **5** in sehr geringem Maße

Des Weiteren sind zwei zusätzliche AO („verweigert“ und „weiß nicht“) vorhanden, die vom Interviewer erst vorgelesen werden, wenn entsprechender Bedarf bei den TN entsteht. Wenn keine unmittelbaren Vorgesetzten vorhanden sind, bleiben sämtliche FQ-Items unbearbeitet. Gemäß den Anwendungsempfehlungen von Pejtersen et al. (2010) werden die fünf zentralen AO auf einen Wertebereich von 0 (geringe Ausprägung) bis 100 (hohe Ausprägung) positiv-linear transformiert. Hieraus ergeben sich die fünf nachstehenden Werte: (1 =) 100, (2 =) 75, (3 =) 50, (4 =) 25 und (5 =) 0.

5.2.5 Hilfe und Unterstützung von Kollegen

Zur Erfassung der subjektiv bewerteten Hilfe und Unterstützung von Arbeitskollegen wurde die Struktur der Skala „social support from colleagues“ des „LONG COPSOQ II questionnaire“ (NRCWE, 2007a) herangezogen. Diese Skala besteht aus drei englischsprachigen Items, die in dem deutschen „COPSOQ I Lange Version“ (Nübling et al., 2005) übersetzt sind. Die drei Items, welche im lidA-CAPI die Skala „Hilfe und Unterstützung von Kollegen (HU)“ bilden, sind nachstehend angeführt:

- **HU-1** Wie oft erhalten Sie Hilfe und Unterstützung von Ihren Kollegen?
- **HU-2** Wie oft sind Ihre Kollegen bereit, sich Ihre Arbeitsprobleme anzuhören?
- **HU-3** Wie oft sprechen Ihre Kollegen mit Ihnen über die Qualität Ihrer Arbeit?

Für die Bearbeitung dieser drei HU-Items steht den TN eine mehrfach abgestufte, unipolare Likert-Skala mit fünf verbalen Marken (fünf zentrale AO) zur Verfügung:

- **1** immer
- **2** oft
- **3** manchmal
- **4** selten
- **5** nie oder fast nie

Darüber hinaus können zwei zusätzliche AO („verweigert“ und „weiß nicht“) verwendet werden, die im Gegensatz zu den fünf zentralen AO nicht sofort vom Interviewer vorgelesen, sondern erst angeboten werden, wenn die TN entsprechenden Bedarf anzeigen. Sollten keine Arbeitskollegen vorhanden sein, bleiben sämtliche HU-Items unbearbeitet. Gemäß den Anwendungsempfehlungen von Pejtersen et al. (2010) werden die fünf zentralen AO auf einen Wertebereich von 0 (geringe Ausprägung) bis 100 (hohe Ausprägung) positiv-linear transformiert. Daraus resultieren die fünf nachstehend angeführten Werte: (1 =) 100, (2 =) 75, (3 =) 50, (4 =) 25 und (5 =) 0.

5.2.6 Arbeit-Familie/Privatleben-Konflikt

Für die Erhebung subjektiv bewerteter Konflikte zwischen der Arbeit und der Familie bzw. dem Privatleben wurde auf die Struktur der Skala „work family conflict“ des „SHORT COPSOQ II questionnaire“ (NRCWE, 2007b) zurückgegriffen. Diese Skala setzt sich aus zwei englischsprachigen Items zusammen, die vom lidA-Projektkonsortium ins Deutsche übersetzt wurden. Diejenigen zwei Items, die im lidA-CAPI die Skala „Arbeit-Familie/Privatleben-Konflikt (AK)“ bilden, sind im Folgenden dargestellt:

- **AK-1** Meine Arbeit nimmt soviel Energie in Anspruch, dass dies negative Auswirkungen auf mein Privatleben hat.
- **AK-2** Meine Arbeit nimmt soviel Zeit in Anspruch, dass dies negative Auswirkungen auf mein Privatleben hat.

Zur Bearbeitung dieser zwei AK-Items ist eine mehrfach abgestufte, unipolare Likert-Skala mit vier verbalen Marken (vier zentrale AO) vorgesehen:

- **1** Trifft gar nicht zu
- **2** Trifft eher nicht zu
- **3** Trifft eher zu
- **4** Trifft voll und ganz zu

Des Weiteren können zwei zusätzliche AO („verweigert“ und „weiß nicht“) verwendet werden, die vom Interviewer erst angeboten werden, wenn entsprechender Bedarf bei den TN erkennbar wird. Gemäß den Anwendungsempfehlungen von Pejtersen et al. (2010) werden die vier zentralen AO auf einen Wertebereich von 0 (geringe Ausprägung) bis 100 (hohe Ausprägung) positiv-linear transformiert, woraus sich die vier nachstehenden Werte ergeben: (1 =) 0, (2 =) 33, (3 =) 67 und (4 =) 100.

5.3 Weitere Instrumente

In den folgenden drei Abschnitten sind diejenigen drei Konstrukte sowie ihre Operationalisierungen erläutert, die in der vorliegenden Arbeit im Kontext der Erstellung des empirischen nomologischen Referenznetzwerkes (siehe Abschnitt 4.6) von Bedeutung sind:

- Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen (AF – siehe Abschnitt 5.3.1)
- Gedanke an Berufsaufgabe (GB – siehe Abschnitt 5.3.2)
- Gegenwärtiger Gesundheitszustand (GG – siehe Abschnitt 5.3.3)

Jeder Abschnitt beginnt mit der Benennung der Literaturquelle der Skala bzw. des Einzelitems und ggf. der Übersetzung. Nachfolgend sind die jeweiligen Modifikationen der Items, die ausschließlich durch das lidA-Projektkonsortium erarbeitet wurden, skizziert. Darauf folgen die Darstellung der konkreten Inhalte aller Items sowie ihrer für die vorliegende Arbeit maßgebliche Kennzeichnung. Die Auflistung der inhaltlichen bzw. zentralen sowie der zusätzlichen AO und die Erläuterung der jeweiligen Transformationsregel der Werte der zentralen AO bilden den Abschluss jedes Abschnittes.

5.3.1 Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen

Zur Erfassung der subjektiv bewerteten Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen wurde die Struktur der Skala „Work ability in relation to the demands of the job“ des Work Ability Index (WAI) von Tuomi et al. (1998) herangezogen. Jene Skala besteht aus drei englischsprachigen Items, die in „Der Work Ability Index – ein Leitfaden“ (Hasselhorn & Freude, 2007) übersetzt sind. Diejenigen drei Items, welche im lidA-CAPI die Skala „Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen (AF)“ bilden, sind nachstehend angeführt:

- **AF-1** Und sind Sie bei Ihrer Arbeit...
- **AF-2** Wie schätzen Sie Ihre derzeitige Arbeitsfähigkeit in Bezug auf die körperlichen Arbeitsanforderungen ein?
- **AF-3** Und wie schätzen Sie Ihre derzeitige Arbeitsfähigkeit in Bezug auf die geistigen Arbeitsanforderungen ein?

Zur Bearbeitung des Items AF-1 ist eine mehrfach abgestufte, kategorialskalierte Rating-skala mit drei sprachlichen Marken (drei zentrale AO) vorgesehen:

- **1** ...vorwiegend geistig tätig?
- **2** ...vorwiegend körperlich tätig?
- **3** ...etwa gleichermaßen geistig und körperlich tätig?

Des Weiteren können zwei zusätzliche AO („verweigert“ und „weiß nicht“) verwendet werden, die allerdings erst angeboten werden, wenn der TN hieran Bedarf anzeigt.

Für die Beantwortung der Fragen AF-2 sowie AF-3 steht eine mehrfach abgestufte, unipolare Likert-Skala mit fünf verbalen Marken (fünf zentrale AO) zur Verfügung:

- **1** sehr gut
- **2** eher gut
- **3** mittelmäßig
- **4** eher schlecht
- **5** sehr schlecht

Darüber hinaus können zwei zusätzliche AO („verweigert“ und „weiß nicht“) verwendet werden, die im Gegensatz zu den fünf zentralen AO nicht sofort vom Interviewer vorgelesen, sondern erst angeboten werden, wenn entsprechender Bedarf durch den TN kenntlich gemacht wird. Gemäß den Anwendungsempfehlungen von Tuomi et al. (1998) werden die fünf zentralen AO umgepolt, woraus sich die folgenden fünf Werte ergeben: (1 =) 5, (2 =) 4, (3 =) 3 und (4 =) 2 (5 =) 1.

In einem ersten Schritt zur Ermittlung des Gesamtwertes der AF-Skala wird in Entsprechung der Verrechnungsempfehlungen von Hasselhorn & Freude (2007) unter Heranziehung des Einzelwertes von Item AF-1 eine Gewichtung der Einzelwerte der Items AF-2 sowie AF-3 vorgenommen, die im Folgenden dargestellt ist:

- Wenn das Item AF-1 mit dem Wert 1 bearbeitet wird, dann wird der Wert von AF-2 mit dem Faktor 0.50 und der von AF-3 mit dem Faktor 1.50 multipliziert.
- Wenn das Item AF-1 mit dem Wert 2 bearbeitet wird, dann wird der Wert von AF-2 mit dem Faktor 1.50 und der von AF-3 mit dem Faktor 0.50 multipliziert.
- Wenn das Item AF-1 mit dem Wert 3 bearbeitet wird, dann wird der Wert von AF-2 und der Wert von AF-3 jeweils mit dem Faktor 0.50 multipliziert.

In einem zweiten und zugleich letzten Schritt zur Ermittlung des Gesamtwertes der AF-Skala werden die gewichteten Werte der Items AF-2 und AF-3 summiert. Hieraus resultiert für die AF-Skala eine theoretische Spannweite der Werte von 2 (geringe Ausprägung) bis 10 (hohe Ausprägung). Die Polung der AF-Skala erfolgt im Sinne der Skalenbezeichnung. Dies bedeutet, dass eine quantitativ hohe (bzw. niedrige) Ausprägung der AF-Skala als eine hohe (bzw. niedrige) Ausprägung der Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen bewertet wird.

5.3.2 Gedanke an Berufsaufgabe

Für die Erhebung der gedanklichen Beschäftigung mit der Aufgabe des aktuell ausgeübten Berufes wurde ein deutschsprachiges Einzelitem des Erhebungsinstrumentes der Basisbefragung der NEXT-Studie (NEXT-Studie, 2002) herangezogen. Zur Förderung eines schnellen Verständnisses und für eine bessere inhaltliche Passung an die vielfältigen Berufsfelder wurde dieses Item für das lidA-CAPI formal überarbeitet. Hierzu wurden zunächst die Wörter „haben Sie“ umgestellt und unmittelbar nach den Wörtern „Wie oft“ platziert. Ferner wurde die berufsspezifische Formulierung „Ihren Pflegeberuf aufzugeben und eine andere berufliche Tätigkeit zu beginnen“ durch die allgemeine Formulierung „Ihre derzeitige berufliche Tätigkeit zu wechseln“ ersetzt. Das finale Einzelitem, welches im lidA-CAPI als „Gedanke an Berufsaufgabe (GB)“ bezeichnet wird, ist im Folgenden dargestellt:

- **GB** Wie oft haben Sie im Laufe der letzten zwölf Monate daran gedacht, Ihre derzeitige berufliche Tätigkeit zu wechseln?

Zur Bearbeitung dieses Einzelitems (GB) steht den TN eine mehrfach abgestufte, unipolare Likert-Skala mit fünf verbalen Marken (fünf zentrale AO) zur Verfügung:

- **1** nie
- **2** einige Male im Jahr
- **3** einige Male im Monat
- **4** einige Male pro Woche
- **5** jeden Tag

Des Weiteren sind zwei zusätzliche AO („verweigert“ und „weiß nicht“) verfügbar, die im Gegensatz zu den fünf zentralen AO nicht ungefragt vom Interviewer angeboten, sondern erst vorgelesen werden, wenn entsprechender Bedarf erkennbar wird. Die Polung des Einzelitems GB erfolgt im Sinne der Itembezeichnung. Demnach wird eine quantitativ hohe (bzw. niedrige) Ausprägung des Einzelitems (GB) als eine hohe (bzw. niedrige) Ausprägung des Gedankens an Berufsaufgabe bewertet.

5.3.3 Gegenwärtiger Gesundheitszustand

Zur Erfassung des subjektiv bewerteten gegenwärtigen Gesundheitszustandes wird das deutschsprachige Einzelitem „Gegenwärtiger Gesundheitszustand“ des Erhebungsinstrumentes des „Sozio-Oekonomischen Panels (SOEP – Welle 21)“ (Infratest Sozialforschung, 2012) herangezogen. Dieses Einzelitem wird unverändert und unter der gleichen Bezeichnung, nämlich „Gegenwärtiger Gesundheitszustand (GG)“, im IidA-CAPI eingesetzt:

- **GG** Wie würden Sie ihren gegenwärtigen Gesundheitszustand beschreiben?

Zur Bearbeitung dieses Einzelitems (GG) steht den TN eine mehrfach abgestufte, unipolare Likert-Skala mit fünf verbalen Marken (fünf zentrale AO) zur Verfügung:

- **1** sehr gut
- **2** gut
- **3** zufriedenstellend
- **4** weniger gut
- **5** schlecht

Darüber hinaus können zwei zusätzliche AO („verweigert“ und „weiß nicht“) herangezogen werden, die im Gegensatz zu den fünf zentralen AO nicht sofort vom Interviewer vorgelesen, sondern erst angeboten werden, wenn die TN entsprechenden Bedarf anzeigen. Gemäß den Anwendungsvorgaben von Infratest Sozialforschung (2012) werden die fünf zentralen AO umgepolt, woraus sich die folgenden fünf möglichen Werte ergeben: (1 =) 5, (2 =) 4, (3 =) 3 und (4 =) 2 (5 =) 1. Die Polung dieses Einzelitems (GG) erfolgt im Sinne der Itembezeichnung. Demzufolge wird eine quantitativ hohe (bzw. niedrige) Ausprägung des Einzelitems GG als eine hohe (bzw. niedrige) Ausprägung der subjektiven gegenwärtigen Gesundheit aufgefasst.

5.4 Statistische Analysen

Bis auf die Kennwerte der Hauptkomponentenanalyse entsprechen alle statistischen Kennwerte, die im Rahmen der Überprüfung der zwölf aggregierten Hypothesen ermittelt werden (siehe Kapitel 4), denjenigen, welche die empirischen Referenzwerte darstellen (siehe Abschnitt 3.2). Sämtliche Kennwerte werden mit denselben statistischen Verfahren derselben Statistiksoftware („IBM SPSS Statistics 20“) berechnet, um eine möglichst hohe Vergleichbarkeit sicherzustellen. Hinsichtlich der Überprüfung der Normalverteilungsannahme werden ergänzend zwei im vorliegenden Abschnitt angeführte mathematische Gleichungen verwendet, die oben bereits inhaltlich eingeführt wurden (siehe Abschnitt 3.1.1). Die Skalendignität der in der vorliegenden Arbeit verwendeten COPSOQ-II-Items und -Skalen wird als intervallskaliert vorausgesetzt (vgl. Pejtersen et al., 2010; Bortz & Döring, 2009). Die COPSOQ-II-Items und -Skalen werden vor der Durchführung der statistischen Analysen gemäß den Verrechnungsvorschriften von Pejtersen et al. (2010) positiv-linear auf eine Spannweite von 0 bis 100 transformiert (vgl. Abschnitt 5.2).

Zur Überprüfung der TF.1-H.1 wird die Spannweite der COPSOQ-II-Items sowie -Skalen ermittelt (siehe Abschnitt 4.1). Hierzu wird die niedrigste und höchste von den TN gewählte AO angegeben und die Differenz zwischen beiden berechnet (vgl. Abschnitt 3.1.1).

Um mögliche Boden- und Deckeneffekte zu identifizieren, werden zur Überprüfung der TF.1-H.2 (siehe Abschnitt 4.1) jeweils die relativen Häufigkeiten der niedrigsten und höchsten AO der COPSOQ-II-Items ermittelt (vgl. Abschnitt 3.1.1).

Die Normalverteilungsannahme aller COPSOQ-II-Items und -Skalen wird im Rahmen der TF.1-H.3 (siehe Abschnitt 4.1) mittels des K-S-A sowie des S-W-T überprüft. Zusätzlich zur Verwendung der oben angeführten Statistiksoftware wird, für beide Verfahren separat, ein kritischer Wert ermittelt (K-S-A: D-Kritisch. S-W-T: W-Kritisch). Hierbei werden die exakten Stichprobengrößen berücksichtigt und jeweils ein Signifikanzniveau von $\alpha = .01$ gewählt. Der kritische Wert wird sodann mit einer verfahrensspezifischen Statistik (K-S-A: D-Statistik. S-W-T: W-Statistik), die mit der oben angeführten Statistiksoftware ermittelt wird, verglichen (vgl. Abschnitt 3.1.1). Wenn die D-Statistik kleiner ist als D-Kritisch, wird die Nullhypothese, die Grundgesamtheit der Stichprobe sei normalverteilt, angenommen. Wenn die W-Statistik größer ist als W-Kritisch, wird die Nullhypothese, die Grundgesamtheit der Stichprobe sei normalverteilt, angenommen.

Hinsichtlich des K-S-A wird zur Berechnung des kritischen Wertes (D-Kritisch) die folgende mathematische Gleichung verwendet (Conover, 1971):

$$D - \text{Kritisch} = \frac{1,031}{\sqrt{n}}$$

In Bezug auf den S-W-T wird der kritische Wert (*W-Kritisch*) mit der nachstehenden mathematischen Gleichung ermittelt (Precht et al., 2005):

$$W - \text{Kritisch} \approx 0,9963 - \frac{0,0211}{\sqrt{n}} - \frac{1,4106}{n} + \frac{3,1791}{n^2}$$

Zur Überprüfung der Schwierigkeit der COPSOQ-II-Items und -Skalen wird zum Einen der Md als Lagemaß (TF.2-H.1 – siehe Abschnitt 4.2) sowie die Differenz zwischen dem 1. und 3. Quartil (IQR) als Streuungsmaß (TF.2-H.2 – siehe Abschnitt 4.2) berechnet. Zum Anderen wird das AM als Lagemaß (TF.2-H.3 – siehe Abschnitt 4.2) sowie die SD als Streuungsmaß (TF.2-H.4 – siehe Abschnitt 4.2) ermittelt (vgl. Abschnitt 3.1.2).

Um die Gültigkeit der TF.3-H.1 (siehe Abschnitt 4.3) zu überprüfen, werden die einzelnen COPSOQ-II-Itemwerte mit dem Wert ihrer korrespondierenden Skala jeweils bivariat korreliert (Rangkorrelation nach Spearman). Zur Überprüfung der TF.3-H.2 (siehe Abschnitt 4.3) wird ebenso verfahren, wobei hier die Skalenwerte den Wert des jeweils fokussierten Items nicht beinhalten (vgl. Abschnitt 3.1.3).

Die TF.4-H.1 (siehe Abschnitt 4.4) wird überprüft, indem innerhalb der jeweiligen COPSOQ-II-Skala sämtliche I-I-K (Rangkorrelation nach Spearman) ermittelt und mittels der Fisher-z-Transformation das AM sämtlicher I-I-K berechnet wird (vgl. Abschnitt 3.1.4).

Zur Überprüfung der TF.5-H.1 (siehe Abschnitt 4.5) werden in einem ersten Schritt die Korrelationen der 20 interessierenden COPSOQ-II-Items mittels des KMO überprüft (vgl. Backhaus et al., 2011). In einem zweiten Schritt wird eine Hauptkomponentenanalyse angewendet, wobei in Entsprechung zum Kaiser-Guttman-Kriterium ein Eigenwert der Komponenten von zumindest 1.00 gefordert wird (vgl. Wolff & Bacher, 2010; Bortz, 1999). Um schließlich statistisch voneinander unabhängige Komponenten zu erhalten, wird bei der Berechnung der Hauptkomponentenanalyse die orthogonale Varimaxrotation mit Kaiser-Normalisierung (z. B. Fromm, 2012) gewählt (vgl. Abschnitt 3.1.5).

Die Erstellung eines empirischen nomologischen Netzwerkes aller verwendeten Skalen und Einzelitems (vgl. Abschnitt 3.1.6) erfolgt im Zuge der Überprüfung der TF.6-H.1 (siehe Abschnitt 4.6). Hierzu werden die Koeffizienten von 26 bivariaten Korrelationen (Rangkorrelation nach Spearman) kalkuliert.

Die statistischen Analysen werden zunächst anhand der Daten der lidA-Machbarkeitsstudie vollzogen. Aufgrund der geringen Stichprobengröße dieses Datensatzes wird allerdings keine „uneingeschränkte Bestätigung“ (siehe Abschnitt 6) sämtlicher Hypothesen gefordert. Vielmehr soll hier geprüft werden, ob hypothesenkonforme Tendenzen identifiziert werden können, auf deren Grundlage die Durchführung der statistischen Analysen anhand der zwei großen Stichproben (lidA-Hauptstudie: 1. Welle/A+B) als adäquat beurteilt werden können.

6. Ergebnisse

Zur Beantwortung der operationalen Hypothesen (siehe Kapitel 4) wurden auf der Grundlage der im vorherigen Kapitel vorgestellten drei Stichproben (siehe Abschnitt 5.1) statistische Analysen (siehe Abschnitt 5.4) durchgeführt, deren Ergebnisse im vorliegenden Kapitel vorgestellt werden. Die Struktur dieses Kapitels orientiert sich zum Einen an der Chronologie der herangezogenen Stichproben, so dass zuerst sämtliche Ergebnisse der lidA-Machbarkeitsstudie (siehe Abschnitt 6.1) und im Anschluss daran die Ergebnisse der 1. Welle/A+B der lidA-Hauptstudie (siehe Abschnitt 6.2) dargestellt werden. Zum Anderen sind diese beiden Abschnitte jeweils gemäß der sechs Teilfragestellungen (TF) ihrerseits in jeweils sechs Abschnitte untergliedert:

- Verteilung der Werte (siehe Abschnitt 6.1.1 und 6.2.1)
- Item- und Skalenschwierigkeit (siehe Abschnitt 6.1.2 und 6.2.2)
- Trennschärfe (siehe Abschnitt 6.1.3 und 6.2.3)
- Homogenität (siehe Abschnitt 6.1.4 und 6.2.4)
- Dimensionalität (siehe Abschnitt 6.1.5 und 6.2.5)
- nomologisches Netzwerk (siehe Abschnitt 6.1.6 und 6.2.6)

Zur Sicherstellung einer einheitlichen wie eindeutigen Bewertung der Ergebnisse auf der sprachlichen Ebene werden Korrelationen, deren (Irrtums-)Wahrscheinlichkeit (Bortz & Döring, 2009) kleiner als das Signifikanzniveau von $\alpha = .01$ (2-seitig) ist, als „sehr signifikant“, und solche, deren p-Wert kleiner als das Signifikanzniveau von $\alpha = .001$ (2-seitig) ist, als „hoch signifikant“ bezeichnet (vgl. Cowles & Davis, 1982).

Sämtliche operationalen Hypothesen sind zum Zwecke der Übersichtlichkeit als aggregierte Hypothesen formuliert (vgl. Kapitel 4) und vereinen hierdurch stets mehrere Teilhypothesen in sich (z. B. 26 Teilhypothesen hinsichtlich TF.6-H.1). Aus diesem Grund werden die Hypothesen gemäß dem nachfolgenden Sprachsystem bewertet:

- als **„uneingeschränkt bestätigt“** gilt eine aggregierte Hypothese, wenn alle Teilhypothesen den empirischen Ergebnissen vollständig entsprechen
- als **„eingeschränkt bestätigt“** gilt eine aggregierte Hypothese, wenn die Teilhypothesen zu höchstens 10 Prozent nicht den empirischen Ergebnissen entsprechen
- als **„nicht bestätigt“** gilt eine aggregierte Hypothese, wenn mehr als 10 Prozent der Teilhypothesen den empirischen Ergebnissen nicht entsprechen

6.1 lidA-Machbarkeitsstudie

Die folgenden Analyseergebnisse basieren ausschließlich auf dem Datensatz der lidA-Machbarkeitsstudie. Die Ergebnisse werden in Bezug auf die von den Teilfragestellungen (TF) abgeleiteten aggregierten Hypothesen separat bewertet. Gleichsam wird aufgrund der geringen Stichprobengröße keine „uneingeschränkte Bestätigung“ aller Hypothesen erwartet. Vielmehr soll hier geprüft werden, ob hypothesenkonforme Tendenzen identifiziert werden können, auf deren Grundlage die Durchführung der statistischen Analysen anhand der zwei großen Stichproben (lidA-Hauptstudie: 1. Welle/A+B) als adäquat beurteilt werden können.

In dem ersten Abschnitt sind die Ergebnisse der Untersuchung der Verteilung der Werte aller COPSOQ-II-Items und -Skalen dargestellt (siehe Abschnitt 6.1.1). Hierbei werden die Spannweite (vgl. TF.1-H.1), mögliche Boden- und Deckeneffekte (vgl. TF.1-H.2) sowie die Normalverteilungsannahme (vgl. TF.1-H.3) statistisch geprüft.

Im zweiten Abschnitt stehen die Untersuchungsergebnisse im Kontext der Item- und Skalenschwierigkeit im Fokus der Betrachtung (siehe Abschnitt 6.1.2). Die Überprüfungen beziehen sich hier zum Einen auf die Ausprägungen des Md (vgl. TF.2-H.1) sowie des IQR (vgl. TF.2-H.2) und zum Anderen auf die Ausprägungen des AM (vgl. TF.2-H.3) sowie der SD (vgl. TF.2-H.4) sämtlicher COPSOQ-II-Items und -Skalen.

Die Koeffizienten der nicht-korrigierten (vgl. TF.3-H.1) wie auch der korrigierten Trennschärfe (vgl. TF.3-H.2) der 18 relevanten COPSOQ-II-Items sind in dem Abschnitt 6.1.3 ausführlich dargestellt und bewertet. Die AK-Skala bleibt hier unberücksichtigt, da diese lediglich aus zwei Items besteht.

Im vierten Abschnitt sind die Analyseergebnisse der statistischen Überprüfung der Homogenität aller sechs COPSOQ-II-Skalen vorgestellt (siehe Abschnitt 6.1.4). Hier werden die kleinsten und die größten I-I-K-Koeffizienten sowie das AM sämtlicher I-I-K-Koeffizienten innerhalb der jeweiligen Skala (vgl. TF.4-H.1) wiedergegeben.

Die Ergebnisse der Hauptkomponentenanalyse sämtlicher 20 COPSOQ-II-Items (vgl. TF.5-H.1) werden unter dem Stichwort „Dimensionalität“ in dem Abschnitt 6.1.5 vorgestellt und bewertet.

Im sechsten und letzten Abschnitt werden die Analyseergebnisse der statistischen Überprüfung des nomologischen Netzwerkes betrachtet (siehe Abschnitt 6.1.6). Hierbei werden die 26 relevanten bivariaten Korrelationen der sechs COPSOQ-II-Skalen, der AF-Skala sowie der Einzelitems GB und GG dargestellt.

6.1.1 Verteilung der Werte

Aus der deskriptiven Statistik der positiv-linear transformierten Werte geht hervor, dass sämtliche COPSOQ-II-Items und -Skalen der lidA-Machbarkeitsstudie die größtmögliche Spannweite hinsichtlich der zentralen AO (jeweils von 0 bis 100) aufweisen. Auf der Basis dieser Resultate wird die TF.1-H.1 (siehe Abschnitt 4.1) für die lidA-Machbarkeitsstudie als uneingeschränkt bestätigt bewertet. Diese Ergebnisse entsprechen den empirischen Referenzwerten (vgl. Tabelle 2).

Mit Ausnahme der niedrigsten zentralen AO des Items EA-1 (= 53.2 Prozent), sind bei allen 20 COPSOQ-II-Items der lidA-Machbarkeitsstudie in den niedrigsten sowie höchsten zentralen AO jeweils weniger als 50 Prozent der Gesamtzahl gültiger AO zu verzeichnen. Vor dem Hintergrund dieser Ergebnisse wird die TF.1-H.2 (siehe Abschnitt 4.1) für die lidA-Machbarkeitsstudie als eingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten sind allerdings mehr Abweichungen als Übereinstimmungen zu konstatieren (vgl. Tabelle 2).

Insgesamt wird auf der Basis dieser Ergebnisse (Spannweite sowie Boden- und Deckeneffekte) jedoch von einem grundsätzlich akzeptablen Differenzierungspotential der untersuchten COPSOQ-II-Items und -Skalen ausgegangen (vgl. TF.1 – siehe Abschnitt 4.1). Eine hieran anschließende psychometrische Überprüfung anhand größerer Stichproben (lidA-Hauptstudie: 1. Welle/A+B) wird in Bezug auf die Verteilung der Werte als adäquat bewertet.

Die detaillierte Darstellung der empirischen Verteilung der positiv-linear transformierten Werte sämtlicher 20 COPSOQ-II-Items der lidA-Machbarkeitsstudie kann der Tabelle 10 entnommen werden.

Tabelle 10: Verteilung der Werte (lidA-Machbarkeitsstudie)

	0 n (%)	25 n (%)	50 n (%)	75 n (%)	100 n (%)
QA-1	22 (12.9)	32 (18.7)	53 (31.0)	51 (29.8)	13 (7.6)
QA-2	25 (14.6)	33 (19.3)	48 (28.1)	49 (28.7)	16 (9.4)
QA-3	45 (26.3)	56 (32.7)	44 (25.7)	17 (9.9)	9 (5.3)
QA-4	16 (9.4)	42 (24.6)	42 (24.6)	43 (25.1)	28 (16.4)
EA-1	91 (53.2)	27 (15.8)	18 (10.5)	22 (12.9)	13 (7.6)
EA-2	83 (48.5)	27 (15.8)	32 (18.7)	19 (11.1)	10 (5.8)
EA-3	40 (23.4)	18 (10.5)	34 (19.9)	36 (21.1)	43 (25.1)
EM-1	4 (2.3)	8 (4.7)	28 (16.4)	62 (36.3)	69 (40.4)
EM-2	13 (7.6)	12 (7.0)	51 (29.8)	51 (29.8)	44 (25.7)
EM-3	10 (5.8)	3 (1.8)	16 (9.4)	65 (38.0)	77 (45.0)
EM-4	15 (8.8)	22 (12.9)	43 (25.1)	45 (26.3)	46 (26.9)
FQ-1	33 (20.2)	27 (16.6)	40 (24.5)	52 (31.9)	11 (6.7)
FQ-2	17 (9.9)	24 (14.7)	42 (25.8)	56 (32.7)	24 (14.0)
FQ-3	25 (15.3)	24 (14.7)	41 (25.2)	57 (35.0)	16 (9.8)
FQ-4	25 (15.3)	22 (13.5)	51 (31.3)	41 (25.2)	24 (14.7)
HU-1	3 (1.9)	14 (8.7)	35 (21.7)	48 (29.8)	61 (37.9)
HU-2	12 (7.5)	15 (9.3)	24 (14.9)	48 (29.8)	62 (38.5)
HU-3	27 (15.8)	38 (22.2)	40 (24.8)	36 (22.4)	20 (11.7)
	0 n (%)	33 n (%)	67 n (%)	100 n (%)	
AK-1	56 (32.7)	60 (35.1)	44 (25.7)	11 (6.4)	
AK-2	60 (35.1)	65 (38.0)	37 (21.6)	9 (5.3)	

Anmerkungen: **QA**: n = 171; **EA**: n = 171; **EM**: n = 171; **FQ**: n = 163; **HU**: n = 161; **AK**: n = 171.

Die Ergebnisse des K-S-A zeigen, dass der Wert D-Statistik aller 20 COPSOQ-II-Items der lidA-Machbarkeitsstudie wie auch sämtlicher sechs COPSOQ-II-Skalen größer ist (hoch signifikant) als der jeweils korrespondierende Wert D-Kritisch. Im Einzelvergleich mit den empirischen Referenzwerten ist eine hohe Übereinstimmung festzustellen, wobei die wenigen vorhandenen Abweichungen in die Richtung einer Annäherung zur Normalverteilung der beobachteten Werte der lidA-Machbarkeitsstudie gehen (vgl. Tabelle 2).

Konsistent zu diesen Analyseergebnissen ergibt die Anwendung des S-W-T für sämtliche COPSOQ-II-Items und -Skalen, dass der Wert W-Statistik kleiner ist (hoch signifikant), als der jeweils korrespondierende Wert W-Kritisch. Hierbei ist im Einzelvergleich mit den empirischen Referenzwerten in Bezug auf die W-Statistik eine sehr hohe Übereinstimmung festzustellen (vgl. Tabelle 2).

