

Befragungsteilnahme mit dem Smartphone

—

Entstehen durch das Smartphone Moduseffekte in Onlinebefragungen? Eine empirische Analyse des Smartphones als Ausfüllgerät

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft (Dr. rer. oec.)
an der Fakultät der Wirtschaftswissenschaft
– Schumpeter School of Business and Economics –
der Bergischen Universität Wuppertal

vorgelegt von
Mats Schumacher, M.Sc.
aus Wuppertal

Wuppertal, im November 2023

Erstgutachter: Prof. Dr. Dirk Temme
Zweitgutachter: Prof. Dr. Tobias Langner

Danksagungen

Die vorliegende Arbeit ist im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Methoden der empirischen Wirtschafts- und Sozialforschung der Bergischen Universität Wuppertal unter der Leitung von Prof. Dr. Dirk Temme entstanden. Als meinem Betreuer möchte ich insbesondere Herrn Temme herzlich für seine unermüdliche Anleitung und seinen fachlichen Beistand danken. Zu jeder Zeit konnte ich mich an ihn wenden und seine Ratschläge und sein Engagement waren stets eine große Unterstützung für mich. Außerdem gilt mein Dank Prof. Dr. Tobias Langner als Zweitgutachter. Ihm danke ich für viele wertvolle Hinweise und Anregungen während zahlreicher Doktorandenseminare und den fachlichen und stets wertvollen Austausch. Außerdem möchte ich meinen aktuellen und ehemaligen Kolleginnen und Kollegen danken, die dazu beigetragen haben, die Arbeitsatmosphäre jederzeit sowohl produktiv als auch angenehm zu gestalten und zusätzlich immer für Diskussionen und Feedbackprozesse bereitstanden.

Darüber hinaus gilt mein Dank meinem privaten Umfeld, das mir ebenfalls eine große Stütze war und ist. Dies betrifft meine engsten Familienangehörigen – meine Frau, meine Eltern und meine Schwester – sowie alle Freunde und Freundinnen (stellvertretend genannt die Gruppe „Quo Vadis“), auf die ich mich uneingeschränkt verlassen kann und die mich durch alle – auch durch persönlich herausfordernde – Phasen begleitet haben. Ein besonderer Dank gilt dabei meiner Mutter als akribischer Lektorin der Arbeit und meiner Frau, die mich mit ihrer Geduld und Zugewandtheit immer bestärkt hat.

Inhaltsübersicht

Inhaltsübersicht	III
Inhaltsverzeichnis	IV
Abbildungsverzeichnis	VII
Tabellenverzeichnis	VIII
Abkürzungsverzeichnis	IX
1 Einleitung	1
1.1 Relevanz des Themas	1
1.2 Zielsetzung und übergeordnete Forschungsfrage	3
1.3 Aufbau der Arbeit	3
2 Theoretische Grundlagen und Forschungsrahmen	5
2.1 Der Total Survey Error (TSE) als Forschungsrahmen für Moduseffekte	5
2.2 Smartphones als Ausfüllgeräte	9
2.3 Theorien der Smartphone-Nutzung und Bedeutung des Smartphones	16
2.4 Forschungsleitende Fragen und Forschungsplan	22
3 Forschungsstand und theoretische Herleitung	30
3.1 Moduseffekte des Smartphones als Ausfüllgerät	30
3.2 Antwortqualität des Smartphones als Ausfüllgerät	42
3.3 Careless Responding	54
4 Forschungsprogramm	89
4.1 Prüfung möglicher Moduseffekte des Smartphones bei sensiblen Fragen	93
4.2 Gerätspezifische Bedürfnisstrukturen in Onlinebefragungen	95
4.3 Anfälligkeit des Smartphones für Aufmerksamkeitsablenkung	98
4.4 Das Smartphone als Quelle von Careless Responding?	100
5 Empirische Studien	106
5.1 Studie 1 – Laborexperiment zur Aufrichtigkeit bei sensiblen Themenbereichen	106
5.2 Studie 2 – Gerätspezifische Bedürfnisstrukturen der Befragungsteilnahme	123
5.3 Studie 3 – Aufmerksamkeitsstrukturen während der Befragungsteilnahme	141
5.4 Studie 4 – Aufmerksamkeit und Careless Responding am Smartphone	158
6 Abschließende Diskussion	192
6.1 Zusammenfassung der Hauptkenntnisse	192
6.2 Implikationen für die empirische Wirtschafts- und Sozialforschung	197
6.3 Limitationen und Ausblick	199
6.4 Fazit	201
7 Literaturverzeichnis	202
Anhang	234

Inhaltsverzeichnis

Inhaltsübersicht	III
Inhaltsverzeichnis	IV
Abbildungsverzeichnis	VII
Tabellenverzeichnis	VIII
Abkürzungsverzeichnis	IX
1 Einleitung	1
1.1 Relevanz des Themas	1
1.2 Zielsetzung und übergeordnete Forschungsfrage	3
1.3 Aufbau der Arbeit	3
2 Theoretische Grundlagen und Forschungsrahmen	5
2.1 Der Total Survey Error (TSE) als Forschungsrahmen für Moduseffekte	5
2.2 Smartphones als Ausfüllgeräte	9
2.2.1 Begriffsbestimmung des Smartphones	10
2.2.2 Bedeutung und Verbreitung von Smartphones in der Befragungsforschung	12
2.2.3 Umgang mit dem Smartphone als Ausfüllgerät	13
2.2.4 Befragungsrelevante Besonderheiten des Smartphones als Ausfüllgerät	14
2.3 Theorien der Smartphone Nutzung und Bedeutung des Smartphones	16
2.3.1 Ausgewählte Theorien zur Erklärung der Smartphone Nutzung	16
2.3.2 Psychologische Bedeutung des Smartphones	20
2.4 Forschungsleitende Fragen und Forschungsplan	22
2.4.1 Vorüberlegungen zur Identifikation von Moduseffekten des Smartphones	22
2.4.2 Erarbeitung des Forschungsplans	25
2.4.2.1 Voruntersuchungen zur Antwortqualität und Unit-Nonresponse	26
2.4.2.2 Antwortqualität als Hauptbereich des Forschungsprogramms	28
3 Forschungsstand und theoretische Herleitung	30
3.1 Moduseffekte des Smartphones als Ausfüllgerät	30
3.1.1 Moduseffekte des Smartphones auf Stichprobenkennwerte und die Antwortbereitschaft bei sensiblen Fragen	30
3.1.1.1 Forschungsstand und Prüfkonzept	31
3.1.1.2 Herleitung der Hypothese der ersten Voruntersuchung zur Antwortqualität	33
3.1.2 Moduseffekte des Smartphones auf die Unit-Nonresponse	36
3.1.2.1 Forschungsstand und Prüfkonzept	36
3.1.2.2 Herleitung der Hypothese der zweiten Voruntersuchung	38
3.2 Antwortqualität des Smartphones als Ausfüllgerät	42
3.2.1 Grundlagen der Antwortqualität in Befragungen	42
3.2.1.1 Das Survey-Response Modell	43
3.2.1.2 Motivation und Aufmerksamkeit als Grundlagen hoher Antwortqualität	44

3.2.1.3	Bedrohungen der Antwortqualität im Rahmen des Survey-Response Modells	44
3.2.2	Antwortqualität als Prüfkriterium der Smartphoneeignung	46
3.2.3	Zusammenfassung des aktuellen Forschungsstands zur Antwortqualität	48
3.3	Careless Responding	54
3.3.1	Konzeptualisierung von Careless Responding	55
3.3.1.1	Eingrenzung des Begriffs Careless Responding	55
3.3.1.2	Formen von Careless Responding	57
3.3.2	Careless Responding in (Online-)Befragungen	59
3.3.2.1	Ursachen von Careless Responding	59
3.3.2.2	Auswirkungen von Careless Responding	60
3.3.2.3	Umgang mit Careless Responding	61
3.3.3	Aufdeckung von Careless Responding	62
3.3.3.1	Eye-Tracking	63
3.3.3.2	Geplante Detektionsmethoden	68
3.3.3.3	Post-hoc Detektionsmethoden	73
3.3.3.4	Übertragung der CR-Aufdeckung auf die vorliegende Arbeit	79
3.3.4	Das Smartphone als Hintergrundfaktor von Careless Responding	82
3.3.4.1	Das Arbeitsgedächtnis als Grundlage der Aufmerksamkeit	82
3.3.4.2	Herleitung der Hypothesen für die Hauptuntersuchungen	84
4	Forschungsprogramm	89
4.1	Prüfung möglicher Moduseffekte des Smartphones bei sensiblen Fragen	93
4.2	Gerätspezifische Bedürfnisstrukturen in Onlinebefragungen	95
4.3	Anfälligkeit des Smartphones für Aufmerksamkeitsablenkung	98
4.4	Das Smartphone als Quelle von Careless Responding?	100
5	Empirische Studien	106
5.1	Studie 1 – Laborexperiment zur Aufrichtigkeit bei sensiblen Themenbereichen	106
5.1.1	Methode	107
5.1.1.1	Versuchsablauf	107
5.1.1.2	Stichprobe	108
5.1.1.3	Fragebogenbeschreibung	109
5.1.1.4	Operationalisierung der Hypothesen	112
5.1.2	Ergebnisse	115
5.1.2.1	Voranalysen und Voraussetzungsprüfung	115
5.1.2.2	Hypothesenprüfung	117
5.1.2.3	Zusatzbetrachtungen	120
5.1.3	Diskussion	121
5.2	Studie 2 – Gerätspezifische Bedürfnisstrukturen der Befragungsteilnahme	123
5.2.1	Methode	124

5.2.1.1	Versuchsablauf	124
5.2.1.2	Stichprobe.....	125
5.2.1.3	Fragebogenbeschreibung und Operationalisierung der Hypothesen	125
5.2.2	Ergebnisse	130
5.2.2.1	Voranalysen und Voraussetzungsprüfung	130
5.2.2.2	Hypothesenprüfung	132
5.2.2.3	Zusatzbetrachtungen.....	135
5.2.3	Diskussion und Zwischenfazit der Voruntersuchungen.....	136
5.3	Studie 3 – Aufmerksamkeitsstrukturen während der Befragungsteilnahme.....	141
5.3.1	Methode	141
5.3.1.1	Versuchsablauf	142
5.3.1.2	Stichprobe.....	144
5.3.1.3	Fragebogenbeschreibung.....	145
5.3.1.4	Eye-Tracking Maße und Operationalisierungen der Hypothesen	147
5.3.2	Ergebnisse	148
5.3.2.1	Voranalysen und Voraussetzungsprüfung	149
5.3.2.2	Hypothesenprüfung	151
5.3.2.3	Zusatzbetrachtungen.....	152
5.3.3	Diskussion.....	153
5.4	Studie 4 – Aufmerksamkeit und Careless Responding am Smartphone.....	158
5.4.1	Methode	158
5.4.1.1	Versuchsablauf	159
5.4.1.2	Stichprobe.....	160
5.4.1.3	Fragebogenbeschreibung und Operationalisierung der Hypothesen	160
5.4.2	Ergebnisse.....	167
5.4.2.1	Voraussetzungsprüfung und Voranalysen	167
5.4.2.2	Hypothesenprüfung	169
5.4.2.3	Zusatzbetrachtungen.....	178
5.4.3	Diskussion.....	182
6	Abschließende Diskussion.....	192
6.1	Zusammenfassung der Hauptkenntnisse.....	192
6.2	Implikationen für die empirische Wirtschafts- und Sozialforschung	197
6.3	Limitationen und Ausblick	199
6.4	Fazit	201
7	Literaturverzeichnis.....	202
Anhang	234

Abbildungsverzeichnis

Abbildung 1. Total Survey Error als Framework für Moduseffekte angelehnt an Weisberg (2009).....	6
Abbildung 2. Das Technologie-Akzeptanz-Modell (TAM) in Anlehnung an Davis (1993)	17
Abbildung 3. Konsequenzen responsiver Designdarstellungen	23
Abbildung 4. Prüfkonzept möglicher Moduseffekte des Smartphones in der vorliegenden Arbeit.....	25
Abbildung 5. Unterscheidung von Careless Responding und Satisficing	56
Abbildung 6. Übersicht über verschiedene Detektionsmethoden für Careless Responding	69
Abbildung 7. Das Forschungsprogramm der vorliegenden Arbeit.....	92
Abbildung 8. Gerätvergleich hinsichtlich der Dimensionen Zeitaufwand und Privatheit	127
Abbildung 9. Überblick über die Ergebnisse der Hypothesenprüfung der Studie 2	135
Abbildung 10. Versuchsaufbau der dritten Studie	143
Abbildung 11. Areas of Interest (AOI) in Studie 3	149
Abbildung 12. Instruktions-Manipulationscheck (IMC) der Studie 4.....	163
Abbildung 13. Klassifikation als CR über den Long String-Ansatz	175

Tabellenverzeichnis

Tabelle 1. Forschungsstand zu den Forschungslücken der Voruntersuchungen	41
Tabelle 2. Antwortqualitätsunterschiede bei der Teilnahme mit dem Smartphone	52
Tabelle 3. Aufdeckung von Careless Responding in der Studie 4 der vorliegenden Arbeit.....	81
Tabelle 4: Überblick über die erarbeiteten Forschungshypothesen.....	89
Tabelle 5: Übersicht über das Auswertungsmodell von Studie 1	113
Tabelle 6: Übersicht über die Befunde der ersten Studie	120
Tabelle 7. Aufbau der Hypothesenprüfung in Studie 2	130
Tabelle 8. Mittelwerte für die Prüfungen der Hypothesen H2.1a-H2.1d	133
Tabelle 9. Ausgewählte Ergebnisse der logistischen Regression der Studie 2.....	134
Tabelle 10. Ergebnisse der Hypothesenprüfung der Studie 3.....	151
Tabelle 11. Übersicht über die Ergebnisse der Hypothesenprüfung der Studie 4	179
Tabelle 12. Durchfallquoten der Bogus-Items in Studie 4 aufgeteilt nach Itemtyp	182

Abkürzungsverzeichnis

2s/Item	=	2 Sekunden pro Item
"	=	Zoll
ALM	=	Allgemeines Lineares Modell
ANCOVA	=	Kovarianzanalyse
AOI	=	Area of Interest
App	=	mobile Applikation
AV	=	abhängige Variable
BFI	=	Big Five Inventory
BUW	=	Bergische Universität Wuppertal
bzw.	=	beziehungsweise
CA	=	Cronbachs Alpha
cm	=	Zentimeter
DSGVO	=	Datenschutzgrundverordnung
ggf.	=	gegebenenfalls
IMC	=	Instruktions-Manipulationscheck
IRI	=	Instruierte Response-Items
IRV	=	Individuelle Response Validität
ISD	=	Inter-Item Standardabweichung
IT	=	Informationstechnologie
ms	=	Millisekunden
RR	=	Random Responding
RT	=	Response Time
S.	=	Seite
s. o.	=	siehe oben
TAM	=	Technologie-Akzeptanz-Modell
TPB	=	Theory of Planned Behavior
TSE	=	Total Survey Error
UV	=	unabhängige Variable
VIF	=	Varianzinflationsfaktor
vgl.	=	Vergleich
vs.	=	versus
WEdN	=	Wahrgenommene Einfachheit der Nutzung
WN	=	Wahrgenommene Nützlichkeit
z. B.	=	zum Beispiel

1 Einleitung

Die Einleitung der vorliegenden Arbeit führt in das Themenfeld des Smartphones als Ausfüllgerät in Onlinebefragungen ein und umreißt das behandelte Forschungsproblem. Hierfür wird zunächst die Relevanz des Themas dargelegt (Kapitel 1.1). Anschließend werden die Zielsetzung der Arbeit und die übergeordnete Forschungsfrage herausgestellt (Kapitel 1.2). Den Abschluss der Einleitung bildet ein Überblick über den Aufbau der Arbeit (Kapitel 1.3).

1.1 Relevanz des Themas

Onlinebefragungen erweitern die moderne Befragungsforschung seit Beginn des Internetzeitalters (Couper 2000; Evans und Mathur 2018). In der wissenschaftlichen empirischen Wirtschafts- und Sozialforschung sind selbst-administrierte (Online-)Befragungen besonders verbreitet und werden vermehrt von Befragten als Erhebungsform präferiert (Couper 2005, 2008, S. 63-64; Paulhus und Vazire 2007, S. 227). Die wissenschaftliche Onlinebefragungsforschung ist herausgefordert, den sprunghaften technologischen Fortschritt zu berücksichtigen und sich analog zur Gesellschaft dynamisch weiterzuentwickeln (z. B. Couper 2013; Hair Jr., Harrison und Risher 2018). Ein Ausdruck dieser rapiden technologischen Entwicklung ist das Smartphone, dessen Nutzungszahlen seit Jahren kontinuierlich ansteigen (Couper et al. 2018; Wenz 2021a, S. 101; York und Poynter 2017, S. 11). Zu Beginn der aktuellen Dekade besaßen bereits über 90 Prozent der Deutschen ein internetfähiges Smartphone (DeStatis 2021, S. 39). Smartphones sind ständige Begleitung, gestalten den Alltag ihrer Besitzer*innen unmittelbar mit und eröffnen somit der Onlinebefragungsforschung einerseits neue Möglichkeiten, bringen andererseits aber auch neue Herausforderungen mit sich.

Die Einführung des ersten iPhones 2007 bzw. darauffolgend die Vorstellung des Android-Betriebssystems für mobile Geräte markiert einen Wendepunkt in der technologischen Entwicklung (Islam und Want 2014), aus dem auch Konsequenzen für die Befragungsforschung folgen. Smartphones bieten aufgrund des mobilen Netzausbaus eine permanente Vernetzungsmöglichkeit, sodass Smartphone-Besitzer*innen aus Forschungssicht potenziell zu jeder Zeit über verschiedene Situationen und Kontexte hinweg Datenübermittler*innen sein können (Antoun, Couper und Conrad 2017; Mimura et al. 2015). Smartphones besitzen zudem eine große Funktionsbreite, die diverse Möglichkeiten der Datengenerierung und Befragungsgestaltung eröffnet (Buskirk und Andrus 2012; Theobald, Baigger und Knöller 2017, S. 17).

Tatsächlich werden Smartphones in modernen Onlinebefragungen vermehrt als Ausfüllgeräte genutzt (z. B. Antoun et al. 2018; Bosnjak, Bauer und Weyandt 2017, S. 55; Toepoel und Lugtig 2014). Die Teilnahmeraten des Smartphones steigen kontinuierlich (Gummer et al.

2023), während Befragte seltener PCs/Laptops als Ausfüllgeräte verwenden (Gummer, Quöß und Roßmann 2019; Toepoel und Lugtig 2015), die ursprünglich die einzige Teilnahmemöglichkeit bei Onlinebefragungen darstellten. Mittlerweile muss jede Onlinebefragung immer als „mixed-device Studie“ – eine Befragung, an der mit verschiedenen Geräten teilgenommen wird – angesehen werden (Toepoel und Lugtig 2015, S. 155). Smartphones haben sich allmählich als Ausfüllgeräte in Onlinebefragungen etabliert, ohne dass ein eindeutiger Zeitpunkt identifizierbar wäre, ab dem diese für Smartphones gezielt geöffnet wurden (Callegaro 2013; Revilla et al. 2016b). Dies könnte Probleme bergen, wenn durch die Gerätwahl des Smartphones „ungewollte mobil Antwortende“ („unintended mobile respondents“) entstehen (Peterson 2012, S. 1; Wells, Bailey und Link 2014, S. 244). Ungewollte mobil Antwortende sind Befragte, die für die Teilnahme an Onlinebefragungen mobile Geräte nutzen, für deren Einsatz jedoch die Befragungen ursprünglich nicht mitkonzipiert wurden. Dabei könnte durch das Smartphone als Ausfüllgerät ein eigenständiger Befragungsmodus im Vergleich zu PCs oder Laptops entstehen.

Der Befragungsmodus ist ein wesentliches Element des Designs wissenschaftlicher Studien, das beschreibt, wie die Datenerhebung stattfindet. Ein eigenständiger Befragungsmodus entsteht aus den Unterschieden verschiedener Methoden der Datenerhebung und setzt sich beispielsweise zusammen aus dem Befragungsformat, dem Befragungssetting und der Art der Fragevermittlung (Bowling 2005; de Leeuw 2005; Manfreda et al. 2008). Onlinebefragungen sind bei der Datenerhebung mit der Diversität der Ausfüllgeräte und den daraus entstehenden Befragungsbesonderheiten konfrontiert. Es muss daher geprüft werden, ob bei der Erweiterung auf das Smartphone Moduseffekte entstehen, das heißt, ob sich beispielsweise die Ergebnisse von denen unterscheiden, die bei der Verwendung von PCs oder Laptops resultieren, oder die Antwortqualität der Befragten systematisch vom genutzten Ausfüllgerät abhängt.

Die Überprüfung möglicher Moduseffekte bei der Datenerhebung ist aus methodischer Sicht stets angezeigt (de Leeuw 2005). Eine möglichst sorgfältige Prüfung des Smartphones als Ausfüllgerät in Onlinebefragungen ist daher wichtig für die Bewertung der aktuellen Forschungspraxis und relevant für die Einschätzung, ob und inwiefern gerätspezifische Adaptationen notwendig sein könnten, falls das Smartphone Moduseffekte induziert. Dabei muss der Geräteeffekt des Smartphones isoliert und eine differenzierte Betrachtung exklusiver Einflüsse des Smartphones als Ausfüllgerät angestrebt werden, um sowohl eigenständige Geräteinflüsse (z. B. aufgrund der geringen Bildschirmgröße) als auch gerätspezifisches Antwortverhalten der Teilnehmer*innen untersuchen zu können. Die empirische Befundlage unterstreicht die Relevanz der Prüfung möglicher Moduseffekte des Smartphones. Wird das Smartphone als Ausfüllgerät ge-

nutzt, zeigen sich Unterschiede zu anderen Ausfüllgeräten: Befragte benötigen in Onlinebefragungen bei der Teilnahme mit dem Smartphone beispielsweise signifikant mehr Zeit (z. B. Andreadis 2015a, S. 74; Revilla, Toninelli und Ochoa 2016a; Struminskaya, Weyandt und Bosnjak 2015) und die Befragungen werden häufiger abgebrochen (z. B. Buskirk und Andrus 2014; Couper, Antoun und Mavletova 2017, S. 139; Lambert und Miller 2015). Viele Smartphone-Besitzer*innen berichten zudem eine hohe emotionale Bindung an das eigene Smartphone (z. B. Bock et al. 2016a; Fullwood et al. 2017; Thorsteinsson und Page 2014), wodurch ebenfalls Moduseffekte (z. B. bei sensiblen Befragungsthemen) entstehen könnten.

1.2 Zielsetzung und übergeordnete Forschungsfrage

Smartphones sind wesentlicher Bestandteil des Alltags und werden zunehmend als Ausfüllgeräte in Befragungen genutzt. Die mobile Teilnahme an Onlinebefragungen bietet klare Vorteile sowohl für Forscher*innen als auch für Teilnehmer*innen und kann nur schwerlich zurückgefahren werden (Callegaro 2013). Sollte das Smartphone Moduseffekte in Onlinebefragungen induzieren, muss allerdings geprüft werden, welche Folgen dies für seinen Einsatz nach sich zieht. Das vorliegende Dissertationsprojekt widmet sich der Evaluation des Smartphones als Ausfüllgerät in Onlinebefragungen und untersucht die Forschungsfrage:

„Entstehen durch das Smartphone als Ausfüllgerät Moduseffekte in Onlinebefragungen und welche Folgen hat dies ggf. für den Einsatz des Smartphones als Ausfüllgerät?“

Die vorliegende Arbeit strebt die Prüfung möglicher Moduseffekte beim Einsatz des Smartphones als Ausfüllgerät in Onlinebefragungen basierend auf einem breit angelegten Forschungsprogramm an. Als Referenzmaß für mögliche Moduseffekte dienen die etablierten Ausfüllgeräte PC und Laptop, da ursprünglich ausschließlich mit diesen an Onlinebefragungen teilgenommen wurde. Der PC und der Laptop werden im Rahmen der vorliegenden Arbeit literaturkonform als gemeinsamer, etablierter Ausfüllmodus behandelt (z. B. Peterson et al. 2017, S. 204; Sendelbah et al. 2016).

1.3 Aufbau der Arbeit

Zu Beginn (Kapitel 2) entwickelt die vorliegende Arbeit den wissenschaftlichen Bezugsrahmen (Kapitel 2.1) für die antizipierte Forschungsreihe und diskutiert die Bedeutung des Smartphones als Ausfüllgerät. Unter Einbezug relevanter wissenschaftlicher Modelle und Theorien wird ein gemeinsames Begriffssystem entwickelt und eine definitorische Engführung vorgenommen, wobei auch die Besonderheiten des Smartphones als Gerät diskutiert werden (Kapitel 2.2 und

Kapitel 2.3). Daraus resultiert die Vorstrukturierung des Forschungsplans für die Prüfung möglicher Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen, in dem die Hauptbereiche der Prüfung (Antwortqualität und Unit-Nonresponse) festlegt werden (Kapitel 2.4). Anschließend wird durch die Strukturierung der Forschungsliteratur die aktuelle Befundlage zu möglichen Moduseffekten des Smartphones als Ausfüllgerät in Onlinebefragungen für diese Hauptbereiche aufbereitet (Kapitel 3). Die Kapitel 3.1 bis 3.3 identifizieren maßgebliche Forschungslücken und füllen das Forschungsvorhaben sukzessive. Zunächst werden erste, kritische Ansatzpunkte für die Prüfung möglicher Moduseffekte des Smartphones identifiziert (Kapitel 3.1), für die Voruntersuchungen angelegt werden. Als Schwerpunkt der Arbeit wird die Betrachtung der Hintergründe der Antwortqualität (Kapitel 3.2) herausgearbeitet, wobei sich das Themenfeld Careless Responding (Kapitel 3.3) als optimal für die robuste Prüfung möglicher Moduseffekte beim Einsatz des Smartphones als Ausfüllgerät erweist.

Kapitel 4 entwickelt für die Forschungsfrage ein integratives Forschungsprogramm, anhand dessen in insgesamt vier empirischen Studien mögliche Moduseffekte und die Konsequenzen für den Einsatz des Smartphones als Ausfüllgerät kritisch geprüft werden können. Die empirische Umsetzung wird für alle Studien ausführlich dokumentiert und die Befunde werden diskutiert (Kapitel 5). Die Einordnung der Ergebnisse in die aktuelle Erkenntnislage und der Transfer aus anverwandten Wissenschaftsbereichen fördern dabei das Erreichen einer höheren Abstraktionsebene. Kapitel 6 interpretiert die Befunde übergreifend und bildet eine Synthese (Kapitel 6.1). Es werden der Smartphoneeinsatz in Onlinebefragungen und die Bedeutung der Forschungsergebnisse für die Praxis der empirischen Wirtschafts- und Sozialforschung bewertet (Kapitel 6.2). Im Anschluss bereiten die kritische Reflektion des eigenen Vorgehens und der gesammelten Erkenntnisse sowie ein Ausblick auf künftige Forschungsvorhaben (Kapitel 6.3) das abschließende Fazit (Kapitel 6.4) vor.

2 Theoretische Grundlagen und Forschungsrahmen

Grundlage für die Beantwortung der generellen Forschungsfrage bezüglich der Moduseffekte, die das Smartphone als Ausfüllgerät induzieren könnte, ist die Erarbeitung eines gemeinsamen Begriffssystems und die thematische Verortung der vorliegenden Arbeit im aktuellen Forschungsdiskurs. Zunächst wird der Total Survey Error (TSE) (vgl. Bautista 2012; Biemer 2010; Groves und Lyberg 2010; Weisberg 2009) als forschungsleitender Bezugsrahmen (nachfolgend „Framework“) der vorliegenden Arbeit vorgestellt und konkret auf Moduseffekte bezogen (Kapitel 2.1). Gleichfalls erfolgt eine Standortbestimmung der Relevanz von Onlinebefragungen und es wird dargelegt, wieso Modusunterschiede in Befragungen problembehaftet sein können.

Der TSE als Hintergrund der Untersuchung möglicher Moduseffekte des Smartphones verlangt eine präzise Betrachtung des Smartphones als Internet- und Kommunikationsmedium. Dafür wird der Smartphonebegriff für die vorliegende Arbeit eindeutig definiert und die Verbreitung und Rolle des Smartphones als Ausfüllgerät in Onlinebefragungen diskutiert (Kapitel 2.2). Zudem werden unter Einbezug ausgewählter wissenschaftlicher Theorien die Nutzungshintergründe sowie die psychologische Bedeutung des Smartphones für Nutzer*innen betrachtet (Kapitel 2.3). Abschließend wird die forschungsleitende Fragestellung ausdifferenziert und ein Forschungsplan für die vorliegende Arbeit entwickelt (Kapitel 2.4). Dieser bereitet die Auseinandersetzung mit dem aktuellen Forschungsstand zum Smartphone als Ausfüllgerät in Onlinebefragungen vor und legt fest, welche Aspekte in der Arbeit mit Blick auf das TSE-Framework fokussiert werden sollen, um mögliche Moduseffekte aufdecken zu können, die durch das Smartphone entstehen könnten.

2.1 Der Total Survey Error (TSE) als Forschungsrahmen für Moduseffekte

Befragungen sind eine wichtige Methode der Datenerhebung in der empirischen Wirtschafts- und Sozialforschung. Sie sammeln an einer Stichprobe systematisch Informationen, die möglichst valide die Einstellungen, Bedürfnisse, Meinungen, Verhaltensweisen und soziodemografischen Eigenschaften der Befragten abbilden, und ihre Befunde sollen darüber hinaus möglichst repräsentativ für die Zielpopulation sein (z. B. Groves et al. 2009, S. 2). Wissenschaftliche Befragungen sollten das Ziel verfolgen, Daten möglichst unverzerrt zu erheben, und müssen daher sorgsam konzipiert und durchgeführt werden. Das Smartphone als relativ neuartiges Ausfüllgerät in Onlinebefragungen könnte den Erhebungsprozess vielfältig beeinflussen und dadurch Moduseffekte ausüben. Für die Evaluation des Smartphones als Ausfüllgerät sollen die Untersuchungen deshalb breit angelegt werden und verschiedene Problemfelder einbeziehen, die zu Fehlereinflüssen führen könnten.

Das Konzept des „Total Survey Error“ (TSE) stellt ein Framework dar, das alle möglichen Fehlereinflüsse in einer Befragung integriert, die potenziell verhindern, dass die wahren Merkmalsausprägungen und tatsächlichen Datenstrukturen unverzerrt abgebildet werden (Bautista 2012, S. 39; Biemer 2010; Faulbaum 2022, S. 567; Weisberg 2009, S. 16-22). Auch wenn es beispielsweise aus Zeit- und Kostengründen nicht möglich ist, perfekte Befragungen zu entwickeln (Bradburn und Sudman 2004, S. 40; MacKenzie und Podsakoff 2012), sollte der TSE aber dennoch stets möglichst gering ausfallen (Oberski 2012, S. 480; Weisberg 2009, S. 24-25). Bei der Analyse der Auswirkungen methodischer Herausforderungen in Befragungen ist der ganzheitliche TSE-Ansatz einer isolierten Betrachtung einzelner Qualitätsmerkmale einer Befragung (z. B. Stichprobenbesonderheiten oder Abbruchraten) vorzuziehen (Biemer 2010; Groves und Lyberg 2010; Weisberg 2009, S. 330-331). Anhand des TSE als Framework können zuverlässig die bedeutsamsten Fehlerquellen in Befragungen erkannt werden (Biemer 2010; Groves und Lyberg 2010). In der vorliegenden Arbeit bildet der TSE daher die Grundlage für die Identifikation derjenigen TSE-Komponenten, die für die Prüfung möglicher Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen besonders geeignet sind.

Das TSE-Konzept ist bisher noch nicht abschließend einheitlich definiert (Groves und Lyberg 2010). Für die vorliegende Arbeit wird das TSE-Framework daher literaturbasiert hergeleitet und orientiert sich vornehmlich an den Ausführungen von Weisberg (2009), die durch weitere relevante Literaturbezüge ergänzt werden. Um dem aufgestellten Forschungsziel gerecht zu werden, richtet sich die TSE-Konzeptualisierung speziell auf Moduseffekte und trifft dabei eine Vorauswahl, welche Aspekte fokussiert werden sollen und welche vernachlässigt werden können. *Abbildung 1* stellt das TSE-Framework, das in dieser Arbeit angewendet wird,

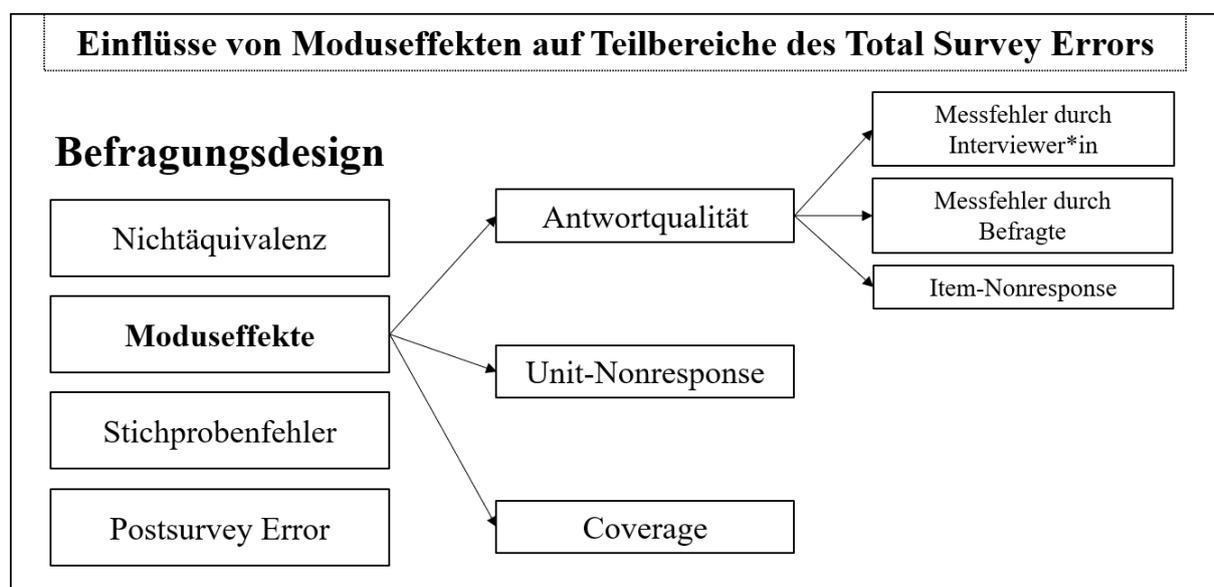


Abbildung 1. Total Survey Error als Framework für Moduseffekte angelehnt an Weisberg (2009)

konkret bezogen auf die Betrachtung der Auswirkungen von Moduseffekten in Anlehnung an Weisberg (2009) kompakt dar.

Die Grafik veranschaulicht zum einen, dass einige Einflüsse auf den TSE vom Befragungsmodus entkoppelt sind und zum anderen, welche Fehlereinflüsse auf den TSE durch Moduseffekte beeinflusst werden können. Generell können TSE-Einflüsse aufgrund des der Untersuchung zugrundeliegenden Befragungsdesigns 1) aus der eingeschränkten Studienvergleichbarkeit (Nichtäquivalenz), 2) aus dem Postsurvey Error oder 3) direkt aus Moduseffekten entstehen (Weisberg 2009, S. 261-312). Zudem kann für die vorliegende Arbeit 4) der Stichprobenfehler als weitere Komponente auf der Designebene ergänzt werden. Diese Designeinflüsse sind im Kontext des Untersuchungsziels der vorliegenden Arbeit, das mögliche Moduseffekte des Smartphones als Ausfüllgerät prüfen möchte, voneinander zu trennen. Nichtäquivalenz kennzeichnet die fehlende Vergleichbarkeit von Studienergebnissen zu thematisch übereinstimmenden Befragungen (Weisberg 2009, S. 297-299). Der TSE-Einfluss ist hier unabhängig von Modusbesonderheiten und folgt stattdessen aus übergeordneten Aspekten wie verschiedenen Forschungsinstitutionen, Erhebungsländern oder Zeitpunkten der Datenerhebung (Weisberg 2009, S. 297-299). Der Postsurvey Error beschreibt alle Fehler, die nach der Datenerhebung im Rahmen der Datenauswertung und Datenverarbeitung auftreten (Biemer 2010; Weisberg 2009, S. 19) und ist somit ebenfalls vom Ausfüllmodus entkoppelt, der die Datenerhebung betrifft. Der Stichprobenfehler resultiert daraus, dass eine Stichprobe nur eine Teilmenge der Population erfasst, und tritt also immer auf, wenn keine Vollerhebung der Population vorliegt (Biemer 2010; Faulbaum 2022, S. 568; Weisberg 2009, S. 225-237). Ist keine Vollerhebung geplant, entstammt auch der Stichprobenfehler somit der Anlage einer wissenschaftlichen Untersuchung und ist unabhängig vom Befragungsmodus der Datenerhebung, der lediglich beeinflussen kann, ob eine geplante Stichprobe realisiert werden kann.

Moduseffekte resultieren stattdessen aus dem Format und den Rahmenbedingungen der Datenerhebung und treten auf, wenn verschiedene Erhebungsmethoden oder -formen eingesetzt werden (Biemer 2010; Bowling 2005; de Leeuw 2005; Weisberg 2009, S. 278). Die Analyse modusübergreifender Daten erfordert große Sorgfalt, da Moduseffekte in den erhobenen Daten systematische Verzerrungen erzeugen können (z. B. Bowling 2005; Gunter et al. 2002; Lynn 1998). Moduseffekte wirken vielfältig auf den TSE, indem sie sowohl beeinflussen, wer in einer Befragung antwortet, als auch, wie die Antworten gegeben werden (Schouten et al. 2013; Tourangeau 2017, S. 122; Weisberg 2009, S. 282-287). Sie sind besonders brisant, wenn sie die Messung von Konstrukten beeinflussen, die Ergebnismuster verändern oder zu Unterschieden zwischen den antwortenden Personengruppen führen (Biemer 2010; Jäckle, Roberts und

Lynn 2010; Tourangeau 2017, S. 129; Weisberg 2009, S. 282-287). Das TSE-Framework (siehe *Abbildung 1*) unterscheidet drei Fehlerdimensionen, die prinzipiell durch Moduseffekte beeinflusst werden können: 1) Antwortqualität, 2) Unit-Nonresponse und 3) Coverage.

Die vorliegende Arbeit verwendet für die erste Fehlerdimension des TSE, auf die Moduseffekte wirken können, den deutschen Terminus Antwortqualität (englisch: Response Accuracy). Diese gliedert sich in drei Unterbereiche und umspannt alle Einflüsse, die die adäquate Konstruktmessung in einer Befragung erschweren, da explizit das Antwortverhalten der Befragten durch den Modus beeinflusst wird (Faulbaum 2022, S. 578; Weisberg 2009, S. 18-19). Item-Nonresponse als eine Fehlerebene der Antwortqualität tritt auf, wenn Befragte einzelne Fragen gar nicht oder nicht-substanziell (z. B. Wahl von Ausweichkategorien wie „weiß nicht“) beantworten (Bautista 2012, S. 43; Biemer 2010; Weisberg 2009, S. 130-138). Zudem kann die Konstruktmessung auch mit Messfehlern behaftet sein, die dem Antwortverhalten entspringen (z. B. Biemer 2010; Faulbaum 2022, S. 578; Jäckle et al. 2010): Zum einen können Messfehler durch den*die Interviewer*in entstehen, die auftreten, wenn die Messung durch die Person, die die Befragung leitet, verzerrt wird (Weisberg 2009, S. 45). Zum anderen hemmen Messfehler durch Befragte die Antwortqualität, bei denen die Verzerrung auf Ebene der Teilnehmer*innen selbst angesiedelt ist, deren Antwortverhalten dafür sorgt, dass die Antwort nicht den wahren Wert abbildet (Faulbaum 2022, S. 578; Weisberg 2009, S. 72). Die Einflussmöglichkeiten von Moduseffekten auf den Messfehler sind dabei auf systematische Verzerrungen (Bias) begrenzt.

Coverage und Unit-Nonresponse als weitere Fehlerdimensionen des TSE, auf die Moduseffekte wirken können, umfassen alle Fehlereinflüsse des Auswahlrahmens und der Zusammensetzung der letztlich realisierten Stichprobe der Befragungsteilnehmer*innen (Faulbaum 2022, S. 574-577; Tourangeau 2017, S. 129; Weisberg 2009, S. 18). Die Coverage ist gefährdet, wenn konkret schon der Auswahlrahmen der Stichprobe bestimmte Populationsmitglieder ausschließt (Weisberg 2009, S. 205-212). Unit-Nonresponse tritt auf, wenn angesprochene Populationsmitglieder sich z. B. gegen die Teilnahme an einer Befragung entscheiden (Bautista 2012, S. 43; Biemer 2010; Weisberg 2009, S. 165-167). Moduseffekte wirken auf die Coverage und Unit-Nonresponse, wenn die Stichprobe kein zutreffendes Abbild der Zielpopulation ist und Rückschlüsse auf Basis der Stichprobendaten daher verzerrt werden (Weisberg 2009, S. 12-13).

Die Literatur der Befragungsforschung unterstreicht die Rolle des Datenerhebungsmodus. Tatsächlich dokumentieren mehrere Studien Moduseffekte, wenn verschiedene Erhebungsmodi angewendet werden (z. B. Al Baghal 2019; Ward et al. 2014; Zhang et al. 2017). Insbesondere die Bereitschaft, sensible Informationen preiszugeben bzw. aufrichtig zu antworten, kann in Befragungen vom Ausfüllmodus beeinflusst werden (Booth-Kewley, Edwards und Rosenfeld

1992; Gerich 2009, S. 116; Tourangeau et al. 1997a). Sensible Informationen adressieren Themen, die sehr persönlich, stressbehaftet und kontrovers sind (Barnett 1998; Wolter 2012, S. 35). Auch im Framework des TSE werden sensible Themenbereiche zur Aufdeckung von Moduseffekten als besonders geeignet beschrieben (Jäckle et al. 2010; Weisberg 2009, S. 289-292). Die vorliegende Arbeit grenzt die Möglichkeiten der Internetnutzung für wissenschaftliche Zwecke ein und nimmt explizit Onlinebefragungen in den Blick. In vielen Forschungsfeldern und insbesondere in der Wirtschafts- und Sozialforschung ist die Onlinebefragung bereichsübergreifend schon länger eine der verbreitetsten Datenerhebungsmethoden (z. B. Couper 2008, S. 62; Evans und Mathur 2018; Hulland, Baumgartner und Smith 2018; Liu et al. 2013). Trotz des gemeinsamen selbst-administrativen Charakters sind etwa Onlinebefragungen nicht zwingend deckungsgleich mit schriftlichen Paper-Pencil Befragungen. Der Vorteil der Selbst-Administration gegenüber dem Interview fällt bei der Untersuchung sensibler Themenbereiche im Onlineformat noch größer aus als bei schriftlichen Befragungen (Gerich 2009, S. 116; Kreuter, Presser und Tourangeau 2008; Taddicken 2009, S. 98; Tourangeau und Smith 1996).