Ergänzend zu den beiden zuvor erwähnten Verfahren (K-S-A und S-W-T) lässt auch eine Augenscheinkontrolle der Verteilung der Werte anhand der oben dargestellten Tabelle 10 deutlich erkennen, dass die (Item-)Werte nicht normalverteilt sind.

Die einzelnen Werte sowohl des K-S-A als auch des S-W-T sind für alle 20 COPSOQ-II-Items und sämtliche sechs COPSOQ-II-Skalen der lidA-Machbarkeitsstudie in der nachfolgenden dargestellt (siehe Tabelle 11).

Tabelle 11: Tests auf Normalverteilung (lidA-Machbarkeitsstudie)

	Kolmogorov-Smirnov-Anpassungstest				Shapiro-Wilk-Test			
	D-Stat.	p-Wert	D-Krit.	α	W-Stat.	p-Wert	W-Krit.	α
QA-1	.18	.000	.079	.01	.91	.000	.987	.01
QA-2	.18	.000	.079	.01	.91	.000	.987	.01
QA-3	.21	.000	.079	.01	.88	.000	.987	.01
QA-4	.17	.000	.079	.01	.91	.000	.987	.01
QA-Skala	.09	.000	.079	.01	.98	.000	.987	.01
EA-1	.31	.000	.079	.01	.76	.000	.987	.01
EA-2	.29	.000	.079	.01	.80	.000	.987	.01
EA-3	.18	.000	.079	.01	.86	.000	.987	.01
EA-Skala	.15	.000	.079	.01	.94	.000	.987	.01
EM-1	.24	.000	.079	.01	.81	.000	.987	.01
EM-2	.19	.000	.079	.01	.88	.000	.987	.01
EM-3	.28	.000	.079	.01	.74	.000	.987	.01
EM-4	.19	.000	.079	.01	.89	.000	.987	.01
EM-Skala	.11	.000	.079	.01	.94	.000	.987	.01
FQ-1	.20	.000	.081	.01	.89	.000	.986	.01
FQ-2	.22	.000	.081	.01	.90	.000	.986	.01
FQ-3	.22	.000	.081	.01	.90	.000	.986	.01
FQ-4	.18	.000	.081	.01	.90	.000	.986	.01
FQ-Skala	.11	.000	.081	.01	.96	.000	.986	.01
HU-1	.22	.000	.081	.01	.85	.000	.986	.01
HU-2	.24	.000	.081	.01	.83	.000	.986	.01
HU-3	.16	.000	.081	.01	.91	.000	.986	.01
HU-Skala	.09	.000	.081	.01	.96	.000	.986	.01
AK-1	.21	.000	.079	.01	.85	.000	.987	.01
AK-2	.22	.000	.079	.01	.84	.000	.987	.01
AK-Skala	.18	.000	.079	.01	.89	.000	.987	.01

Anmerkungen: **QA**: n = 171; **EA**: n = 171; **EM**: n = 171; **FQ**: n = 163; **HU**: n = 161; **AK**: n = 171.

Unter Heranziehung dieser Analyseergebnisse wird die TF.1-H.3 (siehe Abschnitt 4.1) für die lidA-Machbarkeitsstudie als uneingeschränkt bestätigt bewertet. Vor dem Hintergrund dieser Resultate des K-S-A, des S-W-T sowie der Augenscheinkontrolle wird davon ausgegangen, dass die Ausprägungen der mit den fokussierten COPSOQ-II-Items und -Skalen erfassten Konstrukte in der Grundgesamtheit mit hoher Wahrscheinlichkeit nicht normalverteilt sind. Dementsprechend werden für die nachfolgenden statistischen Analysen hinsichtlich der lidA-Machbarkeitsstudie ausschließlich verteilungsfreie Verfahren herangezogen.

6.1.2 Item- und Skalenschwierigkeit

Von den insgesamt 20 COPSOQ-II-Items der lidA-Machbarkeitsstudie ist der Md bei 19 Items innerhalb eines Spektrums von 25.00 bis 75.00 lokalisiert. Das Item EA-1 (= 0.00) befindet sich außerhalb dieses Spektrums. Der Md aller sechs COPSOQ-II-Skalen befindet sich wiederum jeweils innerhalb eines Spektrums von 25.00 bis 75.00. Auf der Grundlage dieser Analyseergebnisse wird die TF.2-H.1 (siehe Abschnitt 4.2) für die lidA-Machbarkeitsstudie als eingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen (vgl. Tabelle 3).

Alle 20 COPSOQ-II-Items der lidA-Machbarkeitsstudie weisen einen IQR von mindestens 25.00 auf. Bei 18 der Items beträgt der IQR 50.00 oder mehr. Ausschließlich der IQR der beiden Items EM-1 (= 25.00) und EM-3 (= 25.00) ist kleiner als der zuletzt genannte Wert. Der IQR aller sechs COPSOQ-II-Skalen erreicht jeweils den geforderten Mindestwert von 25.00. Dementsprechend wird die TF.2-H.2 (siehe Abschnitt 4.2) für die lidA-Machbarkeitsstudie als uneingeschränkt bestätigt bewertet. Auch im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen (vgl. Tabelle 3).

Das AM von 18 der insgesamt 20 COPSOQ-II-Items der lidA-Machbarkeitsstudie befindet sich innerhalb eines Spektrums von 25.00 bis 75.00. Das AM des Items EM-1 (= 76.90) und dasjenige des Items EM-3 (= 78.66) befindet sich außerhalb dieses Spektrums. Das AM sämtlicher sechs COPSOQ-II-Skalen ist wiederum jeweils innerhalb eines Spektrums von 25.00 bis 75.00 lokalisiert. Vor dem Hintergrund dieser Analyseergebnisse wird die TF.2-H.3 (siehe Abschnitt 4.2) für die lidA-Machbarkeitsstudie als eingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen (vgl. Tabelle 3).

Eine SD von mindestens 20.00 ist bei sämtlichen 20 COPSOQ-II-Items sowie allen sechs COPSOQ-II-Skalen zu verzeichnen. Demgemäß wird die TF.2-H.4 (siehe Abschnitt 4.2) für die lidA-Machbarkeitsstudie als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine niedrige Übereinstimmung festzustellen, wobei alle Abweichungen einheitlich in die Richtung einer größeren Streuung der Werte der lidA-Machbarkeitsstudie gehen (vgl. Tabelle 3).

Das erste, zweite (Md) und dritte Quartil sowie das AM und die SD aller COPSOQ-II-Items und COPSOQ-II-Skalen der lidA-Machbarkeitsstudie sind nachstehend tabellarisch aufgelistet (siehe Tabelle 12).

Tabelle 12: Item- und Skalenschwierigkeit (IidA-Machbarkeitsstudie)

	1. Quartil	Md	3. Quartil	AM	SD
QA-1	25.00	50.00	75.00	50.15	28.63
QA-2	25.00	50.00	75.00	49.71	30.07
QA-3	.00	25.00	50.00	33.77	28.24
QA-4	25.00	50.00	75.00	46.35	30.76
QA-Skala	31.25	43.75	62.50	44.99	24.20
EA-1	0.00	0.00	50.00	26.46	33.95
EA-2	0.00	25.00	50.00	27.49	32.05
EA-3	25.00	50.00	100.00	53.51	37.51
EA-Skala	16.67	33.33	50.00	35.82	26.23
EM-1	75.00	75.00	100.00	76.90	24.56
EM-2	50.00	75.00	100.00	64.77	29.16
EM-3	75.00	75.00	100.00	78.66	26.52
EM-4	50.00	75.00	100.00	62.43	31.43
EM-Skala	62.50	75.00	87.50	70.69	21.46
FQ-1	25.00	50.00	75.00	47.09	31.23
FQ-2	25.00	50.00	75.00	57.06	29.85
FQ-3	25.00	50.00	75.00	52.30	30.66
FQ-4	25.00	50.00	75.00	52.61	31.50
FQ-Skala	37.50	56.25	75.00	52.26	24.68
HU-1	50.00	75.00	100.00	73.29	26.39
HU-2	50.00	75.00	100.00	70.65	31.20
HU-3	25.00	50.00	75.00	47.52	31.89
HU-Skala	50.00	66.67	83.33	63.82	24.23
AK-1	0.00	33.00	67.00	35.25	30.71
AK-2	0.00	33.00	67.00	32.30	29.55
AK-Skala	0.00	33.00	50.00	33.78	28.28

Anmerkungen: QA: n = 171; EA: n = 171; EM: n = 171; FQ: n = 163; HU: n = 161; AK: n = 171.

Insgesamt weisen die verschiedenen Lagemaße (Md und AM) akzeptable Ausprägungen auf, wobei zwei COPSOQ-II-Skalen als eher schwierig (EA-Skala und AK-Skala) und zwei als eher leicht (EM-Skala und HU-Skala) zu bewerten sind. Diese Abweichungen vom Idealwert der Schwierigkeit (50 Prozent der Spannweite) wirken sich tendenziell ungünstig auf das Differenzierungspotential der benannten COPSOQ-II-Items und -Skalen aus. Allerdings kann auf der Grundlage der ermittelten Streuungsmaße (IQR und SD) von einem akzeptablen Differenzierungspotential der fokussierten COPSOQ-II-Items und -Skalen ausgegangen werden (vgl. TF.2 – siehe Kapitel 4). Hinsichtlich der Item- und Skalenschwierigkeit wird eine anschließende psychometrische Überprüfung anhand größerer Stichproben (IidA-Hauptstudie: 1. Welle/A+B) als adäquat bewertet.

6.1.3 Trennschärfe

Alle nicht-korrigierten Trennschärfekoeffizienten der 18 hier relevanten COPSOQ-II-Items der lidA-Machbarkeitsstudie sind größer als .50 und hoch signifikant. Darüber hinaus sind die Koeffizienten von 16 Items größer als .70. Die Koeffizienten der Items EA-1 (= .69) und EM-1 (= .56) unterschreiten den zuletzt genannten Wert. Demgemäß wird die TF.3-H.1 (siehe Abschnitt 4.3) für die lidA-Machbarkeitsstudie als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung festzustellen, wobei die meisten Abweichungen durch niedrigere Koeffizienten für die lidA-Machbarkeitsstudie gekennzeichnet sind (vgl. Tabelle 4).

Sämtliche Koeffizienten der korrigierten Trennschärfe der 18 relevanten COPSOQ-II-Items der lidA-Machbarkeitsstudie sind größer als .30 und hoch signifikant. Ferner sind die Koeffizienten von 14 Items größer als .50. Ausschließlich die Koeffizienten der vier Items EA-1 (= .43), EA-2 (= .41), EA-3 (= .38) und EM-1 (= .34) erreichen diesen Wert nicht. Auf der Grundlage jener Analyseergebnisse wird die TF.3-H.2 (siehe Abschnitt 4.3) für die lidA-Machbarkeitsstudie als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung festzustellen, wobei die meisten Koeffizienten in der lidA-Machbarkeitsstudie nach unten abweichen, d. h. eine geringere Trennschärfe aufweisen (vgl. Tabelle 4).

Tabelle 13: Trennschärfe (lidA-Machbarkeitsstudie)

	nicht-korrigierte Trennschärfe		korrigierte Trennschärfe	
	ρ	p-Wert	ρ	p-Wert
QA-1	.78 ***	.000	.62 ***	.000
QA-2	.87 ***	.000	.75 ***	.000
QA-3	.79 ***	.000	.64 ***	.000
QA-4	.82 ***	.000	.65 ***	.000
EA-1	.69 ***	.000	.43 ***	.000
EA-2	.71 ***	.000	.41 ***	.000
EA-3	.79 ***	.000	.38 ***	.000
EM-1	.56 ***	.000	.34 ***	.000
EM-2	.85 ***	.000	.68 ***	.000
EM-3	.68 ***	.000	.51 ***	.000
EM-4	.86 ***	.000	.66 ***	.000
FQ-1	.78 ***	.000	.58 ***	.000
FQ-2	.78 ***	.000	.61 ***	.000
FQ-3	.76 ***	.000	.59 ***	.000
FQ-4	.84 ***	.000	.69 ***	.000
HU-1	.78 ***	.000	.56 ***	.000
HU-2	.85 ***	.000	.64 ***	.000
HU-3	.81 ***	.000	.53 ***	.000

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 171; **EA**: n = 171; **EM**: n = 171; **FQ**: n = 163; **HU**: n = 161.

Die Werte der nicht-korrigierten ebenso wie der korrigierten Trennschärfekoeffizienten aller 18 hier relevanten COPSOQ-II-Items sind in der Tabelle 13 aufgeführt.

Insgesamt wird auf der Basis vorgenannter Ergebnisse (korrigierte und nicht-korrigierte Trennschärfekoeffizienten) konstatiert, dass die fokussierten COPSOQ-II-Items innerhalb der jeweils korrespondierenden Skala sehr gut zwischen TN mit einem hohen und solchen mit einem niedrigen Skalenwert differenzieren. Die „Itemvalidität“ (Bühner, 2011, S. 171) ist hiernach durchweg als akzeptabel zu bezeichnen (vgl. TF.3 – siehe Kapitel 4). Im Hinblick auf die Trennschärfe wird eine anschließende psychometrische Überprüfung anhand größerer Stichproben (IidA-Hauptstudie: 1. Welle/A+B) als adäquat bewertet.

6.1.4 Homogenität

Die Spannweite der AM der I-I-K-Koeffizienten in Bezug auf die fünf relevanten COPSOQ-II-Skalen der lidA-Machbarkeitsstudie erstreckt sich von .33 bis .57. Diese Werte liegen innerhalb des a priori festgelegten Spektrums von .30 bis .80. Der I-I-K-Koeffizient der beiden AK-Items beträgt .74 und ist hoch signifikant. Vor dem Hintergrund dieser Analyseergebnisse wird die TF.4-H.1 (siehe Abschnitt 4.4) für die lidA-Machbarkeitsstudie als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen (vgl. Tabelle 4).

Nachfolgend sind die kleinsten (Minimum) und die größten (Maximum) I-I-K-Koeffizienten sowie das AM der I-I-K-Koeffizienten der fünf relevanten COPSOQ-II-Skalen angeführt (siehe Tabelle 14). Da die AK-Skala aus zwei Items besteht, ist für diese der bivariate Korrelationskoeffizient ermittelt und in Klammern gesetzt.

Tabelle 14: Homogenität (lidA-Machbarkeitsstudie)

	Inter-Item-Korrelation		
	Minimum	AM	Maximum
QA-Skala	.48 ***	.57	.68 ***
EA-Skala	.29 ***	.33	.37 ***
EM-Skala	.21 ***	.44	.66 ***
FQ-Skala	.45 ***	.52	.58 ***
HU-Skala	.42 ***	.51	.59 ***
AK-Skala	---	(.74 ***)	---

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 171; **EA**: n = 171; **EM**: n = 171; **FQ**: n = 163; **HU**: n = 161; **AK**: n = 171.

Die sechs fokussierten COPSOQ-II-Skalen werden auf der Grundlage der oben dargestellten Ergebnisse als in akzeptablem Maße homogen bewertet, d. h. die Items hängen in dem erwarteten Ausmaß statistisch miteinander zusammen (vgl. TF.4 – siehe Kapitel 4). In Bezug auf den Aspekt der Homogenität wird eine anschließende psychometrische Überprüfung anhand größerer Stichproben (lidA-Hauptstudie: 1. Welle/A+B) als adäquat bewertet.

6.1.5 Dimensionalität

Neben der Homogenität der sechs COPSOQ-II-Skalen der lidA-Machbarkeitsstudie (siehe Abschnitt 6.1.4) verweist die Höhe des KMO mit einem Wert von .79 auf die Adäquanz der Verwendung einer Faktorenanalyse. Dieser Wert des KMO liegt deutlich über dem in der Literatur geforderten Minimum von .50 (vgl. Backhaus et al., 2011) bzw. .60 (vgl. Dziuban & Shirkey) und erreicht fast den von Backhaus et al. (2011) als ideal bezeichneten Bereich, der mit einem Grenzwert von .80 beginnt.

Tabelle 15: Komponentenladungsmatrix (lidA-Machbarkeitsstudie).

	Komponente ¹					
	1	2	3	4	5	6
QA-2	.86					
QA-1	.79					
QA-4	.77					
QA-3	.75					
FQ-3		.79				
FQ-4		.77				
FQ-2		.73				
FQ-1		.70				
EM-2			.83			
EM-4			.80			
EM-3			.76			
EM-1			.58			
HU-2				.81		
HU-3				.79		
HU-1				.71		
AK-1					.86	
AK-2	.33				.85	
EA-1						.79
EA-2						.78
EA-3			.36			.52
Erklärte Varianz in %	14.57	13.19	12.96	9.99	8.85	8.56

¹. Die Rotation ist in 6 Iterationen konvergiert. Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Koeffizienten sind nach der Größe sortiert und kleine Koeffizienten (Absolutwert < 0.30) wurden in der Darstellung unterdrückt. Erklärte Gesamtvarianz = 68.12 %. Listenweises n = 158.

Die Anwendung einer Hauptkomponentenanalyse (Rotationsmethode: Varimax mit Kaiser-Normalisierung) der 20 COPSOQ-II-Items der lidA-Machbarkeitsstudie konvergierte in sechs Iterationen und sechs Komponenten. Jene sechs Komponenten entsprechen den postulierten sechs COPSOQ-II-Skalen. Auf der Grundlage der zuvor angeführten Ergebnisse wird die TF.5-H.1 (siehe Abschnitt 4.5) für die lidA-Machbarkeitsstudie als uneingeschränkt bestätigt bewertet.

Jedes der 20 COPSOQ-II-Items weist ausschließlich eine Hauptladung ($\geq .50$) auf, wobei sich die Werte dieser Hauptladungen über ein Spektrum von .52 bis .86 erstrecken. Darüber hinaus erreichen 18 der 20 Hauptladungen einen Wert mit einer Höhe von .70 oder

mehr. Die Hauptladungen der beiden Items EM-1 (= .58) und EA-3 (= .52) bleiben unter diesem zuletzt genannten Wert. Insgesamt 18 der 20 COPSOQ-II-Items weisen keine Nebenladungen ($\geq .30$ und $< .50$) auf. Eine Nebenladung hat das Item AK-2 (= .33) auf der Komponente 1 (QA) und das Item EA-3 (= .36) auf der Komponente 3 (EM). Die TF.5-H.2 (siehe Abschnitt 4.5) wird vor dem Hintergrund dieser Ergebnisse für die lidA-Machbarkeitsstudie als eingeschränkt bestätigt bewertet.

Das Spektrum der durch die sechs extrahierten Komponenten erklärten Varianz erstreckt sich von 8.56 Prozent bis 14.57 Prozent, wobei die erklärte Gesamtvarianz des Modells 68.12 Prozent beträgt. Die Komponentenstruktur und die erklärte Varianz der jeweiligen Komponenten sowie die Haupt- und Nebenladungen der 20 COPSOQ-II-Items sind in der oben wiedergegebenen Komponentenladungsmatrix dargestellt (siehe Tabelle 15).

Insgesamt werden die berichteten Analyseergebnisse als Bestätigung für die postulierte Dimensionalität, d. h. die Skalenstruktur der COPSOQ-II-Items interpretiert (vgl. Abbildung 3). Die theoretisch zusammengehörenden Items messen hiernach tatsächlich jeweils dasselbe Konstrukt bzw. dieselbe Komponente (vgl. TF.5 – siehe Kapitel 4). Vor diesem Hintergrund wird eine anschließende psychometrische Überprüfung anhand größerer Stichproben (lidA-Hauptstudie: 1. Welle/A+B) als adäquat bewertet.

6.1.6 Nomologisches Netzwerk

Von den insgesamt 26 für das nomologische Netzwerk relevanten bivariaten Korrelationen hinsichtlich der sechs COPSOQ-II-Skalen sowie der AF-Skala und den Einzelitems GB und GG der lidA-Machbarkeitsstudie entsprechen 25 Korrelationen dem empirischen Referenznetzwerk (vgl. Tabelle 6). Eine bivariate Korrelation (EA-GB) weist im Vergleich zum Referenznetzwerk ein abweichendes Vorzeichen auf. Von den 25 erwartungskonformen Korrelationen sind sechs hoch signifikant, acht sehr signifikant, fünf signifikant und sechs nicht signifikant.

Die EA-Skala ist in vier Korrelationen eingebunden, von denen drei Korrelationen bestätigt werden. Sämtliche Korrelationen, in welche die übrigen COPSOQ-II-Skalen eingebunden sind, werden bestätigt: QA-Skala (fünf), EM-Skala (sechs), FQ-Skala (acht), HU-Skala (acht) und AK-Skala (sieben). Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung zu konstatieren, wobei die meisten Abweichungen durch niedrigere Koeffizienten hinsichtlich der lidA-Machbarkeitsstudie gekennzeichnet sind (vgl. Tabelle 40, Tabelle 49, Tabelle 57 und Tabelle 65).

Vor dem Hintergrund dieser Resultate wird die TF.6-H.1 (siehe Abschnitt 4.6) für die lidA-Machbarkeitsstudie als eingeschränkt bestätigt bewertet. Die genauen Koeffizienten der 26 bivariaten Korrelationen der sechs COPSOQ-II-Skalen sowie der AF-Skala und den Einzelitems GB und GG sind in der Tabelle 16 angeführt.

Tabelle 16: Nomologisches Netzwerk (lidA-Machbarkeitsstudie)

	QA	EA	EM	FQ	HU	AK	AF	GB	GG
QA-Skala	---	---	-.36 ***	---	.45 ***	-.24 **	.13	-.20 *	
EA-Skala		---	.41 ***	---	.02	---	---	.04	.13
EM-Skala			---	.23 **	.18 *	---	.30 ***	-.15	.17 *
FQ-Skala				---	.47 ***	-.33 ***	.19 *	-.24 **	.22 **
HU-Skala					---	-.10	.18 *	-.21 **	.09
AK-Skala						---	-.23 **	.24 **	-.27 **
AF-Skala							---	---	---
GB								---	---
GG									---

*. Die Korrelation ist auf dem Niveau von $\alpha = .05$ (2-seitig) signifikant. **. Die Korrelation ist auf dem Niveau von $\alpha = .01$ (2-seitig) sehr signifikant. ***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: Listenweises $n = 158$.

Insgesamt werden diese Ergebnisse als Hinweis dafür gewertet, dass die sechs COPSOQ-II-Skalen im Rahmen des lidA-CAPI dieselben theoretisch postulierten Konstrukte (siehe Abschnitt 5.2) erfassen, die auch im Zuge ihrer Verwendung als „Papier-Bleistift-Verfahren“ erfasst werden (vgl. TF.6 – siehe Kapitel 4). Auf der Grundlage dieser Analyseergebnisse wird eine anschließende psychometrische Überprüfung anhand größerer Stichproben (lidA-Hauptstudie: 1. Welle/A+B) als adäquat bewertet.

6.2 lidA-Hauptstudie: 1. Welle/A+B

Die unten angeführten Ergebnisse basieren auf den Analysen der zwei Datensätze der 1. Welle/A+B der lidA-Hauptstudie. Die Systematisierung der Ergebnisse in sechs Abschnitte folgt der Einteilung der oben dargestellten sechs Teilfragestellung (TF) der handlungsleitenden Fragestellung (F) dieser Arbeit (siehe Abschnitt 4). Sämtliche Ergebnisse werden hinsichtlich der von den Teilfragestellung (TF) abgeleiteten aggregierten Hypothesen (siehe Kapitel 4) sowie der Divergenzen und Konvergenzen bezüglich der Datensätze der 1. Welle/A und 1. Welle/B bewertet.

In dem ersten Abschnitt sind die Ergebnisse der Untersuchung der Verteilung der Werte aller COPSOQ-II-Items und -Skalen dargestellt (siehe Abschnitt 6.2.1). Hierbei werden die Spannweite (vgl. TF.1-H.1), mögliche Boden- und Deckeneffekte (vgl. TF.1-H.2) sowie die Normalverteilungsannahme (vgl. TF.1-H.3) statistisch geprüft.

Im zweiten Abschnitt stehen die Untersuchungsergebnisse zur Item- und Skalenschwierigkeit im Fokus der Betrachtung (siehe Abschnitt 4.2). Die Testungen beziehen sich hier zum Einen auf die Ausprägungen des Md (vgl. TF.2-H.1) und des IQR (vgl. TF.2-H.2). Zum Anderen beziehen sich die Testungen auf die Ausprägungen des AM (vgl. TF.2-H.3) und der SD (vgl. TF.2-H.4) aller COPSOQ-II-Items und -Skalen.

Die Koeffizienten der nicht-korrigierten (vgl. TF.3-H.1) sowie korrigierten Trennschärfe (vgl. TF.3-H.2) aller COPSOQ-II-Items werden in dem Abschnitt 6.2.3 wiedergegeben und bewertet.

Im vierten Abschnitt werden die Ergebnisse der statistischen Überprüfung der Homogenität aller sechs COPSOQ-II-Skalen vorgestellt (siehe Abschnitt 6.2.4). Separiert nach den COPSOQ-II-Skalen sind hier die kleinsten und die größten I-I-K-Koeffizienten sowie das AM sämtlicher I-I-K-Koeffizienten (vgl. TF.4-H.1) dargestellt.

Die Vorstellung und Bewertung der Ergebnisse der Hauptkomponentenanalyse sämtlicher 20 COPSOQ-II-Items (vgl. TF.5-H.1) erfolgt im Abschnitt 6.2.5.

Die Ergebnisdarstellung der statistischen Überprüfung des nomologischen Netzwerkes erfolgt im sechsten und letzten Abschnitt (siehe Abschnitt 6.2.6). Hier werden die 26 relevanten bivariaten Korrelationen der sechs COPSOQ-II-Skalen, der AF-Skala sowie der Einzelitems GB und GG angeführt und bewertet.

6.2.1 Verteilung der Werte

Aus der deskriptiven Statistik der positiv-linear transformierten Werte der 1. Welle/A der lidA-Hauptstudie geht hervor, dass sämtliche COPSOQ-II-Items und -Skalen die größtmögliche Spannweite hinsichtlich der zentralen AO (jeweils von 0 bis 100) aufweisen. Gemäß dieser Analyseergebnisse wird die TF.1-H.1 (siehe Abschnitt 4.1) als uneingeschränkt bestätigt bewertet. Diese Ergebnisse entsprechen den empirischen Referenzwerten (vgl. Tabelle 2).

Bei allen 20 COPSOQ-II-Items der 1. Welle/A der lidA-Hauptstudie sind in den niedrigsten und höchsten zentralen AO jeweils weniger als 50 Prozent der Gesamtzahl gültiger AO zu konstatieren. Demensprechend wird die TF.1-H.2 (siehe Abschnitt 4.1) als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist bezüglich der beobachteten Verteilungen der Werte eine mittlere Übereinstimmung festzustellen (vgl. Tabelle 2).

Die empirische Verteilung der positiv-linear transformierten Werte aller 20 COPSOQ-II-Items der 1. Welle/A der lidA-Hauptstudie ist der Tabelle 17 zu entnehmen.

Tabelle 17: Verteilung der Werte (lidA-Hauptstudie: 1. Welle/A)

	0 n (%)	25 n (%)	50 n (%)	75 n (%)	100 n (%)
QA-1	281 (9.9)	482 (17.0)	975 (34.5)	864 (30.5)	228 (8.1)
QA-2	348 (12.3)	584 (20.6)	875 (30.9)	790 (27.9)	233 (8.2)
QA-3	669 (23.6)	923 (32.6)	764 (27.0)	383 (13.5)	91 (3.2)
QA-4	490 (17.3)	897 (31.7)	572 (20.2)	664 (23.5)	207 (7.3)
EA-1	1.330 (47.0)	412 (14.6)	393 (13.9)	393 (13.9)	302 (10.7)
EA-2	1.258 (44.5)	569 (20.1)	490 (17.3)	318 (11.2)	195 (6.9)
EA-3	710 (25.1)	353 (12.5)	445 (15.7)	715 (25.3)	607 (21.4)
EM-1	73 (2.6)	115 (4.1)	514 (18.2)	1.044 (36.9)	1.084 (38.3)
EM-2	192 (6.8)	276 (9.8)	898 (31.7)	911 (32.2)	553 (19.5)
EM-3	131 (4.6)	107 (3.8)	451 (15.9)	1.169 (41.3)	972 (34.3)
EM-4	207 (7.3)	381 (13.5)	987 (34.9)	825 (29.2)	430 (15.2)
FQ-1	319 (11.5)	521 (18.8)	931 (33.7)	784 (28.4)	210 (7.6)
FQ-2	213 (7.7)	425 (15.4)	750 (27.1)	1.040 (37.6)	337 (12.2)
FQ-3	295 (10.7)	445 (16.1)	835 (30.2)	947 (34.2)	243 (8.8)
FQ-4	301 (11.1)	472 (17.4)	790 (29.1)	870 (32.1)	279 (10.3)
HU-1	58 (2.1)	190 (7.0)	572 (21.0)	1.014 (37.3)	888 (32.6)
HU-2	136 (5.0)	218 (8.0)	532 (19.5)	869 (31.9)	967 (35.5)
HU-3	321 (11.8)	628 (23.1)	792 (29.1)	687 (25.2)	294 (10.8)
	0 n (%)	33 n (%)	67 n (%)	100 n (%)	
AK-1	829 (29.3)	1.189 (42.0)	632 (22.3)	180 (6.4)	
AK-2	848 (30.0)	1.125 (39.8)	689 (24.3)	168 (5.9)	

Anmerkungen: **QA:** n = 2 830; **EA:** n = 2 830; **EM:** n = 2 830; **FQ:** n = 2 712; **HU:** n = 2 722; **AK:** n = 2 830.

Gemäß der deskriptiven Statistik der positiv-linear transformierten Werte der 1. Welle/B der lidA-Hauptstudie weisen sämtliche COPSOQ-II-Items und -Skalen die größtmögliche Spannweite hinsichtlich der zentralen AO (jeweils von 0 bis 100) auf. Auf der Grundlage dieser Ergebnisse wird die TF.1-H.1 (siehe Abschnitt 4.1) als uneingeschränkt bestätigt bewertet. Diese Ergebnisse entsprechen den empirischen Referenzwerten (vgl. Tabelle 2).

In den niedrigsten und höchsten zentralen AO der 20 COPSOQ-II-Items werden jeweils weniger als 50 Prozent der Gesamtzahl gültiger AO der 1. Welle/B der lidA-Hauptstudie festgestellt. Die TF.1-H.2 (siehe Abschnitt 4.1) wird vor diesem Hintergrund als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist in Bezug auf die beobachteten Verteilungen der Werte eine mittlere Übereinstimmung festzustellen (vgl. Tabelle 2).

Eine detaillierte Darstellung der empirischen Verteilung der positiv-linear transformierten Werte aller 20 COPSOQ-II-Items der 1. Welle/B der lidA-Hauptstudie ist in Tabelle 18 vollzogen.

Tabelle 18: Verteilung der Werte (lidA-Hauptstudie: 1. Welle/B)

	0 n (%)	25 n (%)	50 n (%)	75 n (%)	100 n (%)
QA-1	325 (11.5)	489 (17.2)	1.009 (35.6)	820 (28.9)	193 (6.8)
QA-2	390 (13.8)	599 (21.1)	854 (30.1)	789 (27.8)	204 (7.2)
QA-3	708 (25.0)	911 (32.1)	766 (27.0)	365 (12.9)	86 (3.0)
QA-4	509 (17.9)	907 (32.0)	603 (21.3)	632 (22.3)	185 (6.5)
EA-1	1.335 (47.1)	407 (14.4)	384 (13.5)	376 (13.3)	334 (11.8)
EA-2	1.252 (44.1)	556 (19.6)	472 (16.6)	335 (11.8)	221 (7.8)
EA-3	745 (26.3)	310 (10.9)	452 (15.9)	689 (24.3)	640 (22.6)
EM-1	71 (2.5)	116 (4.1)	517 (18.2)	1.047 (36.9)	1.085 (38.3)
EM-2	224 (7.9)	312 (11.0)	863 (30.4)	891 (31.4)	546 (19.3)
EM-3	149 (5.3)	126 (4.4)	445 (15.7)	1.145 (40.4)	971 (34.2)
EM-4	256 (9.0)	388 (13.7)	951 (33.5)	856 (30.2)	385 (13.6)
FQ-1	310 (11.3)	530 (19.2)	927 (33.6)	792 (28.7)	196 (6.9)
FQ-2	208 (7.5)	419 (15.2)	751 (27.3)	1.046 (38.0)	331 (12.0)
FQ-3	294 (10.7)	445 (16.2)	845 (30.7)	954 (34.6)	217 (7.9)
FQ-4	283 (10.5)	472 (17.4)	838 (31.0)	853 (31.5)	259 (9.6)
HU-1	66 (2.4)	188 (6.9)	533 (19.6)	1.016 (37.4)	917 (33.7)
HU-2	120 (4.4)	219 (8.1)	521 (19.2)	892 (32.8)	968 (35.6)
HU-3	313 (11.5)	611 (22.5)	806 (29.6)	674 (24.8)	316 (11.6)
	0 n (%)	33 n (%)	67 n (%)	100 n (%)	
AK-1	879 (31.0)	1.046 (36.9)	732 (25.8)	179 (6.3)	
AK-2	911 (32.1)	1.050 (37.0)	705 (24.9)	170 (6.0)	

Anmerkungen: **QA:** n = 2 836; **EA:** n = 2 836; **EM:** n = 2 836; **FQ:** n = 2 705; **HU:** n = 2 720; **AK:** n = 2 836.