Das Ziel der vorliegenden Arbeit, eine fundierte Prüfung auf mögliche Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen durchzuführen und daraus mögliche Konsequenzen für die Smartphonennutzung in Befragungen abzuleiten, deckt sich mit dem Ansatz des TSE, übergeordnet nach Fehlerquellen zu suchen und diese dann differenziert zu evaluieren (Groves und Lyberg 2010). Anhand des TSE kann ein wissenschaftliches Forschungsprogramm entwickelt werden, das methodische Herausforderungen deduktiv prüft (Weisberg 2009, S. 330). Für die angestrebte breite Prüfung des Smartphones als Ausfüllgerät werden gezielt die Teilbereiche des TSE (siehe *Abbildung 1*) adressiert, die in Onlinebefragungen durch (ungewollte) mobil Antwortende mit dem Smartphone beeinflusst werden. Der weitere Verlauf des Kapitels richtet sich am TSE-Framework aus und zielt darauf, geeignete Prüfbedingungen für Moduseffekte des Smartphones zu bestimmen. Diese werden jeweils adäquat ausdifferenziert, weiterentwickelt und bezogen auf das zugrundeliegende Forschungsproblem strukturiert.

2.2 Smartphones als Ausfüllgeräte

Mobile Endgeräte und insbesondere das Smartphone sind aus dem heutigen Lebensalltag nicht mehr wegzudenken und stets präsent. Weltweit ist der Anteil der Menschen, die ein Smartphone besitzen, mittlerweile größer als der Anteil der PC- und/oder Laptop-Besitzer*innen (Statcounter 2023). Den Ausgangspunkt der vorliegenden wissenschaftlichen Untersuchung bilden eine präzise Begriffsbestimmung des Smartphones (Kapitel 2.2.1) und die Bewertung seiner Relevanz als Ausfüllgerät in Onlinebefragungen (Kapitel 2.2.2). Das Smartphone wird hinsichtlich der breiten Palette technologischer Eigenschaften und Einsatzmöglichkeiten definiert und von

anderen technischen Geräten abgegrenzt. Diese definitorische sowie zeitliche Verortung erlaubt, auch zukünftig die Relevanz der vorliegenden Befunde zeitadäquat einzuordnen. Um das Forschungsziel der Prüfung möglicher Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen adäquat verfolgen zu können, werden anhand des TSE-Frameworks die grundsätzlichen Umgangsformen mit dem Smartphone als Ausfüllgerät diskutiert (Kapitel 2.2.3) und die befragungsrelevanten Besonderheiten des Smartphones betrachtet (Kapitel 2.2.4).

2.2.1 Begriffsbestimmung des Smartphones

Mit der Einführung des iPhones im Jahr 2007 läutete Apple das Zeitalter der Smartphones ein und brachte das erste Gerät auf den Markt, das als Ursprung moderner Smartphones angesehen werden kann (Islam und Want 2014). Herkömmliche Mobiltelefone oder Handys, ursprünglich als reine Alternativen zur Festnetztelefonie entwickelt, unterscheiden sich immer mehr vom Nutzungsschwerpunkt des Smartphones. Die Telefonfunktion ist mittlerweile bei Smartphones nur noch von untergeordneter Bedeutung (Deng et al. 2019). Heutige Kernelemente liegen im mobilen Internetzugang und der applikationsbasierten Umgebung, die es den Nutzer*innen erlaubt, die eigenen Geräte durch die manuelle Installation mobiler Applikationen (Apps) an individuelle Bedürfnisstrukturen anzupassen (z. B. de Reuver, Nikou und Bouwman 2016). Damit unterscheiden sich moderne Smartphones insgesamt stark von den telefoniezentrierten Vorgängermodellen und verkörpern handliche, vollfunktionsfähige Computer (Wang, Xiang und Fesenmaier 2014). Während frühe Handys und Mobiltelefone eher sperrig waren und nur sehr kleine Displays besaßen (Jones, Buchanan und Thimbleby 2003), wird bei der Gestaltung moderner Geräte das Nutzungserleben immer weiter optimiert und die Interaktion mit dem Smartphone erleichtert. Melumad und Pham (2020) charakterisieren Smartphones als 1) tragbare Geräte, die 2) einen gewissen privaten Rahmen bieten und dabei 3) hochgradig persönliche Besitzobjekte mit 4) klaren haptischen Vorzügen darstellen.

Für die Teilnahme an Onlinebefragungen ist bei der Charakterisierung des Smartphones zunächst die Bildschirmgröße relevant. Diese bestimmt unmittelbar, inwiefern eine gut lesbare und klare Darstellung von Befragungsinhalten möglich ist. Einschlägige Smartphone-Datenbanken zeigen, dass sich Bildschirmgrößen zwischen 4" (10,16 cm) und 7" (17,78 cm) bei gängigen Smartphones finden lassen (z. B. DeviceAtlas 2019; Type/Code 2023). Die Bildschirme von Smartphones dienen einerseits der Darstellung visuellen Outputs, andererseits fungieren sie durch eine integrierte Touchfunktion zugleich als Interaktionsschnittstelle. Zudem ist die technische Ausstattung trotz gerät- und anbieterspezifischer Variationen insgesamt sehr breit und integriert mehrere Sensoren sowie Input- und Outputmodule (z. B. Spracheingabe und

Sprachausgabe, Kamerafunktionen, WLAN, GPS, Bluetooth), die zudem über Zusatzhardware nahezu beliebig erweitert werden können (z. B. Miller 2012).

Wesentlich für die vorliegende Arbeit ist die eindeutige Unterscheidung zwischen Smartphones und PCs bzw. Laptops, die als Vergleichsbedingung für Moduseffekte angelegt werden. Smartphones sind erheblich kleiner und handlicher und ermöglichen auch ohne WLAN-Zugang Internetzugang (Schlosser und Silber 2020, S. 266). Aufgrund dieser Eigenschaften werden Smartphones deutlich häufiger unterwegs und in anderen, diverseren Kontexten genutzt als PCs/Laptops (z. B. Deng et al. 2019; Kim et al. 2017; Kim, Kim und Lee 2005; Skierkowski und Wood 2012). Hedonistische Nutzungsziele spielen zudem am Smartphone eine größere Rolle als am PC/Laptop (Kim et al. 2005; Skierkowski und Wood 2012). Die Interaktionsfrequenz mit dem Smartphone ist zwar recht groß, die einzelnen Nutzungserlebnisse sind jedoch von kürzerer Dauer als am PC/Laptop (Kim et al. 2017; Theobald et al. 2017, S. 17).

Für die Begriffsbestimmung ist ebenfalls die Abgrenzung des Smartphones von anderen internetfähigen mobilen Geräten hilfreich. Neuartige Geräte wie Smart-Watches und Digital-Glasses verfügen über viel kleinere Bildschirme und erfordern eine andere Handhabung als Smartphones (Poynter 2015, S. 12). Weniger eindeutig ist die Abgrenzung zu Tablets, die in ihrem Charakter, ihrer Ausstattung und ihrem Leistungsumfang Smartphones stärker ähneln. Tablets sind wie Smartphones tragbar und in diversen Kontexten nutzbar. In Bezug auf die Bildschirmgröße und den Nutzungskomfort gleichen sie aber eher PCs und Laptops (Gummer und Roßmann 2015). Mit Blick auf die Besonderheiten von Ausfüllgeräten, die für Onlinebefragungen relevant sind, werden Tablets häufig mit PCs/Laptops gleichgesetzt und somit von Smartphones abgegrenzt (z. B. Antoun et al. 2017; Peterson et al. 2017, S. 204; Wells, Bailey und Link 2013) oder sogar von der Betrachtung komplett ausgeschlossen (Schlosser und Silber 2020, S. 266). Dies ist sinnvoll, da für Bereiche, bei denen keine Unterschiede zwischen Smartphones und PCs/Laptops auftreten, auch keine Unterschiede für Tablets zu erwarten sind. Smartphones stellen in Onlinebefragungen unter anderem aufgrund der geringeren Bildschirmgröße die größeren Herausforderungen für Befragte dar (z. B. Antoun et al. 2018; Kim und Sundar 2014; Smith und Kim 2015). Erst wenn tatsächlich Modusunterschiede auftreten, ist die separate Betrachtung von Tablets angezeigt, um in anschließenden Untersuchungen die Befunde einordnen zu können. Für die vorliegende Arbeit werden Tablets daher ausgeschlossen und Smartphones mit PCs/Laptops verglichen. Der Ausschluss von Tablets ist überdies auch für die Analyse von Moduseffekten des Smartphones praxisnäher, da Tablets als Ausfüllgeräte in Onlinebefragungen nur vergleichsweise selten verwendet werden (z. B. Gummer et al. 2023; Mavletova und Couper 2014; Poulová, Klímová und Pulkrábková 2019, S. 305; Revilla 2017).

2.2.2 Bedeutung und Verbreitung von Smartphones in der Befragungsforschung

Smartphones werden für vielfältige Aufgaben genutzt, für die es zum Teil ursprünglich andere Lösungen gab (Poynter 2015, S. 13). Neuere Smartphones und der Blick auf die prognostizierte technische Entwicklung offenbaren, beispielsweise durch starke Prozessorleistungen, steigenden Speicherplatz sowie eine Vielzahl interner und externer Sensoren, auch für die wissenschaftliche Forschung multiple Einsatzmöglichkeiten (z. B. Miller 2012; Pinter, Toninelli und de Pedraza 2015, S. 2). Blondel, Decuyper und Krings (2015) konstatieren, dass auch die Einsatzmöglichkeiten für die empirische Wirtschafts- und Sozialforschung reichhaltig sind. Durch ihre Sensorenvielfalt (z. B. Mikrofone, GPS) könnten Smartphones schon mittelfristig die Verhaltensmessung revolutionieren, da sie als Erhebungsinstrument keine externe Belastung induzieren, sondern in den Alltag der Proband*innen integriert sind. Somit ermöglichen sie – selbst im Längsschnitt – weitestgehend stressfreie und einfache Erhebungen (Cottrill et al. 2013; Mimura et al. 2015). Das Datenpotenzial ist nicht nur reichhaltig, dynamisch und multimodal (z. B. haptische, auditive und visuelle Inputs/Outputs), sondern auch die Güte der Sensoren und die Aufzeichnungsqualität steigen kontinuierlich an (Miller 2012; Mimura et al. 2015). Selbst komplexe und leistungsstarke Programme werden einschränkungsfrei von modernen Smartphones ausgeführt. Zudem kann die Datenübermittlung augenblicklich erfolgen. Nutzer*innen können „live“ untersucht und befragt werden und zu jeder Zeit und an (fast) jedem Ort an Untersuchungen teilnehmen (Cottrill et al. 2013; Poynter 2015, S. 16-17; Raento, Oulasvirta und Eagle 2009; Theobald 2017, S. 305-306). Der erleichterte Zugang zu Befragten und die diskrete Erhebung erhöhen die Kosteneffizienz bei hoher Detailgenauigkeit (Raento et al. 2009). Aus Forschungssicht ist es somit interessant, das Smartphone aufgrund seiner Vorzüge als Erhebungsinstrument zu erschließen (z. B. Dufau et al. 2011; Miller 2012; Raento et al. 2009).

Die vorliegende Arbeit begrenzt sich auf die Analyse von Moduseffekten des Smartphones als Ausfüllgerät in Onlinebefragungen. Kurz nach ihrer Einführung wurden Smartphones nur vereinzelt als Ausfüllgeräte in Onlinebefragungen verwendet (Antoun et al. 2018). Seitdem entwickeln sich jedoch die Teilnahmezahlen von Befragten, die das Smartphone nutzen, stetig nach oben (z. B. Antoun et al. 2018; Gummer et al. 2019; Gummer et al. 2023; Struminskaya et al. 2015). Im Gegenzug sinkt der Anteil der Personen, die PCs oder Laptops für die Befragungsteilnahme einsetzen (Gummer et al. 2019; Gummer et al. 2023). Schon in der letzten Dekade waren Raten von Smartphone-Teilnehmer*innen von bis zu 25 Prozent nicht ungewöhnlich (Bosnjak et al. 2017, S. 55; Revilla 2017). Während in vielen Befragungen der PC oder Laptop noch das häufigste Ausfüllgerät darstellt, kann sich dies für einzelne Befragungen sogar schon länger umkehren (Krämer 2017, S. 93; Toepoel und Lugtig 2014). Diese Entwicklung ist

möglicherweise kritisch, da Onlinebefragungen nicht immer bewusst für den Einsatz des Smartphones als Ausfüllinstrument entwickelt werden, die Proband*innen aber in der Regel trotzdem selbstständig entscheiden können, mit welchem Gerät sie an Onlinebefragungen teilnehmen. Werden Smartphones genutzt, kann dies zu einer Inkongruenz aus Planung der Studie und letztllicher Teilnahme in Form von „ungewollten mobil Antwortenden“ führen (Peterson 2012, S. 1; Wells et al. 2014, S. 244).

2.2.3 Umgang mit dem Smartphone als Ausfüllgerät

Eine möglichst umfassende Betrachtung des Smartphones als Ausfüllgerät in Onlinebefragungen ist notwendig, da das Smartphone elementarer Bestandteil der Befragungspraxis geworden ist. Die vorliegende Arbeit untersucht, ob das Smartphone als Ausfüllgerät in Onlinebefragungen den Befragungsrahmen grundlegend verändert und eigenständige Moduseffekte ausübt. Der TSE als Grundlage dieser Prüfung ermöglicht es, über die Auseinandersetzung mit dem Smartphone als Gerät und seinen Eigenschaften mögliche kritische Einsatzbereiche in Onlinebefragungen zu identifizieren. Konkret werden die Einflüsse des Smartphones als Ausfüllgerät und seine befragungsrelevanten Besonderheiten auf die TSE-Bereiche der Antwortqualität, Coverage und Unit-Nonresponse sowie ggf. zugehörige Unterbereiche (siehe *Abbildung 1*) bezogen. Sollten durch das Smartphone Moduseffekte entstehen, verlangt dies ggf. Anpassungen bei seinem Einsatz als Ausfüllgerät in Onlinebefragungen. In Vorbereitung des Forschungsplans strukturiert die vorliegende Arbeit daher die drei herausgearbeiteten TSE-Bereiche (Antwortqualität, Unit-Nonresponse und Coverage) hinsichtlich der Anfälligkeit für mögliche Moduseinflüsse durch das Smartphone als Ausfüllgerät.

Herzing (2019, S. 3) unterscheidet prinzipiell vier Möglichkeiten, mit dem Smartphone als Ausfüllgerät in Onlinebefragungen umzugehen. Die erste Option, die Verschiedenheit von Ausfüllgeräten einfach zu ignorieren, ist nicht empfehlenswert, da die Konsequenzen unerkannt bleiben. Aus wissenschaftlicher Sicht kommen nur die drei anderen Maßnahmen in Betracht: 1) den Einsatz von Smartphones zu verhindern, 2) Onlinebefragungen auf mobile Geräte anzupassen oder 3) spezifische Apps zu entwickeln, über die Fragebogen exklusiv am Smartphone ausgefüllt werden. Herzing (2019, S. 3) selbst konstatiert, dass der Fokus der Forschung auf der Anpassung von Befragungen an das Smartphone liegen sollte. Die appbasierte Lösung ist aufgrund der flexiblen Einsatzmöglichkeiten und des Datenpotenzials zwar interessant, erfordert von Forschenden aber fortgeschrittene Programmierkenntnisse und unterscheidet sich in der Erhebungsform stark von browserbasierten Onlinebefragungen. Dass auch der Ausschluss des Smartphones als Ausfüllgerät eher keine sinnvolle Strategie darstellt, kann aus dem TSE als

Framework der vorliegenden Arbeit geschlossen werden. Die weltweite Verbreitung internetfähiger Smartphones reduziert ursprüngliche Coverage-Probleme von Onlinebefragungen aufgrund fehlenden stationären Internetzugangs durch die Öffnung für die Teilnahme mit dem Smartphone (Couper et al. 2018; Wenz 2021a, S. 101). Nur ein geringer Prozentsatz der Bevölkerung verfügt weltweit und gerade in westlichen Kulturen nicht über ein internetfähiges Smartphone (Couper et al. 2018; York und Poynter 2017, S. 11). Die Smartphoneteilnahme in Onlinebefragungen zuzulassen, ist insbesondere wichtig, da immer mehr Personen der Kategorie „Mobile-Only“ (Internetzugang ausschließlich mit dem Smartphone ohne Besitz anderer internetfähiger Geräte) zuzuordnen sind (Busse und Fuchs 2012; Revilla et al. 2015, S. 128). Somit würde der Ausschluss des Smartphones als Ausfüllgerät neue Coverage-Probleme in Onlinebefragungen erzeugen. Die Befundlage demonstriert zudem, dass es den Bedürfnissen einzelner Teilnehmer*innen zuwiderhandeln würde, die Teilnahme an Onlinebefragungen mit dem Smartphone zu verhindern. Während die Teilnahmeraten in Befragungen weltweit häufig sinken (z. B. Brick und Williams 2013; Groves 2006; Stedman et al. 2019) und in Onlinebefragungen zumindest stark schwanken (Wu, Zhao und Fils-Aime 2022), wird die Relevanz des Smartphones als Ausfüllgerät dadurch betont, dass ein wesentlicher Teil der Befragten bevorzugt mit diesem an Onlinebefragungen teilnimmt (Krämer 2017, S. 100; Revilla et al. 2016b).

2.2.4 Befragungsrelevante Besonderheiten des Smartphones als Ausfüllgerät

Die Vorüberlegungen des vorherigen Abschnitts dokumentieren, dass Smartphones als Ausfüllgeräte für adäquate Coverage in Onlinebefragungen notwendig sind. Systematische Einflüsse des Smartphones auf die Unit-Nonresponse, die zum TSE beitragen, könnten hingegen kritisch sein. Auch Unterschiede in der Antwortqualität sind denkbar, wenn das Smartphone Moduseffekte ausübt. Da sich das Smartphone als Gerät definitorisch markant von anderen Geräten abgrenzt (siehe Kapitel 2.2.1), kann anhand des TSE-Ansatzes genau der Frage nachgegangen werden, ob und inwiefern seine Besonderheiten es als eigenen Befragungsmodus kennzeichnen könnten, der Moduseffekte ausübt und somit den TSE entweder erhöht oder sogar im Vergleich zum Basismodus (PC/Laptop) senkt.

Tourangeau und Kolleg*innen (2018) führen drei Gründe an, weshalb aus dem Einsatz des Smartphones in Onlinebefragungen wesentliche, eigenständige Einflüsse entstehen könnten, die auf den TSE wirken. Das kritischste Argument liegt in der reduzierten Bildschirmgröße, da browserbasierte Inhalte (wie Onlinebefragungen) eine ausreichende Screengröße benötigen (Ghose, Goldfarb und Han 2013; Jones et al. 1999). Durch die geringere Screengröße von Smartphones sind Smartphone-Teilnehmer*innen jedoch in Onlinebefragungen mit besonderen

Herausforderungen konfrontiert (Antoun et al. 2018; Kim und Sundar 2014; Schlosser und Silber 2020, S. 266). Die Leistung von Personen, die an kleinen Bildschirmen arbeiten, ist deutlich reduziert im Vergleich zur Leistung der Personen, die dieselben Aufgaben an größeren Bildschirmen bearbeiten (z. B. Chae und Kim 2004; Dillon, Richardson und McKnight 1990; Jones et al. 1999; Jones et al. 2003; Raptis et al. 2013, S. 133). Insbesondere die Leseleistung und -effizienz leiden (Dillon et al. 1990; Duchnicky und Kolars 1983). Hier könnten also in Onlinebefragungen Moduseffekte des Smartphones entstehen. Die erforderliche Screengröße korrespondiert zusätzlich auch mit der Komplexität der Inhalte. Sind die Inhalte komplex, müssen Bildschirme größer sein, um optimale Leistung zu erreichen (Chae und Kim 2004). Onlinebefragungen können zwar unterschiedlich komplex ausgestaltet sein, erfordern jedoch immer ein relativ hohes Maß kognitiver Anstrengung, sodass in Befragungen die vergleichsweise geringe Bildschirmgröße von Smartphones zu Problemen bei der Teilnahme führen könnte.

In Onlinebefragungen wirken auch die Art der Fragedarstellung und der Fragevermittlung auf das Antwortverhalten von Proband*innen und tragen schon abseits von Moduseffekten zum TSE bei (Bowling 2005; Weisberg 2009, S. 100-120). So beeinflusst etwa die eingesetzte Antwortskala (z. B. Skalenbreite und Skalenrichtung) die Befundmuster (Couper et al. 2004; Krämer 2017, S. 97-99; Toepoel et al. 2009). Aufgrund der reduzierten Screengröße sind am Smartphone somit auch gerätspezifische Einflüsse der Skalierungen bzw. der Darstellung auf den TSE denkbar. Die eher geringe Bildschirmgröße des Smartphones bedingt Darstellungsunterschiede von Fragebogeninhalten. Scrolling wird wahrscheinlicher und der Sichtbarkeitsbereich ist verglichen mit anderen Ausfüllgeräten eingeschränkt (Couper et al. 2004; Jones et al. 1999). Nicht alle Elemente des Fragebogens oder einzelner Skalen sind am Smartphone im gleichen Maße sichtbar, wodurch Sichtbarkeitsprinzipien (konkret Primacy- and Recency-Effekte¹) das Antwortverhalten anders beeinflussen können als bei ganzheitlicher Darstellung am PC/Laptop (Couper et al. 2004; Couper et al. 2017, S. 143; Ghose et al. 2013; Tourangeau et al. 2017). Werden Befragten nicht alle Inhalte gleichzeitig angezeigt, können der Darstellungszeitpunkt (z. B. zuerst vs. zuletzt) und der Bildausschnitt (z. B. alles ist direkt sichtbar vs. Sichtbarkeit aller Elemente nur durch Scrolling) dargestellter Elemente auf die Antwortgabe wirken.

Zusätzlich zur Bildschirmgröße birgt als zweiter Aspekt nach Tourangeau und Kolleg*innen (2018) auch das Eingabeformat des Smartphones Verzerrungspotenzial. Durch die Nutzung von Touchscreens unterscheidet sich die Eingabe am Smartphone von der Interaktion mit PCs/Lap-

¹ Elemente die zuerst (Primacy) und zuletzt (Recency) wahrgenommen werden, werden besser verarbeitet und erinnert (Miller und Campbell 1959).

tops, da bei Letzteren primär Computermäuse (bzw. Touchpads) und Tastaturen eingesetzt werden. Eingabeformat und Bildschirmgröße könnten zusätzlich interagieren, falls das Touchformat die präzise Antwortgabe an kleinen Screens erschwert (vgl. Tourangeau et al. 2018).

Abseits technischer Geräteigenschaften könnte sich als dritter Aspekt der Befragungskontext, in dem sich Befragte bei der Teilnahme befinden, je nach Ausfüllgerät unterscheiden und/oder gerätweise unterschiedlich auswirken (Tourangeau et al. 2018). Die große Funktionsbreite erlaubt Smartphones eine überaus flexible Nutzung (z. B. Miller 2012; Theobald et al. 2017, S. 17). Entsprechend werden sie über diverse Situationen (z. B. in der Bahn, im Bett oder auf dem Sofa vor dem Fernseher) hinweg eingesetzt (z. B. Deng et al. 2019; Kim et al. 2017). Die ohnehin hohe Diversität des Teilnahmekontexts von Onlinebefragungen (z. B. Wharton et al. 2003) wird durch den Einsatz von Smartphones noch einmal potenziert (z. B. Höhne und Schlosser 2019; Kim et al. 2017). Genau solche Kontext- und Umgebungseinflüsse tragen zum TSE bei (Biemer 2010) und müssen daher in der vorliegenden Arbeit bei der Prüfung des Smartphones als Ausfüllgerät adäquat berücksichtigt werden.

2.3 Theorien der Smartphonennutzung und Bedeutung des Smartphones

Die Hintergründe der Nutzung des Smartphones als Ausfüllgerät in Onlinebefragungen bestimmen, inwiefern durch den Einsatz von Smartphones Moduseffekte entstehen könnten. Folgend werden ausgewählte Theorien der generellen Smartphonennutzung vorgestellt (Kapitel 2.3.1), die versuchen, Nutzungsentscheidungen für das Smartphone zu erklären. Daran anknüpfend werden die persönliche und psychologische Bedeutung des Smartphones für die Nutzer*innen diskutiert (Kapitel 2.3.2). In Vorbereitung der Prüfung möglicher Moduseffekte werden die nachfolgenden Ausführungen stets auf etwaige Einflüsse des Smartphones auf den TSE und seine Teilbereiche bezogen, um wesentliche Prüfbedingungen systematisch herauszuarbeiten.

2.3.1 Ausgewählte Theorien zur Erklärung der Smartphonennutzung

Die Technologienutzung im Allgemeinen und die Smartphonennutzung im Besonderen basieren auf individuellen Überzeugungen der Nutzer*innen. Insbesondere Selbstwirksamkeitserwartungen bilden eine wichtige Grundlage der Verhaltensplanung (vgl. Bandura 2012). Die „Theory of Planned Behavior“ (TPB) (vgl. Ajzen 1985, 1991) deklariert Nutzungsintentionen als den Ursprung späteren Verhaltens und wesentlichen Prädiktor der Verhaltensausführung.

Das „Technologie-Akzeptanz-Modell“ (TAM) erweitert die TPB auf technische Geräte und Systeme und konzeptualisiert wesentliche Ursachen der Technologienutzung (Davis 1993; Davis, Bagozzi und Warshaw 1989). Der Ursprung des TAM liegt im Arbeitskontext. Gemäß

des TAM werden Technologien von Anwender*innen genutzt, wenn ihr Einsatz als nützlich und einfach wahrgenommen wird (z. B. Davis 1993). Wahrgenommene Nützlichkeit (WN) lässt sich definieren als: „*the degree to which a person believes that using a particular system would enhance his or her job performance*“ (Davis 1989, S. 320). Je hilfreicher eine Technologie für Arbeitsinhalte oder zu bewältigende Aufgaben empfunden wird, desto größer ist die WN. Wahrgenommene Einfachheit der Nutzung (WEdN) definiert das TAM als: „*the degree to which a person believes that using a particular system would be free of effort*“ (Davis 1989, S. 320). Die WEdN ist maximal, wenn der antizipierte Aufwand der Nutzung möglichst gering ist. Das TAM postuliert, dass die Designaspekte eines Systems bzw. einer Technologie die tatsächliche Nutzung nicht direkt vorhersagen, sondern WN, WEdN und Nutzungsintention als kognitive Instanzen zwischengeschaltet werden (siehe *Abbildung 2*). Der Einfluss der WN und der WEdN auf die tatsächliche Nutzung wird vollkommen mediiert durch die Nutzungsintention (Einstellung zur Nutzung des Systems/der Technologie). WEdN und WN wirken einerseits direkt auf die Nutzungsintention, und die WEdN übt über die WN zusätzlich einen indirekten Effekt auf die Nutzungsintention aus (Davis 1993). Die Forschung zum TAM zeigt, dass WN und WEdN stark miteinander zusammenhängen, wobei WN einen stärkeren prädiktiven Charakter für die letztliche Nutzung besitzt als WEdN (z. B. Davis 1989; Davis et al. 1989). WN ist daher für die Beurteilung der Nutzung von Technologien das etwas wichtigere Element. Die ein wenig geringere Aussagekraft von WEdN lässt sich mit Blick auf die Verbreitung und Verankerung einer Technologie im gesellschaftlichen Leben nachvollziehen. Je länger Technologien schon bekannt sind und je mehr Erfahrungen mit ihnen gesammelt werden, desto geringer werden die Unterschiede in der WEdN bei (potenziellen) Nutzer*innen (Venkatesh und Bala 2008).

Aktuelle Sichtweisen auf die Nutzung neuer Technologien lösen sich vom ursprünglichen Arbeitskontext und berücksichtigen die rapide technologische Entwicklung. Über den PC als Bezugskonzept hinaus, auf den die Ursprünge des TAM zurückgehen (Davis et al. 1989), kann moderne Technologienutzung allgemein über das TAM nachvollzogen werden. Für

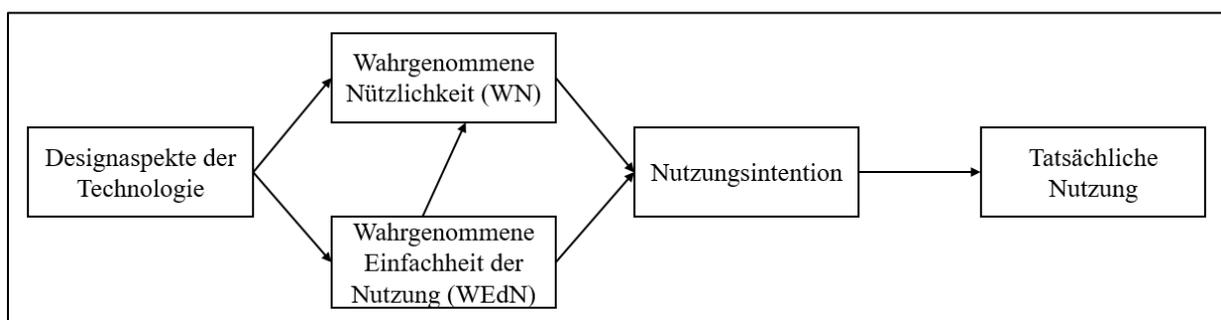


Abbildung 2. Das Technologie-Akzeptanz-Modell (TAM) in Anlehnung an Davis (1993)

Smartphones als alltägliche Begleittechnologie hat sich das TAM bereits mehrfach bewährt (z. B. Joo und Sang 2013; Liu und Li 2011; Park und Chen 2007). Auch für die Integration des Smartphones in den Lebensalltag stellen WN und WEdN wichtige Determinanten dar (Joo und Sang 2013; Park und Chen 2007). Theoriekonform wird die tatsächliche Smartphonennutzung ebenfalls durch die Selbstwirksamkeit der Nutzer*innen beeinflusst (Park und Chen 2007). Das TAM und vor allem die Konzepte der WN und der WEdN verdeutlichen, dass Moduseffekte im TSE-Framework durch das Smartphone entstehen könnten, wenn Befragte das Smartphone als Ausfüllgerät nicht als nützlich oder die Nutzung in Onlinebefragungen nicht als einfach wahrnehmen und sie daher nicht mit dem Smartphone teilnehmen oder an diesem anders antworten. Das TAM-Grundmodell wurde in der Weiterentwicklung um weitere relevante Determinanten der Technologienutzung ergänzt (Park und Chen 2007; Venkatesh et al. 2003). Bei der Anwendung auf neuere Technologien, wie zum Beispiel auf das Smartphone, werden stets mögliche bedeutsame Moderatoren wie Alter, Geschlecht und Erfahrung als Einflussfaktoren der Nutzung integriert (Venkatesh et al. 2003).

Darüber hinaus leisten auch der Stellenwert des Smartphones im Leben seiner Nutzer*innen und die Beziehung des Smartphones zum eigenen Selbst einen Erklärungsbeitrag zu den Nutzungshintergründen. Objekte können aus identitätstheoretischer Sicht Teil des Selbst werden, wenn sie als bedeutsam wahrgenommen werden und eine persönliche Relevanz besitzen – insbesondere, wenn ihnen zudem eine soziale Funktion zukommt (Stets und Biga 2003). Der identitätstheoretische Hintergrund basiert auf der Forschung zur allgemeinen Bedeutung von Informationstechnologien (IT) für Nutzer*innen. IT als Oberbegriff für elektronische Datenverarbeitung umfasst sowohl verwendete Hardware als auch Software. Somit muss auch das Smartphone als IT-Anwendungsbereich behandelt werden. Generell können IT-Systeme und IT-Produkte zur Identitätsbildung ihrer Nutzer*innen beitragen (z. B. Jones und Karsten 2008). Bei der Nutzung von Informations- und Kommunikationstechnologien wächst die Bedeutung der genutzten IT-Systeme über die eines einfachen Werkzeugs hinaus und die Systeme sind an Konstruktionsprozessen der eigenen Identität beteiligt (Lamb und Kling 2003). Nutzer*innen binden sich teilweise so intensiv an die genutzten Systeme/Produkte, dass diese eine erhebliche Bedeutung für das eigene Leben erhalten und essenzielle Handlungsrelevanz erwerben (z. B. Schwarz und Chin 2007). Carter und Grover (2015) formulieren das Konzept der IT-Identität, die das Ausmaß definiert, in dem ein Individuum die Nutzung eines IT-Systems/Produktes als integralen Bestandteil des eigenen Selbst sieht. Dieser neue Identitätstyp setzt sich aus drei verschiedenen Dimensionen zusammen: 1) Verbundenheit, 2) Affekt (emotionale Energie) und 3) Abhängigkeit. Verbundenheit reflektiert das Ausmaß, in dem Individuen eine Verbindung

zwischen dem eigenen Selbst und der jeweiligen IT erleben. Affekt umfasst das (positive) emotionale Erleben im Zusammenhang mit der IT-Nutzung und beruht konkret auf Zuversicht, Enthusiasmus und Antrieb. Abhängigkeit erweitert das Konzept der IT-Identität um das Erleben von starkem Vertrauen in die genutzte IT (vgl. Carter und Grover 2015). Während das TAM primär den Fokus auf die initiale Aneignung und Nutzung von Technologien legt, setzt das Konzept der IT-Identität bei IT-Systemen/Produkten an, die für die Nutzer*innen bereits besonders salient und bedeutsam sind. Der theoretische Anwendungsbereich des TAM wird demnach auf die dauerhafte Technologienutzung erweitert. Die IT-Identität determiniert einerseits das Verhalten der Nutzer*innen und zugleich konsolidiert die IT-Nutzung selbst die IT-Identität (Carter und Grover 2015).

In der Forschungsliteratur wird schon frühen Mobiltelefonen eine Ausdrucksmöglichkeit der eigenen Identität attestiert (Katz und Sugiyama 2005, S. 77; Mannetti, Pierro und Livi 2002). Mobile Technologien und insbesondere das Smartphone gestalten die eigene Identität aktiv mit und tragen zur Identitätsbildung bei (z. B. Campbell und Park 2008; Srivastava 2005; Walsh und White 2007; Walsh, White und Young 2010). Das Konzept der IT-Identität wird im Forschungsdiskurs daher auch konkret auf Smartphones in Form einer „Smartphone-Identität“ angewendet (Carter, Grover und Thatcher 2016, S. 1438-1442). Die Nutzer*innen verändern die Selbstwahrnehmung dergestalt, dass Smartphones als Teil des eigenen Selbst angesehen werden (z. B. Carter et al. 2016, S. 1440; Davel 2017, S. 183-189). Dies geschieht zum einen funktional, indem Smartphones die intellektuellen und physischen Fähigkeiten der Nutzer*innen fördern, aber auch anthropomorphisch dahingehend, dass das eigene Smartphone personifiziert werden kann (Fullwood et al. 2017; Park und Kaye 2019). Individuelle Persönlichkeitseigenschaften werden auf das Smartphone projiziert und das Selbst über den eigenen Körper hinaus auf das Gerät erweitert (Park und Kaye 2019). Als Teil der eigenen Identität kann in der Folge bei den Nutzer*innen das Bedürfnis steigen, mit dem Smartphone zu interagieren (Carter und Grover 2015; Walsh, White und Young 2008). Bei intensiver Nutzung entsteht ein sich selbst erhaltender Prozess, bei dem die wiederholte Nutzung die Bedeutung des Smartphones für das eigene Selbst zunächst weiter vertieft und letztlich die Verankerung in der eigenen Identität aufrechterhält. In Konsequenz könnten Smartphones z. B. auch als Ausfüllgeräte in Onlinebefragungen bevorzugt werden. Die Theorien zur Erklärung der Smartphone-Nutzung und insbesondere das Konzept der Smartphone-Identität deuten an, dass das Smartphone für Nutzer*innen auch eine psychologische Bedeutung einnehmen kann. Dies wäre für die Beurteilung möglicher Moduseffekte als Ausfüllgerät in Onlinebefragungen relevant, weshalb sich das folgende Unterkapitel explizit der psychologischen Bedeutung des Smartphones widmet.

2.3.2 Psychologische Bedeutung des Smartphones

Für die Einordnung der möglichen psychologischen Bedeutung, die Smartphones für ihre Besitzer*innen erlangen können, muss die Nutzer*innenperspektive eingenommen werden. Smartphones sind im Alltag für ihre Besitzer*innen omnipräsent, sodass diese sich der Geräte immer wieder bewusst werden (Melumad und Pham 2020). Insbesondere, da das Smartphone in der Lage ist, verschiedenste Funktionen zu übernehmen und die Nutzungsmuster entsprechend heterogen sind, ist eine differenzierte Betrachtung grundlegend. Die Nutzungsmuster unterscheiden sich dabei je nach Altersgruppe und Persönlichkeitseigenschaften. Beispielsweise nutzen jüngere Menschen das Smartphone verstärkt für die Freizeitgestaltung, extravertiertere Menschen legen größeren Wert auf den Smartphonebesitz, während verträglichere Nutzer*innen unter anderem die tatsächliche Anruhfunktion höher gewichten (z. B. Lane und Manner 2011). Die letztliche Smartphone-Nutzung ist dementsprechend recht individuell und hängt vom Profil der Nutzer*innen ab (z. B. Billieux 2012). Übergreifend dominieren bei den Nutzungsmotiven von Smartphones hedonistische Beweggründe wie Bequemlichkeit, Vergnügen oder Entspannung sowie die sozial-kommunikative Funktion des Smartphones, um Kontakt mit Befreundeten und Familie aufzunehmen bzw. zu halten (z. B. Eastman et al. 2014; North, Johnston und Ophoff 2014; Park und Han 2013).

Eine Besonderheit des Smartphones liegt in der empfundenen Privatheit des Geräts. Diese beschreibt, inwiefern das Smartphone als persönlicher Gegenstand wahrgenommen wird, mit dem Nutzer*innen vertrauensvoll interagieren. Das eigene Smartphone wird von Besitzer*innen fast ausschließlich exklusiv genutzt und nicht geteilt oder weitergegeben (Keusch, Wenz und Conrad 2022). Nutzer*innen, die ihr Smartphone stärker personalisieren und entsprechend Dritten gegenüber mehr von sich preisgeben (z. B. Ernährungsstile und Körpermaße bei Fitness-Apps), weisen paradoxerweise eine sehr hohe Intimitätswahrnehmung im Umgang mit ihrem Gerät auf (Sutanto et al. 2013). Es entsteht eine enge Vertrautheit mit dem und Bindung an das Gerät. Dabei ist der tatsächliche Besitzfaktor von wesentlicher Bedeutung. Die Bindung an das eigene Smartphone ist stärker als an fremde Geräte, selbst wenn sich die Modelle ähneln (z. B. Brasel und Gips 2014; Melumad und Pham 2020; Thorsteinsson und Page 2014). Die Touchfunktion moderner Smartphones erhöht das Besitzgefühl zusätzlich und verstärkt die Bindung an das eigene Gerät (Brasel und Gips 2014; Thorsteinsson und Page 2014).

Zudem kann durch den individuellen, oft positiv besetzten Nutzungskontext dem Smartphone sogar ein emotionaler Wert zugeschrieben werden (Fullwood et al. 2017). Positive Emotionen, die mit Objekten verknüpft sind, vertiefen die Bindung daran (Fredrickson 2001), sodass die emotionale Komponente des Smartphones Bindungsgefühle verstärken kann. Liegt eine

emotionale Bindung an das Gerät vor, fundiert dies die theoretische Annahme, dass Smartphones durchaus als Erweiterung des eigenen Selbst wahrgenommen werden können (z. B. Carter et al. 2016, S. 1438; Park und Kaye 2019).

Die psychologische Bedeutung von Smartphones für ihre Besitzer*innen manifestiert sich außerdem in Empfindungen, die mit dem Smartphone zusammenhängen und unmittelbar in den Lebensalltag hineinwirken. Smartphones kommt im Alltag häufig eine beruhigende Funktion zu: Melumad und Pham (2020, S. 237) bezeichnen das Smartphone als „Schnuller für Erwachsene“. In stressigen Situationen setzen manche Besitzer*innen die Nutzung ihres Smartphones als Coping-Strategie ein, wodurch die Nutzungsintensität dann nochmals deutlich steigt (Augner und Hacker 2012; Melumad und Pham 2020). Bei Stresserleben kann das Smartphone ein Gefühl von Kontrolle vermitteln, das hilft, besser mit dem Stress umzugehen, ohne ihn aber tatsächlich abzubauen (Rieger, Hefner und Vorderer 2017). Zum (eigenen) Smartphone besteht dementsprechend auch eine klare mental-affektive Verbindung, die direkt in das alltägliche Erleben wirken kann. Einige Smartphone-Besitzer*innen erleben z. B. Phantom-Vibrationen, bei denen sie eine Vibration des Smartphones wahrnehmen, obwohl dieses eine solche nicht ausendet (Drouin, Kaiser und Miller 2012; Rosenberger 2015; Sauer et al. 2015). Die starke und mitunter sogar identitätsstiftende Bedeutung (siehe Kapitel 2.3.1) des Smartphones besitzt selbstverstärkenden Charakter, da das Smartphone zum Teil so intensiv in das eigene Leben integriert wird, dass ein Leben „ohne“ nicht mehr vorstellbar ist. Dies kann in der Folge zirkulär die Nutzungsintentionen der Besitzer*innen noch weiter erhöhen (Walsh et al. 2008, 2010). Mitunter resultieren gar auf das Gerät bezogene Angstzustände. Die Angst, vom eigenen Smartphone getrennt zu sein bzw. keinen Zugriff darauf zu haben, wird als „Nomophobia“ (No mobile phone phobia) bezeichnet (z. B. Bhattacharya et al. 2019, S. 1297; King et al. 2013, S. 141). Nomophobe Nutzer*innen streben an, ihr Gerät immer möglichst griffbereit zu halten (Stald 2008, S. 151; Vorderer, Krömer und Schneider 2016) und berichten Separierungsängste (Carter et al. 2016, S. 1440; Walsh et al. 2010). Die Trennung vom Smartphone und sogar die bloße Vorstellung, vom Gerät getrennt zu sein, können Zustände ausgeprägten negativen Affekterlebens und Stressempfindungen hervorrufen (z. B. Konok, Pogány und Miklósi 2017; North et al. 2014; Thorsteinsson und Page 2014; Vorderer et al. 2016). Solche starken Bindungen zum Smartphone zeigen sich vornehmlich bei jüngeren Menschen (Bock et al. 2016b, S. 3406; Lepp, Barkley und Karpinski 2014; Walsh et al. 2008; Walsh et al. 2011).