Im Vergleich der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie hinsichtlich der relativen Höhe der positiv-linear transformierten Werte sind ausschließlich geringfügige paarweise Abweichungen zu konstatieren, die höchstens 5.10 Prozentpunkte betragen. Insgesamt kann auf der Basis jener Resultate (Spannweite sowie Boden- und Deckeneffekte) von einem grundsätzlich akzeptablen Differenzierungspotential der fokussierten COPSOQ-II-Items und -Skalen ausgegangen werden (vgl. TF.1 – siehe Kapitel 4).

Hinsichtlich der 1. Welle/A der lidA-Hauptstudie zeigen die Ergebnisse des K-S-A, dass der Wert D-Statistik sämtlicher COPSOQ-II-Items und -Skalen größer ist (hoch signifikant) als der jeweils korrespondierende Wert D-Kritisch. Im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen, wobei die wenigen Abweichungen in die Richtung einer Annäherung zur Normalverteilung der positiv-linear transformierten Werte der 1. Welle/A der lidA-Hauptstudie gehen (vgl. Tabelle 2).

Konsistent zu diesen Ergebnissen ergibt die Anwendung des S-W-T ebenfalls für alle COPSOQ-II-Items und -Skalen, dass der Wert W-Statistik kleiner ist (hoch signifikant) als der jeweils korrespondierende Wert W-Kritisch. Auch im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen, wobei die wenigen vorhandenen Abweichungen nicht einheitlich in die Richtung einer Annäherung zur Normalverteilung der positiv-linear transformierten Werte der 1. Welle/A der lidA-Hauptstudie gehen (vgl. Tabelle 2).

Zusätzlich zu den zwei zuvor genannten statistischen Verfahren ergibt eine Augenscheinkontrolle der Verteilung der positiv-linear transformierten Werte (vgl. Tabelle 18), dass sämtliche (Item-)Werte nicht normalverteilt sind.

Unter Heranziehung dieser Ergebnisse wird die TF.1-H.3 (siehe Abschnitt 4.1) in Bezug auf die 1. Welle/A der lidA-Hauptstudie als uneingeschränkt bestätigt bewertet. Die einzelnen Werte sowohl des K-S-A als auch diejenigen des S-W-T für alle COPSOQ-II-Items und -Skalen, die auf der Grundlage des Datensatzes der 1. Welle/A ermittelt sind, werden in der folgenden Tabelle 19 dargestellt.

Die Anwendung des K-S-A hinsichtlich aller COPSOQ-II-Items und -Skalen der 1. Welle/B der lidA-Hauptstudie ergibt, dass sämtliche D-Statistiken größer sind (hoch signifikant) als die jeweils korrespondierenden Werte D-Kritisch. Im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen, wobei die wenigen vorhandenen Abweichungen in die Richtung einer Annäherung zur Normalverteilung der beobachteten Werte gehen (vgl. Tabelle 2).

Analog hierzu zeigen die Ergebnisse des S-W-T für sämtliche COPSOQ-II-Items und -Skalen, dass der Wert W-Statistik kleiner ist (hoch signifikant), als der jeweils korrespondierende Wert W-Kritisch. Ebenfalls im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen, wobei auch hier die wenigen

vorhandenen Abweichungen nicht einheitlich in die Richtung einer Annäherung zur Normalverteilung der beobachteten Werte gehen (vgl. Tabelle 2).

Auch eine ergänzende Augenscheinkontrolle der Verteilung der Werte anhand der oben dargestellten Tabelle 18 ergibt, dass alle (Item-)Werte nicht normalverteilt sind.

Entsprechend der vorgenannten Ergebnisse wird die TF.1-H.3 (siehe Abschnitt 4.1) hinsichtlich der 1. Welle/B der lidA-Hauptstudie als uneingeschränkt bestätigt bewertet. Die einzelnen Werte des K-S-A sowie des S-W-T für sämtliche COPSOQ-II-Items und -Skalen der 1. Welle/B sind in der Tabelle 20 dargestellt.

Tabelle 19: Tests auf Normalverteilung (lidA-Hauptstudie: 1. Welle/A)

	Kolmogorov-Smirnov-Anpassungstest				Shapiro-Wilk-Test			
	D-Stat.	p-Wert	D-Krit.	α	W-Stat.	p-Wert	W-Krit.	α
QA-1	.20	.000	.019	.01	.91	.000	.995	.01
QA-2	.17	.000	.019	.01	.91	.000	.995	.01
QA-3	.21	.000	.019	.01	.89	.000	.995	.01
QA-4	.21	.000	.019	.01	.90	.000	.995	.01
QA-Skala	.07	.000	.019	.01	.98	.000	.995	.01
EA-1	.28	.000	.019	.01	.80	.000	.995	.01
EA-2	.26	.000	.019	.01	.81	.000	.995	.01
EA-3	.20	.000	.019	.01	.86	.000	.995	.01
EA-Skala	.10	.000	.019	.01	.95	.000	.995	.01
EM-1	.24	.000	.019	.01	.82	.000	.995	.01
EM-2	.20	.000	.019	.01	.90	.000	.995	.01
EM-3	.27	.000	.019	.01	.81	.000	.995	.01
EM-4	.18	.000	.019	.01	.91	.000	.995	.01
EM-Skala	.10	.000	.019	.01	.96	.000	.995	.01
FQ-1	.19	.000	.020	.01	.91	.000	.995	.01
FQ-2	.23	.000	.020	.01	.90	.000	.995	.01
FQ-3	.21	.000	.020	.01	.90	.000	.995	.01
FQ-4	.20	.000	.020	.01	.91	.000	.995	.01
FQ-Skala	.10	.000	.020	.01	.97	.000	.995	.01
HU-1	.23	.000	.020	.01	.85	.000	.995	.01
HU-2	.23	.000	.020	.01	.85	.000	.995	.01
HU-3	.16	.000	.020	.01	.92	.000	.995	.01
HU-Skala	.11	.000	.020	.01	.96	.000	.995	.01
AK-1	.24	.000	.019	.01	.85	.000	.995	.01
AK-2	.23	.000	.019	.01	.85	.000	.995	.01
AK-Skala	.18	.000	.019	.01	.91	.000	.995	.01

Anmerkungen: **QA**: n = 2 830; **EA**: n = 2 830; **EM**: n = 2 830; **FQ**: n = 2 712; **HU**: n = 2 722; **AK**: n = 2 830.

Im paarweisen Vergleich der 1. Welle/A und der 1. Welle/B der lidA-Hauptstudie hinsichtlich der COPSOQ-II-Items und -Skalen besteht bei 17 von 26 D-Statistiken eine perfekte Übereinstimmung. Bei neun D-Statistiken ist im paarweisen Vergleich eine geringfügige Abweichung von maximal .02 festzustellen. Alle 26 D-Statistiken weisen in beiden Datensätzen jeweils einen p-Wert von .000 auf.

In Bezug auf die fokussierten COPSOQ-II-Items und -Skalen der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie besteht bei 19 von 26 W-Statistiken im paarweisen Vergleich eine perfekte Übereinstimmung. Bei sieben W-Statistiken liegt im Paarvergleich eine geringfügige Abweichung von maximal .01 vor. Alle 26 W-Statistiken weisen in beiden Datensätzen jeweils einen p-Wert von .000 auf.

Tabelle 20: Tests auf Normalverteilung (lidA-Hauptstudie: 1. Welle/B)

	Kolmogorov-Smirnov-Anpassungstest				Shapiro-Wilk-Test			
	D-Stat.	p-Wert	D-Krit.	α	W-Stat.	p-Wert	W-Krit.	α
QA-1	.20	.000	.019	.01	.91	.000	.995	.01
QA-2	.17	.000	.019	.01	.91	.000	.995	.01
QA-3	.20	.000	.019	.01	.89	.000	.995	.01
QA-4	.21	.000	.019	.01	.90	.000	.995	.01
QA-Skala	.07	.000	.019	.01	.98	.000	.995	.01
EA-1	.28	.000	.019	.01	.79	.000	.995	.01
EA-2	.26	.000	.019	.01	.82	.000	.995	.01
EA-3	.20	.000	.019	.01	.85	.000	.995	.01
EA-Skala	.10	.000	.019	.01	.95	.000	.995	.01
EM-1	.23	.000	.019	.01	.82	.000	.995	.01
EM-2	.20	.000	.019	.01	.90	.000	.995	.01
EM-3	.27	.000	.019	.01	.82	.000	.995	.01
EM-4	.18	.000	.019	.01	.91	.000	.995	.01
EM-Skala	.10	.000	.019	.01	.96	.000	.995	.01
FQ-1	.19	.000	.020	.01	.91	.000	.995	.01
FQ-2	.23	.000	.020	.01	.90	.000	.995	.01
FQ-3	.21	.000	.020	.01	.90	.000	.995	.01
FQ-4	.19	.000	.020	.01	.91	.000	.995	.01
FQ-Skala	.09	.000	.020	.01	.98	.000	.995	.01
HU-1	.24	.000	.020	.01	.85	.000	.995	.01
HU-2	.23	.000	.020	.01	.84	.000	.995	.01
HU-3	.16	.000	.020	.01	.92	.000	.995	.01
HU-Skala	.12	.000	.020	.01	.96	.000	.995	.01
AK-1	.22	.000	.019	.01	.85	.000	.995	.01
AK-2	.22	.000	.019	.01	.85	.000	.995	.01
AK-Skala	.16	.000	.019	.01	.90	.000	.995	.01

Anmerkungen: **QA**: n = 2 836; **EA**: n = 2 836; **EM**: n = 2 836; **FQ**: n = 2 705; **HU**: n = 2 720; **AK**: n = 2 836.

Vor dem Hintergrund der Analyseergebnisse des K-S-A, des S-W-T sowie der Augenscheinkontrolle wird davon ausgegangen, dass die Ausprägungen der mit den fokussierten COPSOQ-II-Items und -Skalen erfassten Konstrukte in der Grundgesamtheit mit ausreichend hoher Wahrscheinlichkeit nicht normalverteilt sind. Dementsprechend werden für die nachfolgenden statistischen Analysen hinsichtlich der 1. Welle/A+B der lidA-Hauptstudie ausschließlich verteilungsfreie Verfahren herangezogen.

6.2.2 Item- und Skalenschwierigkeit

Bei allen COPSOQ-II-Items und -Skalen der 1. Welle/A der lidA-Hauptstudie ist der Md innerhalb eines Spektrums von 25.00 bis 75.00 lokalisiert. Entsprechend dieser Ergebnisse wird die TF.2-H.1 (siehe Abschnitt 4.2) für die 1. Welle/A als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen (vgl. Tabelle 3).

Sämtliche 20 COPSOQ-II-Items der 1. Welle/A der lidA-Hauptstudie weisen einen IQR von mindestens 25.00 auf. Bei 14 Items beträgt der IQR 50.00 oder mehr. Bei den sechs Items QA-3, EM-1, EM-2, EM-3, EM-4 und FQ-2 beträgt der IQR genau 25.00. Der IQR sämtlicher sechs COPSOQ-II-Skalen der 1. Welle/A beträgt mindestens 25.00. Dementsprechend wird die TF.2-H.2 (siehe Abschnitt 4.2) für die 1. Welle/A als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine sehr hohe Übereinstimmung festzustellen (vgl. Tabelle 3).

Das AM von 19 der insgesamt 20 COPSOQ-II-Items der 1. Welle/A der lidA-Hauptstudie befindet sich innerhalb eines Spektrums von 25.00 bis 75.00. Das AM des Items EM-1 (76.07) befindet sich außerhalb dieses Spektrums. Das AM sämtlicher sechs COPSOQ-II-Skalen der 1. Welle/A ist wiederum jeweils innerhalb eines Spektrums von 25.00 bis 75.00 lokalisiert. Auf der Grundlage dieser Analyseergebnisse wird die TF.2-H.3 (siehe Abschnitt 4.2) für die 1. Welle/A als eingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine hohe Übereinstimmung festzustellen (vgl. Tabelle 3).

Weil die SD sämtlicher 20 COPSOQ-II-Items sowie aller sechs COPSOQ-II-Skalen der 1. Welle/A der lidA-Hauptstudie jeweils mindestens 20.00 beträgt, wird die TF.2-H.4 (siehe Abschnitt 4.2) für die 1. Welle/A als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung festzustellen, wobei fast alle Abweichungen in die Richtung einer größeren Streuung der positiv-linear transformierten Werte der 1. Welle/A der lidA-Hauptstudie gehen (vgl. Tabelle 3).

Das erste, zweite (Md) und dritte Quartil sowie das AM und die SD aller COPSOQ-II-Items und -Skalen der 1. Welle/A der lidA-Hauptstudie sind nachstehend tabellarisch aufgelistet (siehe Tabelle 21).

Tabelle 21: Item- und Skalenschwierigkeit (lidA-Hauptstudie: 1. Welle/A)

	1. Quartil	Md	3. Quartil	AM	SD
QA-1	25.00	50.00	75.00	52.44	27.23
QA-2	25.00	50.00	75.00	49.79	28.58
QA-3	25.00	25.00	50.00	35.02	27.12
QA-4	25.00	50.00	75.00	42.94	30.18
QA-Skala	31.25	43.75	62.50	45.05	22.83
EA-1	0.00	25.00	50.00	31.67	35.83
EA-2	0.00	25.00	50.00	29.00	32.23
EA-3	0.00	50.00	75.00	51.38	37.39
EA-Skala	16.67	33.33	58.33	37.35	27.46
EM-1	75.00	75.00	100.00	76.07	24.47
EM-2	50.00	75.00	75.00	61.99	27.87
EM-3	75.00	75.00	100.00	74.24	25.86
EM-4	50.00	50.00	75.00	57.86	27.70
EM-Skala	56.25	68.75	81.25	67.54	20.67
FQ-1	25.00	50.00	75.00	50.41	27.81
FQ-2	50.00	50.00	75.00	57.80	27.71
FQ-3	25.00	50.00	75.00	53.60	28.08
FQ-4	25.00	50.00	75.00	53.26	28.87
FQ-Skala	37.50	56.25	68.75	53.71	23.00
HU-1	50.00	75.00	100.00	72.81	25.00
HU-2	50.00	75.00	100.00	71.24	28.49
HU-3	25.00	50.00	75.00	50.05	29.45
HU-Skala	50.00	66.67	83.33	64.70	21.78
AK-1	0.00	33.00	67.00	35.19	29.30
AK-2	0.00	33.00	67.00	35.37	29.48
AK-Skala	16.50	33.00	50.00	35.28	27.61

Anmerkungen: **QA**: n = 2 830; **EA**: n = 2 830; **EM**: n = 2 830; **FQ**: n = 2 712; **HU**: n = 2 722; **AK**: n = 2 830.

Bei allen COPSOQ-II-Items und -Skalen der 1. Welle/B der lidA-Hauptstudie liegt der Md innerhalb eines Spektrums von 25.00 bis 75.00. Auf dieser Grundlage wird die TF.2-H.1 (siehe Abschnitt 4.2) für die 1. Welle/B als uneingeschränkt bestätigt bewertet.

Der IQR aller 20 COPSOQ-II-Items der 1. Welle/B der lidA-Hauptstudie beträgt mindestens 25.00. Bei 15 Items beträgt der IQR 50.00 oder mehr. Fünf Items (QA-3, EM-1, EM-2, EM-4, FQ-2) weisen einen IQR von genau 25.00 auf. Der IQR sämtlicher sechs COPSOQ-II-Skalen der 1. Welle/B erreicht jeweils eine Mindesthöhe von 25.00. Demgemäß wird die TF.2-H.2 (siehe Abschnitt 4.2) für die 1. Welle/B als uneingeschränkt bestätigt bewertet.

Das AM von 19 der insgesamt 20 COPSOQ-II-Items der 1. Welle/B der lidA-Hauptstudie befindet sich innerhalb eines Spektrums von 25.00 bis 75.00. Einzig das AM des Items EM-1 (AM = 76.08) befindet sich außerhalb dieses Spektrums. Das AM sämtlicher sechs COPSOQ-II-Skalen der 1. Welle/B ist innerhalb eines Spektrums von 25.00 bis 75.00 lokalisiert. Vor dem Hintergrund dieser Analyseergebnisse wird die TF.2-H.3 (siehe Abschnitt 4.2) für die 1. Welle/B als eingeschränkt bestätigt bewertet.

Die TF.2-H.4 (siehe Abschnitt 4.2) wird für die 1. Welle/B der lidA-Hauptstudie als uneingeschränkt bestätigt bewertet, da sämtliche 20 COPSOQ-II-Items sowie alle sechs COPSOQ-II-Skalen der 1. Welle/B eine SD von mindestens 20.00 aufweisen. Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung festzustellen, wobei fast alle Abweichungen einheitlich in die Richtung einer größeren Streuung der positiv-linear transformierten Werte der 1. Welle/B der lidA-Hauptstudie gehen (vgl. Tabelle 3).

In der Tabelle 22 sind das erste, zweite (Md) und dritte Quartil sowie das AM und die SD aller COPSOQ-II-Items und -Skalen der 1. Welle/B der lidA-Hauptstudie wiedergegeben.

Tabelle 22: Item- und Skalenschwierigkeit (lidA-Hauptstudie: 1. Welle/B)

	1. Quartil	Md	3. Quartil	AM	SD
QA-1	25.00	50.00	75.00	50.59	27.30
QA-2	25.00	50.00	75.00	48.40	28.76
QA-3	25.00	25.00	50.00	34.22	27.06
QA-4	25.00	50.00	75.00	41.86	29.75
QA-Skala	25.00	43.75	62.50	43.77	22.77
EA-1	0.00	25.00	75.00	32.08	36.38
EA-2	0.00	25.00	50.00	29.88	33.02
EA-3	0.00	50.00	75.00	51.49	37.94
EA-Skala	16.67	33.33	58.33	37.81	27.63
EM-1	75.00	75.00	100.00	76.08	24.40
EM-2	50.00	75.00	75.00	60.78	28.77
EM-3	50.00	75.00	100.00	73.48	26.77
EM-4	50.00	50.00	75.00	56.40	28.26
EM-Skala	56.25	68.75	81.25	66.69	21.30
FQ-1	25.00	50.00	75.00	50.31	27.56
FQ-2	50.00	50.00	75.00	57.92	27.55
FQ-3	25.00	50.00	75.00	53.22	27.77
FQ-4	25.00	50.00	75.00	53.08	28.25
FQ-Skala	37.50	56.25	68.75	53.60	22.96
HU-1	50.00	75.00	100.00	73.25	25.29
HU-2	50.00	75.00	100.00	71.77	27.96
HU-3	25.00	50.00	75.00	50.63	29.55
HU-Skala	50.00	66.67	83.33	65.22	21.73
AK-1	0.00	33.00	67.00	35.78	30.20
AK-2	0.00	33.00	67.00	34.87	30.05
AK-Skala	0.00	33.00	50.00	35.32	28.51

Anmerkungen: **QA**: n = 2 836 ; **EA**: n = 2 836; **EM**: n = 2 836; **FQ**: n = 2 705; **HU**: n = 2 720; **AK**: n = 2 836.

Bezüglich der COPSOQ-II-Items und -Skalen der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie ist im paarweisen Vergleich bei 25 von 26 Md eine perfekte Übereinstimmung der Werte festzustellen. Bei dem Item EA-3 besteht eine Abweichung von 25.00.

Im Vergleich der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie hinsichtlich der COPSOQ-II-Items besteht bei 18 von 20 IQR eine perfekte Übereinstimmung. Bei den Items EA-1 (= 25.00) und EM-3 (= 25.00) ist jeweils eine Abweichung zu verzeichnen. In

Bezug auf die COPSOQ-II-Skalen der 1. Welle/A und 1. Welle/B besteht bei vier von sechs IQR im paarweisen Vergleich eine perfekte Übereinstimmung. Abweichungen liegen hier jeweils bei der QA-Skala (= 6.25) sowie der AK-Skala (= 16.50) vor.

Bei den AM aller COPSOQ-II-Items und -Skalen der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie sind im paarweisen Vergleich sehr hohe Übereinstimmungen erkennbar. Die geringfügigen Abweichungen erreichen bei einer Spannweite der Werte von 29.00 bis 76.07 (1. Welle/A) bzw. von 29.88 bis 76.08 (1. Welle/B) maximal einen Wert von 1.85.

In einem paarweisen Vergleich der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie in Bezug auf die COPSOQ-II-Items und -Skalen besteht bei allen SD eine sehr hohe Übereinstimmung. Die geringfügigen Abweichungen erreichen bei einer Spannweite der Werte von 20.67 bis 35.83 (1. Welle/A) bzw. von 21.30 bis 37.94 (1. Welle/B) maximal einen Wert von 0.91.

Auf der Grundlage der ermittelten Streuungsmaße (IQR und SD) wird von einem akzeptablen Differenzierungspotential der fokussierten COPSOQ-II-Items und -Skalen ausgegangen. Ferner weisen auch die verschiedenen Lagemaße (Md und AM) akzeptable Ausprägungen auf, wobei zwei COPSOQ-II-Skalen als eher schwierig (EA-Skala und AK-Skala) und zwei als eher leicht (EM-Skala und HU-Skala) zu bewerten sind. Diese Abweichungen vom Idealwert der Schwierigkeit (ca. 50 Prozent der Spannweite) wirken sich tendenziell reduzierend auf das Differenzierungspotential der benannten COPSOQ-II-Items und -Skalen aus (vgl. TF.2 – siehe Kapitel 4).

6.2.3 Trennschärfe

Alle nicht-korrigierten Trennschärfekoeffizienten der 18 hier relevanten COPSOQ-II-Items der 1. Welle/A der lidA-Hauptstudie sind erwartungsgemäß größer als .50 und hoch signifikant. Ferner sind die Koeffizienten von 17 Items größer als .70. Ausschließlich der Koeffizient des Items EM-1 (= .64) unterschreitet den zuletzt genannten Wert. Vor diesem Hintergrund wird die TF.3-H.1 (siehe Abschnitt 4.3) für die 1. Welle/A als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung festzustellen, wobei die meisten Abweichungen für die 1. Welle/A der lidA-Hauptstudie niedrigere Koeffizienten darstellen (vgl. Tabelle 4).

Größer als .30 und hoch signifikant sind sämtliche Koeffizienten der korrigierten Trennschärfe der 18 hier relevanten COPSOQ-II-Items der 1. Welle/A der lidA-Hauptstudie. Darüber hinaus sind die Koeffizienten von 14 Items größer als .50. Die Koeffizienten der vier Items EA-1 (= .47), EA-3 (= .49), EM-1 (= .42), HU-3 (= .42) unterschreiten diesen Wert. Die TF.3-H.2 (siehe Abschnitt 4.3) wird auf der Grundlage dieser Analyseergebnisse der 1. Welle/A als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung festzustellen, wobei die meisten Abweichungen niedrigere Koeffizienten für die 1. Welle/A der lidA-Hauptstudie darstellen (vgl. Tabelle 4).

Die einzelnen Werte ebenso der nicht-korrigierten wie der korrigierten Trennschärfekoeffizienten und der korrespondierenden p-Werte der hier relevanten 18 COPSOQ-II-Items der 1. Welle/A der lidA-Hauptstudie sind in der Tabelle 23 dargestellt.

Tabelle 23: Trennschärfe (lidA-Hauptstudie: 1. Welle/A)

	nicht-korrigierte Trennschärfe		korrigierte Trennschärfe	
	ρ	p-Wert	ρ	p-Wert
QA-1	.76 ***	.000	.59 ***	.000
QA-2	.85 ***	.000	.71 ***	.000
QA-3	.81 ***	.000	.67 ***	.000
QA-4	.80 ***	.000	.61 ***	.000
EA-1	.75 ***	.000	.47 ***	.000
EA-2	.75 ***	.000	.50 ***	.000
EA-3	.81 ***	.000	.49 ***	.000
EM-1	.64 ***	.000	.42 ***	.000
EM-2	.82 ***	.000	.64 ***	.000
EM-3	.74 ***	.000	.57 ***	.000
EM-4	.84 ***	.000	.70 ***	.000
FQ-1	.81 ***	.000	.66 ***	.000
FQ-2	.83 ***	.000	.70 ***	.000
FQ-3	.78 ***	.000	.61 ***	.000
FQ-4	.81 ***	.000	.65 ***	.000
HU-1	.77 ***	.000	.53 ***	.000
HU-2	.80 ***	.000	.54 ***	.000
HU-3	.76 ***	.000	.42 ***	.000

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 2 830; **EA**: n = 2 830; **EM**: n = 2 830; **FQ**: n = 2 712; **HU**: n = 2 722.

Größer als .50 und hoch signifikant sind alle Koeffizienten der nicht-korrigierten Trennschärfe der 18 hier relevanten COPSOQ-II-Items der 1. Welle/B der lidA-Hauptstudie. Außerdem sind die Koeffizienten von 17 Items größer als .70. Der nicht-korrigierte Trennschärfekoeffizient des Items EM-1 (= .64) unterschreitet den zuletzt genannten Wert. Diesen Ergebnissen entsprechend wird die TF.3-H.1 (siehe Abschnitt 4.3) für die 1. Welle/B als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung festzustellen, wobei die meisten Koeffizienten in der 1. Welle/B der lidA-Hauptstudie nach unten abweichen, d. h. eine geringere Trennschärfe aufweisen (vgl. Tabelle 4).

Sämtliche Koeffizienten der korrigierten Trennschärfe der 18 hier relevanten COPSOQ-II-Items der 1. Welle/B der lidA-Hauptstudie sind erwartungsgemäß größer als .30 und hoch signifikant. Die Koeffizienten von 13 Items sind darüber hinaus größer als .50. Die Koeffizienten der fünf Items EA-1 (= .44), EA-2 (= .49), EA-3 (= .45), EM-1 (= .43), HU-3 (= .41) unterschreiten diesen Wert. Vor dem Hintergrund dieser Analyseergebnisse wird die TF.3-H.2 (siehe Abschnitt 4.3) für die 1. Welle/B als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine mittlere Übereinstimmung festzustellen, wobei die meisten Koeffizienten in der 1. Welle/B der lidA-Hauptstudie nach unten abweichen, d. h. eine geringere Trennschärfe aufweisen (vgl. Tabelle 4).

Im Folgenden sind sämtliche Werte ebenso der nicht-korrigierten wie der korrigierten Trennschärfekoeffizienten und der korrespondierenden p-Werte aller 18 hier relevanten COPSOQ-II-Items der 1. Welle/B der lidA-Hauptstudie dargestellt (siehe Tabelle 24).

Tabelle 24: Trennschärfe (lidA-Hauptstudie: 1. Welle/B)

	nicht-korrigierte Trennschärfe		korrigierte Trennschärfe	
	ρ	p-Wert	ρ	p-Wert
QA-1	.76 ***	.000	.59 ***	.000
QA-2	.85 ***	.000	.72 ***	.000
QA-3	.82 ***	.000	.67 ***	.000
QA-4	.79 ***	.000	.61 ***	.000
EA-1	.74 ***	.000	.44 ***	.000
EA-2	.75 ***	.000	.49 ***	.000
EA-3	.79 ***	.000	.45 ***	.000
EM-1	.64 ***	.000	.43 ***	.000
EM-2	.84 ***	.000	.67 ***	.000
EM-3	.75 ***	.000	.59 ***	.000
EM-4	.85 ***	.000	.69 ***	.000
FQ-1	.83 ***	.000	.68 ***	.000
FQ-2	.84 ***	.000	.71 ***	.000
FQ-3	.77 ***	.000	.60 ***	.000
FQ-4	.83 ***	.000	.68 ***	.000
HU-1	.78 ***	.000	.54 ***	.000
HU-2	.80 ***	.000	.55 ***	.000
HU-3	.76 ***	.000	.41 ***	.000

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 2 836; **EA**: n = 2 836; **EM**: n = 2 836; **FQ**: n = 2 705; **HU**: n = 2 720.

Hinsichtlich sämtlicher COPSOQ-II-Items der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie ist bei allen nicht-korrigierten Trennschärfekoeffizienten im paarweisen Vergleich eine sehr hohe Übereinstimmung festzustellen. Geringfügige Abweichungen erreichen bei einer Spannweite der Werte von .64 bis .85 (1. Welle/A) bzw. von .64 bis .85 (1. Welle/B) maximal einen Wert von .02.

Eine sehr hohe Übereinstimmung ist ebenfalls in Bezug auf den paarweisen Vergleich der korrigierten Trennschärfekoeffizienten sämtlicher COPSOQ-II-Items der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie festzustellen. Die geringfügigen Abweichungen erreichen bei einer Spannweite der Werte von .42 bis .71 (1. Welle/A) bzw. von .41 bis .72 (1. Welle/B) höchstens einen Wert von .04.

Zusammenfassend wird auf der Basis der korrigierten und nicht-korrigierten Trennschärfekoeffizienten konstatiert, dass die 18 relevanten COPSOQ-II-Items innerhalb der jeweils korrespondierenden Skala sehr gut zwischen TN mit einem hohen und solchen mit einem niedrigen Skalenwert differenzieren. Die „Itemvalidität“ (Bühner, 2011, S. 171) ist demgemäß ausnahmslos als akzeptabel zu bezeichnen (vgl. TF.3 – siehe Kapitel 4).

6.2.4 Homogenität

Die Spannweite der AM der I-I-K-Koeffizienten in Bezug auf die fünf relevanten COPSOQ-II-Skalen der 1. Welle/A der lidA-Hauptstudie erstreckt sich von .34 bis .67. Diese Werte liegen innerhalb des a priori festgelegten Spektrums von .30 bis .80. Der I-I-K-Koeffizient der beiden AK-Items beträgt .77 und ist hoch signifikant. Dementsprechend wird die TF.4-H.1 (siehe Abschnitt 4.4) für die 1. Welle/A als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine hohe Übereinstimmung festzustellen (vgl. Tabelle 4). Die einzelnen Werte sind im Folgenden aufgeführt (siehe Tabelle 25).

Tabelle 25: Homogenität (lidA-Hauptstudie: 1. Welle/A)

	Inter-Item-Korrelation		
	Minimum	AM	Maximum
QA-Skala	.46 ***	.54	.64 ***
EA-Skala	.39 ***	.41	.42 ***
EM-Skala	.34 ***	.47	.67 ***
FQ-Skala	.49 ***	.55	.63 ***
HU-Skala	.36 ***	.43	.55 ***
AK-Skala	---	(.77 ***)	---

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 2 830; **EA**: n = 2 830; **EM**: n = 2 830; **FQ**: n = 2 712; **HU**: n = 2 722; **AK**: n = 2 830.

In Bezug auf die fünf relevanten COPSOQ-II-Skalen der 1. Welle/A der lidA-Hauptstudie erstreckt sich die Spannweite der AM der I-I-K-Koeffizienten von .33 bis .68. Diese Werte liegen innerhalb des im Vorfeld festgelegten Spektrums von .30 bis .80. Der I-I-K-Koeffizient der beiden AK-Items beträgt .79 und ist hoch signifikant. Auf der Grundlage dieser Analyseergebnisse wird die TF.4-H.1 (siehe Abschnitt 4.4) für die 1. Welle/B als uneingeschränkt bestätigt bewertet. Im Einzelvergleich mit den empirischen Referenzwerten ist eine hohe Übereinstimmung festzustellen (vgl. Tabelle 4). Die einzelnen Werte sind nachstehend in der Tabelle 26 dargestellt.