Zusammenfassend verdeutlicht der Forschungsstand, dass Nutzer*innen ihren Smartphones eine hohe psychologische Bedeutung beimessen und das Smartphone tatsächlich auch identi-

tätsstiftende Funktionen übernehmen kann. Das eigene Smartphone wird einerseits zur Stressbewältigung eingesetzt, kann sich aber zum anderen, wie im Beispiel der Nomophobia, negativ auf das eigene Wohlbefinden und Erleben auswirken. Es besteht also eine emotionale und psychologische Verbindung zum Gerät bzw. zu den Funktionen, die das Gerät übernimmt (Panova und Carbonell 2018; Thorsteinsson und Page 2014).

Die psychologische Bedeutung des Smartphones hat diverse Implikationen für die Prüfung von Moduseffekten im TSE-Framework. Da der Gerätkarakter markant wahrgenommen wird und mit bestimmten Nutzungsmustern verknüpft ist, könnte sich dies auf das Smartphone als Ausfüllgerät übertragen, das dann womöglich auch als eigenständiger Modus wahrgenommen wird. In Konsequenz könnte sich das Antwortverhalten, infolge der Verbindung zum Gerät, aufgrund der psychologischen Bedeutung sowie aufgrund der Nutzungsgewohnheiten und -erfahrungen gerätweise systematisch unterscheiden (z. B. Unterschiede in der Offenheit oder Gewissenhaftigkeit), was dann zu Unterschieden in der Antwortqualität führen würde. Möglicherweise schätzen die Nutzer*innen das Smartphone zudem als unterschiedlich geeignet für die Teilnahme an Onlinebefragungen ein, wodurch gerätabhängige Einflüsse auf die Unit-Nonresponse entstehen könnten. Unterscheiden sich die Personengruppen, die verschiedene Ausfüllgeräte nutzen, systematisch voneinander, kann das insbesondere problematisch sein, wenn die Unterschiede mit Zielvariablen der Analysen assoziiert sind. Mögliche Moduseffekte, die aus dem Smartphone als Ausfüllgerät entstehen könnten, müssen daher auch vor dem Hintergrund der Bedeutung des Geräts für das Leben der Teilnehmer*innen beurteilt werden.

2.4 Forschungsleitende Fragen und Forschungsplan

Peterson und Kolleg*innen (2017, S. 204) führen dezidiert aus, dass das Smartphone den TSE vielfältig beeinflussen könnte und daher eine Prüfung, ob dadurch Moduseffekten entstehen, notwendig ist. Für die Aufdeckung von Moduseffekten wird im TSE-Framework ein klares Prüfkonzept benötigt (Weisberg 2009, S. 278-292). Daher wird nachfolgend ein solches Prüfkonzept für Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen entwickelt, um die Forschungsfrage der vorliegenden Arbeit adäquat untersuchen zu können.

2.4.1 Vorüberlegungen zur Identifikation von Moduseffekten des Smartphones

Die Forschungsfrage der vorliegenden Arbeit bezieht sich konkret auf den direkten Geräteinfluss des Smartphones als Ausfüllgerät in Onlinebefragungen. Das zugehörige Prüfkonzept muss demnach diejenigen Einflüsse, die aus unterschiedlicher Darstellung entstehen und zum

TSE beitragen, vom tatsächlichen Effekt des Smartphones als Ausfüllgerät so präzise wie möglich entkoppeln, um den Geräteinfluss zu isolieren. Dies ist bisher in der Forschungsliteratur noch nicht ausreichend geschehen. Einige moderne Onlinebefragungen passen die Befragungsinhalte stattdessen dynamisch an die Geräteigenschaften an. Diese sogenannten „responsiven Designs“ können aber, je nach Gerät, in sehr unterschiedlichen Darstellungsvarianten resultieren (siehe *Abbildung 3*) und sind nicht validierte, rein technische Anpassungen an die Geräteigenschaften des jeweiligen Smartphones (Lütters 2017, S. 72). Münden responsive Designs in starken Darstellungsunterschieden, könnte dies eigenständige Moduseffekte auslösen, die nur indirekt von dem Ausfüllgerät abhängen (Peterson et al. 2017, S. 216; Revilla et al. 2016a, 2017). Damit am Smartphone solche separaten Einflüsse abweichender Darstellung nicht auftreten können, sollten alle Skalen und Befragungsinhalte gerätübergreifend möglichst gleich dargestellt werden (Antoun et al. 2018; Krämer 2017, S. 99; Lütters 2017, S. 73). Im Vergleich mit responsiven Designs führt die gerätübergreifend gleiche Darstellung dazu, dass die Befragten mit den Befragungen zufriedener sind (Antoun et al. 2018), die Vervollständigungsrate deutlich steigt (Mavletova und Couper 2015, S. 89), komplexere Fragetypen wie offene Fragen ähnlich gut beantwortet werden (Arn, Klug und Kolodziejcki 2015; Revilla und Ochoa 2016) und auch möglicherweise ungewollte destruktive responsive Designanpassungen verhindert werden (Schlosser und Silber 2020, S. 275). Der Forderung nach gerätübergreifend möglichst gleicher Darstellung folgt auch die vorliegende Arbeit mit dem Ziel, mögliche Moduseffekte des Smartphones, die eindeutig auf den Einfluss des Geräts zurückzuführen sind, verlässlich

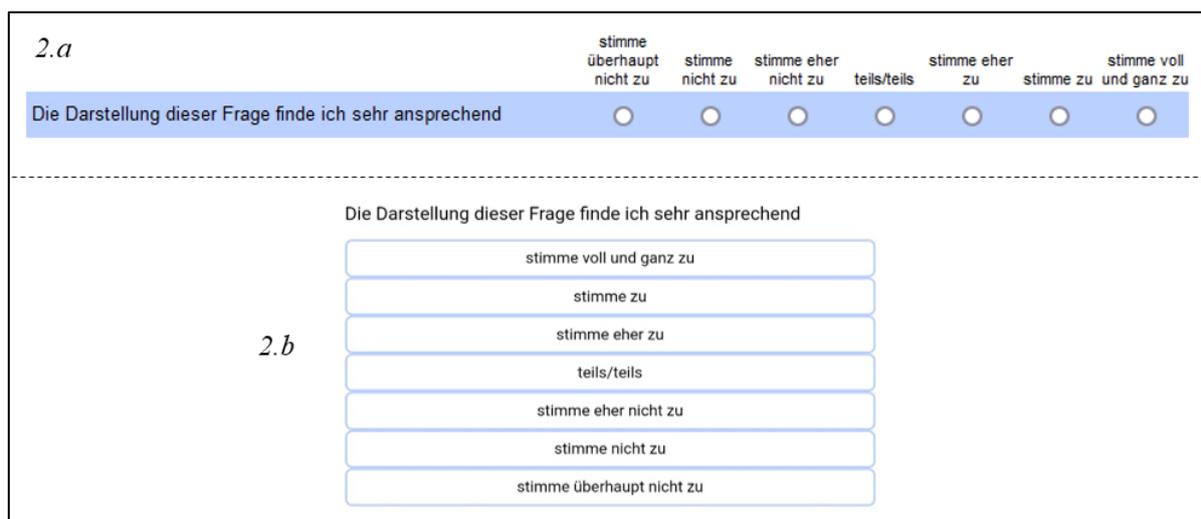


Abbildung 3. Konsequenzen responsiver Designdarstellungen

Darstellungsunterschiede einer Frage in der Befragungssoftware SoSci Survey (Leiner 2023) je nach Ausfüllgerät. 2.a präsentiert die Bildschirmdarstellung am PC/Laptop. Bei 2.b findet sich die responsive Fragedarstellung, die von SoSci Survey automatisch an den Smartphonebildschirm angepasst wird. Quelle: Eigene Darstellung, SoSci Survey (Leiner 2023)

aufzudecken. Insofern werden moderne mobile Forschungs-Apps in das Forschungsprogramm nicht einbezogen, da bei diesen die Darstellung sehr unterschiedlich im Vergleich zu browserbasierten Onlinebefragungen ist (Antoun et al. 2018).

Einen möglichen Einflussbereich von Moduseffekten bildet im TSE-Framework die Antwortqualität. In der Forschungsliteratur findet sich daneben häufig auch der Begriff der „Datenqualität“. Sowohl Antwortqualität als auch Datenqualität beschreiben den Problembereich des TSE, der alle Einflüsse zusammenfasst, die die Konstruktmessung beeinträchtigen, sodass die erhobenen Werte nicht valide sind (Bautista 2012, S. 38; Edwards 2019; Weisberg 2009, S. 18-19; Wiersma 2013). Die Trennung der Termini Antwortqualität und Datenqualität ist in der Forschungsliteratur bisher weder einheitlich noch stets eindeutig. Im Folgenden wird für diese Arbeit eine definitorische Abgrenzung anhand der Hintergründe der Begriffe „Antwortqualität“ und „Datenqualität“ vorgenommen.

Die vorliegende Arbeit behandelt Datenqualität im Vergleich zur Antwortqualität als etwas breiteres Konzept, das auch alle möglichen Fehlerquellen einbezieht, die durch eingesetzte Skalen, Formulierungen der Items oder Darstellungsbesonderheiten auftreten können (z. B. Bautista 2012, S. 45; Edwards 2019; Groves und Lyberg 2010). Auch diese Aspekte tragen zwar zum TSE bei (Weisberg 2009, S. 114-121), für die Prüfung des Smartphones als Ausfüllgerät bieten sie aber wenig Mehrwert und werden daher ausgeschlossen. Hier wirken keine exklusiven Geräteinflüsse, sondern die Validität der eingesetzten Skalen, Inventare und Messinstrumente beeinflusst gerätunabhängig den TSE. Bei der Antwortqualität im hergeleiteten TSE-Framework für Moduseffekte liegt der Fokus – anders als bei der Datenqualität – rein auf Verzerrungen, die dem Antwortprozess der Befragten entstammen (Faulbaum 2022, S. 579). Genau solche potenziellen Einflüsse des Smartphones auf den Prozess der Antwortgabe sind für das Forschungsziel wesentlich. Sie wirken auf die Antwortqualität und tragen somit zum TSE bei, weshalb die vorliegende Arbeit nachfolgend den Begriff „Antwortqualität“ verwendet.

Sollte das Antwortverhalten der Befragten durch das Smartphone als Ausfüllgerät beeinflusst werden und im Vergleich zum PC/Laptop anders ausfallen, spricht dies für Moduseffekte des Smartphones. Jede Datenerhebung gelingt nur valide, wenn die Antwortqualität der Befragten hoch ist (Edwards 2019; MacKenzie und Podsakoff 2012; Weisberg 2009, S. 18-19). Qualitätsunterschiede in den Antworten können die Zuverlässigkeit inferenzstatistischer Hypothesentests verzerren, da sie die Sensitivität und Spezifität der Hypothesenprüfung verringern (z. B. Clark, Girona und Young 2003; Huang, Liu und Bowling 2015b; Maniaci und Rogge 2014). Hohe Sensitivitäts- und Spezifitätsraten garantieren die größtmögliche Fehlerreduktion in der Hypothesenprüfung. Hypothesen sollten weder fälschlicherweise häufiger abgelehnt

(Fehler erster Art) noch akzeptiert (Fehler zweiter Art) werden, als durch einkalkulierte Irrtumswahrscheinlichkeiten vorgegeben (vgl. Lehmann 1993; Neyman und Pearson 1933).

2.4.2 Erarbeitung des Forschungsplans

Um mögliche Moduseffekte des Smartphones als Ausfüllgerät aufzudecken, muss abgewogen werden, auf welche Bereiche des TSE das Smartphone als Ausfüllgerät in welcher Form Einfluss nehmen könnte. Für diese kann dann gezielt untersucht werden, ob das Smartphone Moduseffekte induziert. *Abbildung 4* stellt das Prüfkonzept der vorliegenden Arbeit übersichtlich dar. Das Smartphone könnte systematisch die Antwortqualität und die Unit-Nonresponse beeinflussen und darüber auf den TSE in Onlinebefragungen wirken. Für Coverage als Auswahlrahmen der Befragten entstehen hingegen durch das Smartphone keine kritischen Fehlereinflüsse. Die Möglichkeit der Teilnahme mit dem Smartphone ist aus der Perspektive der Coverage sogar notwendig (siehe Kapitel 2.2.3). Moduseffekte des Smartphones sind zudem nicht für alle Unterasspekte der Antwortqualität plausibel. Bei selbst-administrierten Onlinebefragungen treten keine Fehleranteile durch Messfehler seitens Interviewer*innen auf (Weisberg 2009, S. 45), sodass als wesentliche Prüfbedingungen der Antwortqualität der Messfehler durch Befragte und die Item-Nonresponse in das Forschungsprogramm integriert werden. Für diese Teilbereiche, auf die das Smartphone im TSE-Framework wirken könnte, werden nachfolgend die wesentlichen forschungsleitenden Fragen vorstrukturiert und als Forschungsplan angelegt.

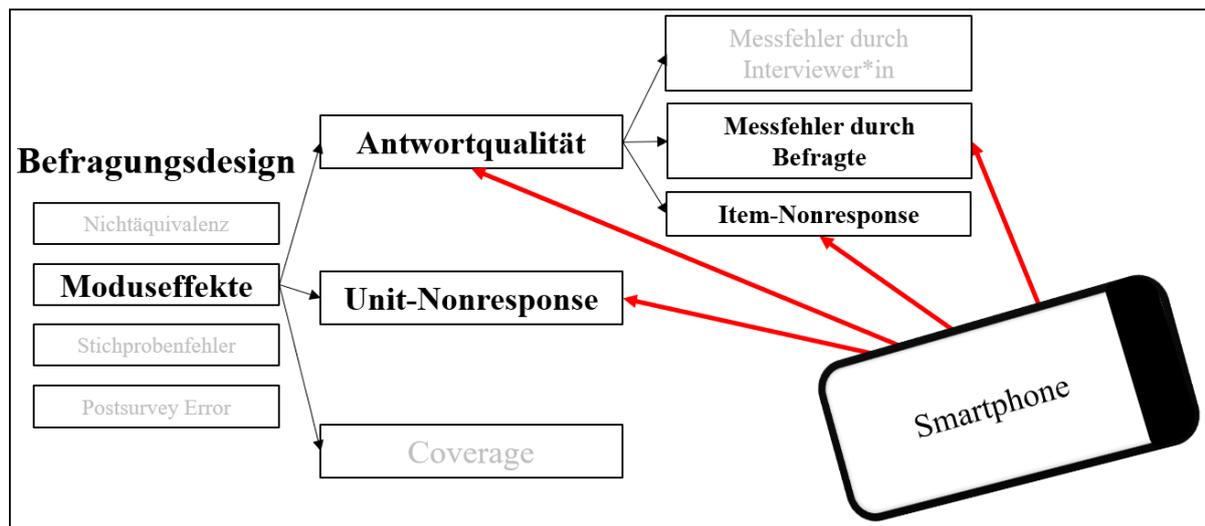


Abbildung 4. Prüfkonzept möglicher Moduseffekte des Smartphones in der vorliegenden Arbeit
Die Abbildung wendet das TSE-Konzept auf die Prüfung des Smartphones als Ausfüllgerät in Onlinebefragungen an. Fett gedruckt sind die Teilbereiche des TSE, die durch mögliche Moduseffekte des Smartphones beeinflusst werden könnten (Einflüsse durch Pfeile verbildlicht). Daraus ergeben sich zwei Hauptbereiche für die Prüfung des Smartphoneeinsatzes: Antwortqualität und Unit-Nonresponse, für die Moduseffekte des Smartphones entstehen könnten. Das Smartphone könnte dabei Einfluss auf den Messfehler durch Befragte und Item-Nonresponse als Unterbereiche der Antwortqualität nehmen.

2.4.2.1 Voruntersuchungen zur Antwortqualität und Unit-Nonresponse

Sollten sich tatsächlich Moduseffekte des Smartphones zeigen, müssen diese analysiert und kritische von unkritischen Einsatzmöglichkeiten getrennt werden (Engel 2015, S. 17), um die Validität des Smartphones als Ausfüllgerät in Onlinebefragungen zu sichern. *Abbildung 4* zeigt, dass sich Moduseffekte des Smartphones im TSE-Framework auf die Antwortqualität und die Unit-Nonresponse auswirken könnten. Tourangeau (2017, S. 117) führt aus, dass Antwortqualität und Unit-Nonresponse in Voruntersuchungen voneinander gelöst untersucht werden sollten, um separat zu ermitteln, wie (am jeweiligen Gerät) geantwortet wird (Antwortqualität) und wer (mit welchem Gerät) teilnimmt (Unit-Nonresponse). Dieser Empfehlung folgt die vorliegende Arbeit, sodass zunächst zwei Voruntersuchungen durchgeführt werden, die mögliche Moduseffekte des Smartphones getrennt für Antwortqualität und Unit-Nonresponse analysieren. Die Voruntersuchungen orientieren sich am bisherigen Forschungsstand, der im nachfolgenden Kapitel 3 (insbesondere 3.1) aufbereitet wird. In Vorbereitung darauf werden in diesem Abschnitt separat diejenigen Anwendungsfälle für die beiden TSE-Bereiche (Antwortqualität und Unit-Nonresponse) vorstrukturiert, die besonders anfällig für mögliche Moduseffekte des Smartphones als Ausfüllgerät sind und daher in den Voruntersuchungen betrachtet werden.

Die erste Voruntersuchung der vorliegenden Arbeit fokussiert die Antwortqualität und untersucht diese speziell für sensible Themenbereiche (siehe auch Kapitel 2.1), da sensible Fragen besonders empfänglich für Moduseffekte der Datenerhebung sind (z. B. Jäckle et al. 2010; Weisberg 2009, S. 278). Durch die psychologische Bedeutung des Smartphones, die es als sehr vertrautes und persönliches Gerät kennzeichnet (siehe Kapitel 2.3), könnten bei sensiblen Fragen möglicherweise Moduseffekte auftreten, die den ehrlicheren Selbstbericht am Smartphone fördern. Ganz eindeutig wären Moduseffekte hier erkennbar, wenn z. B. die erhobenen Daten gerätweise in unterschiedlichen Ergebnissen resultieren (Weisberg 2009, S. 282). Die TSE-Unterbereiche der Antwortqualität (Messfehler durch Befragte und Item-Nonresponse) werden in der ersten Voruntersuchung daher auf Einflüsse des Smartphones im Themenfeld sensibler Fragen geprüft, die mögliche Moduseffekte besonders deutlich signalisieren würden. Sollte das Smartphone Moduseffekte induzieren und damit am Smartphone auf sensible Fragen anders geantwortet werden, könnte sich dies in Unterschieden in den Stichprobenkennwerten (als besonders offensichtlicher Ausdruck des Messfehlers durch Befragte) und der Antwortbereitschaft auf sensiblen Fragen (als passende Bezeichnung der Item-Nonresponse bei sensiblen Themen) manifestieren. Wenn bereits für diese Aspekte klare Antwortqualitätsunterschiede auftreten, müsste dies beim Einsatz des Smartphones als Ausfüllgerät bei sensiblen Fragen und sicherheitshalber auch bei allgemeinen Themen stets berücksichtigt werden, da die Validität

von Analysen, die mehrere Konstrukte in Beziehung setzen (so insbesondere regressionsanalytische Auswertungen) dann ebenfalls gerätübergreifend variieren könnte (z. B. Fuller 2004, S. 617-618; MacKenzie und Podsakoff 2012; Osborne und Waters 2002).

Im nächsten Schritt des Forschungsprogramms betrachtet die zweite Voruntersuchung mögliche Moduseffekte für die Unit-Nonresponse bei der Teilnahme mit dem Smartphone. Verweigern Befragte die Befragungsteilnahme, sind direkte Beurteilungen der Unit-Nonresponse nicht möglich, da dafür Angaben der Nicht-Teilnehmer*innen erforderlich sind. Die zweite Voruntersuchung nähert sich der Untersuchung der Unit-Nonresponse an, indem geprüft wird, ob sich die Personengruppen unterscheiden, die unterschiedliche Ausfüllgeräte nutzen. Für die Identifikation möglicher kritischer Moduseffekte, die den TSE erhöhen würden, müssen dabei zwei Szenarien unterschieden werden: 1. Die entstehenden Gruppenunterschiede (z. B. Jüngere nehmen eher mit dem Smartphone teil und Ältere mit dem PC/Laptop) hängen nicht mit Zielvariablen einer Untersuchung zusammen und 2. Die Hintergrundvariablen der Gerätwahl korrespondieren mit den Zielvariablen einer Studie (z. B. Jüngere nehmen am PC/Laptop in geringerem Ausmaß teil, wodurch eventuell etwa das Einkommen als Zielvariable am PC/Laptop überschätzt werden könnte). Der erste Fall ist eher unkritisch, da hier durch die verschiedenen Ausfüllgeräte eher Repräsentativität erreicht werden kann. Der zweite Fall würde hingegen Moduseffekte signalisieren, die den TSE erhöhen, da gewisse Personengruppen an verschiedenen Ausfüllgeräten unterrepräsentiert sind. Dies eröffnet dann Probleme, sobald der Anteil der Personen, die den PC/Laptop als Ausfüllgerät nutzen, sich vom Anteil der Personen, die ein Smartphone verwenden, unterscheidet, da die resultierenden Schlussfolgerungen durch die gerätspezifischen Unterschiede in den Raten der Unit-Nonresponse verzerrt werden. Die zweite Voruntersuchung prüft daher, ob bestimmte personenspezifische Variablen mit der Gerätwahl der Befragten verknüpft sind.