Tabelle 26: Homogenität (lidA-Hauptstudie: 1. Welle/B)

	Inter-Item-Korrelation		
	Minimum	AM	Maximum
QA-Skala	.45 ***	.55	.64 ***
EA-Skala	.35 ***	.39	.41 ***
EM-Skala	.33 ***	.49	.68 ***
FQ-Skala	.50 ***	.57	.64 ***
HU-Skala	.35 ***	.44	.58 ***
AK-Skala	---	(.79 ***)	---

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 2 836; **EA**: n = 2 836; **EM**: n = 2 836; **FQ**: n = 2 705; **HU**: n = 2 720; **AK**: n = 2 836.

Im Vergleich der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie hinsichtlich der COPSOQ-II-Items ist bei allen AM der I-I-Koeffizienten eine sehr hohe Übereinstimmung festzustellen. Die geringfügigen Abweichungen erreichen bei einer Spannweite der Werte von .41 bis .55 (1. Welle/A) bzw. von .39 bis .57 (1. Welle/B) maximal einen Wert von .02. Die sechs fokussierten COPSOQ-II-Skalen werden auf der Grundlage der oben dargestellten Ergebnisse als in akzeptablem Maße homogen bewertet, d. h. die Items hängen in dem erwarteten Ausmaß statistisch miteinander zusammen (vgl. TF.4 – siehe Kapitel 4).

6.2.5 Dimensionalität

Neben der Homogenität der sechs COPSOQ-II-Skalen der 1. Welle/A der lidA-Hauptstudie (siehe Abschnitt 6.2.4) verweist die Höhe des KMO (= .81) auf die Adäquanz der Verwendung einer Faktorenanalyse. Dieser Wert des KMO erreicht den von Backhaus et al. (2011) als ideal bezeichneten Bereich, welcher dem Grenzwert von .80 beginnt.

Die Anwendung einer Hauptkomponentenanalyse (Rotationsmethode: Varimax mit Kaiser-Normalisierung) der 20 COPSOQ-II-Items der 1. Welle/A der lidA-Hauptstudie konvergiert in fünf Iterationen und insgesamt sechs Komponenten. Jene sechs Komponenten entsprechen den postulierten sechs COPSOQ-II-Skalen. Auf der Grundlage dieser Analyseergebnisse wird die TF.5-H.1 (siehe Abschnitt 4.5) für die 1. Welle/A der lidA-Hauptstudie als uneingeschränkt bestätigt bewertet.

Jedes der 20 COPSOQ-II-Items weist ausschließlich eine Hauptladung ($\geq .50$) auf, wobei sich die Werte dieser Hauptladungen über ein Spektrum von .56 bis .90 erstrecken. Insgesamt 19 der 20 Hauptladungen erreichen einen Wert mit einer Höhe von .70 oder mehr. Die Hauptladungen des Items EM-1 (= .58) bleibt unter diesem zuletzt genannten Wert. Von den 20 COPSOQ-II-Items weisen 19 Items keine Nebenladungen ($\geq .30$ und $< .50$) auf. Eine Nebenladung hat das Item EM-1 (= .31) auf der Komponente 5 (EA).

Die TF.5-H.2 (siehe Abschnitt 4.5) wird gemäß der oben angeführten Ergebnisse für die 1. Welle/A der lidA-Hauptstudie als eingeschränkt bestätigt bewertet.

Das Spektrum der durch die sechs extrahierten Komponenten erklärten Varianz erstreckt sich von 5.96 Prozent bis 20.70 Prozent, wobei die erklärte Gesamtvarianz des Modells 66.64 Prozent beträgt. Die detaillierten Ergebnisse sind in der nachfolgenden Komponentenladungsmatrix dargestellt (siehe Tabelle 27).

Tabelle 27: Komponentenladungsmatrix (lidA-Hauptstudie: 1. Welle/A).

	Komponente ¹					
	1	2	3	4	5	6
FQ-2	.81					
FQ-4	.81					
FQ-1	.79					
FQ-3	.77					
QA-2		.85				
QA-3		.83				
QA-4		.75				
QA-1		.74				
EM-4			.84			
EM-3			.79			
EM-2			.79			
EM-1			.56		.31	
HU-2				.80		
HU-1				.78		
HU-3				.73		
EA-2					.76	
EA-1					.76	
EA-3					.71	
AK-2						.90
AK-1						.90
Erklärte Varianz in %	20.70	17.29	8.81	7.52	6.36	5.96

¹. Die Rotation ist in 5 Iterationen konvergiert. Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Koeffizienten sind nach der Größe sortiert und kleine Koeffizienten (Absolutwert < .30) wurden in der Darstellung unterdrückt. Erklärte Gesamtvarianz = 66.64 %. Listenweises n = 2 640.

Die Adäquanz der Verwendung einer Faktorenanalyse der 20 COPSOQ-II-Items der 1. Welle/B der lidA-Hauptstudie ist nicht nur durch die Homogenität der sechs COPSOQ-II-Skalen (siehe Abschnitt 6.2.4), sondern auch durch die Höhe des KMO mit einem Wert von .81 begründet. Dieser KMO-Wert erreicht den von Backhaus et al. (2011) als ideal bezeichneten Bereich, welcher einen unteren Grenzwert von .80 aufweist.

Die Durchführung einer Hauptkomponentenanalyse (Rotationsmethode: Varimax mit Kaiser-Normalisierung) sämtlicher 20 COPSOQ-II-Items der 1. Welle/B der lidA-Hauptstudie konvergiert in sechs Iterationen und sechs Komponenten. Diese sechs Komponenten entsprechen den postulierten sechs COPSOQ-II-Skalen. Den angeführten Ergebnissen entsprechend wird die TF.5-H.1 (siehe Abschnitt 4.5) für die 1. Welle/B der lidA-Hauptstudie als uneingeschränkt bestätigt bewertet.

Jedes der 20 COPSOQ-II-Items weist ausschließlich eine Hauptladung ($\geq .50$) auf, wobei sich die Werte dieser Hauptladungen über ein Spektrum von .57 bis .91 erstrecken. Insgesamt 18 der 20 Hauptladungen erreichen hierbei einen Wert mit einer Höhe von .70 oder mehr. Die Hauptladungen der beiden Items EM-1 (= .57) und EA-3 (= .65) bleiben unter diesem zuletzt genannten Wert. Von den 20 COPSOQ-II-Items weisen 18 Items keine Nebenladungen ($\geq .30$ und $< .50$) auf. Eine Nebenladung hat das Item EM-1 (= .31)

auf der Komponente 5 (EA) sowie das Item EA-3 (= .39) auf der Komponenten 3 (EM). Gemäß diesen Ergebnissen wird die TF.5-H.2 (siehe Abschnitt 4.5) für die 1. Welle/B der lidA-Hauptstudie als eingeschränkt bestätigt bewertet.

Das Spektrum der durch die sechs extrahierten Komponenten erklärten Varianz der erstreckt sich von 5.73 Prozent bis 21.21 Prozent, und die erklärte Gesamtvarianz des Modells beträgt 67.09 Prozent. In der nachstehenden Komponentenladungsmatrix sind die einzelnen Ergebnisse angeführt (siehe Tabelle 28).

Tabelle 28: Komponentenladungsmatrix (lidA-Hauptstudie: 1. Welle/B).

	Komponente ¹					
	1	2	3	4	5	6
FQ-4	.83					
FQ-2	.82					
FQ-1	.81					
FQ-3	.76					
QA-2		.85				
QA-3		.82				
QA-4		.76				
QA-1		.71				
EM-4			.84			
EM-3			.80			
EM-2			.80			
EM-1			.57		.31	
HU-2				.81		
HU-1				.79		
HU-3				.72		
EA-2					.78	
EA-1					.76	
EA-3			.39		.65	
AK-1						.91
AK-2						.91
Erklärte Varianz in %	21.21	16.92	8.81	7.94	6.48	5.73

¹. Die Rotation ist in 6 Iterationen konvergiert. Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Koeffizienten sind nach der Größe sortiert und kleine Koeffizienten (Absolutwert < 0.30) wurden in der Darstellung unterdrückt. Gesamtvarianz = 67.09 %. Listenweises n = 2 633.

Im Vergleich der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie hinsichtlich der Hauptkomponentenanalyse der 20 COPSOQ-II-Items sind jeweils sechs Komponenten festzustellen, deren konstituierende Hauptladungen den gleichen Items zuzuordnen sind. Ferner ist die Reihenfolge der Komponenten, ausgerichtet an der durch diese jeweils erklärten Varianz, identisch. Die absolute Abweichung der erklärten Varianz pro Komponente beträgt im paarweisen Vergleich maximal 0.51 und die absolute Abweichung der erklärten Gesamtvarianz beträgt 0.45 Prozentpunkte. Darüber hinaus ist die Hierarchie der Hauptladungen, ausgerichtet an deren Absolutwerten, innerhalb der Komponenten nahezu identisch. Ausschließlich die Hierarchie der Items FQ-4 und FQ-2 auf der Komponente 1 (FQ) sowie diejenige der Items AK-1 und AK-2 auf der Komponente 6 (AK) sind

jeweils miteinander vertauscht. Die Differenz der absoluten Werte der 20 Hauptladungen beträgt im paarweisen Vergleich maximal .06. Die Nebenladung des Items EM-1 ist in beiden Komponentenladungsmatrizen jeweils mit einem Absolutwert von .31 auf der gleichen Komponente (5 = EA) lokalisiert. Eine zweite Nebenladung ist ausschließlich in der 1. Welle/B der lidA-Hauptstudie zu verzeichnen. Hierbei handelt es sich um das Item EA-3 (= .39), welches auf der Komponente 3 (EM) lädt. In beiden Komponentenladungsmatrizen sind keine doppelten Hauptladungen ($\geq .50$) zu beobachten.

Insgesamt werden die zuvor angeführten Analyseergebnisse als Bestätigung für die postulierte Dimensionalität, d. h. die Skalenstruktur der COPSOQ-II-Items (siehe Abbildung 3) interpretiert. Die theoretisch zusammengehörenden Items messen also tatsächlich jeweils dasselbe Konstrukt bzw. dieselbe Komponente (vgl. TF.5 – siehe Kapitel 4).

6.2.6 Nomologisches Netzwerk

Von den insgesamt 26 für das nomologische Netzwerk relevanten bivariaten Korrelationen zwischen den sechs COPSOQ-II-Skalen sowie der AF-Skala und den Einzelitems GB und GG für die 1. Welle/A der lidA-Hauptstudie entsprechen 26 Korrelationen dem Referenznetzwerk (siehe Abschnitt 4.6). Sämtliche 26 bivariaten Korrelationen sind zudem hoch signifikant.

Alle bivariaten Korrelationen, in welche die sechs COPSOQ-II-Skalen eingebunden sind, entsprechen den theoretischen Erwartungen: QA-Skala (fünf), EA-Skala (vier), EM-Skala (sechs), FQ-Skala (acht), HU-Skala (acht) und AK-Skala (sieben). Vor dem Hintergrund dieser Analyseergebnisse wird die TF.6-H.1 (siehe Abschnitt 4.6) für die 1. Welle/A der lidA-Hauptstudie als uneingeschränkt bestätigt bewertet.

Die bivariaten Korrelationskoeffizienten der sechs COPSOQ-II-Skalen sowie der AF-Skala und den Einzelitems GB und GG sind in der nachfolgenden Tabelle 29 dargestellt.

Tabelle 29: Nomologisches Netzwerk (lidA-Hauptstudie: 1. Welle/A)

	QA	EA	EM	FQ	HU	AK	AF	GB	GG
QA-Skala	---	---	-.21 ***	---	.39 ***	-.21 ***	.14 ***	-.12 ***	---
EA-Skala		---	.39 ***	---	.12 ***	---	---	-.12 ***	.10 ***
EM-Skala			---	.23 ***	.19 ***	---	.21 ***	-.17 ***	.16 ***
FQ-Skala				---	.29 ***	-.20 ***	.26 ***	-.27 ***	.16 ***
HU-Skala					---	-.12 ***	.14 ***	-.11 ***	.10 ***
AK-Skala						---	-.28 ***	.22 ***	-.21 ***
AF-Skala							---	---	---
GB								---	---
GG									---

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: Listenweises $n = 2\ 640$.

In Bezug auf die für das nomologische Netzwerk relevanten bivariaten Korrelationen zwischen den sechs COPSOQ-II-Skalen sowie der AF-Skala und den Einzelitems GB und GG der 1. Welle/B der lidA-Hauptstudie entsprechen 26 von insgesamt 26 Korrelationen dem Referenznetzwerk (siehe Abschnitt 4.6). Alle 26 Korrelationen sind hoch signifikant.

Demgemäß sind sämtliche bivariaten Korrelationen, in welche die sechs COPSOQ-II-Skalen eingebunden sind, erwartungskonform: QA-Skala (fünf), EA-Skala (vier), EM-Skala (sechs), FQ-Skala (acht), HU-Skala (acht) und AK-Skala (sieben). Die TF.6-H.1 (siehe Abschnitt 4.6) wird auf der Grundlage dieser Ergebnisse für die 1. Welle/B der lidA-Hauptstudie als uneingeschränkt bestätigt bewertet.

Die exakten Koeffizienten in Bezug auf die bivariaten Korrelationen zwischen den sechs COPSOQ-II-Skalen und der AF-Skala und den Einzelitems GB und GG sind in Tabelle 30 aufgeführt.

Tabelle 30: Nomologisches Netzwerk (lidA-Hauptstudie: 1. Welle/B)

	QA	EA	EM	FQ	HU	AK	AF	GB	GG
QA-Skala	---	---	-.26 ***	---	.39 ***	-.22 ***	.16 ***	-.10 ***	
EA-Skala		---	.45 ***	---	.15 ***	---	-.09 ***	.12 ***	
EM-Skala			---	.23 ***	.20 ***	---	.20 ***	-.19 ***	
FQ-Skala				---	.24 ***	-.22 ***	.21 ***	-.26 ***	
HU-Skala					---	-.15 ***	.10 ***	-.10 ***	
AK-Skala						---	-.26 ***	.22 ***	
AF-Skala							---	---	
GB								---	
GG									

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: Listenweises $n = 2\ 633$.

Bezüglich der sechs COPSOQ-II-Skalen sowie der AF-Skala und den Einzelitems GB und GG der 1. Welle/A und 1. Welle/B der lidA-Hauptstudie ist im Paarvergleich bei allen 26 erwartungskonformen Korrelationskoeffizienten eine sehr hohe Übereinstimmung festzustellen. Die geringfügigen Abweichungen erreichen bei einer Spannweite der absoluten Werte von .10 bis .39 (1. Welle/A) bzw. von .09 bis .45 (1. Welle/B) maximal einen Wert von .06. In beiden nomologischen Netzwerken werden jeweils alle bivariaten Korrelationen bestätigt.

Zusammenfassend werden diese Ergebnisse als Bestätigung dafür interpretiert, dass die COPSOQ-II-Skalen im Rahmen des lidA-CAPI dieselben Konstrukte (siehe Abschnitt 5.2) erfassen, die auch im Zuge ihrer Verwendung als „Papier-Bleistift-Verfahren“ erfasst werden (vgl. TF.6 – siehe Kapitel 4).

7. Diskussion

Die handlungsleitende Fragestellung der vorliegenden Arbeit lautete: Weisen die sechs COPSOQ-II-Skalen (QA-Skala, EA-Skala, EM-Skala, FQ-Skala, HU-Skala und AK-Skala) im CAPI der 1. Welle der lidA-Hauptstudie eine für inhaltliche statistische Analysen ausreichend hohe psychometrische Qualität auf? Zur Überprüfung dieser Fragestellung wurden zunächst psychometrische Kriterien bestimmt, deren quantitative Grenzwerte sowohl vor dem Hintergrund theoretischer Maßgaben als auch auf der Grundlage empirischer Vergleichsdaten bestimmt wurden. Die statistischen Analysen wurden in der Folge zunächst im Rahmen einer kleinen Stichprobe $n = 171$ (lidA-Machbarkeitsstudie) und anschließend anhand einer nach dem Zufallsprinzip in zwei etwa gleichgroße Stichproben (lidA-Hauptstudie: 1. Welle/A: $n = 2\,830$ und lidA-Hauptstudie: 1. Welle/B: $n = 2\,836$) geteilten großen Stichprobe vollzogen. Da die Ergebnisse der lidA-Machbarkeitsstudie eindeutig hypothesenkonforme Tendenzen ergaben und hier auch nicht mehr als Tendenzen erwartet wurden, werden nachfolgend ausschließlich die Analyseergebnisse zur der 1. Welle/A+B der lidA-Hauptstudie zusammengefasst. Wenn in diesem Zusammenhang kein besonderer Hinweis erfolgt, bestehen keine bedeutsamen Divergenzen hinsichtlich der Ergebnisse dieser beiden Stichproben (A+B). Die Ergebniszusammenfassung ist zunächst nach den sechs fokussierten COPSOQ-II-Skalen geordnet: (1) „Quantitative Anforderungen“ (siehe Abschnitt 7.1), (2) „Einfluss bei der Arbeit“ (siehe Abschnitt 7.2), (3) „Entwicklungsmöglichkeiten“ (siehe Abschnitt 7.3), (4) „Führungsqualität“ (siehe Abschnitt 7.4), (5) „Hilfe und Unterstützung von Kollegen“ (siehe Abschnitt 7.5) sowie (6) „Arbeit-Familie/Privatleben-Konflikt“ (siehe Abschnitt 7.6). Während in diesen Abschnitten lediglich die sechs Teilfragestellungen separat für die sechs COPSOQ-II-Skalen beantwortet werden, wird im nächstfolgenden Abschnitt (siehe Abschnitt 7.7) die handlungsleitende Fragestellung beantwortet und ein abschließendes Resümee gezogen. Hier wird zudem eine mögliche Maßnahme zur Optimierung der psychometrischen Qualität der sechs COPSOQ-II-Skalen vorgestellt. In dem letzten Abschnitt dieses Kapitels werden die Grenzen der vollzogenen Untersuchung sowie sinnvolle und praktikable Anschlussuntersuchungen diskutiert (siehe Abschnitt 7.8).

Zu den Stärken dieser Untersuchung zählen die Stichprobengrößen, die Heterogenität der Berufsklassen, die hohe Anzahl der Sample-Points und ihre bundesweite Verteilung sowie die Fokussierung auf zwei ältere Geburtsjahrgänge und die Repräsentativität innerhalb dieser Geburtsjahrgänge für die sozialversicherungspflichtig Beschäftigten gemäß der Beschäftigten-Historie des IAB zum 31. Dezember 2009 (vgl. Kersting et al., 2012).

7.1 Quantitative Anforderungen

Die empirische Spannweite der QA-Skala wie auch sämtlicher QA-Items entsprach der theoretischen Spannweite. Überdies wurden weder Boden- noch Deckeneffekte beobachtet, wobei die QA-Items einheitlich eingipflig und rechtsschief verteilt waren. Diese empirische Verteilung der Werte der QA-Skala sowie der QA-Items wird als hinreichender Beleg für ein akzeptables Differenzierungspotential bewertet (vgl. TF.1). Im Vergleich der sechs COPSOQ-II-Skalen untereinander wiesen die empirischen Werte der QA-Skala die größte Nähe zur Normalverteilung auf. Jedoch ergab der K-S-A und der S-W-T (sowie die Augenscheinkontrolle) für die QA-Items und die QA-Skala jeweils eine statistisch hoch signifikante Abweichung von der Normalverteilung (vgl. TF.1). Demgemäß wurden für die folgenden psychometrischen Item- und Skalenanalysen ausschließlich verteilungsfreie Verfahren verwendet.

Das AM der QA-Skala (= 43.77 bzw. 45.05) und der vier QA-Items (= 34.22 bis 52.44) war innerhalb eines Wertebereiches von 25.00 bis 75.00 lokalisiert. Der Md der QA-Skala (= 43.75) und der vier QA-Items (= 25.00 bis 50.00) lag ebenfalls innerhalb dieses Wertebereiches. Sämtliche Lagemaße (AM und Md) waren sehr nah an dem Idealwert von 50.00 und insgesamt eher kleiner, d. h. die Items waren schwieriger. Die SD der QA-Skala (= 22.77 bzw. 22.83) und der QA-Items (= 27.06 bis 30.18) betrug mindestens 20.00. Der IQR der QA-Skala (= 31.25 bzw. 37.50) und der QA-Items (= 25.00 bis 50.00) betrug mindestens 25.00. Diese hohen Ausprägungen der beiden Streuungsmaße (SD und IQR) unterstreichen noch einmal das zuvor benannte hohe Differenzierungspotential dieser Skala. Auf der Grundlage dieser Ergebnisse wird die Schwierigkeit der QA-Skala wie auch der QA-Items insgesamt als akzeptabel bewertet (vgl. TF.2). Ferner sind diese Ergebnisse konsistent zu denen internationaler Studien, in welchen die QA-Skala mit teilweise abweichenden Operationalisierungen im Rahmen anonymisierter schriftlicher Mitarbeiterbefragungen erfasst wurde: COPSOQ-I-Medium in Dänemark: AM = 46.8 und SD = 18.6 (Kristensen et al., 2005), COPSOQ-II-Long in Dänemark: AM = 40.2 und SD = 20.5 (Pejtersen et al., 2010) sowie COPSOQ-II-Short in Frankreich: AM = 53 und SD = 21 (Dupret et al., 2012).

Alle QA-Items erreichten die geforderte Mindesthöhe sowohl für die nicht-korrigierte (= .76 bis .85) als auch für die korrigierte Trennschärfe (= .59 bis .72). Sämtliche vier Items der QA-Skala weisen hiernach eine akzeptable Trennschärfe auf (vgl. TF.3). Außerdem sind diese Ergebnisse konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): korrigierte Trennschärfe (COPSOQ-II-Long) = .58 bis .73 (Produkt-Moment-Korrelationen nach Pearson). Im Kontext der vorliegenden Untersuchung ist dasjenige Item mit der höchsten „Repräsentativität“ (Bortz & Döring, 2009, S. 219) für die QA-Skala das Item QA-2: „Wie oft kommt es vor, dass Sie nicht genügend Zeit haben, alle Ihre Aufgaben zu erledigen?“ (siehe Abschnitt 5.2.1). Dieses Item repräsentiert die QA-Skala

auch in den vier empirischen Vergleichsstudien am besten (vgl. Tabelle 4). Vor diesem Hintergrund und unter Bezugnahme auf die Inhalte der anderen drei QA-Items, misst die QA-Skala Folgendes:

Die QA-Skala erfasst die subjektive Bewertung des Verhältnisses von der Menge der Arbeitsaufgaben zu der für diese zur Verfügung stehenden Arbeitszeit.

Das AM der I-I-K (= .54 bzw. .55) sowie die einzelnen I-I-K (= .45 bis .64) der QA-Items waren innerhalb eines Spektrums von .30 bis .80 lokalisiert. Die QA-Skala weist hiernach eine akzeptable Homogenität auf (vgl. TF.4), so dass von einer hinreichenden inhaltlichen Konvergenz der QA-Items ausgegangen wird. Diese Ergebnisse sind überdies konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): I-I-K (COPSOQ-II-Long) = .45 bis .65 (Produkt-Moment-Korrelationen nach Pearson).

Die Hauptkomponentenanalyse der 20 COPSOQ-II-Items ergab eine Komponente (QA), auf welcher ausschließlich Hauptladungen der QA-Items (= .71 bis .85) zu verzeichnen waren. Zudem wies keines der QA-Items Hauptladungen ($\geq .50$) oder Nebenladungen ($\geq .30$ und $< .50$) auf einer zweiten Komponente auf. Die COPSOQ-II-Items der QA-Skala bilden folglich die postulierte Skalenstruktur ab (vgl. TF.5), d. h. sie messen dasselbe latente Konstrukt bzw. dieselbe Komponente und unterscheiden sich hinreichend von den Items der fünf anderen COPSOQ-II-Skalen. Hiernach besteht eine akzeptable faktorielle Validität der QA-Skala (z. B. Amelang & Schmidt-Atzert, 2006).

Die Vorzeichen der fünf bivariaten Korrelationen, in welche die QA-Skala im Rahmen des nomologischen Netzwerkes eingebunden ist, entsprachen denen des Referenznetzwerkes. In Bezug auf die QA-Skala ist daher ein empirisches nomologisches Referenznetzwerk adäquat replizierbar (vgl. TF.6). Hiernach korreliert die QA-Skala („Quantitative Anforderungen“) positiv mit der AK-Skala („Arbeit-Familie/Privatleben-Konflikt“) sowie dem Einzelitem GB („Gedanke an Berufsaufgabe“) und negativ mit der FQ-Skala („Führungsqualität“), der AF-Skala („Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen“) sowie dem Einzelitem GG („Gegenwärtiger Gesundheitszustand“). Die stärkste positive Korrelation ist mit der AK-Skala und die stärkste negative Korrelation ist mit der FQ-Skala zu verzeichnen. Diese inhaltlichen Zusammenhänge sind erwartungskonform und werden demgemäß als zusätzlicher Hinweis auf eine akzeptable Konstruktvalidität der QA-Skala interpretiert.

7.2 Einfluss bei der Arbeit

Die empirische Spannweite der EA-Skala wie auch sämtlicher EA-Items entsprach der theoretischen Spannweite. Überdies wurden weder Boden- noch Deckeneffekte beobachtet, wobei die Items EA-1 und EA-2 einheitlich eingipflig und rechtsschief verteilt waren und das Item EA-3 zweigipflig (= 0.00 und 75.00) verteilt war. Diese empirische Verteilung der Werte der EA-Skala sowie der EA-Items wird als hinreichender Beleg für ein akzeptables Differenzierungspotential bewertet (vgl. TF.1). Der K-S-A und der S-W-T (sowie die Augenscheinkontrolle) ergaben für die EA-Items und die EA-Skala jeweils eine statistisch hoch signifikante Abweichung von der Normalverteilung (vgl. TF.1). Demnach wurden für die folgenden psychometrischen Item- und Skalenanalysen ausschließlich verteilungsfreie Verfahren verwendet.

Das AM der EA-Skala (= 37.35 bzw. 37.81) und der drei EA-Items (= 29.00 bis 51.49) war innerhalb eines Wertebereiches von 25.00 bis 75.00 lokalisiert. Der Md der EA-Skala (= 33.33) und der drei EA-Items (= 25.00 bis 50.00) lag ebenfalls innerhalb dieses Wertebereiches. Fast alle Lagemaße (AM und Md) wichen deutlich vom Idealwert (= 50.00) ab und waren in der Tendenz eindeutig kleiner, d. h. die Items waren schwieriger. Die SD der EA-Skala (= 27.46 bzw. 27.63) und der EA-Items (= 32.23 bis 37.94) betrug mindestens 20.00. Der IQR der EA-Skala (= 41.66) und der EA-Items (= 50.00 bis 75.00) betrug mindestens 25.00. Diese hohen Ausprägungen der beiden Streuungsmaße (SD und IQR) unterstreichen noch einmal das zuvor benannte hohe Differenzierungspotential dieser Skala. Auf der Grundlage dieser Ergebnisse wird die Schwierigkeit der EA-Skala wie auch der EA-Items insgesamt als akzeptabel bewertet (vgl. TF.2). Allerdings sind diese Ergebnisse nicht konsistent zu denen internationaler Studien, in welchen die EA-Skala mit teilweise abweichenden Operationalisierungen im Rahmen anonymisierter schriftlicher Mitarbeiterbefragungen erfasst wurde: COPSQ-I-Medium in Dänemark: AM = 55.4 und SD = 24.2 (Kristensen et al., 2005), COPSQ-II-Long in Dänemark: AM = 49.8 und SD = 21.2 (Pejtersen et al., 2010) sowie COPSQ-II-Short in Frankreich: AM = 48 und SD = 21 (Dupret et al., 2012).

Alle EA-Items erreichten die geforderte Mindesthöhe sowohl für die nicht-korrigierte (= .74 bis .81) als auch für die korrigierte Trennschärfe (= .44 bis .50). Sämtliche Items der EA-Skala weisen hiernach eine akzeptable Trennschärfe auf (vgl. TF.3). Außerdem sind diese Ergebnisse konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): korrigierte Trennschärfe (COPSQ-II-Long) = .43 bis .59 (Produkt-Moment-Korrelationen nach Pearson). Im Kontext der vorliegenden Untersuchung ist dasjenige Item mit der höchsten „Repräsentativität“ (Bortz & Döring, 2009, S. 219) für die EA-Skala mit Blick auf die korrigierte Trennschärfe das Item EA-2: „Haben Sie Einfluss auf die Menge der Arbeit, die Ihnen übertragen wird?“ (siehe Abschnitt 5.2.2). Hinsichtlich der vier empirischen Vergleichsstudien wird die QA-Skala mit Blick auf die korrigierte Trennschärfe besser durch

das Item EA-1 repräsentiert (vgl. Tabelle 4): „Haben Sie Einfluss darauf, mit wem Sie arbeiten?“. Vor diesem Hintergrund und unter Bezugnahme auf das dritte EA-Item, misst die EA-Skala Folgendes:

Die EA-Skala erfasst die subjektive Bewertung der Häufigkeit des Einflusses auf die Wahl vor allem der Menge und der Kollegen, aber auch des Inhaltes der Arbeit.

Das AM der I-I-K (= .41 bzw. .39) sowie die einzelnen I-I-K (= .35 bis .42) der EA-Items waren innerhalb eines Spektrums von .30 bis .80 lokalisiert. Die EA-Skala weist also eine akzeptable Homogenität auf (vgl. TF.4), so dass von einer hinreichenden inhaltlichen Konvergenz der EA-Items ausgegangen wird. Diese Ergebnisse sind überdies konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): I-I-K (COPSOQ-II-Long) = .31 bis .49 (Produkt-Moment-Korrelationen nach Pearson).

Die Hauptkomponentenanalyse der 20 COPSOQ-II-Items ergab eine Komponente (EA), auf der ausschließlich Hauptladungen der EA-Items (= .65 bis .78) zu verzeichnen waren. Das Item EM-1 wies in beiden Stichproben (A+B) der 1. Welle der lidA-Hauptstudie eine Nebenladung (= .31) auf der Komponente EA auf. Keines der EA-Items war durch eine Hauptladung ($\geq .50$) auf einer zweiten Komponente gekennzeichnet. Das Item EA-3 wies in der 1. Welle/B der lidA-Hauptstudie eine Nebenladung (= .39) auf der Komponente EM auf. Die COPSOQ-II-Items der EA-Skala bilden im Wesentlichen die postulierte Skalenstruktur ab (vgl. TF.5), d. h. sie messen dasselbe latente Konstrukt bzw. dieselbe Komponente und unterscheiden sich weitgehend von den Items der fünf anderen COPSOQ-II-Skalen. Hiernach besteht eine akzeptable faktorielle Validität der EA-Skala (z. B. Amelang & Schmidt-Atzert, 2006).

Die Vorzeichen der vier bivariaten Korrelationen, in welche die EA-Skala im Rahmen des nomologischen Netzwerkes eingebunden ist, entsprachen denen des Referenznetzwerkes. In Bezug auf die EA-Skala ist daher ein empirisches nomologisches Referenznetzwerk adäquat replizierbar (vgl. TF.6). Hiernach korreliert die EA-Skala („Einfluss bei der Arbeit“) positiv mit der EM-Skala („Entwicklungsmöglichkeiten“), der HU-Skala („Hilfe und Unterstützung von Kollegen“) sowie dem Einzelitem GG („Gegenwärtiger Gesundheitszustand“) und negativ mit dem Einzelitem GB („Gedanke an Berufsaufgabe“). Die stärkste positive Korrelation ist mit der EM-Skala und die stärkste negative Korrelation ist mit dem Einzelitem GB zu verzeichnen. Diese inhaltlichen Zusammenhänge sind erwartungskonform und werden demgemäß als weiterer Hinweis auf eine akzeptable Konstruktvalidität der EA-Skala interpretiert.