Wie ausgangs Kapitel 2.2.3 berichtet, ist die Befragungsforschung zunehmend mit einer gewissen Befragungsmüdigkeit konfrontiert, die es Forscher*innen ohnehin schon erschwert, eine adäquate Anzahl motivierter Befragter zu akquirieren (Groves 2006; Wu et al. 2022). Dabei hängt die Teilnahmebereitschaft stark von den Normen, Werten oder Einstellungen der Proband*innen ab (Weisberg 2009, S. 229). Nur wenn potenzielle Teilnehmer*innen ihre psychosozialen Bedürfnisse für die Befragungsteilnahme gewahrt sehen, nehmen sie auch wirklich teil (z. B. Biemer 2010; Groves, Cialdini und Couper 1992). Insbesondere für Mobilgeräte ist dabei die Passung zwischen Gerät und Nutzungskontext wesentlich (Campbell 2007). Personen verwenden Smartphones vorrangig dann, wenn sie den Einsatz als effektiv empfinden (Kim et al. 2017). Auch für Onlinebefragungen sollten Befragte entsprechend des TAM (siehe Kapitel

2.3.1) diejenigen Geräte wählen, deren Einsatz sie bei der Befragungsteilnahme als nützlich (WN) und einfach (WE_dN) wahrnehmen. Die Bedeutung des Smartphones als Ausfüllgerät steigt demnach proportional zu der empfundenen Effektivität des Smartphoneeinsatzes.

Schon jetzt präferieren viele Menschen das Smartphone für die Befragungsteilnahme (Krämer 2017, S. 100; Revilla et al. 2016b). In der zweiten Voruntersuchung wird für die Hintergründe möglicher Moduseffekte des Smartphones auf die Unit-Nonresponse auch gezielt die psychologische Bedeutung von Smartphones (siehe Kapitel 2.3.2) einbezogen. Bei der Untersuchung wird berücksichtigt, ob und unter welchen Umständen die Gerätwahl in Onlinebefragungen durch den Affekt und die psychologischen Bedürfnisse, die mit den verschiedenen Ausfüllgeräten verknüpft sind, beeinflusst wird. Wenn verschiedene Ausfüllgeräte befragungsrelevante Bedürfnisse der Proband*innen unterschiedlich gut saturieren, könnte sich die Gruppe der Smartphone-Teilnehmer*innen also tatsächlich strukturell von der Gruppe der Befragten, die mit dem PC/Laptop teilnehmen, unterscheiden. So könnten gerätspezifische Unterschiede in der Unit-Nonresponse entstehen (Weisberg 2009, S. 160-165), die, wenn sie mit den Zielvariablen einer Studie assoziiert sind, den TSE erhöhen können.

2.4.2.2 Antwortqualität als Hauptbereich des Forschungsprogramms

Im TSE-Framework ist die Antwortqualität, verstanden als umfassenderes Konzept mit verschiedenen Unteraspekten, für Moduseffekte besonders anfällig, weshalb das Forschungsprogramm darauf abgestimmt in der Folge weiter vertieft wird. Der Messfehler durch Befragte und die Item-Nonresponse stellen diejenigen TSE-Unterbereiche der Antwortqualität dar (siehe *Abbildung 4*), die in selbst-administrierten Onlinebefragungen durch das Ausfüllgerät beeinflusst werden könnten (Biemer 2010; Weisberg 2009, S. 73-91). Der Untersuchungsansatz löst sich dafür vom Themenbereich sensibler Fragen und wird breiter angelegt.

Gerade in Onlinebefragungen basiert die Antwortqualität wesentlich auf dem Antwortverhalten der Befragten (z. B. Bautista 2012, S. 38; Edwards 2019; Huang et al. 2012; Meade und Craig 2012; Wiersma 2013). Sollte die Befragung am Smartphone etwa herausfordernder erlebt werden (siehe Kapitel 2.2.4), könnten die abgegebenen Antworten stärker von den wahren Werten abweichen (höherer Messfehler durch Befragte), oder es kommt ggf. dazu, dass Fragen am Smartphone häufiger nicht oder nicht-substanziell beantwortet werden (mehr Item-Nonresponse). Auch die psychologische Bedeutung, die das Smartphone im Vergleich zu PCs/Laptops für Nutzer*innen einnehmen kann (siehe Kapitel 2.3.2), begründet die Vermutung, dass sich das Antwortverhalten am Smartphone vom Antwortverhalten am PC/Laptop unterscheiden

könnte und daraus möglicherweise Antwortqualitätsunterschiede folgen. Smartphones sind omnipräsente, sehr persönliche Geräte, die emotionalen Wert besitzen (z. B. Bock et al. 2016a; Brasel und Gips 2014; Melumad und Pham 2020; Thorsteinsson und Page 2014). Vor dem Hintergrund, dass Smartphones sogar vermenschlicht (Fullwood et al. 2017) und in die eigene Identität integriert werden (Carter et al. 2016, S. 1438; Park und Kaye 2019) – jeweils genauer ausgeführt in den Kapiteln 2.3.1 und 2.3.2 –, könnten Befragte anders antworten, wenn sie ein Smartphone bei der Teilnahme nutzen.

Um mögliche Moduseffekte des Smartphones hinsichtlich der Antwortqualität möglichst gehaltvoll untersuchen zu können, adressieren die Hauptuntersuchungen des Forschungsprogramms explizit die Hintergründe der Antwortqualität. Sowohl der Messfehler durch Befragte als auch die Item-Nonresponse steigen im TSE-Framework, wenn die Motivation und dadurch die Aufmerksamkeit der Teilnehmer*innen reduziert sind (z. B. Jäckle et al. 2010; MacKenzie und Podsakoff 2012; Steinbrecher, Roßmann und Blumenstiel 2015; Weisberg 2009, S. 286-287). Hier könnten also Moduseffekte entstehen, wenn die Aufmerksamkeit als Folge der Motivation durch den Teilnehmerahmen und die Besonderheiten des Smartphones beeinflusst wird. Insbesondere Implikationen, die sich aus dem Smartphone als Ausfüllgerät für mögliche Messfehler durch Befragte ergeben könnten, sind vielschichtig, da der Prozess der Antwortgabe an verschiedenen Stellen und unterschiedlich stark beeinflusst werden kann. Wie dargestellt werden wird (siehe Kapitel 3.2.1 und 3.3), ist die Analyse der gerätassoziierten Aufmerksamkeitsstrukturen für die Prüfung von Antwortqualitätsunterschieden dabei besonders geeignet. Um allgemeine Moduseffekte des Smartphones auf die Antwortqualität (die auch abseits sensibler Themen die Antwortqualität eher hemmen könnten) ausschließen zu können, sollte die Aufmerksamkeit der Teilnehmer*innen gerätübergreifend ähnlich sein. Im besten Fall antworten Teilnehmer*innen stets reflektiert und gewissenhaft. Dieses Antwortverhalten wird als „Optimizing“ definiert und ist verletzt, wenn Befragte den Aufwand der Antwortgabe verringern und sogenanntes „Satisficing“ auftritt (vgl. Krosnick 1991, 1999; Roßmann 2017). Ein Extremfall von Satisficing ist „Careless Responding“, das extrem unaufmerksames Antworten beschreibt (vgl. Curran 2016; Hong, Steedle und Cheng 2020; Meade und Craig 2012). Die Hauptuntersuchungen der vorliegenden Arbeit fokussieren mögliche Moduseffekte des Smartphones als Ausfüllgerät hinsichtlich der Antwortqualität, indem zunächst geprüft wird, ob sich die Aufmerksamkeitsstrukturen während der Teilnahme am Smartphone im Vergleich zum PC/Laptop unterscheiden (Studie 3) und anschließend, ob Antwortqualitätsunterschiede als Folge verringerter Aufmerksamkeit (insbesondere Careless Responding) am Smartphone bei der Befragungsteilnahme tatsächlich häufiger auftreten (Studie 4).

3 Forschungsstand und theoretische Herleitung

Das Kapitel strukturiert für den in Kapitel 2.4 beschriebenen Forschungsplan den Forschungsstand hinsichtlich möglicher Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen und erarbeitet offene Fragen. Entsprechend werden in den nachfolgenden Kapiteln zunächst die wesentlichen Ansatzpunkte für die Voruntersuchungen möglicher Moduseffekte im TSE-Framework literaturbasiert gefestigt und die zugehörige bisherige Befundlage aufbereitet (Kapitel 3.1). Der Forschungsstand wird dann im nächsten Abschnitt für die Antwortqualität deutlich breiter ausgeführt, wobei nun alle möglichen Moduseinflüsse des Smartphones auf die Antwortqualität sowie die theoretischen Hintergründe der Antwortqualität integriert werden (Kapitel 3.2). Das Kapitel mündet in einer umfangreichen Sammlung und Strukturierung der relevanten Literatur zu unaufmerksamem Antworten („Careless Responding“) in Onlinebefragungen (Kapitel 3.3). Etwaige Auswirkungen des Smartphones als Ausfüllgerät in Onlinebefragungen auf die Motivation und dadurch auch auf die Aufmerksamkeit der Befragten, sowie eventuell daraus folgende Konsequenzen für die Antwortqualität, stellen einen kaum beachteten Forschungsbereich dar, sodass diese Perspektivöffnung die Beurteilung von möglichen Moduseffekten des Smartphones im TSE-Framework wesentlich erweitert.

3.1 Moduseffekte des Smartphones als Ausfüllgerät

Die Voruntersuchungen betrachten die Einflüsse möglicher Moduseffekte auf die Antwortqualität und auf die Unit-Nonresponse gemäß den Empfehlungen von Tourangeau (2017, S. 117) getrennt. Eingang stellt das Kapitel die Erkenntnislage zu Unterschieden in den Stichprobenkennwerten und der Antwortbereitschaft begrenzt auf den Themenbereich sensibler Fragen als erste Ansatzpunkte der Prüfung der Antwortqualität dar (Kapitel 3.1.1). Im Anschluss daran wird separat der Forschungsstand zu möglichen Einflüssen des Smartphones auf die Unit-Nonresponse in Onlinebefragungen ohne Begrenzung der thematischen Befragungsinhalte aufbereitet (Kapitel 3.1.2). Dabei werden jeweils, falls existent, maßgebliche Forschungslücken identifiziert und erste übergeordnete Hypothesen für das Forschungsprogramm abgeleitet.

3.1.1 Moduseffekte des Smartphones auf Stichprobenkennwerte und die Antwortbereitschaft bei sensiblen Fragen

Mit Blick auf das Forschungsziel wird die Befundlage zu Moduseffekten des Smartphones bei sensiblen Fragen hinsichtlich der Antwortqualität strukturiert. Unterschiede in Stichprobenkennwerten sind ein besonders offensichtlicher Ausdruck des Messfehlers durch Befragte. Die vorliegende Arbeit bezeichnet zudem die Item-Nonresponse im Kontext der sensiblen Fragen

als Antwortbereitschaft, da hier die Nicht-Antwort (komplette Antwortverweigerung oder nicht-substanzielles Antworten) eher bewusst stattfindet.

3.1.1.1 Forschungsstand und Prüfkonzept

Moduseffekte des Smartphones in Onlinebefragungen, die sich in Unterschieden in den Stichprobenkennwerten für die einzelnen Ausfüllgeräte offenbaren, würden sehr klare Messfehler durch Befragte signalisieren. Die Literaturbasis dazu ist bereits vergleichsweise breit: Eine Vielzahl der Studien findet, dass die Datenerhebung themenübergreifend zu ähnlichen Ergebnissen führt, wenn mit Smartphones teilgenommen wird (z. B. de Bruijne und Wijnant 2013; Liebe et al. 2015; Mavletova 2013; Mavletova und Couper 2013; Peytchev und Hill 2010; Revilla und Ochoa 2016). Andere Untersuchungen zeichnen teilweise ein kontroverseres Bild: Vereinzelt finden sich Unterschiede in den Stichprobenkennwerten (z. B. in Mittelwerten) zwischen Personen, die mit dem Smartphone teilnehmen, und PC/Laptop-Teilnehmer*innen (z. B. Krebs und Höhne 2021; Struminskaya et al. 2015; Tourangeau et al. 2017). Ob dies Moduseffekte signalisiert, ist umstritten: Scheinbare Abweichungen zwischen Befunden auf Basis unterschiedlicher Ausfüllgeräte verschwinden laut Cook (2014) durch die Kontrolle soziodemografischer Unterschiede der Nutzer*innen verschiedener Geräte.

Wie schon bei der Einführung des TSE-Konzepts erläutert (siehe Kapitel 2.1), sollten insbesondere sensible Themen besonders empfänglich für Moduseffekte sein (Barnett 1998; Tourangeau und Yan 2007; Weisberg 2009, S. 278-280). In Onlinebefragungen wird beispielsweise offener auf sensible Fragen als in schriftlichen Befragungen geantwortet (z. B. de Leeuw 2005; Gnams und Kaspar 2015; Kreuter et al. 2008; Langhaug, Sherr und Cowan 2010; Ritter und Sue 2007; Taddicken 2009, S. 98; Tourangeau und Smith 1996; Tourangeau und Yan 2007; Wharton et al. 2003). Die Forschungsliteratur ist sich einig, dass sensible Themenbereiche auch für die Aufdeckung möglicher Moduseffekte des Smartphones als Ausfüllgerät den erfolgversprechendsten Anwendungsfall darstellen (z. B. Antoun et al. 2017; Lee et al. 2019; Mavletova 2013; Mavletova und Couper 2013; Toninelli und Revilla 2016b).

Sensible Fragen werden weniger offen und seltener beantwortet, vor allem wenn die Teilnehmer*innen Angst vor negativen Konsequenzen bzw. Verurteilungen haben und daher Aufdeckung vermeiden wollen (vgl. Gnams und Kaspar 2015; Lee 1993). Die eigene Sexualität (vgl. Langhaug et al. 2010; McCallum und Peterson 2012; Wolter 2012), Delinquenz/Kriminalität (vgl. Barnett 1998; Kleck und Roberts 2012), Alkohol- und Drogenkonsum (vgl. Krumpal und Näher 2012; Tourangeau und Yan 2007), Einkommen (vgl. Barnett 1998; Tourangeau und Yan 2007) und das Wahlverhalten (vgl. Krumpal und Näher 2012; Tourangeau und Yan 2007)

sind klassische Beispiele sensibler Themenbereiche. Bei sensiblen Fragen treten zudem bewusste Antwortverzerrungen in Form von sozial erwünschtem Antworten häufiger auf (Jäckle et al. 2010; Weisberg 2009, S. 289-291). Soziale Erwünschtheit bezeichnet Antwortverhalten, bei dem die Befragten nicht aufrichtig antworten, da ihre wahrheitsgemäße Antwort soziale und gesellschaftliche Normen oder Werte verletzen würde und sie sich gerne positiv präsentieren möchten (z. B. Krumpal 2013; Steenkamp, de Jong und Baumgartner 2010; Tourangeau und Yan 2007; Weisberg 2009, S. 289). Je weniger sozial erwünscht eine Antwort erscheint, desto seltener wird sie in Studien genannt (Jann 2015, S. 102; Stocké 2004).

Während die vorliegende Arbeit Moduseffekte betrachten möchte, die dem möglichen reinen Geräteffekt des Smartphones entspringen, ist auch der Befragungskontext für Moduseffekte höchst relevant und wirkt unmittelbar auf das Antwortverhalten der Befragten (z. B. Höhne, Revilla und Schlosser 2020a; Miller 2008; Mutz und Kämpfer 2011; Pan, Woodside und Meng 2014). Besonders bei sensiblen Themen hängen die Auskunftsbereitschaft und das Ausmaß an sozial erwünschtem Antwortverhalten stark vom Kontext der Befragungsteilnahme ab (Barnett 1998; Tourangeau und Yan 2007; Wolter 2012, S. 62-64). Wenngleich ohnehin bei Onlinebefragungen andere Kontextbedingungen (z. B. Möglichkeit, parallel soziale Medien zu nutzen oder E-Mails zu beantworten) vorherrschen als bei schriftlichen Befragungen (siehe Kapitel 2.1), ist die Diversität des Ausfüllkontexts (z. B. Teilnahme in Bus und Bahn) durch den Einsatz von Smartphones noch zusätzlich potenziert (z. B. Antoun et al. 2017; Höhne und Schlosser 2019). Die Teilnahme an öffentlicheren Orten als dem eigenen Zuhause ist dabei beispielsweise anfälliger für sozial erwünschtes Antworten und geringere Antwortqualität bei sensiblen Themenbereichen (Vicente und Lopes 2013, S. 144). Auch durch die Anwesenheit Dritter (sogenannte Bystander) sinkt meist die Offenheit der Befragten: sensible Fragen werden weniger aufrichtig beantwortet und Verzerrungen durch soziale Erwünschtheit steigen (z. B. Aquilino 1993; Aquilino, Wright und Supple 2000; Becker und Günther 2004; Reuband 1987). Sollte die Teilnahme an Onlinebefragungen außerhalb der eigenen vier Wände mit dem Smartphone häufiger auftreten oder sollten bei der Smartphoneteilnahme häufiger Bystander zugegen sein, kann dies also schon unabhängig vom Geräteffekt des Smartphones direkt auf die Antwortgabe bei sensiblen Fragen wirken. Dies ist erwartbar, da der Kontext der mobilen Internetnutzung auch abseits von Onlinebefragungen tatsächlich diverser ist als bei PCs/Laptops und auch verstärkt an öffentlichen Orten stattfindet (z. B. Deng et al. 2019; Lee, Kim und Kim 2005).

Die Forschungsliteratur zeigt, dass in Onlinebefragungen die Teilnehmeraten im eigenen Zuhause für verschiedene Ausfüllgeräte vereinzelt gleich hoch sind (Toninelli und Revilla 2016b). Viele Befunde deuten aber an, dass mit dem Smartphone tatsächlich häufiger als mit

dem PC/Laptop an öffentlichen Orten oder bei der Arbeit statt im eigenen Zuhause teilgenommen wird (z. B. Antoun et al. 2017; de Bruijne und Oudejans 2015, S. 136; Lütters 2017, S. 68; Mavletova und Couper 2013). Insgesamt befindet sich aber auch die deutliche Mehrheit der Teilnehmer*innen mit dem Smartphone bei der Teilnahme zu Hause (Revilla et al. 2016a). Allerdings sind generell während des Ausfüllens einer Befragung mit dem Smartphone tatsächlich häufiger Dritte anwesend als während der Fragebogenbeantwortung mit dem PC/Laptop (z. B. Bosnjak et al. 2017, S. 62; Mavletova und Couper 2013; Revilla et al. 2016a; Toninelli und Revilla 2016b). Die Bystander bei der Smartphoneteilnahme sind für die Befragten in der Regel aber keine unbekanntenen Personen (Toninelli und Revilla 2016b) und trotz der größeren Häufigkeit von Bystandern wird die Teilnahme mit dem Smartphone im Feld nicht zwingend negativ eingeschätzt: Die Befragten berichten gerätübergreifend ein ähnliches Ausmaß von Wohlbefinden und bewerten trotz Bystandern die Privatheit bei der Smartphoneteilnahme mindestens gleich gut (z. B. Revilla et al. 2016a; Toninelli und Revilla 2016b).

Bisherige Studien fokussieren bei der Untersuchung des Smartphones als Ausfüllgerät bei sensiblen Fragen trotzdem lediglich mögliche Einflüsse des potenziell diverseren Befragungskontexts, der bei der Smartphoneteilnahme entsteht, und werten diesen als Hemmschwelle der Selbstoffenbarung (vgl. Mavletova und Couper 2013; Toninelli und Revilla 2016b). Bei sensiblen Themen werden aufgrund des Kontexts am Smartphone negative Einflüsse auf die Antwortqualität vermutet. Diese könnten sowohl auf die Stichprobenkennwerte (als Ausdruck eines sehr starken Messfehlers durch Befragte) als auch auf die Antwortbereitschaft als ganz bewusste Item-Nonresponse, da bei der Antwortbereitschaft die Nicht-Antwort (anders als bei genereller Item-Nonresponse in anderen Themenbereichen) meist vom adressierten Iteminhalt abhängt, wirken. Unter dieser Forschungsprämisse zeigen sich in der Literatur in Onlinebefragungen bisher weder Unterschiede in den Stichprobenkennwerten noch bei der Antwortbereitschaft bei sensiblen Fragen zwischen verschiedenen Ausfüllgeräten: Befragte antworten am Smartphone im Feld ähnlich aufrichtig und ähnlich häufig auf sensible Fragen wie am PC/Laptop sowie auch ähnlich sozial erwünscht (Antoun et al. 2017; Couper et al. 2017, S. 146; Mavletova 2013; Revilla et al. 2016a; Toninelli und Revilla 2016b).