7.3 Entwicklungsmöglichkeiten

Die empirische Spannweite der EM-Skala wie auch sämtlicher EM-Items entsprach der theoretischen Spannweite. Überdies wurden weder Boden- noch Deckeneffekte beobachtet, wobei die EM-Items einheitlich eingipflig und linksschief verteilt waren. Diese empirische Verteilung der Werte der EM-Skala sowie der EM-Items wird als hinreichender Beleg für ein akzeptables Differenzierungspotential bewertet (vgl. TF.1). Der K-S-A und der S-W-T (sowie die Augenscheinkontrolle) ergaben für die EM-Items und die EM-Skala jeweils eine statistisch hoch signifikante Abweichung von der Normalverteilung (vgl. TF.1). Demgemäß wurden für die nachfolgenden psychometrischen Item- und Skalenanalysen ausschließlich verteilungsfreie Verfahren verwendet.

Das AM der EM-Skala (= 66.69 bzw. 67.54) und drei von vier EM-Items (= 56.40 bis 74.24) war innerhalb eines Wertebereiches von 25.00 bis 75.00 lokalisiert. Das Item EM-1 (= 76.08 bzw. 76.07) lag außerhalb dieses Wertebereiches. Der Md der EM-Skala (= 68.75) und der vier EM-Items (= 50.00 bis 75.00) wiederum lag innerhalb des Wertebereiches von 25.00 bis 75.00. Die deutliche Mehrheit der Lagemaße (AM und Md) waren größer als der Idealwert (= 50.00), d. h. die Items waren leichter. Die SD der EM-Skala (= 20.67 bzw. 21.30) und der EM-Items (= 24.40 bis 28.77) betrug mindestens 20.00. Der IQR der EM-Skala (= 25.00) und der EM-Items (= 25.00 bis 50.00) betrug mindestens 25.00. Diese hohen Ausprägungen der beiden Streuungsmaße (SD und IQR) unterstreichen noch einmal das zuvor benannte hohe Differenzierungspotential dieser Skala. Auf der Grundlage dieser Ergebnisse wird die Schwierigkeit der EM-Skala wie auch der QA-Items insgesamt als akzeptabel bewertet (vgl. TF.2). Ferner sind diese Ergebnisse konsistent zu denen internationaler Studien, in welchen die EM-Skala mit teilweise abweichenden Operationalisierungen im Rahmen anonymisierter schriftlicher Mitarbeiterbefragungen erfasst wurde: COPSOQ-I-Medium in Dänemark: AM = 72.4 und SD = 19.1 (Kristensen et al., 2005), COPSOQ-II-Long in Dänemark: AM = 65.9 und SD = 17.6 (Pejtersen et al., 2010) sowie COPSOQ-II-Short in Frankreich: AM = 70 und SD = 22 (Dupret et al., 2012).

Alle EM-Items erreichten die geforderte Mindesthöhe sowohl für die nicht-korrigierte (= .64 bis .85) als auch für die korrigierte Trennschärfe (= .42 bis .70). Die vier Items der EM-Skala weisen hiernach eine akzeptable Trennschärfe auf (vgl. TF.3). Außerdem sind diese Ergebnisse konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): korrigierte Trennschärfe (COPSOQ-II-Long) = .47 bis .70 (Produkt-Moment-Korrelationen nach Pearson). Im Kontext der vorliegenden Untersuchung ist dasjenige Item mit der höchsten „Repräsentativität“ (Bortz & Döring, 2009, S. 219) für die EM-Skala das Item EM-4: „Ermöglicht es Ihre Arbeit, Ihre Fertigkeiten weiterzuentwickeln?“ (siehe Abschnitt 5.2.3). Dieses Item repräsentiert die EM-Skala auch in den vier empirischen Vergleichsstudien am besten (vgl. Tabelle 4). Vor diesem Hintergrund und unter Bezugnahme auf die Inhalte der drei anderen EM-Items misst die EM-Skala Folgendes:

Die EM-Skala erfasst die subjektive Bewertung der Möglichkeit, infolge der Durchführung der Arbeit eigene Fertigkeiten und Wissen weiterzuentwickeln und neue Dinge zu erlernen.

Das AM der I-I-K (= .47 bzw. .49) sowie die einzelnen I-I-K (= .33 bis .68) der EM-Items waren innerhalb eines Spektrums von .30 bis .80 lokalisiert. Die EM-Skala weist hiernach eine akzeptable Homogenität auf (vgl. TF.4), so dass von einer hinreichenden inhaltlichen Konvergenz der EM-Items ausgegangen wird. Diese Ergebnisse sind überdies konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): I-I-K (COPSOQ-II-Long) = .34 bis .70 (Produkt-Moment-Korrelationen nach Pearson).

Die Hauptkomponentenanalyse der 20 COPSOQ-II-Items ergab eine Komponente (EM), auf welcher ausschließlich Hauptladungen der EM-Items (= .56 bis .84) zu verzeichnen waren. Das Item EA-3 wies in der 1. Welle/B der lidA-Hauptstudie eine Nebenladung (= .39) auf der Komponente EM auf. Keines der EM-Items war durch eine Hauptladung ($\geq .50$) auf einer zweiten Komponente gekennzeichnet. Das Item EM-1 wies allerdings in beiden Stichproben (A+B) der 1. Welle der lidA-Hauptstudie eine Nebenladung (= .31) auf der Komponente EA auf. Die COPSOQ-II-Items der EM-Skala bilden im Wesentlichen die postulierte Skalenstruktur ab (vgl. TF.5), d. h. sie messen dasselbe latente Konstrukt bzw. dieselbe Komponente und unterscheiden sich weitgehend von den Items der fünf anderen COPSOQ-II-Skalen. Es besteht eine akzeptable faktorielle Validität der EM-Skala (z. B. Amelang & Schmidt-Atzert, 2006).

Die Vorzeichen der sechs bivariaten Korrelationen, in welche die EM-Skala im Rahmen des nomologischen Netzwerkes eingebunden ist, entsprachen denen des Referenznetzwerkes. In Bezug auf die EM-Skala ist daher ein empirisches nomologisches Referenznetzwerk adäquat replizierbar (vgl. TF.6). Hiernach korreliert die EM-Skala („Entwicklungsmöglichkeiten“) positiv mit der EA-Skala („Einfluss bei der Arbeit“), der FQ-Skala („Führungsqualität“), der HU-Skala („Hilfe und Unterstützung von Kollegen“), der AF-Skala („Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen“) sowie dem Einzelitem GG („Gegenwärtiger Gesundheitszustand“) und negativ mit dem Einzelitem GB („Gedanke an Berufsaufgabe“). Die stärkste positive Korrelation ist mit der EA-Skala und die stärkste negative Korrelation ist mit dem Einzelitem GB zu verzeichnen. Diese inhaltlichen Zusammenhänge sind erwartungskonform und werden demgemäß als zusätzlicher Hinweis auf eine akzeptable Konstruktvalidität der EM-Skala interpretiert.

7.4 Führungsqualität

Die empirische Spannweite der FQ-Skala wie auch sämtlicher FQ-Items entsprach der theoretischen Spannweite. Überdies wurden weder Boden- noch Deckeneffekte beobachtet, wobei die FQ-Items einheitlich eingipflig und linksschief verteilt waren. Diese empirische Verteilung der Werte der FQ-Skala sowie der FQ-Items wird als hinreichender Beleg für ein akzeptables Differenzierungspotential bewertet (vgl. TF.1). Der K-S-A und der S-W-T (sowie die Augenscheinkontrolle) ergaben für die FQ-Items und die FQ-Skala jeweils eine statistisch hoch signifikante Abweichung von der Normalverteilung (vgl. TF.1). Demgemäß wurden für die nachfolgenden psychometrischen Item- und Skalenanalysen ausschließlich verteilungsfreie Verfahren verwendet.

Das AM der FQ-Skala (= 53.60 bzw. 53.71) sowie der vier FQ-Items (= 50.31 bis 57.92) war innerhalb eines Wertebereiches von 25.00 bis 75.00 lokalisiert. Der Md der FQ-Skala (= 56.25) und der vier FQ-Items (= 50.00) lag ebenfalls innerhalb dieses Wertebereiches. Sämtliche Lagemaße (AM und Md) waren sehr nah an dem Idealwert von 50.00 und insgesamt eher größer, d. h. die Items waren leichter. Die SD der FQ-Skala (= 22.96 bzw. 23.00) und der FQ-Items (= 27.55 bis 28.87) betrug mindestens 20.00. Der IQR der FQ-Skala (= 31.25) und der FQ-Items (= 25.00 bis 50.00) betrug mindestens 25.00. Diese hohen Ausprägungen der beiden Streuungsmaße (SD und IQR) unterstreichen noch einmal das zuvor benannte hohe Differenzierungspotential dieser Skala. Auf der Grundlage dieser Ergebnisse wird die Schwierigkeit der FQ-Skala wie auch der FQ-Items insgesamt als akzeptabel bewertet (vgl. TF.2). Ferner sind diese Ergebnisse konsistent zu denen internationaler Studien, in welchen die FQ-Skala mit teilweise abweichenden Operationalisierungen im Rahmen anonymisierter schriftlicher Mitarbeiterbefragungen erfasst wurde: COPSOQ-I-Medium in Dänemark: AM = 55.5 und SD = 21.4 (Kristensen et al., 2005), COPSOQ-II-Long in Dänemark: AM = 55.3 und SD = 21.1 (Pejtersen et al., 2010) sowie COPSOQ-II-Short in Frankreich: AM = 62 und SD = 28 (Dupret et al., 2012).

Alle FQ-Items erreichten die geforderte Mindesthöhe sowohl für die nicht-korrigierte (= .77 bis .84) als auch für die korrigierte Trennschärfe (= .60 bis .71). Alle vier Items der FQ-Skala weisen hiernach eine akzeptable Trennschärfe auf (vgl. TF.3). Außerdem sind diese Ergebnisse konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): korrigierte Trennschärfe (COPSOQ-II-Long) = .73 bis .79 (Produkt-Moment-Korrelationen nach Pearson). Im Kontext der vorliegenden Untersuchung ist dasjenige Item mit der höchsten „Repräsentativität“ (Bortz & Döring, 2009, S. 219) für die FQ-Skala das Item FQ-2: „In welchem Maße misst Ihr/e unmittelbare/r Vorgesetzte/r der Arbeitszufriedenheit einen hohen Stellenwert bei?“ (siehe Abschnitt 5.2.4). Dieses Item repräsentiert die FQ-Skala auch in den vier empirischen Vergleichsstudien am besten (vgl. Tabelle 4). Vor diesem Hintergrund und unter Bezugnahme auf die Inhalte der anderen FQ-Items, misst die FQ-Skala Folgendes:

Die FQ-Skala erfasst die subjektive Bewertung des Ausmaßes, in welchem der unmittelbare Vorgesetzte für gute Entwicklungsmöglichkeiten sowie gute Konfliktlösungen und vor allem für eine hohe Arbeitszufriedenheit sorgt.

Das AM der I-I-K (= .55 bzw. .57) sowie die einzelnen I-I-K (= .49 bis .64) der FQ-Items waren innerhalb eines Spektrums von .30 bis .80 lokalisiert. Die FQ-Skala weist also eine akzeptable Homogenität auf (vgl. TF.4), so dass von einer hinreichenden inhaltlichen Konvergenz der FQ-Items ausgegangen werden kann. Diese Ergebnisse sind überdies weitestgehend konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): I-I-K (COPSOQ-II-Long) = .60 bis .71 (Produkt-Moment-Korrelationen nach Pearson).

Die Hauptkomponentenanalyse der 20 COPSOQ-II-Items ergab eine Komponente (FQ), auf welcher ausschließlich Hauptladungen der FQ-Items (= .76 bis .83) zu verzeichnen waren. Zudem wies keines der FQ-Items Hauptladungen ($\geq .50$) oder Nebenladungen ($\geq .30$ und $< .50$) auf einer zweiten Komponente auf. Die COPSOQ-II-Items der FQ-Skala bilden folglich die postulierte Skalenstruktur ab (vgl. TF.5), d. h. sie messen dasselbe latente Konstrukt bzw. dieselbe Komponente und unterscheiden sich hinreichend von den Items der fünf anderen COPSOQ-II-Skalen. Hiernach besteht akzeptable faktorielle Validität der FQ-Skala (z. B. Amelang & Schmidt-Atzert, 2006).

Die Vorzeichen der sieben bivariaten Korrelationen, in welche die FQ-Skala im Rahmen des nomologischen Netzwerkes eingebunden ist, entsprachen denen des Referenznetzwerkes. In Bezug auf die FQ-Skala ist daher ein empirisches nomologisches Referenznetzwerk adäquat replizierbar (vgl. TF.6). Hiernach korreliert die FQ-Skala („Führungsqualität“) positiv mit der EM-Skala („Entwicklungsmöglichkeiten“), der HU-Skala („Hilfe und Unterstützung von Kollegen“), der AF-Skala („Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen“) sowie dem Einzelitem GG („Gegenwärtiger Gesundheitszustand“) und negativ mit der QA-Skala („Quantitative Anforderungen“), der AK-Skala („Arbeit-Familie/Privatleben-Konflikt“) sowie dem Einzelitem GB („Gedanke an Berufsaufgabe“). Die stärkste positive Korrelation ist mit der HU-Skala und die stärkste negative Korrelation ist mit dem Einzelitem GB zu verzeichnen. Diese inhaltlichen Zusammenhänge sind erwartungskonform und werden demgemäß als weiterer Hinweis auf eine akzeptable Konstruktvalidität der FQ-Skala interpretiert.

7.5 Hilfe und Unterstützung von Kollegen

Die empirische Spannweite der HU-Skala wie auch sämtlicher HU-Items entsprach der theoretischen Spannweite. Überdies wurden weder Boden- noch Deckeneffekte beobachtet, wobei die HU-Items einheitlich eingipflig und linksschief verteilt waren. Diese empirische Verteilung der Werte der HU-Skala sowie der HU-Items wird als hinreichender Beleg für ein akzeptables Differenzierungspotential bewertet (vgl. TF.1). Der K-S-A und der S-W-T (sowie die Augenscheinkontrolle) ergaben für die HU-Items und die HU-Skala jeweils eine statistisch hoch signifikante Abweichung von der Normalverteilung (vgl. TF.1). Demgemäß wurden für die nachfolgenden psychometrischen Item- und Skalenanalysen ausschließlich verteilungsfreie Verfahren verwendet.

Das AM der HU-Skala (= 64.70 bzw. 65.22) und der drei HU-Items (= 50.05 bis 73.25) war innerhalb eines Wertebereiches von 25.00 bis 75.00 lokalisiert. Der Md der HU-Skala (= 66.67) und der drei HU-Items (= 50.00 bis 75.00) lag ebenfalls innerhalb dieses Wertebereiches. Sämtliche Lagemaße (AM und Md) waren oberhalb des Idealwertes von 50.00, d. h. die Items waren leichter. Die SD der HU-Skala (= 21.73 bzw. 21.78) und der HU-Items (= 25.00 bis 29.55) betrug mindestens 20.00. Der IQR der HU-Skala (= 33.33) und der HU-Items (= 50.00) betrug mindestens 25.00. Diese hohen Ausprägungen der beiden Streuungsmaße (SD und IQR) unterstreichen noch einmal das zuvor benannte hohe Differenzierungspotential dieser Skala. Auf der Grundlage dieser Ergebnisse wird die Schwierigkeit der HU-Skala wie auch der HU-Items insgesamt als akzeptabel bewertet (vgl. TF.2). Ferner sind diese Ergebnisse konsistent zu denen internationaler Studien, in welchen die HU-Skala mit teilweise abweichenden Operationalisierungen im Rahmen anonymisierter schriftlicher Mitarbeiterbefragungen erfasst wurde: COPSOQ-I-Medium in Dänemark: AM = 68.3 und SD = 20.4 (Kristensen et al., 2005) und COPSOQ-II-Long in Dänemark: AM = 57.3 und SD = 19.7 (Pejtersen et al., 2010).

Alle HU-Items erreichten die geforderte Mindesthöhe sowohl für die nicht-korrigierte (= .76 bis .80) als auch für die korrigierte Trennschärfe (= .41 bis .55). Alle drei Items der HU-Skala weisen hiernach eine akzeptable Trennschärfe auf (vgl. TF.3). Außerdem sind diese Ergebnisse konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): korrigierte Trennschärfe (COPSOQ-II-Long) = .48 bis .56 (Produkt-Moment-Korrelationen nach Pearson). Im Kontext der vorliegenden Untersuchung ist dasjenige Item mit der höchsten „Repräsentativität“ (Bortz & Döring, 2009, S. 219) für die HU-Skala das Item HU-2: „Wie oft sind Ihre Kollegen bereit, sich Ihre Arbeitsprobleme anzuhören?“ (siehe Abschnitt 5.2.1). Hinsichtlich der vier empirischen Vergleichsstudien repräsentiert dieses Item die HU-Skala in vergleichbarer Weise wie das Item HU-1: „Wie oft erhalten Sie Hilfe und Unterstützung von Ihren Kollegen?“ (vgl. Tabelle 4). Vor diesem Hintergrund und unter Bezugnahme auf die Inhalte des dritten HU-Items, misst die HU-Skala Folgendes:

Die HU-Skala erfasst die subjektive Bewertung der Häufigkeit der Hilfe und Unterstützung sowie der Bereitschaft von Kollegen, sich die eigenen Arbeitsprobleme anzuhören und zu reflektieren.

Das AM der I-I-K (= .43 bzw. .44) sowie die einzelnen I-I-K (= .35 bis .58) der HU-Items waren innerhalb eines Spektrums von .30 bis .80 lokalisiert. Die HU-Skala weist hiernach eine akzeptable Homogenität auf (vgl. TF.4), so dass von einer hinreichenden inhaltlichen Konvergenz der HU-Items ausgegangen wird. Diese Ergebnisse sind überdies konsistent zu denen der dänischen Validierungsstudie (NRCWE, 2007a): I-I-K (COPSOQ-II-Long) = .39 bis .49 (Produkt-Moment-Korrelationen nach Pearson).

Die Hauptkomponentenanalyse der 20 COPSOQ-II-Items ergab eine Komponente (HU), auf der ausschließlich Hauptladungen der HU-Items (= .72 bis .81) zu verzeichnen waren. Zudem wies keines der HU-Items Hauptladungen ($\geq .50$) oder Nebenladungen ($\geq .30$ und $< .50$) auf einer zweiten Komponente auf. Die COPSOQ-II-Items der HU-Skala bilden folglich die postulierte Skalenstruktur ab (vgl. TF.5), d. h. sie messen dasselbe latente Konstrukt bzw. dieselbe Komponente und unterscheiden sich hinreichend von den Items der fünf anderen COPSOQ-II-Skalen. Es besteht akzeptable faktorielle Validität der HU-Skala (z. B. Amelang & Schmidt-Atzert, 2006).

Die Vorzeichen der sieben bivariaten Korrelationen, in welche die HU-Skala im Rahmen des nomologischen Netzwerkes eingebunden ist, entsprachen denen des Referenznetzwerkes. In Bezug auf die HU-Skala ist daher ein empirisches nomologisches Referenznetzwerk adäquat replizierbar (vgl. TF.6). Hiernach korreliert die HU-Skala („Hilfe und Unterstützung von Kollegen“) positiv mit der EA-Skala („Einfluss bei der Arbeit“), der EM-Skala („Entwicklungsmöglichkeiten“), der FQ-Skala („Führungsqualität“), der AF-Skala („Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen“) sowie dem Einzelitem GG („Gegenwärtiger Gesundheitszustand“) und negativ mit der AK-Skala („Arbeit-Familie/Privatleben-Konflikt“) sowie dem Einzelitem GB („Gedanke an Berufsaufgabe“). Die stärkste positive Korrelation ist mit der FQ-Skala und die stärkste negative Korrelation ist mit dem Einzelitem GB zu verzeichnen. Diese inhaltlichen Zusammenhänge sind erwartungskonform und werden demgemäß als ein zusätzlicher Hinweis auf eine akzeptable Konstruktvalidität der HU-Skala interpretiert.

7.6 Arbeit-Familie/Privatleben-Konflikt

Die empirische Spannweite der AK-Skala und aller AK-Items entsprach der theoretischen Spannweite. Überdies wurden weder Boden- noch Deckeneffekte beobachtet, wobei die AK-Items einheitlich eingipflig und rechtsschief verteilt waren. Diese empirische Verteilung der Werte der AK-Skala sowie der AK-Items wird als hinreichender Beleg für ein akzeptables Differenzierungspotential bewertet (vgl. TF.1). Im Vergleich der sechs COPSOQ-II-Skalen untereinander wiesen die empirischen Werte der AK-Skala die größte Abweichung von der Normalverteilung auf. Demgemäß ergab der K-S-A und der S-W-T (sowie die Augenscheinkontrolle) für die AK-Items und die AK-Skala jeweils eine statistisch hoch signifikante Abweichung von der Normalverteilung (vgl. TF.1). Aufgrund dessen wurden für die folgenden psychometrischen Item- und Skalenanalysen ausschließlich verteilungsfreie Verfahren verwendet.

Das AM der AK-Skala (= 35.28 bzw. 35.32) und der zwei AK-Items (= 34.87 bis 35.78) war innerhalb eines Wertebereiches von 25.00 bis 75.00 lokalisiert. Der Md der AK-Skala (= 33.00) und der zwei AK-Items (= 33.00) lag ebenfalls innerhalb dieses Wertebereiches. Sämtliche Lagemaße (AM und Md) waren kleiner als der Idealwert von 50.00, d. h. die Items waren schwieriger. Die SD der AK-Skala (= 27.61 bzw. 28.51) und der AK-Items (= 29.30 bis 30.20) betrug mindestens 20.00. Der IQR der AK-Skala (= 33.50 bzw. 50.00) und der AK-Items (= 67.00) betrug mindestens 25.00. Diese hohen Ausprägungen der beiden Streuungsmaße (SD und IQR) unterstreichen noch einmal das zuvor benannte hohe Differenzierungspotential dieser Skala. Auf der Grundlage dieser Ergebnisse wird die Schwierigkeit der AK-Skala wie auch der AK-Items insgesamt als akzeptabel bewertet (vgl. TF.2). Ferner sind diese Ergebnisse konsistent zu denen internationaler Studien, in welchen die AK-Skala mit teilweise abweichenden Operationalisierungen im Rahmen anonymisierter schriftlicher Mitarbeiterbefragungen erfasst wurde: COPSOQ-II-Long in Dänemark: AM = 33.5 und SD = 24.3 (Pejtersen et al., 2010) sowie COPSOQ-II-Short in Frankreich: AM = 37 und SD = 22 (Dupret et al., 2012).

Die Korrelation der beiden AK-Items erreichte die geforderte Mindesthöhe (= .77 bzw. .79). Von einer Trennschärfe der AK-Items kann in diesem Kontext aufgrund der zu geringen Itemanzahl nicht gesprochen werden (vgl. TF.3). Vielmehr fällt die Trennschärfe hier mit der Homogenität (vgl. TF.4) zusammen, so dass von einer hinreichenden inhaltlichen Konvergenz der AK-Items ausgegangen wird. Diese Ergebnisse differieren leicht von denen der dänischen Validierungsstudie (NRCWE, 2007a): I-I-K (COPSOQ-II-Long) = .39 bis .69 (Produkt-Moment-Korrelationen nach Pearson). Unter Bezugnahme auf die Inhalte beider AK-Items, misst die AK-Skala Folgendes:

Die AK-Skala erfasst die subjektive Bewertung des Ausmaßes negativer Auswirkungen auf das Privatleben infolge der Inanspruchnahme von Energie und Zeit durch die Arbeit.

Die Hauptkomponentenanalyse der 20 COPSOQ-II-Items ergab eine Komponente (AK), auf welcher ausschließlich Hauptladungen der AK-Items (= .90 bis .91) zu verzeichnen waren. Zudem wies keines der AK-Items Hauptladungen ($\geq .50$) oder Nebenladungen ($\geq .30$ und $< .50$) auf einer zweiten Komponente auf. Die COPSOQ-II-Items der AK-Skala bilden folglich die postulierte Skalenstruktur ab (vgl. TF.5), d. h. sie messen dasselbe latente Konstrukt bzw. dieselbe Komponente und unterscheiden sich hinreichend von den Items der fünf anderen COPSOQ-II-Skalen. Hiernach besteht akzeptable faktorielle Validität der AK-Skala (z. B. Amelang & Schmidt-Atzert, 2006).

Die Vorzeichen der sechs bivariaten Korrelationen, in welche die AK-Skala im Rahmen des nomologischen Netzwerkes eingebunden ist, entsprachen denen des Referenznetzwerkes. In Bezug auf die AK-Skala ist daher ein empirisches nomologisches Referenznetzwerk adäquat replizierbar (vgl. TF.6). Hiernach korreliert die AK-Skala („Arbeit-Familie/Privatleben-Konflikt“) positiv mit der QA-Skala („Quantitative Anforderungen“) sowie dem Einzelitem GB („Gedanke an Berufsaufgabe“) und negativ mit der FQ-Skala („Führungsqualität“), der HU-Skala („Hilfe und Unterstützung von Kollegen“), der AF-Skala („Arbeitsfähigkeit in Bezug auf die Arbeitsanforderungen“) sowie dem Einzelitem GG („Gegenwärtiger Gesundheitszustand“). Die stärkste positive Korrelation ist mit der AK-Skala und die stärkste negative Korrelation ist mit der AF-Skala zu verzeichnen. Diese inhaltlichen Zusammenhänge sind erwartungskonform und werden demgemäß als ein weiterer Hinweis auf eine akzeptable Konstruktvalidität der AK-Skala interpretiert.

7.7 Allgemeines

Auf der Grundlage der ausführlich dargestellten statistischen Analyseergebnisse können keine bedeutsamen Unterschiede zwischen den sechs COPSOQ-II-Skala im lidA-CAPI der 1. Welle der lidA-Hauptstudie und den sechs COPSOQ-II-Skalen, wie sie in den vier empirischen Vergleichsstudien eingesetzt wurden, berichtet werden. Darüber hinaus konnte keine der sechs Teilfragestellungen (TF) im Rahmen der vorliegenden Untersuchung verworfen werden. Außerdem sind die Ergebnisse im Wesentlichen konsistent zu denen internationaler Studien. **Daher wird die psychometrische Qualität sämtlicher sechs COPSOQ-II-Skalen (QA-Skala, EA-Skala, EM-Skala, FQ-Skala, HU-Skala, AK-Skala) abschließend als eindeutig akzeptabel und dementsprechend für inhaltliche statistische Analysen als uneingeschränkt geeignet bewertet.**

Eine optionale Optimierungsmaßnahme betrifft die EM-Skala. Hintergrund dieser Maßnahme sowie das entsprechende Vorgehen sind im Folgenden dargestellt. Das AM des Items EM-1 (= 76.07 bzw. 76.08) war in beiden Stichproben (A+B) der 1. Welle der lidA-Hauptstudie außerhalb des in der TF.2-H.3. geforderten Spektrums von 25.00 bis 75.00 lokalisiert. Ferner wies der Koeffizient der nicht-korrigierten Trennschärfe (= .64) desselben Items in Relation zu den anderen COPSOQ-II-Items den niedrigsten absoluten Wert auf. Zudem zeichnete sich der Koeffizient der korrigierten Trennschärfe des Items EM-1 in der 1. Welle/A der lidA-Hauptstudie durch den niedrigsten (= .42) und in der 1. Welle/B der lidA-Hauptstudie durch den zweitniedrigsten absoluten Wert aus (= .43). Schließlich fiel dieses Item als einziges der 20 COPSOQ-II-Items in beiden Stichproben (A+B) der 1. Welle der lidA-Hauptstudie durch eine Nebenladung auf, wobei beide Nebenladungen mit dem gleichen absoluten Wert (= .31) auf derselben Komponente (EA) lokalisiert waren. Die einzige weitere Nebenladung wies das Item EA-3 (= .39) in der 1. Welle/B der lidA-Hauptstudie auf der EM-Komponente auf. Eine genaue Betrachtung dieser beiden Items macht eine mögliche inhaltliche Nähe bzw. Interdependenz deutlich (EM-1: „Verlangt es Ihre Arbeit, dass Sie die Initiative ergreifen?“. EA-3: „Haben Sie Einfluss darauf, was Sie bei Ihrer Arbeit tun?“). Es wird gemeinhin schwer fallen, einen Arbeitsplatz zu finden, der zum Einen „Initiative“ verlangt und zum Anderen keinen Raum für inhaltlichen „Einfluss“ bietet (und vice versa). Vor diesem Hintergrund könnte die Elimination des Items EM-1 und die Bildung der EM-Skala ohne dieses Item eine Verbesserung der faktoriellen Konstruktvalidität und darüber hinaus auch eine Verbesserung der Gestalt des nomologischen Netzwerkes zur Folge haben. Allerdings ist im Vorfeld einer solchen Maßnahme die erneute Überprüfung der psychometrischen Qualität der EM-Skala mit nur drei Items (EM-2, EM-3 und EM-4) anhand der folgenden Aspekte der psychometrischen Item- und Skalenanalyse geboten: (1) Verteilung der Werte, (2) Skalenschwierigkeit, (3) Trennschärfe, (4) Homogenität, (5) Dimensionalität sowie (6) nomologisches Netzwerk.

7.8 Grenzen und Ausblick

Die im Rahmen dieser Arbeit verwendeten lidA-Stichproben (lidA-Machbarkeitsstudie, lidA-Hauptstudie: 1. Welle/A+B) sowie die Stichproben der vier Vergleichsstudien (NEXT-Studie, COPSOQ-I-Studie, IPSO-Studie, 3Q-Studie) wurden ausschließlich im Querschnitt untersucht. Infolgedessen konnten für alle Items und Skalen keine adäquate Reliabilitätsprüfung durchgeführt und ausschließlich konkurrente Zusammenhänge ermittelt werden.

Verhältnismäßig viele fehlende Werte wurden sowohl in den zwei lidA-Datensätzen bzw. drei lidA-Stichproben (= 14.0 bis 14.5 Prozent) als auch in den vier Datensätzen der Vergleichsstudien (= 7.5 bis 20.4 Prozent) ausgeschlossen. Hierdurch reduzierten sich die jeweiligen Stichprobengrößen in nicht unerheblichem Maße, wodurch sich ebenfalls die Aussagekraft der Analyseergebnisse potentiell reduziert. Zusätzlich können durch den Ausschluss fehlender Werte systematische Verzerrungen in den Daten erfolgt sein, weil sich die TN mit fehlenden Werten möglicherweise systematisch von den TN ohne fehlende Werte unterscheiden. Eine solche Analyse im Anschluss an diese Arbeit kann hier Klarheit schaffen.

Im Vorfeld der statistischen Analysen der lidA-Stichproben wurde nicht kontrolliert, ob ein bedeutsamer Einfluss der Interviewer auf das Antwortverhalten der TN vorlag. Zum Zwecke einer kritischen Betrachtung der Objektivität der sechs COPSOQ-II-Skalen im lidA-CAPI der 1. Welle der lidA-Hauptstudie könnte im Anschluss an diese Arbeit geprüft werden, ob sich die Antworten der TN systematisch je nach Interviewer unterscheiden. Eine solche Prüfung kann mit der zuvor thematisierten Analyse der TN mit fehlenden Werten verknüpft werden, da potentiell auch diese systematisch mit den Interviewern variieren.

Hinsichtlich einzelner Items sowie spezifischer Skalenzusammensetzungen bestanden zahlreiche Divergenzen zwischen den vier Vergleichsstudien und den zwei lidA-Datensätzen (siehe Abschnitt 3.2.5), so dass für die vorliegende Arbeit keine optimalen Referenzwerte vorlagen. Auch die in der deutsch- und englischsprachigen Literatur publizierten Kennwerte aus Studien weisen oftmals Divergenzen entweder in Bezug auf die ihnen zugrunde liegenden statistischen Verfahren oder den Operationalisierungen einzelner COPSOQ-Konstrukte im Vergleich zur lidA-Studie auf.

Die Stärke der verschiedenen Zusammenhänge, die in der vorliegenden Arbeit berichtet wurden, können sowohl hinsichtlich der lidA-Stichproben als auch in Bezug auf die Vergleichsstudien in nicht unerheblichem Maße durch die so genannte Einheitsmethodenvarianz („common-method variance“, z. B. Podsakoff & Organ, 1986) beeinflusst worden sein. Diese resultiert aus der Verwendung ausschließlich eines Verfahrens zur Datenerhebung in einem einzigen Erhebungsmodus. In diesem Kontext ist auch das Streben der TN nach inhaltlicher Konsistenz zu erwähnen (Laux & Weber, 1993; Mummendey, 1990), das, über subjektive semantische und logische Zusammenhänge seitens der TN, statisti-

sche Zusammenhänge beeinflussen kann (Syrek, Bauer-Emmel, Antoni & Klusemann, 2011). Solche Einflussgrößen sind ohne eine umfängliche Berücksichtigung derselben im Vorfeld, d. h. im Zuge der Planung einer Datenerhebung, nachträglich nicht mehr zu identifizieren bzw. zu eliminieren (z. B. Richardson, Simmering & Sturman, 2009).

Das nomologische (Referenz-)Netzwerk beinhaltet ausschließlich diskriminante Konstrukte, konvergente Konstrukte fehlen aufgrund der ökonomischen Gestaltung der Datenerhebungsverfahren in den lidA-Datensätzen sowie denjenigen der Vergleichsstudien, so dass die hiervon abgeleiteten Aussagen hinsichtlich der Konstruktvalidität zwangsläufig als eingeschränkt beurteilt werden müssen (vgl. Cronbach & Meehl, 1955). Die Erstellung einer systematischen Übersicht inhaltlich konvergenter (Kurz-)Skalen zur ergänzenden Überprüfung der Konstruktvalidität und eine separate (Querschnitts-)Datenerhebung im Anschluss an diese Arbeit sicherten die Analyseergebnisse zusätzlich ab.

Bei der Ermittlung der statistischen Kennwerte wurden weder für die lidA-Stichproben noch für die Stichproben der Vergleichsstudien die Berufsklassen der TN berücksichtigt. Aus mehreren empirischen Studien ist gleichsam bekannt, dass sich die verschiedenen Ausprägungen der einzelnen COPSOQ-II-Skalen in Abhängigkeit von den Berufsklassen z. T. signifikant voneinander unterscheiden (Dupret et al., 2012; Nübling & Hasselhorn, 2010). Vor diesem Hintergrund wird aktuell eine deutsche COPSOQ-Datenbank aufgebaut (Nübling et al., 2011). Unter Heranziehung der in jener Datenbank verwendeten Klassifikation der Berufe ist eine Überprüfung der oben dargestellten Analyseergebnisse im Anschluss an diese Arbeit zur Absicherung der berichteten Ergebnisse möglich. Dabei ist allerdings zu beachten, dass sich auch innerhalb einzelner Berufsklassen die tatsächlichen beruflichen Tätigkeiten und hierdurch die verschiedenen Ausprägungen der einzelnen COPSOQ-II-Skalen signifikant voneinander unterscheiden können.

Keine Berücksichtigung in den statistischen Analysen der vorliegenden Arbeit fanden zudem die Persönlichkeitsmerkmale der TN. Diese spielen in der Beschreibung und Bewertung der verschiedenen Aspekte der Arbeitstätigkeit und Arbeitsumgebung allerdings mitunter eine wichtige Rolle (Schallenberg, 1995). Im Anschluss an diese Arbeit kann daher eine Reanalyse der sechs COPSOQ-II-Skalen unter Berücksichtigung bspw. der „negativen Affektivität“ (vgl. Schaubroek, Ganster & Fox, 1992), welche in der lidA-Studie mittels des „Positive and Negative Affect Schedule (PANAS)“ (deutsche Adaptation von Krohne, Egloff, Kohlmann & Tausch, 1996) erhoben wird, vorgenommen werden.

Abschließend ist darauf hinzuweisen, dass die hier psychometrisch analysierten COPSOQ-II-Skalen auf der Grundlage weiterführender inhaltlicher Analysen keine Ableitungen konkreter Arbeitsgestaltungsmaßnahmen zulassen, da diese lediglich allgemeine und zudem ausschließlich subjektive Bewertungen des eigenen Arbeitsplatzes darstellen. Nichtsdestotrotz können entsprechende Analysen Hinweise zutage fördern, welche wiederum den Ausgangspunkt psychologischer Arbeitsplatzanalysen darstellen.

Literaturverzeichnis

- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (4. Auflage). Heidelberg: Springer.
- Asendorpf, J. (2007). *Psychologie der Persönlichkeit* (4. Auflage). Heidelberg: Springer.
- Atteslander, P. (1999). Epidemiologie als Demoskopie. Die Verwendung von Fragebögen in epidemiologischen Studien. In C. Maeder, C. Burton-Jeangros & M. Haour-Knipe (Hrsg.), *Gesundheit, Medizin und Gesellschaft. Beiträge zur Soziologie der Gesundheit* (361-383). Zürich: Seismo Verlag.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2008). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (12. Auflage). Berlin: Springer.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2011). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (13. Auflage). Berlin: Springer.
- Badawi, G., Pagé, V., Smith, K. J., Gariépy, G., Malla, A., Wang, J. L. et al. (in Druck). Self-rated health: A predictor for the three year incidence of major depression in individuals with type ii diabetes. *Journal of Affective Disorders*.
- BAuA – Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Hrsg.) (2011). Why WAI? Der Work Ability Index im Einsatz für Arbeitsfähigkeit und Prävention. Erfahrungsberichte aus der Praxis (4. Auflage). Dortmund: BAuA.
- BAuA – Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Hrsg.) (2002). *Europäische Erfahrungen mit dem Arbeitsbewältigungsindex (Work Ability Index). Erfahrungsaustausch vom 13. Mai 2000 in Berlin* (1. Auflage). Bremerhaven: NW Verlag.
- Bellach, B. M., Knopf, H. & Thefeld, W. (1998). Der Bundes-Gesundheitssurvey 1997/98. *Gesundheitswesen*, 60 (Sonderheft 2), 59-68.
- Beermann, B. (2011). Zwölf Thesen für eine alternsgerechte Arbeitswelt – Demografischer Wandel und Arbeitsgestaltung. In BAuA (Hrsg.), *baua: Aktuell. Amtliche Mitteilungen der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin*, 2/2011 (3-4). Dortmund: BAuA.
- Blickle, G. (2011). Berufswahl und berufliche Entwicklung. In F. W. Nerdinger, G. Blickle & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (173-193). Berlin: Springer-Verlag.
- BMAS – Bundesministerium für Arbeit und Soziales (Hrsg.) (2011). *Strategie Europa 2020*. Verfügbar unter: <http://www.bmas.de/DE/Themen/Soziales-Europa-und-Internationales/Europa/strategie-eu-2010-beschaefigungspolitik.html> [16.05.2012]
- BMBF – Bundesministerium für Bildung und Forschung (Hrsg.) (2009). *Verbundprojekt: lidA – leben in der Arbeit – Kohortenstudie zu Gesundheit und Älterwerden in der Arbeit*. Verfügbar unter: <http://www.gesundheitsforschung-bmbf.de/de/2101.php#lidA> [04.09.2012]

- BMWi – Bundesministerium für Wirtschaft und Technologie (Hrsg.) (2011). *Nationales Reformprogramm Deutschland 2011 (Dokumentation 596)*. Verfügbar unter: <http://www.bmwi.de/BMWi/Redaktion/PDF/Publikationen/Dokumentationen/dokumentation-596-nationales-reformprogramm,property=pdf,bereich=bmwi,sprache=de,rwb=true.pdf> [16.05.2012]
- Bode, S. & Richter, G. (2011). Strategien für eine demografiefeste Personalpolitik – Mit Prävention die Zukunft gewinnen. In BAuA (Hrsg.), *baua: Aktuell. Amtliche Mitteilungen der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin*, 2/2011 (5). Dortmund: BAuA.
- Borgetto, B. & Troschke, J. (2001). Interdependenzen zwischen Sozial-Epidemiologie und Medizinischer Soziologie. In A. Mielck & K. Bloomfield (Hrsg.), *Sozial-Epidemiologie. Eine Einführung in die Grundlagen, Ergebnisse und Umsetzungsmöglichkeiten* (231-245). Weinheim: Juventa-Verlag.
- Börsch-Supan, A. (2011). Ökonomische Auswirkungen des demografischen Wandels. *APuZ – Aus Politik und Zeitgeschichte*, 10-11/2011, 19-26.
- Bortz, J. (2005). *Statistik* (6. Auflage). Berlin: Springer.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Auflage). Berlin: Springer.
- Bortz, J. & Döring, N. (2009). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Auflage). Heidelberg: Springer Medizin Verlag.
- Bortz, J., Lienert, G. A. & Boehnke, K. (2008). *Verteilungsfreie Methoden in der Biostatistik* (3. Auflage). Heidelberg: Springer.
- Briggs, S. R. & Cheek, J. M. (1986). The Role of Factor Analysis in the Development and Evaluation of Personality Scales. *Journal of Personality and Social Psychology*, 54 (1), 106-148.
- de Bruin, A., Picavet, H. S. J. & Nossikoy, A. (Hrsg.) (1996). *Health interview surveys: Towards international harmonization of methods and instruments*. Copenhagen, WHO Regional Publication.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Auflage). München: Pearson Studium.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Auflage). München: Pearson Studium.
- Buunk, B. P.; de Jonge, J.; Ybema, J. F. & de Wolff, C. J. (1998). Psychosocial aspects of occupational stress. In P. J. D. Drenth, H. Thierry & C. J. de Wolff (Hrsg.), *Handbook of work and organizational psychology* (145-182). East Sussex: Psychology Press.
- Chelune, G. J. (Hrsg.) (1979). *Self-Disclosure: origins, patterns, and implications of openness in interpersonal relationships*. San Francisco: Jossey-Bass.
- Coin, D. (2008). Testing normality in the presence of outliers. *Statistical Methods and Applications*, 17, 3-12.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale: Lawrence Erlbaum Associates.

- Conover, W. J. (1971). *Practical nonparametric statistics*. New York: Wiley.
- Cowles, M. & Davis, C. (1982). On the Origins of the .05 Level of Significance. *American Psychologist*, 37, 553-558.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Destatis – Statistisches Bundesamt. Statistische Ämter des Bundes und der Länder (Hrsg.). (2011). *Demografischer Wandel in Deutschland. Bevölkerungs- und Haushaltsentwicklung in Bund und in den Ländern, Heft 1*. Wiesbaden: Statistische Ämter des Bundes und der Länder.
- Destatis – Statistisches Bundesamt (2009). *Bevölkerung Deutschlands bis 2060. 12. koordinierte Bevölkerungsvorausberechnung. Begleitmaterial zur Pressekonferenz am 18. November 2009 in Berlin*. Wiesbaden: Destatis.
- Dichter, M., Schmidt, S., Müller, B. H., & Hasselhorn, H.M. (2007). *Qualität in deutschen Altenpflegeeinrichtungen – die 3Q-Studie. 6. Internationale Konferenz Pflege und Pflegewissenschaft, Nürnberg/Fürth*. Verfügbar unter: www.3q.uni-wuppertal.de/download.php?f=4d44b64a08fb5a62844028097900a426&target=0 [09.05.2012]
- Diehl, J. M. & Staufenbiel, T. (2002). *Statistik mit SPSS. Version 10 + 11*. Eschborn: Dietmar Klotz.
- Dietz, M. & Walwei, U. (2008). Erwerbsarbeit heute. Auf neuen Spuren unterwegs. *IAB-Forum*, 1/08, 4-9.
- DIN EN ISO 10075-1. *Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung – Teil 1: Allgemeines und Begriffe (ISO 10075:1991), Deutsche Fassung EN ISO 10075-1:2000*.
- DIN EN ISO 10075-2. *Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung – Teil 2: Gestaltungsansätze (ISO 10075-2:1996), Deutsche Fassung EN ISO 10075-2:2000*.
- DIN EN ISO 10075-3. *Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung – Teil 3: Grundsätze und Anforderungen an Verfahren zur Messung und Erfassung psychischer Arbeitsbelastung (ISO 10075-3:2004), Deutsche Fassung EN ISO 10075-3:2004*.
- Duller, C. (2008). *Einführung in die nichtparametrische Statistik mit SAS und R. Ein anwendungsorientiertes Lehr- und Arbeitsbuch*. Heidelberg: Physica-Verlag.
- Dunckel, H. (1999). Psychologische Arbeitsanalyse: Verfahrensüberblick und Auswahlkriterien. In H. Dunckel (Hrsg.), *Handbuch psychologischer Arbeitsanalyseverfahren* (9-30). Zürich: vdf Hochschulverlag.
- Dunckel, H. & Volpert, W. (1997). Aufgaben- und kriterienbezogene Gestaltung von Arbeitsstrukturen. In H. Luczak & W. Volpert (Hrsg.), *Handbuch Arbeitswissenschaft* (791-795). Stuttgart: Schäffer-Poeschel.

- Dupret, E., Bocéréan, C., Teherani, M., Feltrin, M. & Pejtersen, J. H. (2012). Psychosocial risk assessment: French validation of the Copenhagen Psychosocial Questionnaire (COPSOQ). *Scandinavian Journal of Public Health*, 40, 482-490.
- Dziuban, C. D. & Shirkey, E. C. (1974). When is a Correlation Matrix Appropriate for Factor Analysis? *Psychological Bulletin*, 81, 358-361.
- Ebener, M., Hardt, J., Galatsch, M. & Hasselhorn, H.-M. (2011). Prädiktive Validität der Dimensionen des Work Ability Index (WAI) in verschiedenen Berufsgruppen im Quer- und Längsschnitt. *Arbeitsmedizin Sozialmedizin Umweltmedizin*, 46, 185.
- FFAS – Freiburger Forschungsstelle Arbeits- und Sozialmedizin GbR (2007). *Mitarbeiterbefragung zu psychosozialen Belastungen am Arbeitsplatz – COPSOQ: deutsche Standard-Version, Stand 1/2007*. Verfügbar unter http://www.copsoq.de/data/copsoq_fb_verkuerzt_und_skalenkung_220107.pdf [12.04.2011]
- Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik. Mit Hinweisen zur Intervention* (2. Auflage). Göttingen: Hogrefe-Verlag.
- Frei, F. (1981). Psychologische Arbeitsanalyse: Eine Einführung zum Thema. In F. Frei & E. Ulich (Hrsg.), *Beiträge zur psychologischen Arbeitsanalyse* (11-36). Bern: Verlag Hans Huber.
- Friedmann, G. (1959). *Grenzen der Arbeitsteilung*. Frankfurt: Europäische Verlagsanstalt.
- Frieling, E. & Sonntag, K. (1999). *Lehrbuch Arbeitspsychologie*. Bern: Huber.
- Fromm, S. (2012). *Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten*. Wiesbaden: Springer.
- Garloff, A.; Pohl, C. & Schanne, N. (2012). *Demografischer Wandel der letzten 20 Jahre: Alterung der Bevölkerung hat sich kaum auf die Arbeitslosigkeit ausgewirkt*. Nürnberg: IAB-Kurzbericht, 10/212.
- Gauggel, B., Tielsch, R. & Portuné, R. (2011). *Gesunde Justiz – Schritt für Schritt: Betriebliches Gesundheitsmanagement in der Justiz*. Düsseldorf: Unfallkasse Nordrhein-Westfalen (PIN 47).
- Gauggel, B., Portuné, R., Tielsch, R., Müller, B. H. & Hasselhorn, H.-M. (2009). *Auswertung der Mitarbeiterbefragung 2008: Mitarbeiterrückmeldung*. Verfügbar unter <http://www.djg-nrw.de/Allgemein/Infotzentrum/Aktuelles/2009/IPSO-%20z0401-Auswertung.pdf> [12.04.2011]
- Gauggel, B. & Tielsch, R. (2008). *IPSO-Justiz Studie: Projekt der Bergischen Universität Wuppertal in Kooperation mit der Unfallkasse NRW*. Verfügbar unter <http://www.ipso.uni-wuppertal.de/download.php?f=92f4411df1deb57a64e4ef8b0708d1a7> [12.04.2011]
- Gauggel, B. (2008). *Das Work Ability – Netzwerk Deutschland*. Verfügbar unter: <http://www.arbeitsfaehigkeit.uni-wuppertal.de/picture/upload/file/netzwerk2.pdf> [12.04.2011]
- Gilberg, R.; Schröder, H. & Trappmann, M. (2010). *Stichprobenkonzept*. Nürnberg: Internes Dokument.

-
- Gorsuch, R. L. (1983). *Factor Analysis* (2nd edition). Hillsdale: Erlbaum.
- Gregersen, S., Dulong, M., Nienhaus, A. & Nübling, M. (2010). Psychosoziale Arbeitsbelastungen in der Pflege – Vergleich verschiedener Pflegegruppen mit anderen Berufsgruppen. In A. Nienhaus (Hrsg.), *Gefährdungsprofile. Unfälle und arbeitsbedingte Erkrankungen in Gesundheitsdienst und Wohlfahrtspflege* (113-128). Landsberg/Lech: ecomed Medizin.
- Greif, S. (1991). Stress in der Arbeit. Einführung und Grundbegriffe. In S. Greif, E. Bamberg & N. Semmer (Hrsg.), *Psychischer Stress am Arbeitsplatz* (1-28). Göttingen: Hogrefe.
- Greiner, B. (1998). Der Gesundheitsbegriff. In E. Bamberg, A. Ducki & A.-M. Metz (Hrsg.), *Handbuch betriebliche Gesundheitsförderung. Arbeits- und organisationspsychologische Methoden und Konzepte* (39-55). Göttingen: Verlag für Angewandte Psychologie.
- Grice, J. W. (2001). Computing and Evaluating Factor Scores. *Psychological Methods*, 6, 430-450.
- Handl, A. (2010). *Multivariate Analysemethoden. Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS* (2. Auflage). Wiesbaden: Springer.
- Hardt, J., Iskenius, M., du Prel, J.-B., Rose, U., Rauch, A., Schröder, H., et al. (2010). *Methodological of the lidA cohort study (leben in der Arbeit) – study design and psychometric analyses*. Gemeinsamer Kongress der DGSM, DGEpi und EUMASS „Individualisierte Prävention und Epidemiologie: Die moderne Medizin“, Berlin 21.-25.09.2010 (Poster).
- Hasselhorn, H.-M. (2011). Forschung für altersgerechte Arbeitsgestaltung im Dreiklang. Arbeit, Alter und Gesundheit. In BAuA (Hrsg.), *baua: Aktuell. Amtliche Mitteilungen der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin*, 2/2011 (4). Dortmund: BAuA.
- Hasselhorn, H.-M. & Freude G. (2007). *Der Work Ability Index – ein Leitfaden*. Dortmund: Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (S 87).
- Hasselhorn, H.-M., Tackenberg, P., Müller, H. B. & NEXT-Studiengruppe (2005). Warum will Pflegepersonal in Europa die Pflege verlassen? In BAuA (Hrsg.), *Berufsausstieg bei Pflegepersonal. Arbeitsbedingungen und beabsichtigter Berufsausstieg bei Pflegepersonal in Deutschland und Europa* (124-134). Dortmund: Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Ü 15).
- Hasselhorn, H.-M., Tackenberg, P., Büscher, A., Stelzig, S., Kümmerling, A. & Müller, B. H. (2005). Wunsch nach Berufsausstieg bei Pflegepersonal in Deutschland. In BAuA (Hrsg.), *Berufsausstieg bei Pflegepersonal. Arbeitsbedingungen und beabsichtigter Berufsausstieg bei Pflegepersonal in Deutschland und Europa* (135-146). Dortmund: Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Ü 15).
- Hasselhorn, H.-M. & Müller, B. H. (2004). Arbeitsbelastungen und -beanspruchungen bei Pflegepersonal in Europa. Ergebnisse der NEXT-Studie. In B. Bandura, H. Schellschmidt & C. Vetter (Hrsg.), *Fehlzeiten-Report 2004. Gesundheitsmanagement in Krankenhäusern und Pflegeeinrichtungen* (21-47). Berlin: Springer Verlag.

- Hasselhorn, H.-M., Tackenberg, P., Büscher, A., Stelzig, S., Kümmerling, A. & Müller, H. B. (2003). Intent to leave nursing in Germany. In H.-M. Hasselhorn, P. Tackenberg & H. B. Müller (Hrsg.), *Working conditions and intent to leave the profession among nursing staff in Europe. Working Life Research Report 7* (136-145). Stockholm: National Institute for Working Life.
- Höfler, M. (2004). *Statistik in der Epidemiologie psychischer Störungen*. Berlin: Springer.
- IBM Corporation (2011). *IBM SPSS Statistics 20 Core-System-Benutzerhandbuch*. Verfügbar unter ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/20.0/de/client/Manuals/IBM_SPSS_Statistics_Core_System_Users_Guide.pdf [09.01.2012]
- Idler, E. L. & Benyamini, Y. (1997). Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *Journal of Health and Social Behavior*, 38, 21-37.
- Ilmarinen, J. (2006). Towards a Longer Worklife! Ageing and the quality of worklife in the European Union. Verfügbar unter: http://www.stm.fi/c/document_library/get_file?folderId=39503&name=DLFE-8602.pdf [04.09.2012]
- Ilmarinen, J. & Tuomi, K. (2004). Past present and future of work ability. In J. Ilmarinen & K. Tuomi (Hrsg.), *Past present and future of work ability – People and Work Research Report 65*. Helsinki: Finnish Institute of Occupational Health.
- Infratest Sozialforschung (2012). *SOEP 2004 – Erhebungsinstrumente 2004 (Welle 21) des Sozio-ökonomischen Panels. SOEP Survey Papers 102: Series A*. Berlin: DIW/SOEP.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* (6. Auflage). Weinheim: Beltz Verlag.
- Jahoda, M. (1981). Work, employment, and unemployment. Values, theories, and approaches in social research. *American Psychologist*, 36, 184-191.
- Janssen, J. & Laatz, W. (2007). *Statistische Datenanalyse mit SPSS für Windows* (6. Auflage). Berlin: Springer.
- Kaiser, H. F. (1974). An Index of Factorial Simplicity. *Psychometrika*, 39, 31-36.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd edition). New York: Holt, Rinehart and Winston.
- Kersting, A., Schröder, H. & Steinwede, J. (2012). *Methodenbericht zur Haupterhebung lidA – leben in der Arbeit*. Unveröffentlichtes Dokument.
- Kersting, A., Schröder, H. & Steinwede, J. (2011). *Pretestbericht. lidA – leben in der Arbeit*. Unveröffentlichtes Dokument.

- Klotter, C. (1999). Historische und aktuelle Entwicklungen der Prävention und Gesundheitsförderung – Warum Verhaltensprävention nicht ausreicht. In R. Oesterreich & W. Volpert (Hrsg.), *Psychologie gesundheitsgerechter Arbeitsbedingungen* (23-61). Schriften zur Arbeitspsychologie, Band 59. Bern: Huber.
- Knapp, T. R. & Brown, J. K. (1995). Ten Measurement Commandments That Often Should Be Broken. *Research in Nursing & Health*, 18, 465-469.
- Konrad, K. (2001). *Mündliche und schriftliche Befragung: Ein Lehrbuch* (2. Auflage). Landau: Verlag Empirische Pädagogik.
- Kreienbrock, L.; Pigeot, I. & Ahrens, W. (2012). *Epidemiologische Methoden* (5. Auflage). Berlin: Springer-Verlag.
- Kristensen, T. S.; Hannerz, H.; Høgh, A. & Borg, V. (2005). The Copenhagen Psychosocial Questionnaire – a tool for the assessment and improvement of the psychosocial work environment. *Scandinavian Journal of Work, Environment & Health*, 31, 305-322.
- Kristensen, T. S. (2000). *A new tool for assessing psychosocial factors at work: The Copenhagen Psychosocial Questionnaire*. Verfügbar unter <http://www.arbejdsmiljoforskning.dk/~media/Spoergeskemaer/copsoq/engelsk-artikel-om-det-mellemlange-skema-pdf.pdf> [03.05.2012]
- Krohne, H. W., Egloff, B., Kohlmann, C.-W. & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der „Positive and Negative Affect Schedule (PANAS)“. *Diagnostica*, 42 (2), 139-156.
- Kümmerling, A., Hasselhorn, H.-M. & Tackenberg, P. (2005). Psychometrische Eigenschaften der in der NEXT-Studie verwendeten Skalen. In BAuA (Hrsg.), *Berufsausstieg bei Pflegepersonal. Arbeitsbedingungen und beabsichtigter Berufsausstieg bei Pflegepersonal in Deutschland und Europa* (147-166). Dortmund: Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Ü 15).
- Kury, H. (1994). Zum Einfluß der Art der Datenerhebung auf die Ergebnisse von Umfragen. *Monatsschrift für Kriminologie und Strafrechtsreform*, 77 (1), 22-33.
- Laux, L. & Weber, H. (1993). *Emotionsbewältigung und Selbstdarstellung*. Stuttgart: Kohlhammer.
- Li, C., Ford, E. S., Zhao, G., Tsai, J. & Balluz, L. S. (2012). A comparison of depression prevalence estimates measured by the Patient Health Questionnaire with two administration modes: computer-assisted telephone interviewing versus computer-assisted personal interviewing. *International Journal of Public Health*, 57, 225-233.
- lidA-Projektkonsortium (2010). *Die lidA-Studie*. Verfügbar unter: <http://www.arbeit.uni-wuppertal.de/index.php?die-lida-studie-1> [04.09.2012]
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Psychologie Verlags Union.
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, 62, 399-402.

- Metzner, H. & Mann, F. (1952). A Limited Comparison of two Methods of Data Collection: The Fixed Alternative Questionnaire and the Open-Ended Interview. *American Sociological Review*, 17 (4), 486-491.
- Mohr, G. & Udris, I. (1997). *Gesundheit und Gesundheitsförderung in der Arbeitswelt*. In R. Schwarzer (Hrsg.), *Gesundheitspsychologie. Ein Lehrbuch* (2. Auflage) (553-573). Göttingen: Hogrefe.
- Moller, L., Kristensen, T. S. & Hollnagel, H. (1996). Self rated health as a predictor of coronary heart disease in Copenhagen, Denmark. *Journal of Epidemiology and Community Health*, 50(4), 423-428.
- Müller, B. H. (1997). *Ergonomie. Bestandteil der Sicherheitswissenschaft* (3. Auflage). Berlin: Beuth Verlag.
- Mummendey, H. D. (1990). *Psychologie der Selbstdarstellung*. Göttingen: Hogrefe.
- Naplava, T. & Oberwittler, D. (2002). Methodeneffekte bei der Messung selbstberichteter Delinquenz von männlichen Jugendlichen. Ein Vergleich zwischen schriftlicher Befragung in der Schule und mündlicher Befragung im Haushalt. *Monatsschrift für Kriminologie und Strafrechtsreform*, 85 (6), 401-423.
- Netemeyer, R. G., Boles, J. S. & McMurrian, R. (1996). Development and validation of work-family conflict and family-work conflict scales. *Journal of Applied Psychology*, 81 (4), 400-410.
- NEXT-Studie – Nurses Early Exit Study (2002). *Fragebogen. Basisbefragung Herbst 2002*. Unveröffentlichtes Dokument.
- Nübling, M., Vomstein, M., Nübling, T., Stößel, U., Hasselhorn H.-M. & Hofmann, F. (2011). *Erfassung psychischer Belastungen anhand eines erprobten Fragebogens – Aufbau der COPSOQ-Datenbank*. Dortmund: Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Fb 2031).
- Nübling, M., Stößel, U. & Michaelis, M. (2010). Messung von Führungsqualität und Belastungen am Arbeitsplatz: Die deutsche Standardversion des COPSOQ (Copenhagen Psychosocial Questionnaire). In B. Badura, H. Schröder, J. Klose & K. Macco (Hrsg.) *Fehlzeiten-Report 2009. Arbeit und Psyche: Belastungen reduzieren – Wohlbefinden fördern* (253-261). Berlin: Springer-Verlag.
- Nübling, M. & Hasselhorn, H.-M. (2010). The Copenhagen Psychosocial Questionnaire in Germany: From the validation of the instrument to the formation of a job-specific database of psychosocial factors at work. *Scandinavian Journal of Public Health*, 38, 120-124.
- Nübling, M., Stößel, U., Hasselhorn, H.-M., Michaelis, M. & Hofmann, F. (2005). *Methoden zur Erfassung psychischer Belastungen – Erprobung eines Messinstrumentes (COPSOQ)*. Dortmund: Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Fb 1058).
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd edition). New York: McGraw-Hill.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd edition). New York: McGraw-Hill.

- NRCWE – National Research Centre of the Working Environment (2007a). *The Scales of the LONG COPSOQ II questionnaire*. Verfügbar unter <http://www.arbejdsmiljoforskning.dk/en/publikationer/spoergeskemaer/~media/Spoergeskemaer/copsoq/uk/scales-of-the-long-copsoq-ii-questionnaire.pdf#> [20.10.2011].
- NRCWE – National Research Centre of the Working Environment (2007b). *The Scales of the SHORT COPSOQ II questionnaire*. Verfügbar unter <http://www.arbejdsmiljoforskning.dk/en/publikationer/spoergeskemaer/~media/Spoergeskemaer/copsoq/uk/scales-of-the-short-copsoq-ii-questionnaire.pdf#> [20.10.2011].
- NRCWE – National Research Centre of the Working Environment (2007c). *The construction of the scales in COPSOQ II*. Verfügbar unter <http://www.arbejdsmiljoforskning.dk/en/publikationer/spoergeskemaer/~media/Spoergeskemaer/copsoq/uk/construction-of-the-scales-of-copsoq-ii-22-6.pdf#> [20.10.2011]
- Oppolzer, A. (2010). Psychische Belastungsrisiken aus Sicht der Arbeitswissenschaft und Ansätze für Prävention. In B. Badura, H. Schröder, J. Klose & K. Macco (Hrsg.), *Fehlzeiten-Report 2009. Arbeit und Psyche: Belastungen reduzieren – Wohlbefinden fördern. Zahlen, Daten, Analysen aus allen Branchen der Wirtschaft*. Berlin: Springer-Verlag.
- Pejtersen, J. H.; Kristensen, T. S.; Borg, V. & Bjorner, J. B. (2010). The second version of the Copenhagen Psychosocial Questionnaire. *Scandinavian Journal of Public Health*, 38 (Suppl 3), 8-24.
- Plath, H.E. (2000). *Arbeitsanforderungen im Wandel, Kompetenzen für die Zukunft – Eine folgenreichere Auseinandersetzung mit aktuellen Positionen*. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung (MittAB), 33/2000, 581-593.
- Precht, M., Kraft, R. & Bachmaier, M. (2005). *Angewandte Statistik 1* (7. Auflage). München: Oldenbourg Verlag.
- du Prel, J.-B. & Peter, R. (2011). *Erste Ergebnisse aus der Machbarkeitsstudie zum Zusammenhang von Alter, Geschlecht und ERI auf den subjektiv eingeschätzten Gesundheitszustand*. Gemeinsame Jahrestagung der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSM) sowie der Deutschen Gesellschaft für Medizinische Soziologie (DGMS), Bremen 21.-23.09.2011 (Vortrag).
- du Prel, J.-B., Hardt, J., Rauch, A., Rose, U., Schröder, H., Steinwede, J., et al. (2010). *Investigating work-related determinants of physical and psychological health in an aging workforce – The prospective lidA-Study (leben in der Arbeit)*. 55. GMDS Jahrestagung, Mannheim 05.-09.09.2010 (Poster).
- Podsakoff, P. M. & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, 12 (4), 531-544.
- Rahman, M. M. & Govindarajulu, Z. (1997). A Modification of the test of Shapiro and Wilk for Normality. *Journal of Applied Statistics*, 24, 219-235.