3.1.1.2 Herleitung der Hypothese der ersten Voruntersuchung zur Antwortqualität

Die Einheitlichkeit der oben zitierten Befunde überrascht nicht, da allen gemein ist, dass sie aus dem Feld stammen und die Trennung der Gerät- und Kontexteffekte nicht bei der Beurteilung

der Modusbesonderheiten des Smartphones berücksichtigen². Die erste Voruntersuchung der vorliegenden Arbeit zielt jedoch genau darauf, den Geräteffekt des Smartphones als Ausfüllgerät für die Einschätzung möglicher Moduseffekte des Smartphones zu isolieren. Da die Kontexteffekte der Teilnahme sowie Selbstselektionseffekte bei der Wahl des Ausfüllgeräts wesentlich für die Antwortqualität sind (Schouten et al. 2013; Tourangeau 2017, S. 124), soll der mögliche Einfluss des Smartphones auf die resultierenden Stichprobenkennwerte und auf die Antwortbereitschaft bei sensiblen Fragen davon entkoppelt evaluiert werden. Daher werden die Besonderheiten des Smartphones als Ausfüllgerät in Onlinebefragungen für die Bereitschaft zum ehrlichen Selbstbericht bei sensiblen Themen in die Herleitung der ersten Hypothese des Forschungsprogramms der vorliegenden Arbeit einbezogen.

Die Geräteigenschaften und die psychologische Bedeutung des Smartphones lassen vermuten, dass mögliche Moduseffekte des Smartphones bei sensiblen Themen eher vorteilhaft für den ehrlichen Selbstbericht in Onlinebefragungen sein könnten, wenn Kontexteffekte konstant gehalten werden. Gerade die psychologische Bedeutung des Smartphones, die in dieser Arbeit bereits eingeführt wurde (siehe Kapitel 2.3.2), findet in der Forschung für mögliche Moduseffekte noch kaum Berücksichtigung. Die Interaktion mit Smartphones kann bei geübten Nutzer*innen ein Gefühl von Komfort und Zufriedenheit während der Nutzung auslösen (siehe Kapitel 2.3), wobei Wohlbefinden während der Befragungsteilnahme zugleich zu aufrichtigeren Antworten auf sensible Fragen führt (Krumpal 2013). Die Nutzung des eigenen Smartphones birgt zudem als Coping-Strategie für Stress, den sensible Themen induzieren können, eine beruhigende Wirkung für die Nutzer*innen (Melumad und Pham 2020). Für das Smartphone spricht ebenfalls, dass es als sehr privates Gerät wahrgenommen wird (z. B. Fullwood et al. 2017; Vincent 2006, S. 135) und dass hohe erlebte Privatheit ehrliche Antworten auf sensible Fragen begünstigt (de Leeuw 2005; Ong und Weiss 2000). Bei der Smartphonennutzung entsteht ein Paradoxon der Privatheit: Moderne Informationstechnologien können über subjektives Kontrollerleben den Bericht sensibler Informationen fördern, obwohl sie ein objektiv höheres Entdeckungsrisiko in der digitalen Welt bergen (Brandimarte, Acquisti und Loewenstein 2013). Tatsächlich zeigt sich eine hohe Bereitschaft zur Selbstoffenbarung und Preisgabe privater Informationen am Smartphone in diversen Settings und für verschiedene Formen der Antwortgabe (Melumad und Meyer 2020). Smartphones sind außerdem für ihre Besitzer*innen sehr präsent und häufig emotional besetzt (Melumad, Inman und Pham 2019; Melumad und Pham

² In der Forschungspraxis ist der Gesamteffekt (Geräteffekt plus Kontexteffekt) vorrangig relevant. Zum isolierten Geräteffekt des Smartphones als Ausfüllgerät bei sensiblen Fragen zeigt sich in der Literatur aber eine klare Forschungslücke, für die eine Voruntersuchung angelegt wird, um das Forschungsprogramm adäquat auf die Prüfung des Smartphones als Ausfüllgerät und mögliche geräteexklusive Moduseffekte ausrichten zu können.

2020). Dies überträgt sich etwa auch auf Onlinerezensionen (z. B. Produktbewertungen) von Kundinnen und Kunden, die mit dem Smartphone verfasst werden: Diese weisen eine höhere Valenz auf als Bewertungen, die an anderen Geräten geschrieben werden (Melumad et al. 2019; Piccoli und Ott 2014). Während die Kontexteffekte der Smartphoneteilnahme potenziell eher negativ auf das Antworten bei sensiblen Fragen wirken könnten, spricht der reine Geräteeffekt aufgrund der psychologischen Bedeutung und der Nutzungsgewohnheiten also eher dafür, dass sensible Fragen am Smartphone offener und häufiger beantwortet werden.

Selbst-administrierte Onlinebefragungen besitzen bereits im Vergleich zu schriftlichen Befragungen generelle Modusvorteile bei der Aufdeckung sensibler Informationen (siehe Kapitel 2.1 und weiter oben 3.1.1.1). Wenn mit dem Smartphone teilgenommen wird, könnten die Vorteile selbst-administrierter Onlinebefragungen bezüglich höherer Anonymität (Buchanan 2000, S. 123; Joinson 1999; Trau, Härtel und Härtel 2013; Ward et al. 2014) und somit für den aufrichtigen Selbstbericht bei sensiblen Themen (Bowling 2005; Newman et al. 2002) zunächst also fortbestehen. Darüber hinaus könnte das Smartphone als Gerät, wie beschrieben, auch noch zusätzlich eigenständige Moduseffekte induzieren, die den Selbstbericht weiter fördern.

Bisherige Untersuchungen zum Einfluss des Smartphones auf Selbstberichtstendenzen für sensible Themenbereiche in Onlinebefragungen haben diese potenziell positive Rolle des Smartphones jedoch vernachlässigt (vgl. Antoun et al. 2017; Mavletova 2013; Mavletova und Couper 2013; Toninelli und Revilla 2016b). Im Feld könnten sich aber die Wirkung von Kontexteffekten der Smartphoneteilnahme durch die größere Diversität (z. B. in Form von Ortsunabhängigkeit und Bystandern) und die möglichen positiven Geräteeffekte (z. B. größere wahrgenommene Privatheit und Kontrolle) möglicherweise aufheben. Daher werden in der ersten Voruntersuchung zur Abschätzung des reinen Geräteeffekts Kontexteffekte ausgeschlossen. Zur Maximierung der internen Validität bei der Evaluation potenzieller Moduseffekte in Befragungen sind Laborexperimente das Mittel der Wahl (Bowling 2005; Schouten et al. 2013; Tourangeau 2017, S. 124). In einer experimentellen Überprüfung des reinen Geräteinflusses des Ausfüllgeräts ermöglichen 1) die Konstanzhaltung des Ausfüllkontexts 2) die Fremdzuweisung des Ausfüllgeräts, mit dem die Befragten teilnehmen, die Untersuchung des isolierten Geräteeffekts des Smartphones. Da dieser Geräteinfluss des Smartphones für die Untersuchung sensibler Verhaltensweisen bei Konstanzhaltung möglicher negativer Kontexteffekte für die Antwortqualität förderlich sein könnte, wird die übergeordnete erste Hypothese formuliert:

H1: *Personen, die an einer Onlinebefragung mit dem eigenen Smartphone teilnehmen, beantworten sensible Fragen aufrichtiger als Personen, die unter gleichen Bedingungen mit dem eigenen PC/Laptop teilnehmen.*

Die **H1** prüft mögliche Moduseffekte hinsichtlich der Antwortqualität zunächst für den Anwendungsfall sensibler Fragen über Stichprobenkennwerte und Antwortbereitschaft. Für sensible Themenbereiche wird eher ein Vorteil des Smartphones als Ausfüllgerät postuliert. Wird die **H1** bestätigt, spricht dies für Moduseffekte des Smartphones, die positiv für den Selbstbericht bei sensiblen Fragen sind und somit den TSE sogar senken. Wird die **H1** verworfen, sollte dies ein Indiz für das Fehlen von Moduseffekten hinsichtlich der Antwortqualität bei sensiblen Fragen sein, da negative Effekte des Smartphones auf die Antwortbereitschaft und die Aufrichtigkeit bei sensiblen Fragen unter Konstanthaltung der Kontextbedingungen aufgrund des Forschungsstands nicht zu erwarten sind. Der Ausschluss von Moduseffekten bei sensiblen Themen würde in der Folge erlauben, das Forschungsprogramm für die Antwortqualität zu vertiefen und die Prüfung entkoppelt vom Themenkomplex sensibler Fragen breiter anzulegen.

3.1.2 Moduseffekte des Smartphones auf die Unit-Nonresponse

Die Unit-Nonresponse verkörpert im Forschungsplan (siehe Kapitel 2.4) den zweiten TSE-Bereich, auf den Moduseffekte des Smartphones wirken könnten und für den eine zweite Voruntersuchung potenzieller Einflüsse des Smartphones als Ausfüllgerät angelegt wird. Dementsprechend wird die Literatur zu möglichen Besonderheiten der Befragten, die mit verschiedenen Geräten an Befragungen teilnehmen, in den Blick genommen, um auf gerätspezifische Unterschiede in den Raten der Unit-Nonresponse schließen zu können. Aus der Forschungsliteratur werden mögliche Ausfallmechanismen und Gründe der Teilnahmeverweigerung erarbeitet, die mit dem Ausfüllgerät zusammenhängen könnten, um ein Prüfkonzept zu entwickeln.

3.1.2.1 Forschungsstand und Prüfkonzept

Kritische Unterschiede in der Unit-Nonresponse, die durch das Smartphone als Ausfüllgerät entstehen, könnten sich ergeben, wenn die Teilnahmeentscheidung in Onlinebefragungen gerätweise variiert und von Variablen abhängt, die mit den Zielvariablen einer Studie korrelieren. Dies erhöht den TSE, wenn dabei die Teilnehmeraten verschiedener Ausfüllgeräte nicht gleich sind (siehe Kapitel 2.4.2.1). Die direkte Untersuchung der Unit-Nonresponse ist in Onlinebefragungen kaum möglich, da nur Personen befragt werden können, die auch tatsächlich teilnehmen. Daher bildet die zweite Voruntersuchung in der vorliegenden Arbeit eine Vorstufe der Prüfung der Unit-Nonresponse, um Hinweise auf mögliche gerätspezifische Unit-Nonresponse-raten zu erhalten, die Moduseffekte des Smartphones signalisieren würden. Unterschiede zwischen den Personengruppen, die verschiedene Ausfüllgeräte nutzen, können darauf hindeuten, welche Variablen ggf. mit gerätspezifischer Unit-Nonresponse verknüpft sind. Wenn sich

Variablen identifizieren lassen, die mit der Gerätwahl zusammenhängen und auch tatsächlich dazu führen, dass bestimmte Geräte von verschiedenen Personen bevorzugt genutzt werden, könnten für Studien, die Konzepte untersuchen, die mit diesen Variablen zusammenhängen, Moduseffekte hinsichtlich der Unit-Nonresponse entstehen. Die zweite Voruntersuchung betreibt daher Grundlagenforschung und analysiert, ob bestimmte Personengruppen bevorzugt mit verschiedenen Geräten teilnehmen und auf welche Variablen bezogen sich Befragte unterscheiden, die verschiedene Ausfüllgeräte nutzen.

Tatsächlich zeigen sich in Onlinebefragungen beispielsweise soziodemografische Unterschiede zwischen Teilnahmegruppen, die das Smartphone oder den PC/Laptop nutzen. Besonders deutlich lassen sich in Onlinebefragungen Altersunterschiede und ein Geschlechtseffekt erkennen: Smartphone-Teilnehmer*innen sind im Durchschnitt jünger als Befragte, die einen PC/Laptop nutzen (z. B. Bosnjak et al. 2017, S. 57; Buskirk, Saunders und Michaud 2015; de Bruijne und Wijnant 2014; Erens et al. 2019; Krämer 2017, S. 93; Mavletova 2013; Sommer, Diederhofen und Musch 2017; Toepoel und Lugtig 2014), und häufiger weiblich (z. B. Bosnjak et al. 2017, S. 57; Cook 2014; de Bruijne und Wijnant 2014; Erens et al. 2019; Keusch und Yan 2017; Sommer et al. 2017). Für soziodemografische Variablen zeigen sich also Hinweise auf gerätweise Unterschiede in der Unit-Nonresponse, wodurch allerdings nur dann kritische Moduseffekte entstehen könnten, die den TSE erhöhen würden, wenn die Soziodemografika mit Zielvariablen einer Untersuchung zusammenhängen. Andernfalls kann durch die Teilnahmemöglichkeit mit dem Smartphone eher Repräsentativität erreicht werden, da über die Geräte hinweg die Altersstrukturen und Geschlechtshäufigkeiten der Gesamtbevölkerung besser abgebildet werden können.

Lediglich auf die Teilnahmezahlen bestimmter soziodemografisch definierter Personengruppen zu schauen und daraus Schlussfolgerungen zu ziehen, reicht für die Analyse der Unit-Nonresponse in Befragungen nicht aus (z. B. Bautista 2012, S. 40-43; Peytchev 2013). Da Alters- und Geschlechtseffekte durch die zunehmende Verbreitung des Smartphones kontinuierlich abnehmen (z. B. Antoun 2015, S. 104-110; Deng et al. 2019; Mavletova 2013; Poulová et al. 2019, S. 304-308; Toepoel und Lugtig 2014) und soziodemografischen Unterschieden gut begegnet werden kann (z. B. Cook 2014; de Bruijne und Wijnant 2014; Heerwegh und Loosveldt 2008; Schlosser und Silber 2020, S. 270-271; Sommer et al. 2017), wird für die zweite Voruntersuchung nach weiteren Variablen gesucht, die die Teilnahmeentscheidung mit einem bestimmten Ausfüllgerät in Befragungen beeinflussen könnten. Die Nutzung technischer Geräte hängt von persönlichen Werten und individuellen Erwartungen (z. B. hedonistisch oder utilitaristisch) ab (Wang 2016; Yang et al. 2012). Auch für das Smartphone präzisieren individuelle

Bedürfnisstrukturen die letztlich Nutzung (Park 2015). Für eine Annäherung an das Verständnis der gerätabhängigen Teilnahmeentscheidung in Onlinebefragungen können bei der Suche nach relevanten Variablen die theoretischen Hintergründe des TAM herangezogen werden. Sind WN und WEdN für das Smartphone (oder den PC/Laptop) hoch, steigt auch die Nutzungsintention (Joo und Sang 2013; Park und Chen 2007), sodass es als bevorzugtes Ausfüllgerät in Onlinebefragungen eingeschätzt werden könnte. Auf die Teilnahmeentscheidung mit verschiedenen Geräten könnten also etwa solche Variablen wirken, durch die Personen die WN und WEdN des Smartphones als Ausfüllgerät in Onlinebefragungen unterschiedlich einschätzen. Diese könnten dann gerätspezifische Unit-Nonresponse fördern und zu Moduseffekten führen.

Aufgrund der Geräteigenschaften und des Nutzungskontexts des Smartphones könnte das Gerät je nach Teilnehmer*in als unterschiedlich geeignet für Onlinebefragungen wahrgenommen werden, da die Interaktion mit dem Smartphone sehr individuell ist (Keusch et al. 2022) und die persönliche Verbindung zum Gerät (siehe Kapitel 2.3.2) sich interindividuell unterschiedlich auf die Nutzungsfrequenz auswirkt. Heterogenität in befragungsspezifischen Bedürfnisstrukturen zwischen Personen, die verschiedene Ausfüllgeräte nutzen, wird nach Kenntnisstand des Autors in die Prüfung von Moduseffekten des Smartphones bisher nicht einbezogen, um Hinweise auf mögliche gerätweise Unterschiede der Unit-Nonresponse zu erhalten. Zusätzlich wirken auch Geräterfahrungen und Bedürfnisse für die auszuführende Tätigkeit (wie z. B. der Wunsch nach Mobilität) auf die Gerätwahl (Venkatesh et al. 2003; Yang et al. 2012). Befriedigt ein Gerät die eigenen Bedürfnisse besser (z. B. bezogen auf Onlinebefragungen), wird es eher genutzt (Wang 2016). Auch wenn eine starke Bindung an das Gerät besteht, kann dies die Nutzungsintention fördern, wenn Geräte identitätsstiftend (IT-Identität) wirken und somit wahl dominant werden (Carter und Grover 2015). Solche befragungsspezifischen Werte und Einstellungen der Proband*innen tragen systematisch zum TSE in Befragungen bei (Weisberg 2009, S. 229) und sollten daher möglichst gerätübergreifend balanciert sein oder zumindest kontrolliert werden, falls sie für die Zielvariablen der eigenen Untersuchung bedeutsam sind.

3.1.2.2 Herleitung der Hypothese der zweiten Voruntersuchung

Damit sich Personen für die Teilnahme an Onlinebefragungen entscheiden und die Fragen sorgfältig beantworten, sollten die psychosozialen Bedürfnisse, Wünsche und Erwartungen der Teilnehmer*innen an die Befragung erfüllt werden (z. B. Biemer 2010; Groves et al. 1992). Das umfasst sowohl gerätunabhängige Bedürfnisse (z. B., ob die Dauer der Befragung akzeptabel ist) als auch womöglich mit dem Gerät assoziierte Bedürfnisse (z. B. Teilnahmeort bzw. -zeitpunkt oder gerätbezogene Teilnahmedauer). Für mögliche Moduseffekte des Smartphones

als Ausfüllgerät auf die Unit-Nonresponse sind diejenigen Bedürfnisse relevant, die das Teilnahmesetting betreffen und mit dem Gerät verknüpft sein könnten.

Onlinebefragungen am Smartphone wirken für Befragte zeitaufwendiger (Bosnjak et al. 2017, S. 61; de Bruijne und Wijnant 2013) und dauern studienübergreifend auch objektiv signifikant länger (z. B. Andreadis 2015a, S. 74; Antoun und Cernat 2020; Couper und Peterson 2017; Keusch und Yan 2017; Liebe et al. 2015; Mavletova 2013; Revilla et al. 2016a; Revilla und Ochoa 2016; Toninelli und Revilla 2020, S. 361). Solche längeren Bearbeitungszeiten können bei der Entscheidung zur Befragungsteilnahme jedoch abschreckend wirken (Galesic und Bosnjak 2009; Heberlein und Baumgartner 1978; Revilla und Ochoa 2017; Yammarino, Skinner und Childers 1991). Die Gerätesonderheiten des Smartphones ermöglichen andererseits eine flexiblere Teilnahme an selbstgewählten Orten und zu selbstgewählten Zeiten (siehe Kapitel 2.2 und 3.1.1.1). Auch dies könnte für die Gerätwahl in Onlinebefragungen bedeutsam sein, wenn Befragte möglichst selbstbestimmt teilnehmen möchten und sie diese Selbstbestimmung durch das Smartphone eher gewahrt sehen. Befragte, die besonders großen Wert auf flexible Teilnahmebedingungen legen oder eher unterwegs Befragungen beantworten möchten, könnten beispielsweise das Smartphone präferieren und sich häufiger gegen die Teilnahme entscheiden, wenn die Smartphoneteilnahme (gerade) nicht möglich ist. Im Gegenzug könnte der PC/Laptop von Teilnehmer*innen, die möglichst konzentriert teilnehmen möchten, ohne von anderen Gerätefunktionen unterbrochen zu werden, als bevorzugtes Gerät eingeschätzt werden. Wenn solche Bedürfnisse mit Zielvariablen einer Studie verknüpft sind (z. B. Wirkung verschiedener Führungsstile auf Mitarbeiter*innen, die Flexibilität schätzen), könnten kritische Moduseffekte der Unit-Nonresponse auftreten, etwa wenn weniger (oder mehr) Befragte mit dem Smartphone als mit dem PC/Laptop teilnehmen.

Da Befragungen möglichst belastungsarm gestaltet werden sollten, könnte zudem die psychologische Bedeutung des Smartphones die Teilnahmeentscheidung beeinflussen. Smartphones sind emotional besetzt und haben für die Besitzer*innen eine persönliche Bedeutung (siehe Kapitel 2.3.2). Sie fungieren eher als Komfortquelle und wirken stressreduzierend (Melumad und Pham 2020). In Onlinebefragungen zeigt sich analog dazu eine höhere Teilnahmezufriedenheit bei Befragten, die das Smartphone nutzen (Mavletova, Couper und Lebedev 2018). Erzeugt die Verbindung zum eigenen Smartphone (oder zum eigenen PC/Laptop) bei Personen das Bedürfnis, dieses eher für die Teilnahme zu nutzen, überträgt sich das womöglich auf die Unit-Nonresponse. Auch hier könnten Moduseffekte resultieren, wenn die Gerätebindung mit Zielvariablen einer Studie korrespondiert (z. B. Untersuchung des Kaufverhaltens im Internet, das mit der Smartphonebindung zusammenhängen könnte).