- Richardson, H. A., Simmering, M. J., Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, 12, 762–800.
- Richter, P. & Hacker, W. (1998). *Belastung und Beanspruchung. Stress, Ermüdung und Burnout im Arbeitsleben*. München: Asanger.
- Robert Koch-Institut (Hrsg.) (2011). *Beiträge zur Gesundheitsberichterstattung des Bundes. Daten und Fakten: Ergebnisse der Studie „Gesundheit in Deutschland aktuell 2009“*. Berlin: Oktoberdruck.
- Roethlisberger, F. J. & Dickson, W. J. (1964). *Management and the Worker*. Cambridge: Harvard University Press.
- Rohmert, W. (1984). Das Belastungs-Beanspruchungs-Konzept. *Zeitschrift für Arbeitswissenschaft*, 38 (4), 193-200.
- Rohmert, W. & Rutenfranz, J. (1975). *Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen industriellen Arbeitsplätzen. Forschungsbericht*. Bonn: Bundesministerium für Arbeit und Sozialordnung.
- Rothe, I. (2011). Beschäftigungsfähigkeit erhalten und verbessern. In BAuA (Hrsg.), *baua: Aktuell. Amtliche Mitteilungen der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin*, 2/2011 (2). Dortmund: BAuA.
- Sachs, L. & Hedderich, J. (2006). *Angewandte Statistik. Methodensammlung mit R* (12. Auflage). Berlin: Springer-Verlag.
- de Salvo, K. B., Bloser, N., Reynolds, K., He, J. & Muntner, P. (2006). Mortality Prediction with a Single General Self-Rated Health Question. *Journal of General Internal Medicine*, 21 (3), 267-275.
- Schallenger, U. (1995). Die Persönlichkeitsabhängigkeit von Beschreibungen der eigenen Arbeitssituation. *Zeitschrift für Experimentelle Psychologie*, 42 (1), 111-131.
- Schaper, N. (2011a). Arbeitsanalyse und -bewertung. In F. W. Nerdinger, G. Blickle & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (327-348). Berlin: Springer-Verlag.
- Schaper, N. (2011b). Arbeitsgestaltung in Produktion und Verwaltung. In F. W. Nerdinger, G. Blickle & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (349-367). Berlin: Springer-Verlag.
- Schaper, N. (2011c). Wirkungen der Arbeit. In F. W. Nerdinger, G. Blickle & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (475-495). Berlin: Springer-Verlag.
- Schaper, N. (2011d). Neue Formen der Arbeit: Das Beispiel Telekooperation. In F. W. Nerdinger, G. Blickle & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (497-516). Berlin: Springer-Verlag.
- Schaubroek, J., Ganster, D. C. & Fox, M. L. (1992). Dispositional affect and work-related stress. *Journal of applied psychology*, 77, 322-335.
- Schendera, C. F. G. (2007). *Datenqualität mit SPSS*. München: Oldenbourg Wissenschaftsverlag.

- Schimany, P. (2007). *Migration und Demographischer Wandel*. BAMF Forschungsbericht 5. Nürnberg: BAMF.
- Schmidt, S., Hasselhorn, H.-M. & Müller, B. (2007). *Sachbericht. Zeitraum: 01. Juni bis 30. November 2007*. Verfügbar unter www.3q.uni-wuppertal.de/download.php?f=4217b7238eb92bde67d49338882edbbc&target=0 [09.05.2012]
- Schmidt, S. & Hasselhorn, H.-M. (2007). *4-seitige Information zum Hintergrund und Zielen der 3Q-Studie*. Verfügbar unter www.3q.uni-wuppertal.de/download.php?f=868a268aaa0b3cdb689be332692d9c2b&target=0 [09.05.2012]
- Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha, *Psychological Assessment*, 8(4), 350-353.
- Schneider, N. F. & Dorbritz, J. (2011). Wo bleiben die Kinder? Der niedrigen Geburtenrate auf der Spur. *APuZ – Aus Politik und Zeitgeschichte*, 10-11/2011, 26-34.
- Schnell, R., Hill, P. B. & Esser, E. (2008). *Methoden der empirischen Sozialforschung* (8. Auflage). München: Oldenbourg Wissenschaftsverlag.
- Schröder, H., Gilberg, R., Tisch, A., Peter, R., du Prel, J. B. & Trappmann, M. (2011). Kohortenstudie „lidA – leben in der Arbeit“ – Stichprobendesign. *Gesundheitswesen*, 73, 8-9.
- Schumann, H. (2012). *Rettungsdienst am Limit: Gesundheit von Einsatzkräften im Rettungsdienst (GERD®). Ein Vergleich zwischen der Berufsfeuerwehr und den Hilfsorganisationen*. Hamburg: Diplomica Verlag.
- Graf von der Schulenburg, J. M., Claes, C., Greiner, W. & Uber, A. (1998). Die deutsche Version des EuroQol-Fragebogens. *Zeitschrift für Gesundheitswissenschaft*, 6, 3-20.
- Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591-611.
- Seier, E. (2002). *Comparison of Tests for Univariate Normality*. Verfügbar unter: <http://interstat.statjournals.net/YEAR/2002/articles/0201001.pdf> [29.05.2012]
- Siegrist, J. (2001). Zukünftige Aufgaben der Sozial-Epidemiologie. In A. Mielck & K. Bloomfield (Hrsg.), *Sozial-Epidemiologie. Eine Einführung in die Grundlagen, Ergebnisse und Umsetzungsmöglichkeiten* (358-370). Weinheim: Juventa-Verlag.
- Simon, M., Tackenberg, P., Hasselhorn, H.-M., Kümmerling, A., Büscher, A., & Müller, B. H. (2005). *Auswertung der ersten Befragung der NEXT-Studie in Deutschland*. Verfügbar unter www.next.uni-wuppertal.de/download.php?f=67c55b82536b145ec6a7faf17db66dff&target=0 [03.05.2012]
- Syrek, C., Bauer-Emmel, C., Antoni, C. & Klusemann, J. (2011). Entwicklung und Validierung der Trier Kurzskala zur Messung von Work-Life-Balance (TKS-WLB). *Diagnostica*, 57 (3), 134-145.

- Tuomi, K., Ilmarinen, J., Jahkola, A., Katajarinne, L. & Tulkki, A. (1998). *Work Ability Index* (2nd edition). Helsinki: Finnish Institute of Occupational Health.
- Ulich, E. (2005). *Arbeitspsychologie* (6. Auflage). Stuttgart: vdf Hochschulverlag und Schäffer-Poeschel Verlag.
- Warr, P. (1987). *Work, unemployment, and mental health*. Oxford: Oxford University Press.
- Weise, G. (1975). Psychologische Leistungstests. Ein Handbuch für Studium und Praxis. Göttingen: Hogrefe.
- Wolff, H.-G. & Bacher, J. (2010). Hauptkomponentenanalyse und explorative Faktorenanalyse. In C. Wolf & H. Best (Hrsg.) *Handbuch der sozialwissenschaftlichen Datenanalyse* (333-365). Wiesbaden: Springer.
- Zeller, R. A. & Carmines, E. G. (1980). *Measurement in the social sciences*. New York: Cambridge University Press.
- Zielhuis, G. A. & Kiemeny, L. A. L. M. (2001). Social epidemiology? No way. *International Journal of Epidemiology*, 30, 43-44.

Anhang

lidA-Studie

Tabelle 31: Konstrukte des lidA-CAPI (lidA-Hauptstudie: 1. Welle)

Arbeit	Fragen zur gegenwärtigen Erwerbstätigkeit inhaltlich und formal, Erwerbstätigkeit (j/n)
	Gedanke an Berufsaufgabe Arbeitgeber, Beruf, Erwerbstätigkeit
	Quantitative Anforderungen und Arbeitstempo Zeit, Geschwindigkeit
	Geforderte Verausgabung bei der Arbeitstätigkeit Zeitdruck, Unterbrechung, Verantwortung, Überstunden, körperliche Belastung, Quantität
	Physische Arbeitsexposition Stehen, Sitzen, Kälte, Hitze, Nässe, Feuchtigkeit, Zugluft, Lärm, Bücken, Hocken, Knien, Liegen, über Kopf, Heben, Tragen, einseitige Belastung
	Entwicklungsmöglichkeiten bei der Arbeit Initiative, Neues, Anwendung, Entwicklung
	Einfluss bei der Arbeit Kollegen, Menge, Inhalt
	Arbeitsmotivation Bedeutung, intrinsische Motivation, Verantwortung
	Wichtige Aspekte des Arbeitslebens Einkommen, Aufstieg, Interesse, Selbständigkeit, Helfen, Nützlichkeit, Zeiteinteilung, soziale Kontakte
	Veränderungen im Unternehmen Umstrukturierung, Entlassung
	Erfahrene oder erwartete Gratifikationen Wertschätzung, Gehalt/beruflicher Aufstieg, Arbeitsplatzsicherheit
	Hilfe und Unterstützung von Kollegen Unterstützung, Probleme, Rückmeldung
	Führungsqualität: Entwicklung, Zufriedenheit, Planung, Konflikte
	Altersdiskriminierung
Mobbing: Führung, Kollegen	
Arbeit-Familie/Privatleben-Konflikt Zeit, Energie	
Relevanz Stuserhalt Lohn/Gehalt, Anfahrtsweg, Wohnortwechsel, Überqualifikation	
Person	Geburtsjahr
	Geschlecht
	Migrationshintergrund
	Bildung Schule, Ausbildung
	Familienstand/Haushalt Partner, Haushaltsgröße, Kinder
	Einkommen Einzelperson und Haushalt
	Belastungen im Haushaltskontext Arbeitsweg, Haushalt und Kinder, Pflege
	Persönliche Hilfeleistungen Sorgen/Probleme, Aufgaben/Arbeiten, Ämter/Behörden, Geld, Besuch/Einladung
	Rauchen
	Größe
	Gewicht
	Körperliche Bewegung in der Freizeit
	Übersteigerte berufliche Verausgabungsneigung
	Selektion, Optimierung und Kompensation Arbeitsziele, Hindernisse, Coping
Positive und Negative Affektivität (Eigenschaft)	
Berufliche Weiterbildungsteilnahme	
Gesundheit	Krankheiten ärztliche Diagnosen (Gegenwart, Vergangenheit), Krankenkasse
	Schmerzen Körperregion, Einfluss Arbeit, Einfluss Freizeit
	Depressivität (Selbstausfüllfragebogen)
	Gesundheitliche Einschränkungen
	Inanspruchnahme Gesundheitsförderung und Rehabilitation
	Schlaf Einschlafen, Durchschlafen
Arbeitsfähigkeit	
Handgreifkraft (physische Messung)	

Tabelle 32: Soziodemografie (IidA: ursprüngliche Nettostichprobe)

		n	%
Gesamtzahl		6 585	100.0
Geburtsjahr	1959	2 908	44.2
	1965	3 677	55.8
Geschlecht	männlich	3 063	46.5
	weiblich	3 522	53.5
Familienstand	verheiratet, eingetragene Lebenspartnerschaft	4 798	72.9
	ledig	812	12.3
	geschieden	856	13.0
	verwitwet	106	1.6
	keine Angabe	15	0.2
Haushaltsgröße	1 Person	798	12.1
	2 Personen	1 885	28.6
	3 Personen	1 666	25.3
	4 Personen und mehr	2 229	33.8
	keine Angabe	7	0.1
Kinder im Haushalt unter 14 Jahren	keine Kinder	4 061	61.7
	1 Kind	1 108	16.8
	2 Kinder und mehr	611	9.3
	keine Angabe	805	12.2
Schulabschluss	ohne Schulabschluss	101	1.5
	Volks-, Hauptschulabschluss	1 680	25.5
	Realschulabschluss, polytechnische Oberschule	2 718	41.3
	Fach-/Hochschulreife, erweiterte Oberschule	2 042	31.0
	anderer Schulabschluss	44	0.7
Ausbildungsabschluss	ohne Ausbildungsabschluss	370	5.6
	beruflich-betrieblicher Ausbildungsabschluss	3 228	49.0
	beruflich-schulischer Ausbildungsabschluss	623	9.5
	Fach-, Meister-, Technikerschule etc.	992	15.1
	Fach-/Hochschulabschluss	1 315	20.0
	sonstiges	57	0.9
Stellung im Beruf	un-, angelernte Arbeiter	700	10.6
	Facharbeiter	617	9.4
	Vorarbeiter, Meister etc.	156	2.4
	Angestellte ohne Personalverantwortung	2 302	35.0
	Angestellte mit Personalverantwortung	2 363	35.9
	keine Angabe	447	6.7
Wochenarbeitszeit inkl. Überstunden	unter 20 Stunden	644	9.8
	20 bis unter 30 Stunden	823	12.5
	30 bis unter 40 Stunden	1 592	24.2
	40 Stunden und mehr	3 154	47.9
	schwankt beträchtlich	124	1.9
	keine Angabe	248	3.8
Nettoeinkommen persönlich	unter 1.000 EUR	1 586	24.1
	1.000 bis unter 2.000 EUR	2 614	39.7
	2.000 bis unter 3.000 EUR	1 306	19.8
	3.000 bis unter 4.000 EUR	370	5.6
	4.000 EUR und mehr	307	4.7
	keine Angabe	402	6.1

NEXT-Studie

Tabelle 33: Iteminhalt (NEXT-Studie)

Item-Nr.	Iteminhalt	AO
EM-1	Erfordert Ihre Arbeit Eigeninitiative?	1
EM-2	Haben Sie durch Ihre Arbeit die Möglichkeiten, etwas Neues zu lernen?	1
EM-3	Können Sie Ihre Fähigkeiten oder Sachkenntnisse bei Ihrer Arbeit einsetzen?	1
FQ-1	Bitte schätzen Sie ein, in welchem Maß Ihr direkter Vorgesetzter sicherstellt, dass jedes einzelne Teammitglied gute Entwicklungsmöglichkeiten hat.	1
FQ-2	Bitte schätzen Sie ein, in welchem Maß Ihr direkter Vorgesetzter hohen Wert auf Arbeitszufriedenheit legt.	1
FQ-3	Bitte schätzen Sie ein, in welchem Maß Ihr direkter Vorgesetzter die Arbeit gut planen kann.	1
FQ-4	Bitte schätzen Sie ein, in welchem Maß Ihr direkter Vorgesetzter Konflikte gut lösen kann.	1
AK-a^N	<i>Die Anforderungen meiner Arbeit behindern mein Privat-/Familienleben.</i>	2
AK-b^N	<i>Die Zeit, die meine Arbeit einnimmt, macht es schwer, meinen privaten oder familiären Pflichten nachzukommen.</i>	2
AK-c^N	<i>Dinge, die ich zu Hause tun möchte, schaffe ich nicht, weil mich meine Arbeit so fordert.</i>	2
AK-d^N	<i>Die Arbeit belastet mich so sehr, dass es schwierig ist, meinen privaten oder familiären Pflichten nachzukommen.</i>	2
AK-e^N	<i>Aufgrund von Verpflichtungen durch die Arbeit muss ich Pläne für private oder familiäre Aktivitäten ändern.</i>	2
AF-1	Ist Ihre Arbeit...	3
AF-2	Wie schätzen Sie Ihre derzeitige Arbeitsfähigkeit in Bezug auf die körperlichen Arbeitsanforderungen ein?	4
AF-3	Wie schätzen Sie Ihre derzeitige Arbeitsfähigkeit in Bezug auf die psychischen Arbeitsanforderungen ein?	4
GB	Wie oft im Laufe der letzten 12 Monate haben Sie daran gedacht, Ihren Pflegeberuf aufzugeben und eine andere berufliche Tätigkeit zu beginnen?	5
GG	Wie würden Sie Ihren Gesundheitszustand im Allgemeinen beschreiben?	6

^N. Dieses Item wurde nicht im lidA-CAPI verwendet. Anmerkungen: **1-AO**: (1) in sehr geringem Maß, (2) in geringem Maß, (3) etwas, (4) in gewissem Maß und (5) in großem Maß. **2-AO**: (1) stimme nicht zu (2) 2, (3) 3, (4) 4 und (5) stimme völlig zu. **3-AO** (dieses Item wurde indirekt über die Abfrage des Arbeitsbereiches bearbeitet): (1) vorwiegend psychisch (geistig) belastend, (2) vorwiegend physisch (körperlich) belastend? und (3) psychisch (geistig) und physisch (körperlich) gleichermaßen belastend oder unbelastend? **4-AO**: (1) sehr schlecht, (2) eher schlecht, (3) mittelmäßig, (4) eher gut und (5) sehr gut. **5-AO**: (1) nie, (2) einige Male im Jahr, (3) einige Male im Monat, (4) einige Male pro Woche und (5) jeden Tag. **6-AO**: (1) schlecht, (2) weniger gut, (3) gut, (4) sehr gut und (5) ausgezeichnet.

Tabelle 34: Soziodemografie (NEXT-Studie)

		n	%
Gesamtzahl		3 298	100.0
Alter	18 bis unter 30 Jahre	627	19.0
AM = 39.1 Jahre	30 bis unter 40 Jahre	1 047	31.7
SD = 9.3 Jahre	40 bis unter 50 Jahre	1 138	34.5
	50 Jahre und älter	477	14.5
	keine Angabe	9	0.3
Geschlecht	männlich	554	16.8
	weiblich	2 741	83.1
	keine Angabe	3	0.1
Wohnsituation	allein	687	20.8
	mit Erwachsenem, ohne Kind	1 026	31.1
	mit Erwachsenem, mit Kind/ern	1 339	40.6
	ohne Erwachsenem, mit Kind/ern	221	6.7
	keine Angabe	25	0.8
Kinder im Haushalt	keine Kinder	1 726	52.3
	1 Kind und mehr	1 548	47.0
	keine Angabe	27	0.7
Schulabschluss	ohne Abschluss	8	.2
	Volks-, Hauptschulabschluss	380	11.5
	Berufsfachschule/mittlere Reife	2 051	62.2
	Fach-/Abitur	832	25.2
	keine Angabe	27	0.8
Art der Einrichtung	Krankenhaus	2 451	74.3
	Pflegeheim	477	14.5
	Ambulante Pflege	293	8.9
	keine Angabe	77	2.3
Stellung im Beruf	ohne leitende Position	2 638	80.0
	stellvertretende leitende Position	295	8.9
	leitende Position	339	10.3
	keine Angabe	26	0.8
Wochenarbeitszeit inkl. Überstunden	unter 20 Stunden	397	12.0
	20 bis unter 30 Stunden	446	13.5
	30 bis unter 40 Stunden	1 259	38.2
	40 Stunden und mehr	1 028	31.2
	keine Angabe	168	5.1

Anmerkungen: Alle TN haben zum Zeitpunkt der Datenerhebung eine sozialversicherungspflichtige Beschäftigung.

Tabelle 35: Verteilung der Werte (NEXT-Studie)

	RW 0 n (%)	RW 25 n (%)	RW 50 n (%)	RW 75 n (%)	RW 100 n (%)
EM-1	51 (1.5)	115 (3.5)	215 (6.5)	1 304 (39.5)	1 613 (48.9)
EM-2	113 (3.4)	320 (9.7)	462 (14.0)	1 592 (48.3)	811 (24.6)
EM-3	55 (1.7)	154 (4.7)	274 (8.3)	1 571 (47.6)	1 244 (37.7)
FQ-1	280 (8.6)	496 (15.3)	556 (17.2)	1 348 (41.6)	561 (17.3)
FQ-2	231 (7.1)	421 (13.0)	460 (14.2)	1 231 (38.0)	898 (27.7)
FQ-3	200 (6.2)	386 (11.9)	565 (17.4)	1 399 (43.2)	691 (21.3)
FQ-4	355 (10.8)	512 (15.5)	592 (18.0)	1 240 (37.6)	542 (16.4)

Anmerkungen: EM: n = 3 298; FQ: n = 3 241.

Tabelle 36: Tests auf Normalverteilung (NEXT-Studie)

	Kolmogorov-Smirnov-Anpassungstest				Shapiro-Wilk-Test			
	D-Stat.	p-Wert	D-Krit.	α	W-Stat.	p-Wert	W-Krit.	α
EM-1	.28	.000	.018	.01	.73	.000	.996	.01
EM-2	.30	.000	.018	.01	.84	.000	.996	.01
EM-3	.29	.000	.018	.01	.78	.000	.996	.01
EM-Skala	.17	.000	.018	.01	.89	.000	.996	.01
FQ-1	.27	.000	.018	.01	.87	.000	.995	.01
FQ-2	.27	.000	.018	.01	.86	.000	.995	.01
FQ-3	.28	.000	.018	.01	.86	.000	.995	.01
FQ-4	.25	.000	.018	.01	.88	.000	.995	.01
FQ-Skala	.14	.000	.018	.01	.94	.000	.995	.01

Anmerkungen: **EM**: n = 3 298; **FQ**: n = 3 241.

Tabelle 37: Item- und Skalenschwierigkeit (NEXT-Studie)

	1. Quartil	Md	3. Quartil	AM	SD
EM-1	75.00	75.00	100.00	82.69	21.48
EM-2	50.00	75.00	75.00	70.22	25.57
EM-3	75.00	75.00	100.00	78.77	22.00
EM-Skala	66.67	83.33	91.67	77.23	18.24
FQ-1	50.00	75.00	75.00	60.91	29.76
FQ-2	50.00	75.00	100.00	66.54	30.27
FQ-3	50.00	75.00	75.00	65.39	28.20
FQ-4	25.00	75.00	75.00	58.50	30.95
FQ-Skala	43.75	68.75	81.25	62.83	26.45

Anmerkungen: **EM**: n = 3 298; **FQ**: n = 3 241.

Tabelle 38: Trennschärfe (NEXT-Studie)

	nicht-korrigierte Trennschärfe		korrigierte Trennschärfe	
	ρ	p-Wert	ρ	p-Wert
EM-1	.70 ***	.000	.42 ***	.000
EM-2	.81 ***	.000	.49 ***	.000
EM-3	.76 ***	.000	.49 ***	.000
FQ-1	.86 ***	.000	.76 ***	.000
FQ-2	.90 ***	.000	.81 ***	.000
FQ-3	.87 ***	.000	.78 ***	.000
FQ-4	.89 ***	.000	.80 ***	.000

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **EM**: n = 3 298; **FQ**: n = 3 241.

Tabelle 39: Homogenität (NEXT-Studie)

	Minimum	Inter-Item-Korrelation AM	Maximum
EM-Skala	.36 ***	.39	.45 ***
FQ-Skala	.65 ***	.71	.75 ***

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **EM**: n = 3 298; **FQ**: n = 3 241.

Tabelle 40: Nomologisches Netzwerk (NEXT-Studie)

	EM	FQ	AK	AF	GB	GG
EM-Skala		.27 ***	-.06 ***	.18 ***	-.21 ***	.11 ***
FQ-Skala			-.22 ***	.20 ***	-.24 ***	.13 ***
AK-Skala				-.38 ***	.30 ***	-.28 ***
AF-Skala					-.27 ***	.56 ***
GB						-.15 ***
GG						

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **EM**: n = 3 298; **FQ**: n = 3 241; **AK**: n = 3 298; **AF**: n = 3 298; **GB**: n = 3 298; **GG**: n = 3 298.

COPSOQ-I-Studie

Tabelle 41: Iteminhalt 1/2 (COPSOQ-I-Studie)

Item-Nr.	Iteminhalt	AO
QA-1	Ist Ihre Arbeit ungleich verteilt, so dass sie sich auftürmt?	1
QA-2	Wie oft kommt es vor, dass Sie nicht genügend Zeit haben, alle Ihre Aufgaben zu erledigen?	1
QA-3	Kommen Sie mit Ihrer Arbeit in Rückstand?	1
QA-4	Haben Sie genug Zeit für Ihre Arbeitsaufgaben? (invertiert)	1
EA-1	Haben Sie Einfluss darauf, mit wem Sie arbeiten?	1
EA-2	Haben Sie Einfluss auf die Menge der Arbeit, die Ihnen übertragen wird?	1
EA-3	Haben Sie Einfluss darauf, was Sie bei Ihrer Arbeit tun?	1
EM-1	Verlangt es Ihre Arbeit, dass Sie die Initiative ergreifen?	2
EM-2	Haben Sie die Möglichkeit, durch Ihre Arbeit neue Dinge zu erlernen?	2
EM-3	Können Sie Ihre Fertigkeiten oder Ihr Fachwissen bei Ihrer Arbeit anwenden?	2
EM-4	Ermöglicht es Ihrer Arbeit, Ihre Fertigkeiten weiter zu entwickeln?	2
FQ-1	Bitte schätzen Sie ein, in welchem Maß Ihr unmittelbarer Vorgesetzter für gute Entwicklungsmöglichkeiten der einzelnen Mitarbeiter sorgt?	2
FQ-2	Bitte schätzen Sie ein, in welchem Maß Ihr unmittelbarer Vorgesetzter der Arbeitszufriedenheit einen hohen Stellenwert beimisst?	2
FQ-3	Bitte schätzen Sie ein, in welchem Maß Ihr unmittelbarer Vorgesetzter die Arbeit gut plant?	2
FQ-4	Bitte schätzen Sie ein, in welchem Maß Ihr unmittelbarer Vorgesetzter Konflikte gut löst?	2
HU-1	Wie oft erhalten Sie Hilfe und Unterstützung von Ihren Kollegen?	1
HU-2	Wie oft sind Ihre Kollegen bereit, sich Ihre Arbeitsprobleme anzuhören?	1
HU-3	Wie oft sprechen Ihre Kollegen mit Ihnen über die Qualität Ihrer Arbeit?	1
AK-a ^N	<i>Die Anforderungen meiner Arbeit stören mein Privat- und Familienleben.</i>	3
AK-b ^N	<i>Der Zeitaufwand meiner Arbeit macht es schwierig für mich, meinen Pflichten in der Familie oder im Privatleben nachzukommen.</i>	3
AK-c ^N	<i>Dinge, die ich zu Hause machen möchte, bleiben wegen der Anforderungen meiner Arbeit liegen.</i>	3
AK-d ^N	<i>Meine Arbeit erzeugt Stress, der es schwierig macht, privaten oder familiären Verpflichtungen nachzukommen.</i>	3
AK-e ^N	<i>Wegen beruflicher Verpflichtungen muss ich Pläne für private oder Familienaktivitäten ändern.</i>	3

^N. Dieses Item wurde nicht im lidA-CAPI verwendet. Anmerkungen: **1-AO**: (1) nie oder fast nie, (2) selten, (3) manchmal, (4) oft und (5) immer. **2-AO**: (1) in sehr geringem Maße, (2) in geringem Maße, (3) zum Teil, (4) in hohem Maße und (5) in sehr hohem Maße. **3-AO**: (1) stimme nicht zu, (2) stimme eher nicht zu, (3) unentschieden, (4) stimme eher zu und (5) stimme voll zu.

Tabelle 42: Iteminhalt 2/2 (COPSOQ-I-Studie)

Item-Nr.	Iteminhalt	AO
AF-1	Ist Ihre Arbeit ...	4
AF-2	Wie schätzen Sie Ihre jetzige Arbeitsfähigkeit im Verhältnis zu den physischen (körperlichen) Arbeitsanforderungen ein?	5
AF-3	Wie schätzen Sie Ihre jetzige Arbeitsfähigkeit im Verhältnis zu den psychischen (geistigen) Arbeitsanforderungen ein?	5
GB	Wie oft im Laufe der letzten 12 Monate haben Sie daran gedacht Ihren Beruf aufzugeben?	6
GG	Ihr Gesundheitszustand: Wenn Sie den besten denkbaren Gesundheitszustand mit 10 Punkten bewerten und den schlechtesten denkbaren mit 0 Punkten: Wie viele Punkte vergeben Sie dann für Ihren derzeitigen Gesundheitszustand? Bitte kreuzen Sie die entsprechende Zahl an.	7

4-AO: (1) vorwiegend psychisch (geistig) belastend, (2) vorwiegend physisch (körperlich) belastend? und (3) psychisch (geistig) und physisch (körperlich) gleichermaßen belastend oder unbelastend? **5-AO:** (1) sehr schlecht, (2), eher schlecht, (3) mittelmäßig, (4) eher gut und (5) sehr gut. **6-AO:** (1) nie, (2) einige Male im Jahr, (3) einige Male im Monat, (4) einige Male in der Woche und (5) jeden Tag. **7-AO:** (1) 0 = schlechtesten, denkbaren Gesundheitszustand, (2) 1, [...], (10) 9 und (11) 10 = bester denkbaren Gesundheitszustand.

Tabelle 43: Soziodemografie (COPSOQ-I-Studie)

		n	%
Gesamtzahl		2 561	100.0
Alter	AM: 42.6 und SD: 10.6		
Geschlecht	männlich	1 077	42.1
	weiblich	1 455	56.8
	keine Angabe	29	1.1
Familienstand	verheiratet	1 333	52.1
	ledig	631	24.6
	geschieden	232	9.1
	verwitwet	22	0.9
	keine Angabe	343	13.4
Wohnsituation	mit Partner	1 719	67.1
	ohne Partner	612	23.9
	keine Angabe	230	9.0
Kinder im Haushalt unter 15 Jahren	ja	715	27.9
	nein	1 284	50.1
	keine Angabe	562	21.9
Schulabschluss	ohne Schulabschluss	17	0.7
	Hauptschulabschluss	251	9.8
	Mittlere Reife, POS	790	30.9
	Fachhochschulreife, Abitur	1 347	52.6
	anderer Schulabschluss	30	1.2
	keine Angabe	126	4.9
Ausbildungsabschluss	ohne Ausbildungsabschluss	112	4.3
	beruflich-betriebliche Ausbildungsabschluss	834	32.6
	beruflich-schulische Ausbildungsabschluss	332	13.0
	Fach-, Meister-, Technikerschule etc.	138	5.4
	Fach-/Hochschulabschluss	1 059	41.3
	anderer Ausbildungsabschluss	154	6.0
	keine Angabe	342	13.4
Stellung im Beruf	Auszubildende/r	45	1.8
	un-/angelernte/r	33	1.3
	Facharbeiter/in, Vorarbeiter/in	21	0.8
	Angestellte/r	1 699	66.3
	Beamte/r	662	25.9
	Selbständige/r	12	0.4
	keine Angabe	89	3.5
Wochenarbeitszeit inkl. Überstunden	Ausbildung, Lehre, Umschulung	50	2.0
	15 bis unter 34 Stunden	621	24.2
	35 Stunden und mehr	1 797	70.2
	keine Angabe	17	0.7

Anmerkungen: Alle TN gehen zum Zeitpunkt der Datenerhebung einer regelmäßigen Erwerbstätigkeit nach.

Tabelle 44: Verteilung der Werte (COPSOQ-I-Studie)

	RW 0 n (%)	RW 25 n (%)	RW 50 n (%)	RW 75 n (%)	RW 100 n (%)	k. A. n (%)
QA-1	152 (6.1)	414 (16.5)	979 (39.1)	822 (32.8)	136 (5.4)	58 (2.3)
QA-2	130 (5.1)	532 (21.0)	910 (36.0)	810 (32.0)	147 (5.8)	32 (1.2)
QA-3	379 (15.0)	750 (29.7)	978 (38.7)	339 (13.4)	78 (3.1)	37 (1.4)
QA-4	149 (5.9)	630 (25.1)	832 (33.1)	774 (30.8)	127 (5.1)	49 (1.9)
EA-1	802 (31.6)	812 (32.0)	586 (23.1)	269 (10.6)	67 (2.6)	25 (1.0)
EA-2	786 (31.0)	902 (35.6)	601 (23.7)	211 (8.3)	36 (1.4)	25 (1.0)
EA-3	154 (6.1)	315 (12.5)	605 (24.0)	997 (39.5)	450 (17.9)	40 (1.6)
EM-1	37 (1.5)	110 (4.3)	674 (26.6)	1 071 (42.3)	641 (25.3)	28 (1.1)
EM-2	66 (2.6)	222 (8.8)	1 158 (45.7)	788 (31.1)	298 (11.8)	29 (1.1)
EM-3	35 (1.4)	69 (2.7)	541 (21.3)	1 234 (48.7)	656 (25.9)	26 (1.0)
EM-4	109 (4.3)	302 (11.9)	1 105 (43.7)	766 (30.3)	248 (9.8)	31 (1.2)
FQ-1	259 (10.5)	481 (19.5)	941 (38.1)	614 (24.8)	177 (7.2)	89 (3.5)
FQ-2	220 (9.0)	398 (16.2)	828 (33.7)	747 (30.4)	262 (10.7)	106 (4.1)
FQ-3	195 (8.1)	428 (17.7)	897 (37.1)	711 (29.4)	187 (7.3)	143 (5.6)
FQ-4	350 (14.3)	470 (19.2)	828 (33.9)	604 (24.7)	194 (7.9)	115 (4.5)
HU-1	41 (1.6)	156 (6.2)	635 (25.2)	1 146 (45.4)	544 (21.6)	39 (1.5)
HU-2	34 (1.4)	188 (7.5)	574 (22.9)	1 159 (46.3)	550 (22.0)	56 (2.2)
HU-3	295 (11.8)	712 (28.4)	900 (35.9)	525 (21.0)	73 (2.9)	56 (2.2)

Anmerkungen: k. A. = keine Angabe. n = 2 561.

Tabelle 45: Tests auf Normalverteilung (COPSOQ-I-Studie)

	Kolmogorov-Smirnov-Anpassungstest				Shapiro-Wilk-Test			
	D-Stat.	p-Wert	D-Krit.	α	W-Stat.	p-Wert	W-Krit.	α
QA-1	.21	.000	.022	.01	.89	.000	.995	.01
QA-2	.20	.000	.022	.01	.90	.000	.995	.01
QA-3	.21	.000	.022	.01	.90	.000	.995	.01
QA-4	.19	.000	.022	.01	.90	.000	.995	.01
QA-Skala	.07	.000	.022	.01	.99	.000	.995	.01
EA-1	.21	.000	.022	.01	.87	.000	.995	.01
EA-2	.22	.000	.022	.01	.87	.000	.995	.01
EA-3	.25	.000	.022	.01	.89	.000	.995	.01
EA-Skala	.09	.000	.022	.01	.98	.000	.995	.01
EM-1	.24	.000	.022	.01	.86	.000	.995	.01
EM-2	.25	.000	.022	.01	.88	.000	.995	.01
EM-3	.27	.000	.022	.01	.84	.000	.995	.01
EM-4	.23	.000	.022	.01	.89	.000	.995	.01
EM-Skala	.10	.000	.022	.01	.97	.000	.995	.01
FQ-1	.21	.000	.022	.01	.91	.000	.995	.01
FQ-2	.19	.000	.022	.01	.91	.000	.995	.01
FQ-3	.20	.000	.022	.01	.91	.000	.995	.01
FQ-4	.19	.000	.022	.01	.91	.000	.995	.01
FQ-Skala	.08	.000	.022	.01	.98	.000	.995	.01
HU-1	.26	.000	.022	.01	.86	.000	.995	.01
HU-2	.27	.000	.022	.01	.86	.000	.995	.01
HU-3	.20	.000	.022	.01	.91	.000	.995	.01
HU-Skala	.12	.000	.022	.01	.97	.000	.995	.01

Anmerkungen: Fehlende Werte sind listenweise ausgeschlossen. n = 2 232.

Tabelle 46: Item- und Skalenschwierigkeit (COPSOQ-I-Studie)

	1. Quartil	Md	3. Quartil	AM	SD
QA-1	50.00	50.00	75.00	53.72	24.04
QA-2	25.00	50.00	75.00	53.11	24.25
QA-3	25.00	50.00	50.00	39.78	24.75
QA-4	25.00	50.00	75.00	48.80	24.71
QA-Skala	---	50.00	---	48.85	20.05
EA-1	0.00	25.00	50.00	29.72	26.61
EA-2	0.00	25.00	50.00	28.16	24.58
EA-3	50.00	75.00	75.00	61.86	27.41
EA-Skala	---	41.67	---	39.91	19.26
EM-1	50.00	75.00	75.00	71.28	22.18
EM-2	50.00	50.00	75.00	60.44	22.26
EM-3	50.00	75.00	100.00	73.81	20.84
EM-4	50.00	50.00	75.00	57.66	23.53
EM-Skala	---	68.75	---	65.80	17.35
FQ-1	25.00	50.00	75.00	50.02	26.62
FQ-2	50.00	50.00	75.00	54.70	27.53
FQ-3	25.00	50.00	75.00	52.84	26.08
FQ-4	25.00	50.00	75.00	48.23	28.71
FQ-Skala	---	50.00	---	51.45	23.72
HU-1	50.00	75.00	75.00	70.44	22.07
HU-2	50.00	75.00	75.00	70.72	22.13
HU-3	25.00	50.00	50.00	43.87	24.94
HU-Skala	---	66.67	---	61.68	17.83

Anmerkungen: Fehlende Werte sind listenweise ausgeschlossen. n = 2 232.

Tabelle 47: Trennschärfe (COPSOQ-I-Studie)

	nicht-korrigierte Trennschärfe		korrigierte Trennschärfe	
	ρ	p-Wert	ρ	p-Wert
QA-1	.78 ***	.000	.62 ***	.000
QA-2	.87 ***	.000	.75 ***	.000
QA-3	.85 ***	.000	.72 ***	.000
QA-4	.78 ***	.000	.60 ***	.000
EA-1	.77 ***	.000	.42 ***	.000
EA-2	.73 ***	.000	.41 ***	.000
EA-3	.67 ***	.000	.27 ***	.000
EM-1	.68 ***	.000	.43 ***	.000
EM-2	.78 ***	.000	.59 ***	.000
EM-3	.72 ***	.000	.53 ***	.000
EM-4	.82 ***	.000	.63 ***	.000
FQ-1	.85 ***	.000	.74 ***	.000
FQ-2	.88 ***	.000	.78 ***	.000
FQ-3	.84 ***	.000	.72 ***	.000
FQ-4	.88 ***	.000	.77 ***	.000
HU-1	.81 ***	.000	.57 ***	.000
HU-2	.79 ***	.000	.53 ***	.000
HU-3	.72 ***	.000	.34 ***	.000

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: Fehlende Werte sind innerhalb der Skalen listenweise ausgeschlossen. **QA:** n = 2 463; **EA:** n = 2 511; **EM:** n = 2 511; **FQ:** n = 2 374; **HU:** n = 2 480.

Tabelle 48: Homogenität (COPSOQ-I-Studie)

	Inter-Item-Korrelation		
	Minimum	AM	Maximum
QA-Skala	.46 ***	.57	.71 ***
EA-Skala	.22 ***	.30	.43 ***
EM-Skala	.33 ***	.44	.62 ***
FQ-Skala	.61 ***	.67	.71 ***
HU-Skala	.29 ***	.43	.63 ***

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: Fehlende Werte sind innerhalb der Skalen listenweise ausgeschlossen. **QA:** n = 2 463; **EA:** n = 2 511; **EM:** n = 2 511; **FQ:** n = 2 374; **HU:** n = 2 480.

Tabelle 49: Nomologisches Netzwerk (COPSOQ-I-Studie)

	QA	EA	EM	FQ	HU	AK	AF	GB	GG
QA-Skala		-.09 ***	-.00	-.19 ***	-.16 ***	.43 ***	-.20 ***	.24 ***	-.11 ***
EA-Skala			.28 ***	.17 ***	.11 ***	-.09 ***	.14 ***	-.13 ***	.10 ***
EM-Skala				.20 ***	.22 ***	.10 ***	.20 ***	-.17 ***	.16 ***
FQ-Skala					.27 ***	-.22 ***	.20 ***	-.23 ***	.19 ***
HU-Skala						-.15 ***	.16 ***	-.17 ***	.17 ***
AK-Skala							-.23 ***	.32 ***	-.26 ***
AF-Skala								-.38 ***	.52 ***
GB									-.29 ***
GG									

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: Fehlende Werte sind innerhalb der Skalen listenweise und zwischen den Skalen fallweise ausgeschlossen. **QA:** n = 2.463; **EA:** n = 2 511; **EM:** n = 2 511; **FQ:** n = 2 374; **HU:** n = 2 480; **AK:** n = 2 534; **AF:** n = 2 405; **GB:** n = 1 948; **GG:** n = 2 427.

IPSO-Studie

Tabelle 50: Iteminhalt (IPSO-Studie)

Item-Nr.	Iteminhalt	AO
QA-1	Ist Ihre Arbeit ungleich verteilt, so dass sie sich auftürmt?	1
QA-2	Wie oft kommt es vor, dass Sie nicht genügend Zeit haben, alle Ihre Aufgaben zu erledigen?	1
QA-3	Geraten Sie mit Ihrer Arbeit in Rückstand?	1
QA-4	Haben Sie genug Zeit für Ihre Arbeitsaufgaben? (AO invertiert)	1
EA-1	Haben Sie Einfluss darauf, mit wem Sie arbeiten?	1
EA-2	Haben Sie Einfluss auf die Menge der Arbeit, die Ihnen übertragen wird?	1
EA-3	Haben Sie Einfluss darauf, was Sie bei Ihrer Arbeit tun?	1
EM-1	Verlangt es Ihre Arbeit, dass Sie die Initiative ergreifen?	2
EM-2	Haben Sie die Möglichkeit, durch Ihre Arbeit neue Dinge zu erlernen?	2
EM-3	Können Sie Ihre Fertigkeiten oder Ihr Fachwissen bei Ihrer Arbeit anwenden?	2
EM-4	Gibt Ihnen Ihre Arbeit die Möglichkeit, Ihre Fertigkeiten weiter zu entwickeln?	2
HU-1	Wie oft erhalten Sie Hilfe und Unterstützung von Ihren Kollegen?	1
HU-2	Wie oft sind Ihre Kollegen bereit, sich Ihre Arbeitsprobleme anzuhören?	1
HU-3	Wie oft sprechen Ihre Kollegen mit Ihnen über die Qualität Ihrer Arbeit?	1
AK-a ^N	<i>Die Anforderungen meiner Arbeit stören mein Privat- und Familienleben.</i>	3
AK-b ^N	<i>Der Zeitaufwand meiner Arbeit macht es schwierig für mich, meinen Pflichten in der Familie oder im Privatleben nachzukommen.</i>	3
AK-c ^N	<i>Dinge, die ich zu Hause machen möchte, bleiben wegen der Anforderungen meiner Arbeit liegen.</i>	3
AK-d ^N	<i>Meine Arbeit erzeugt Stress, der es schwierig macht, privaten oder familiären Verpflichtungen nachzukommen.</i>	3
AK-e ^N	<i>Wegen beruflicher Verpflichtungen muss ich Pläne für private oder Familienaktivitäten ändern.</i>	3
GB	Wie oft haben Sie im Laufe der letzten 12 Monate daran gedacht, Ihren Beruf aufzugeben?	4
GG	Wenn Sie den besten denkbaren Gesundheitszustand mit 10 Punkten bewerten und den schlechtesten denkbaren mit 0 Punkten: Wie viele Punkte vergeben Sie dann für Ihren derzeitigen Gesundheitszustand?	5

^N. Dieses Item wurde nicht im lidA-CAPI verwendet. Anmerkungen: **1-AO**: (1) nie oder fast nie, (2) selten, (3) manchmal, (4) oft und (5) immer. **2-AO**: (1) in sehr geringem Maß, (2) in geringem Maß, (3) zum Teil, (4) in hohem Maß und (5) in sehr hohem Maß. **3-AO**: (1) stimme nicht zu (2) stimme eher nicht zu, (3) unentschieden, (4) stimme eher zu und (5) stimme voll zu. **4-AO**: (1) nie, (2) einige Male im Jahr, (3) einige Male im Monat, (4) einige Male in der Woche, (5) jeden Tag. **5-AO**: (1) 0 = schlechtesten denkbaren Gesundheitszustand, (2) 1, [...], (10) 9 und (11) 10 = besten denkbaren Gesundheitszustand.

Tabelle 51: Soziodemografie (IPSO-Studie)

		n	%
Gesamtzahl		1 238	100.0
Alter	18 bis unter 30 Jahre	93	7.5
	30 bis unter 45 Jahre	517	41.8
	45 Jahre und älter	614	49.6
	keine Angabe	14	1.1
Geschlecht	männlich	430	34.7
	weiblich	792	64.0
	keine Angabe	16	1.3
Berufsbezeichnung	Richter	172	13.9
	Staatsanwalt	61	4.9
	Rechtspfleger	179	14.5
	Amtsanwalt	19	1.5
	Justizfachwirt, Beamter mittlerer Justizdienst	248	20.0
	Justizfachangestellter	352	28.4
	Justizwachtmeister, Justizhelfer	75	6.1
	Gerichtsvollzieher	46	3.7
	Fachkraft ambulanter sozialer Dienst	50	4.0
	Sonstige	27	2.2
	keine Angabe	9	0.7
Wochenarbeitszeit	unter 20 Stunden	107	8.6
Arbeitsvertrag	20 bis 35 Stunden	224	18.1
	mehr als 35 Stunden	883	71.3
	keine Angabe	24	1.9

Anmerkungen: Alle TN haben zum Zeitpunkt der Datenerhebung eine sozialversicherungspflichtige Beschäftigung.

Tabelle 52: Verteilung der Werte (IPSO-Studie)

	RW 0 n (%)	RW 25 n (%)	RW 50 n (%)	RW 75 n (%)	RW 100 n (%)
QA-1	105 (8.5)	228 (18.4)	557 (45.0)	293 (23.7)	55 (4.4)
QA-2	54 (4.4)	217 (17.5)	431 (34.8)	429 (34.7)	107 (8.6)
QA-3	138 (11.1)	330 (26.7)	501 (40.5)	206 (16.6)	63 (5.1)
QA-4	138 (11.1)	339 (27.4)	332 (26.8)	353 (28.5)	76 (6.1)
EA-1	711 (57.4)	301 (24.3)	133 (10.7)	73 (5.9)	20 (1.6)
EA-2	828 (66.9)	262 (21.2)	99 (8.0)	41 (3.3)	8 (0.8)
EA-3	258 (20.8)	253 (20.4)	247 (20.0)	293 (23.7)	187 (15.1)
EM-1	94 (7.6)	195 (15.8)	389 (31.4)	327 (26.4)	233 (18.8)
EM-2	155 (12.5)	234 (18.9)	521 (42.1)	229 (18.5)	99 (8.0)
EM-3	53 (4.3)	75 (6.1)	341 (27.5)	472 (38.1)	297 (24.0)
EM-4	175 (14.1)	261 (21.1)	493 (39.8)	228 (18.4)	81 (6.5)
HU-1	100 (8.2)	193 (15.8)	407 (33.4)	338 (27.7)	182 (14.9)
HU-2	73 (6.0)	153 (12.5)	305 (25.0)	395 (32.4)	294 (24.1)
HU-3	249 (20.4)	363 (29.8)	358 (29.3)	210 (17.0)	40 (3.2)

Anmerkungen: **QA:** n = 1 238; **EA:** n = 1 238; **EM:** n = 1 238; **HU:** n = 1 220.

Tabelle 53: Tests auf Normalverteilung (IPSO-Studie)

	Kolmogorov-Smirnov-Anpassungstest				Shapiro-Wilk-Test			
	D-Stat.	p-Wert	D-Krit.	α	W-Stat.	p-Wert	W-Krit.	α
QA-1	.24	.000	.029	.01	.90	.000	.995	.01
QA-2	.21	.000	.029	.01	.90	.000	.995	.01
QA-3	.21	.000	.029	.01	.91	.000	.995	.01
QA-4	.18	.000	.029	.01	.91	.000	.995	.01
QA-Skala	.07	.000	.029	.01	.99	.000	.995	.01
EA-1	.34	.000	.029	.01	.72	.000	.995	.01
EA-2	.39	.000	.029	.01	.65	.000	.995	.01
EA-3	.17	.000	.029	.01	.89	.000	.995	.01
EA-Skala	.12	.000	.029	.01	.93	.000	.995	.01
EM-1	.17	.000	.029	.01	.91	.000	.995	.01
EM-2	.22	.000	.029	.01	.91	.000	.995	.01
EM-3	.23	.000	.029	.01	.87	.000	.995	.01
EM-4	.21	.000	.029	.01	.91	.000	.995	.01
EM-Skala	.08	.000	.029	.01	.98	.000	.995	.01
HU-1	.17	.000	.030	.01	.91	.000	.995	.01
HU-2	.21	.000	.030	.01	.89	.000	.995	.01
HU-3	.19	.000	.030	.01	.90	.000	.995	.01
HU-Skala	.10	.000	.030	.01	.98	.000	.995	.01

Anmerkungen: QA: n = 1 238; EA: n = 1 238; EM: n = 1 238; HU: n = 1 220.

Tabelle 54: Item- und Skalenschwierigkeit (IPSO-Studie)

	1. Quartil	Md	3. Quartil	AM	SD
QA-1	25.00	50.00	75.00	49.29	24.21
QA-2	50.00	50.00	75.00	56.42	24.71
QA-3	25.00	50.00	50.00	44.47	25.42
QA-4	25.00	50.00	75.00	52.22	27.88
QA-Skala	37.50	50.00	68.75	50.60	21.81
EA-1	.00	.00	25.00	17.49	24.67
EA-2	.00	.00	25.00	12.42	20.72
EA-3	25.00	50.00	75.00	47.94	34.22
EA-Skala	8.33	25.00	33.33	25.95	19.97
EM-1	50.00	50.00	75.00	58.28	29.26
EM-2	25.00	50.00	75.00	47.64	27.23
EM-3	50.00	75.00	75.00	67.87	25.77
EM-4	25.00	50.00	50.00	45.54	27.29
EM-Skala	43.75	56.25	68.75	54.83	22.20
HU-1	50.00	50.00	75.00	56.33	28.47
HU-2	50.00	75.00	75.00	64.02	28.93
HU-3	25.00	25.00	50.00	38.30	27.38
HU-Skala	33.33	58.33	66.67	52.88	22.73

Anmerkungen: QA: n = 1 238; EA: n = 1 238; EM: n = 1 238; HU: n = 1 220.

Tabelle 55: Trennschärfe (IPSO-Studie)

	nicht-korrigierte Trennschärfe		korrigierte Trennschärfe	
	ρ	p-Wert	ρ	p-Wert
QA-1	.79 ***	.000	.65 ***	.000
QA-2	.90 ***	.000	.82 ***	.000
QA-3	.85 ***	.000	.74 ***	.000
QA-4	.86 ***	.000	.73 ***	.000
EA-1	.67 ***	.000	.40 ***	.000
EA-2	.62 ***	.000	.40 ***	.000
EA-3	.84 ***	.000	.34 ***	.000
EM-1	.77 ***	.000	.57 ***	.000
EM-2	.83 ***	.000	.68 ***	.000
EM-3	.76 ***	.000	.59 ***	.000
EM-4	.84 ***	.000	.71 ***	.000
HU-1	.84 ***	.000	.62 ***	.000
HU-2	.84 ***	.000	.62 ***	.000
HU-3	.70 ***	.000	.39 ***	.000

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 1 238; **EA**: n = 1 238; **EM**: n = 1 238; **HU**: n = 1 220.

Tabelle 56: Homogenität (IPSO-Studie)

	Inter-Item-Korrelation		
	Minimum	AM	Maximum
QA-Skala	.55 ***	.64	.74 ***
EA-Skala	.28 ***	.33	.42 ***
EM-Skala	.47 ***	.54	.69 ***
HU-Skala	.35 ***	.47	.66 ***

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 1 238; **EA**: n = 1 238; **EM**: n = 1 238; **HU**: n = 1 220.

Tabelle 57: Nomologisches Netzwerk (IPSO-Studie)

	QA	EA	EM	HU	AK	GB	GG
QA-Skala		-.11 ***	.03	-.09 ***	.51 ***	.30 ***	-.28 ***
EA-Skala			.50 ***	.22 ***	-.04	-.18 ***	.15 ***
EM-Skala				.30 ***	.12 ***	-.18 ***	.14 ***
HU-Skala					-.15 ***	-.21 ***	.24 ***
AK-Skala						.37 ***	-.40 ***
GB							-.38 ***
GG							

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 1 238; **EA**: n = 1 238; **EM**: n = 1 238; **HU**: n = 1 220; **AK**: n = 1 238; **GB**: n = 1 238; **GG**: n = 1 238.

3Q-Studie

Tabelle 58: Iteminhalt (3Q-Studie)

Item-Nr.	Iteminhalt	AO
QA-1	Ist Ihre Arbeit ungleich verteilt, so dass sie sich auftürmt?	1
QA-2	Wie oft kommt es vor, dass Sie nicht genügend Zeit haben, Ihre Aufgaben zu erledigen?	1
QA-3	Kommen Sie mit Ihrer Arbeit in Rückstand?	1
QA-4	Haben Sie genug Zeit für Ihre Arbeitsaufgaben? (AO invertiert)	1
EM-1	Erfordert Ihre Arbeit Eigeninitiative?	2
EM-2	Haben Sie durch Ihre Arbeit die Möglichkeiten, etwas Neues zu lernen?	2
EM-3	Können Sie Ihre Fähigkeiten oder Sachkenntnisse bei Ihrer Arbeit einsetzen?	2
FQ-1	Bitte schätzen Sie ein, in welchem Maß Ihr direkter Vorgesetzter sicherstellt, dass jedes einzelne Teammitglied gute Entwicklungsmöglichkeiten hat.	2
FQ-2	Bitte schätzen Sie ein, in welchem Maß Ihr direkter Vorgesetzter hohen Wert auf Arbeitszufriedenheit legt.	2
FQ-3	Bitte schätzen Sie ein, in welchem Maß Ihr direkter Vorgesetzter die Arbeit gut planen kann.	2
FQ-4	Bitte schätzen Sie ein, in welchem Maß Ihr direkter Vorgesetzter Konflikte gut lösen kann.	2
HU-1	Wie oft erhalten Sie Hilfe und Unterstützung von Ihren Kollegen?	1
HU-2	Wie oft sind Ihre Kollegen bereit, sich Ihre Arbeitsprobleme anzuhören?	1
HU-3	Wie oft sprechen Ihre Kollegen mit Ihnen über die Qualität Ihrer Arbeit?	1
AK-a^N	<i>Die Anforderungen meiner Arbeit behindern mein Privat-/Familienleben.</i>	3
AK-b^N	<i>Die Zeit, die meine Arbeit einnimmt, macht es schwer, meinen privaten oder familiären Pflichten nachzukommen.</i>	3
AK-c^N	<i>Dinge, die ich zu Hause tun möchte, schaffe ich nicht, weil mich meine Arbeit so fordert.</i>	3
AK-d^N	<i>Die Arbeit belastet mich so sehr, dass es schwierig ist, meinen privaten oder familiären Pflichten nachzukommen.</i>	3
AK-e^N	<i>Aufgrund von Verpflichtungen durch die Arbeit muss ich Pläne für private oder familiäre Aktivitäten ändern.</i>	3
AF-1	Ist Ihre Arbeit...	4
AF-2	Wie schätzen Sie Ihre derzeitige Arbeitsfähigkeit in Bezug auf die körperlichen Arbeitsanforderungen ein?	5
AF-3	Wie schätzen Sie Ihre derzeitige Arbeitsfähigkeit in Bezug auf die psychischen Arbeitsanforderungen ein?	5
GB	Wie oft im Laufe der letzten 12 Monate haben Sie daran gedacht, Ihren derzeitigen Beruf aufzugeben?	6
GG	Wenn Sie den besten denkbaren Gesundheitszustand mit 10 Punkten bewerten und den schlechtesten denkbaren mit 0 Punkten: Wie viele Punkte vergeben Sie dann für Ihren derzeitigen Gesundheitszustand? Bitte kreuzen Sie die entsprechende Zahl an.	7

^N. Dieses Item wurde nicht im lidA-CAPI verwendet. Anmerkungen: **1-AO**: (1) nie/fast nie, (2) selten, (3) manchmal, (4) häufig und (5) immer. **2-AO**: (1) in sehr geringem Maß, (2) in geringem Maß, (3) etwas, (4) in gewissem Maß und (5) in großem Maß. **3-AO**: (1) stimme nicht zu, (2) 2, (3) 3, (4) 4 und (5) stimme völlig zu. **4-AO** (dieses Item wurde indirekt über die Abfrage des Arbeitsbereiches bearbeitet): (1) vorwiegend psychisch (geistig) belastend, (2) vorwiegend physisch (körperlich) belastend? und (3) psychisch (geistig) und physisch (körperlich) gleichermaßen belastend oder unbelastend? **5-AO**: (1) sehr schlecht, (2) eher schlecht, (3) mittelmäßig, (4) eher gut und (5) sehr gut. **6-AO**: (1) nie, (2) einige Male im Jahr, (3) einige Male im Monat, (4) einige Male pro Woche und (5) jeden Tag. **7-AO**: (1) 0 = denkbar schlechtest, (2) 1, [...], (10) 9 und (11) 10 = denkbar bester.

Tabelle 59: Soziodemografie (3Q-Studie)

		n	%
Gesamtzahl		1 827	100.0
Alter AM = 44.1 Jahre SD = 10.9 Jahre	18 bis unter 30 Jahre	212	11.6
	30 bis unter 40 Jahre	294	16.1
	40 bis unter 50 Jahre	564	30.9
	50 Jahre und älter	590	32.3
	keine Angabe	167	9.1
Geschlecht	männlich	248	13.6
	weiblich	1 485	81.3
	keine Angabe	94	5.1
Wohnsituation	allein	253	13.8
	mit Erwachsenem, ohne Kind	693	37.9
	mit Erwachsenem, mit Kind/ern	684	37.4
	ohne Erwachsenem, mit Kind/ern	153	8.4
	keine Angabe	44	2.4
Kinder im Haushalt	keine Kinder	747	40.9
	1 Kind und mehr	832	45.5
	keine Angabe	248	13.6
Arbeitsbereich	Pflege	1 282	70.2
	sozialer/betreuender Dienst	190	10.4
	Küche	83	4.5
	Wäscherei/Haustechnik	97	5.3
	Reinigung	10	0.5
	Verwaltung	165	9.0
Stellung im Beruf	ohne leitende Position	1 408	77.1
	stellvertretende leitende Position	107	5.9
	leitende Position	273	14.9
	keine Angabe	39	2.1
Wochenarbeitszeit Arbeitsvertrag	unter 20 Stunden	285	15.6
	20 bis unter 30 Stunden	332	18.2
	30 bis unter 40 Stunden	869	47.6
	40 Stunden und mehr	341	18.7

Anmerkungen: Alle TN haben zum Zeitpunkt der Datenerhebung eine sozialversicherungspflichtige Beschäftigung.

Tabelle 60: Verteilung der Werte (3Q-Studie)

	RW 0 n (%)	RW 25 n (%)	RW 50 n (%)	RW 75 n (%)	RW 100 n (%)
QA-1	243 (13.3)	548 (30.0)	703 (38.5)	278 (15.2)	55 (3.0)
QA-2	67 (3.7)	228 (12.5)	617 (33.8)	717 (39.2)	198 (10.8)
QA-3	322 (17.6)	616 (33.7)	624 (34.2)	218 (11.9)	47 (2.6)
QA-4	131 (7.2)	543 (29.7)	619 (33.9)	420 (23.0)	114 (6.2)
EM-1	74 (4.1)	124 (6.8)	113 (6.2)	612 (33.5)	904 (49.5)
EM-2	93 (5.1)	151 (8.3)	239 (13.1)	796 (43.6)	548 (30.0)
EM-3	84 (4.6)	100 (5.5)	150 (8.2)	765 (41.9)	728 (39.8)
FQ-1	101 (5.7)	208 (11.7)	305 (17.1)	717 (40.2)	453 (25.4)
FQ-2	118 (6.6)	204 (11.4)	277 (15.5)	611 (34.2)	574 (32.2)
FQ-3	112 (6.3)	194 (10.9)	294 (16.5)	696 (39.0)	488 (27.4)
FQ-4	137 (7.7)	203 (11.4)	298 (16.7)	674 (37.8)	472 (26.5)
HU-1	29 (1.6)	131 (7.2)	419 (23.1)	704 (38.8)	530 (29.2)
HU-2	38 (2.1)	139 (7.7)	440 (24.3)	689 (38.0)	507 (28.0)
HU-3	66 (3.6)	242 (13.3)	525 (29.0)	792 (43.7)	188 (10.4)

Anmerkungen: **QA:** n = 1 827; **EM:** n = 1 827; **FQ:** n = 1 784; **HU:** n = 1 813.

Tabelle 61: Tests auf Normalverteilung (3Q-Studie)

	Kolmogorov-Smirnov-Anpassungstest				Shapiro-Wilk-Test			
	D-Stat.	p-Wert	D-Krit.	α	W-Stat.	p-Wert	W-Krit.	α
QA-1	.21	.000	.024	.01	.90	.000	.995	.01
QA-2	.23	.000	.024	.01	.89	.000	.995	.01
QA-3	.20	.000	.024	.01	.90	.000	.995	.01
QA-4	.18	.000	.024	.01	.91	.000	.995	.01
QA-Skala	.06	.000	.024	.01	.99	.000	.995	.01
EM-1	.27	.000	.024	.01	.74	.000	.995	.01
EM-2	.29	.000	.024	.01	.83	.000	.995	.01
EM-3	.29	.000	.024	.01	.77	.000	.995	.01
EM-Skala	.20	.000	.024	.01	.85	.000	.995	.01
FQ-1	.27	.000	.024	.01	.86	.000	.995	.01
FQ-2	.25	.000	.024	.01	.85	.000	.995	.01
FQ-3	.27	.000	.024	.01	.86	.000	.995	.01
FQ-4	.26	.000	.024	.01	.86	.000	.995	.01
FQ-Skala	.16	.000	.024	.01	.92	.000	.995	.01
HU-1	.23	.000	.024	.01	.86	.000	.995	.01
HU-2	.23	.000	.024	.01	.87	.000	.995	.01
HU-3	.26	.000	.024	.01	.88	.000	.995	.01
HU-Skala	.14	.000	.024	.01	.96	.000	.995	.01

Anmerkungen: **QA:** n = 1 827; **EM:** n = 1 827; **FQ:** n = 1 784; **HU:** n = 1 813.

Tabelle 62: Item- und Skalenschwierigkeit (3Q-Studie)

	1. Quartil	Md	3. Quartil	AM	SD
QA-1	25.00	50.00	50.00	41.16	24.75
QA-2	50.00	75.00	75.00	60.28	24.10
QA-3	25.00	25.00	50.00	37.03	24.95
QA-4	25.00	50.00	75.00	47.85	25.70
QA-Skala	31.25	43.75	62.50	46.58	20.05
EM-1	75.00	75.00	100.00	79.39	26.95
EM-2	50.00	75.00	100.00	71.28	27.36
EM-3	75.00	75.00	100.00	76.72	26.33
EM-Skala	66.67	83.33	91.67	75.80	23.10
FQ-1	50.00	75.00	100.00	67.00	28.49
FQ-2	50.00	75.00	100.00	68.48	30.23
FQ-3	50.00	75.00	100.00	67.57	29.06
FQ-4	50.00	75.00	100.00	65.99	30.09
FQ-Skala	50.00	75.00	87.50	67.26	27.11
HU-1	50.00	75.00	100.00	71.72	24.24
HU-2	50.00	75.00	100.00	70.52	24.83
HU-3	50.00	75.00	75.00	60.95	24.23
HU-Skala	58.33	66.67	83.33	67.73	20.09

Anmerkungen: **QA:** n = 1 827; **EM:** n = 1 827; **FQ:** n = 1 784; **HU:** n = 1 813.

Tabelle 63: Trennschärfe (3Q-Studie)

	nicht-korrigierte Trennschärfe		korrigierte Trennschärfe	
	ρ	p-Wert	ρ	p-Wert
QA-1	.79 ***	.000	.63 ***	.000
QA-2	.81 ***	.000	.66 ***	.000
QA-3	.83 ***	.000	.69 ***	.000
QA-4	.77 ***	.000	.59 ***	.000
EM-1	.76 ***	.000	.54 ***	.000
EM-2	.83 ***	.000	.58 ***	.000
EM-3	.83 ***	.000	.62 ***	.000
FQ-1	.88 ***	.000	.79 ***	.000
FQ-2	.93 ***	.000	.87 ***	.000
FQ-3	.91 ***	.000	.86 ***	.000
FQ-4	.92 ***	.000	.86 ***	.000
HU-1	.84 ***	.000	.62 ***	.000
HU-2	.83 ***	.000	.60 ***	.000
HU-3	.77 ***	.000	.51 ***	.000

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA:** n = 1 827; **EM:** n = 1 827; **FQ:** n = 1 784; **HU:** n = 1 813.

Tabelle 64: Homogenität (3Q-Studie)

	Minimum	Inter-Item-Korrelation AM	Maximum
QA-Skala	.43 ***	.54	.64 ***
EM-Skala	.46 ***	.51	.56 ***
FQ-Skala	.73 ***	.79	.84 ***
HU-Skala	.45 ***	.51	.59 ***

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 1 827; **EM**: n = 1 827; **FQ**: n = 1 784; **HU**: n = 1 813.

Tabelle 65: Nomologisches Netzwerk (3Q-Studie)

	QA	EM	FQ	HU	AK	AF	GB	GG
QA-Skala		-.12 ***	-.36 ***	-.23 ***	.51 ***	-.35 ***	.38 ***	-.29 ***
EM-Skala			.39 ***	.25 ***	-.17 ***	.29 ***	-.21 ***	.23 ***
FQ-Skala				.34 ***	-.39 ***	.34 ***	-.33 ***	.29 ***
HU-Skala					-.20 ***	.20 ***	-.20 ***	.21 ***
AK-Skala						-.51 ***	.43 ***	-.41 ***
AF-Skala							-.42 ***	.65 ***
GB								-.33 ***
GG								

***. Die Korrelation ist auf dem Niveau von $\alpha = .001$ (2-seitig) hoch signifikant. Anmerkungen: **QA**: n = 1 827; **EM**: n = 1 827; **FQ**: n = 1 784; **HU**: n = 1 813; **AK**: n = 1 827; **AF**: n = 1 827; **GB**: n = 1 827; **GG**: n = 1 827.