Die gerätspezifische Unit-Nonresponse als Indikator für Moduseffekte könnte, wie dokumentiert, mit psychosozialen Bedürfnissen und Zuständen (z. B. Flexibilitätsbedürfnisse, Zeitbedürfnisse bzw. Smartphone- oder PC/Laptopbindung) zusammenhängen. Die Besonderheiten des Smartphones, die auf die Rahmenbedingungen der Teilnahme an Onlinebefragungen wirken, und die mögliche persönliche Bedeutung des Geräts könnten dazu führen, dass das Smartphone als geeigneter oder weniger gut geeignet für bestimmte Bedürfnisdimensionen der Befragungsteilnahme eingeschätzt wird. Dies würde dazu führen, dass Befragte, die bestimmte derartige Bedürfnisse besonders schätzen, bevorzugt mit dem Gerät teilnehmen, das für ihre Bedürfnisse überlegen ist. Die zweite Voruntersuchung der vorliegenden Studie liefert Hinweise, ob gerätspezifische Unterschiede in der Unit-Nonresponse plausibel sind, indem untersucht wird, 1) ob das Smartphone bezogen auf bestimmte Bedürfnisdimensionen von Befragten insgesamt besser oder schlechter eingeschätzt wird als der PC/Laptop und, 2) wie groß die Relevanz dieser befragungsbezogenen Bedürfnisse für die Entscheidung zur Teilnahme mit dem Smartphone vs. PC/Laptop in Onlinebefragungen ist. Wenn Befragte mit bestimmten Geräten bevorzugt oder sogar ausschließlich teilnehmen, und dies davon abhängt, dass ihnen bestimmte Bedürfnisse bei ihrer Entscheidung wichtig sind, liefert das Hinweise auf mögliche Variablen, die gerätspezifische Unit-Nonresponse bedingen könnten. Dafür wird folgende Hypothese formuliert:

H2: *Proband*innen, entscheiden sich eher für die Teilnahme mit dem Smartphone und gegen die Teilnahme mit dem PC/Laptop, wenn ihnen bestimmte Bedürfnisse für die Befragungsteilnahme wichtig sind.*

Die Hypothese **H2** kann aufzeigen, ob sich unterschiedliche Bedürfnisse von Befragten, wenn sie diese für wichtig halten, auf die Teilnahmeentscheidung mit einem bestimmten Ausfüllgerät übertragen. Dies kann dann auf Unit-Nonresponse hindeuten, wenn ein Ausfüllgerät pauschal von allen potenziellen Befragten für bestimmte Bedürfnisse als überlegen bewertet wird. Im zugehörigen Kapitel des Forschungsprogramms (Kapitel 4.2) wird die **H2** aufgefächert und auf mögliche, mit den Geräteigenschaften verknüpfte, einzelne Bedürfnisse bezogen.

Tabelle 1 fasst den Forschungsstand zusammen, der zur Formulierung der Forschungshypothesen **H1** und **H2** der vorliegenden Arbeit geführt hat. Den Erkenntnissen zu möglichen Moduseffekten des Smartphones bei sensiblen Themen sowie für die Annäherung an die Unit-Nonresponse, die die Voruntersuchungen fundieren, werden noch einmal die jeweils wesentlichen Quellen zugeordnet. Zusätzlich werden die Forschungslücken illustriert, die in der vorliegenden Arbeit in den Voruntersuchungen zu Moduseffekten des Smartphones als Ausfüllgerät aufgegriffen werden.

Befundlage für die Voruntersuchungen zu Moduseffekten des Smartphones

Stichprobenkennwerte und Antwortbereitschaft	Vergleichbarkeit der Stichprobenkennwerte zwischen dem Smartphone und dem PC/Laptop	<u>Unterschiede gefunden:</u> Krebs und Höhne 2021; Struminskaya et al. 2015; Tourangeau et al. 2017 <u>Keine Unterschiede gefunden:</u> z. B. Cook 2014; de Bruijne und Wijnant 2013; Liebe et al. 2015; Mavletova 2013; Mavletova und Couper 2013; Peytchev und Hill 2010; Revilla und Ochoa 2016; Wells et al. 2013
	Sensible Fragen als Prüffeld von Moduseffekten des Smartphones	Antoun et al. 2017; Barnett 1998; Lee et al. 2019; Mavletova 2013; Mavletova und Couper 2013; Toninelli und Revilla 2016b; Tourangeau und Yan 2007; Weisberg 2009;
	Teilnahmekontext am Smartphone diverser als am PC/Laptop	z. B. Antoun et al. 2017; Bosnjak et al. 2017; de Bruijne und Oudejans 2015; Deng et al. 2019; Höhne und Schlosser 2019; Kim et al. 2017; Lütters 2017; Mavletova und Couper 2013; Revilla et al. 2016a; Toninelli und Revilla 2016b
	Keine Unterschiede in den Stichprobenkennwerten und in der Antwortbereitschaft bei sensiblen Fragen	Antoun et al. 2017; Couper et al. 2017; Mavletova und Couper 2013; Revilla et al. 2016a; Toninelli und Revilla 2016b
	Mögliche Vorteile des Smartphones für den Selbstbericht bei sensiblen Fragen	<u>Emotionale Bindung an das Gerät:</u> Brasel und Gips 2014; Fullwood et al. 2017; Melumad et al. 2019; Melumad und Pham 2020; Thorsteinsson und Page 2014 <u>Hohe Anonymität/Privatheit am Smartphone erlebt:</u> Fullwood et al. 2017; Keusch et al. 2022; Sutanto et al. 2013; Vincent 2006 <u>Private Informationen werden bereitwillig mit dem Smartphone geteilt:</u> Brandimarte et al. 2013; Melumad & Meyer 2020
	Mögliche Vorteile des Smartphones für Moduseffekte bei sensiblen Fragen berücksichtigt	Forschungslücke <i>Ansatzpunkt der Voruntersuchung zu möglichen Moduseffekten bei sensiblen Fragen</i>
Annäherung an Unit-Nonresponse	Soziodemografische Unterschiede zwischen Befragten, die das Smartphone vs. den PC/Laptop nutzen	<u>Unterschiede gefunden:</u> z. B. Bosnjak et al. 2017; Buskirk et al. 2015; Cook 2014; de Bruijne und Wijnant 2014; Erens et al. 2019; Keusch und Yan 2017; Krämer 2017; Mavletova 2013; Sommer et al. 2017 <u>Unterschiede nehmen ab:</u> Antoun 2015; Deng et al. 2019; Mavletova 2013; Toepoel und Lugtig 2014
	Smartphones als Ausfüllgeräte von einigen Befragten bevorzugt	z. B. Krämer 2017; Mavletova et al. 2018; Revilla et al. 2016b
	Am Smartphone dauert die Befragungsteilnahme länger	z. B. Andreadis 2015a; Antoun und Cernat 2020; Couper und Peterson 2017; Keusch und Yan 2017; Liebe et al. 2015; Mavletova 2013; Revilla et al. 2016a; Revilla und Ochoa 2016; Toninelli und Revilla 2020
	Teilnahme mit dem Smartphone flexibel in diversen Kontexten möglich	z. B. Antoun et al. 2017; Bosnjak et al. 2017; Deng et al. 2019; Höhne und Schlosser 2019; Kim et al. 2017; Lütters 2017; Mavletova und Couper 2013; Mimura et al. 2015; Revilla et al. 2016a; Toninelli und Revilla 2016b
	Emotionale Verbindung mit dem Smartphone, die auf die Teilnahmeentscheidung wirken könnte	Brasel und Gips 2014; Fullwood et al. 2017; Melumad et al. 2019; Melumad und Pham 2020; Thorsteinsson und Page 2014
	Relevanz gerätassoziierter Bedürfnisstrukturen bei der Teilnahmeentscheidung und der Wahl des Ausfüllmodus	<u>Relevanz bestätigt:</u> Biemer 2010; Groves et al. 1992; Weisberg 2009 Forschungslücke, da bisher nicht auf die Besonderheiten des Smartphones bezogen <i>Ansatzpunkt der Voruntersuchung zur Unit-Nonresponse</i>

Tabelle 1. Forschungsstand zu den Forschungslücken der Voruntersuchungen
Die Tabelle gibt einen Überblick über die bisherige Erkenntnislage, auf der die Voruntersuchungen der vorliegenden Arbeit aufbauen. Für die beiden Prüffelder 1) Stichprobenkennwerte und Antwortbereitschaft als Ausdruck der Antwortqualität sowie 2) Unit-Nonresponse werden entscheidende Forschungslücken identifiziert, über die sich der Prüfung möglicher Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen angenähert wird.

3.2 Antwortqualität des Smartphones als Ausfüllgerät

Die erste Voruntersuchung der Antwortqualität ist auf den Themenbereich sensibler Fragen begrenzt. Dort werden Moduseffekte untersucht, die sich in Stichprobenkennwerten (als Ausdruck sehr starker Messfehler durch Befragte) und der Antwortbereitschaft (Item-Nonresponse) zeigen würden. Dabei wird für den isolierten Geräteffekt des Smartphones geprüft, ob etwaige Modusbesonderheiten die Antwortqualität sogar positiv beeinflussen können.

Das Smartphone als Ausfüllgerät könnte aber auch abseits sensibler Themenbereiche in vielfältiger Weise Einfluss auf die Antwortqualität (bzw. Elemente der Antwortqualität) nehmen und somit zum TSE beitragen. Für die anvisierte, möglichst umfassende Beurteilung etwaiger Moduseffekte des Smartphones als Ausfüllgerät wird das Prüfkonzept der Antwortqualität im TSE-Framework in den Hauptuntersuchungen daher breiter angelegt. Für die Hauptuntersuchungen wird vorab literaturbasiert ergründet, ob Moduseffekte des Smartphones für die Antwortqualität unabhängig von sensiblen Themen plausibel sind, und in welcher Form sich dadurch die Antwortqualität verändern könnte. Antwortqualitätsunterschiede des Messfehlers durch Befragte können beispielsweise auch vorliegen, ohne sich in Unterschieden der Stichprobenkennwerte niederzuschlagen (Bautista 2012, S. 38; Biemer 2010). Ausgangspunkt dieser vertiefenden Betrachtungen ist die Aufbereitung der Hintergründe der Antwortqualität (Kapitel 3.2.1). Dabei wird das Survey-Response Modell eingeführt, das den Antwortprozess strukturiert und verdeutlicht, dass Aufmerksamkeitsprozesse die wesentliche Grundlage der Antwortqualität bilden. In Bezug auf das Survey-Response Modells werden Prüfbedingungen für die Hauptuntersuchungen der vorliegenden Arbeit entwickelt, anhand derer mögliche Antwortqualitätsunterschiede, die durch das Smartphone als Ausfüllgerät entstehen könnten, analysiert werden. Kapitel 3.2.2 stellt die wesentlichen Einflussmöglichkeiten des Smartphones auf die Antwortqualität heraus und im Anschluss wird der Forschungsstand zu den differenziellen Effekten des Smartphones als Ausfüllgerät auf die Antwortqualität aufbereitet (Kapitel 3.2.3).

3.2.1 Grundlagen der Antwortqualität in Befragungen

Wissenschaftliche Befragungen sind darauf angewiesen, dass wirklich das gemessen wird, was gemessen werden soll und die Messung möglichst zuverlässig gelingt (Wiersma 2013). Hohe Antwortqualität ist wichtig, da andernfalls Korrelationsstrukturen, Verteilungen sowie statistische Hypothesentests verzerrt werden (z. B. Groves und Lyberg 2010; Maniaci und Rogge 2014; Peytchev 2013). Modusunterschiede können zu Abweichungen in der Reliabilität und Validität der erhobenen Konstrukte führen (Podsakoff, MacKenzie und Podsakoff 2012) und dadurch die adäquate Messung latenter Konstrukte in Befragungen erschweren. Insbesondere

die Methode der Datenerhebung wirkt unmittelbar auf die Antwortqualität (Bautista 2012, S. 45; Bowling 2005; Wiersma 2013), weshalb mögliche systematische Moduseffekte eines Ausfüllgeräts kritisch wären. Aus diesem Grund muss die Antwortqualität als TSE-Bereich bei der Evaluation des Smartphones als Ausfüllgerät umfassend analysiert werden. Die Antwortqualität könnte aufgrund der Gerätesonderheiten sowie aufgrund des üblichen Nutzungskontext bzw. der Erfahrungen der Nutzer*innen mit dem Smartphone von diesem beeinflusst werden, sodass Moduseffekte entstehen. Da die Antwortqualität in Befragungen, wie sie in der vorliegenden Arbeit definiert wird, primär auf dem Antwortverhalten der Proband*innen basiert (siehe Kapitel 2.4.1), wird das Survey-Response Modell, das den Prozess der Antwortgabe strukturiert, als Beurteilungsgrundlage eingeführt.

3.2.1.1 Das Survey-Response Modell

Das „Survey-Response Model“ charakterisiert den idealtypischen Antwortprozess von Befragten (Tourangeau 1987, S. 152-158; Tourangeau, Rips und Rasinski 2000, S. 7-17). Eine Antwort mit hoher Qualität erfordert eine gewissenhafte Auseinandersetzung mit den Frageinhalten³, wobei mehrere kognitive Prozesse sorgfältig durchlaufen werden müssen. Konkret werden für valides und qualitativ hochwertiges Antworten vier obligatorische Schritte unterschieden: 1) Verständnis (Comprehension), 2) Abruf (Retrieval), 3) Urteil (Judgement) sowie 4) Antwortgabe (Response) (Tourangeau et al. 2000, S. 7-17). Den ersten Schritt bildet das Verständnis der Frage, das heißt, welcher Sachverhalt adressiert wird und worauf die Frage abzielt (Tourangeau et al. 2000, S. 9). Ausgehend vom Verständnis findet der Abruf relevanter Informationen aus dem Langzeitgedächtnis statt. Dafür werden ein hohes Maß an Aufmerksamkeit und eine systematische Abrufstrategie benötigt, da ansonsten die episodische Gedächtnisspur zerfällt, die den Abruf steuert (Tourangeau et al. 2000, S. 9-10). Im dritten Schritt der Urteilsbildung generieren die Antwortenden anhand der abgerufenen Informationen mental eine möglichst vollständige und zutreffende Antwort auf die Frage (Tourangeau et al. 2000, S. 10-12). Der vierte Schritt der Antwortgabe wird in zwei weitere Schritte unterteilt: Zunächst muss das gebildete Urteil im Fragebogen berichtet werden, zum Beispiel in Form der Auswahl der bestmöglichen Antwortoption oder in selbstständigen Antwortformulierungen (Response Mapping). Die final resultierende Antwort wird abschließend noch auf ihre Korrektheit geprüft. Hält die jeweils ausgewählte/abgegebene Antwort dieser Prüfung nicht stand, wird sie ggf. editiert (Response Editing) und iterativ optimiert (Tourangeau et al. 2000, S. 13-14).

³ Der Begriff „Frage“ ist als Befragungsinhalt nicht auf Fragen als Satzart eingeschränkt, sondern inkludiert auch Statements und Items, zu denen sich Befragte positionieren müssen, bzw. Inhalte, auf die sie reagieren.

3.2.1.2 Motivation und Aufmerksamkeit als Grundlagen hoher Antwortqualität

Die Voraussetzung hoher Antwortqualität ist, dass Proband*innen bei der Beantwortung alle Schritte des oben skizzierten Antwortprozesses gewissenhaft durchlaufen. Dieses Verhalten ist als „Optimizing“ definiert, da die Befragten eine zutreffende Beantwortung der Fragen zum Ziel haben, und stellt den Idealfall der Antwortgabe dar (vgl. Krosnick 1991). Weisberg (2009, S. 73-76) richtet im Framework des TSE die Antwortqualität ganz konkret am Survey-Response Modell aus. Das Survey-Response Modell demonstriert, dass Proband*innen motiviert sein müssen, aufmerksam teilzunehmen, um mit hoher Qualität zu antworten (Tourangeau et al. 2000, S. 14-15). Daher bildet Aufmerksamkeit als Ausdruck der Motivation der Befragten während der Befragungsteilnahmen im TSE-Framework die wichtigste Grundlage hoher Antwortqualität (z. B. Edwards 2019; Johnson 2005; Lyberg und Stukel 2017, S. 8-9; MacKenzie und Podsakoff 2012; Roberts et al. 2019; Schmidt, Gummer und Roßmann 2020; Steinbrecher et al. 2015; Tourangeau et al. 2000, S. 8; Ward und Meade 2018). Ist die Aufmerksamkeit gering, kann die Antwortqualität sinken, da sowohl der Messfehler durch Befragte größer werden kann als auch ggf. gar keine Antworten gegeben werden und Item-Nonresponse entsteht.

3.2.1.3 Bedrohungen der Antwortqualität im Rahmen des Survey-Response Modells

Reduzierte Aufmerksamkeit als Folge reduzierter Motivation kann dazu führen, dass der Antwortprozess vereinfacht wird, indem Befragte einzelne Schritte verkürzen oder sogar komplett auslassen. Wenn Befragte versuchen, den persönlichen Aufwand bei der Befragungsteilnahme zu reduzieren, tritt sogenanntes „Satisficing“ auf (vgl. Krosnick 1991, 1999). Satisficing kann die Antwortqualität in Onlinebefragungen deutlich senken (z. B. Heerwegh und Loosveldt 2008; MacKenzie und Podsakoff 2012), die Intensität der Qualitätsreduktion hängt aber davon ab, wie gravierend der Antwortprozess abgekürzt oder gestört wird. Das Ausmaß von Satisficing basiert direkt auf dem Aufwand der Befragten und ist immer auf einem Kontinuum von optimalem bis zu minimalem Aufwand verortet (z. B. Baumgartner und Weijters 2019, S. 306; Krosnick 1991). Schwaches Satisficing zeichnet sich dadurch aus, dass der kognitive Aufwand zwar nicht optimal ist, aber zumindest alle obligatorischen Schritte des Antwortprozesses grundsätzlich durchlaufen werden (Krosnick 1991). Starkes Satisficing tritt auf, wenn Proband*innen ihre Anstrengungen noch weiter reduzieren und einzelne oder sogar mehrere Schritte des Antwortprozesses auslassen (Krosnick 1991, 1999). In der jüngeren Forschungsliteratur wird der Satisficing-Ansatz erweitert durch Careless Responding als Extremform von Satisficing. Careless Responding (CR) bezeichnet gänzlich unaufmerksames Antworten Befragter während der Teilnahme, bei dem die Antwortgabe erfolgt, ohne dass die verschiedenen

Schritte des Survey-Response Modells durchlaufen werden (vgl. Arias et al. 2020; Hong et al. 2020; Meade und Craig 2012). Die Teilnehmer*innen beantworten die Fragen und Items zwar, allerdings ohne kognitiven Aufwand zu investieren, sodass die Antwortqualität minimal ist. Die vorliegende Arbeit differenziert in der Analyse des Einflusses des Smartphones auf die Antwortqualität zur Aufdeckung möglicher Moduseffekte die Hintergründe der Qualitätsreduktion, indem CR und verschiedene Grade des Satisficings in der Folge, wenn möglich, voneinander abgegrenzt werden.

Je stärker einzelne Schritte verkürzt oder gar komplett übersprungen werden, desto größer kann der Bias ausfallen. Zur Vereinfachung der Antwortgabe können Befragte bei Satisficing oder Careless Responding, anstatt die Schritte des Survey-Response Modells konzentriert zu durchlaufen, beispielsweise auf Antwortstile zurückgreifen, die die differenzierte Auseinandersetzung mit der Frage und/oder die Antwortfindung verkürzen. Antwortstile beschreiben behaviorale Tendenzen der Befragten, bei denen die Antwortgabe auf Befragungselementen systematisiert erfolgt und nicht ausschließlich am eigentlichen Iteminhalt ausgerichtet wird (z. B. Baumgartner und Steenkamp 2001; Paulhus 1991, S. 17-20). Besonders verbreitete Antwortstile liegen in Akquieszenz (Zustimmungstendenz), der Tendenz zur Mitte (oder auch „Zentrale Tendenz“) oder der Tendenz zu den Extremen (z. B. Dolnicar, Laesser und Matus 2009; van Vaerenbergh und Thomas 2013). Während Akquieszenz die Neigung der Proband*innen beschreibt, bei Fragen und Statements unabhängig vom Iteminhalt eher zuzustimmen (Baumgartner und Steenkamp 2001; Bentler, Jackson und Messick 1971; Weijters, Baumgartner und Schillewaert 2013), kennzeichnet die Tendenz zur Mitte den Hang, eher neutrale oder wenig extreme Antworten zu geben und mittlere Kategorien von Antwortskalen übermäßig häufig zu nutzen (van Vaerenbergh und Thomas 2013; Wilcox und Keselman 2003). Die Tendenz zu den Extremen charakterisiert das gegensätzliche Antwortverhalten, vorrangig eher die äußeren Extremwerte einer Skala zu wählen und moderate Antworten zu vermeiden (Greenleaf 1992b; Liu, Lee und Conrad 2015). Neben diesen klassischen Antwortstilen kann die Minimierung des eigenen Aufwands bei der Befragungsteilnahme auch monotones Antwortverhalten hervorrufen. Dabei entstehen monotone Antwortmuster wie Non-Differentiation (fehlende Differenziertheit der Antworten), wobei nur enge Ausschnitte der zugrundeliegenden Skala genutzt werden (Fricke et al. 2005; Heerwegh und Loosveldt 2008), oder Straightlining, bei dem Proband*innen dieselbe Antwort immer wieder auswählen, auch wenn sie nicht dem wahren Wert entspricht (Kim et al. 2019; Zhang und Conrad 2014). Auch Primacy- und Recency-Effekte, die ausdrücken, dass bevorzugt zuerst oder zuletzt betrachtete Antwortoptionen gewählt werden, sind ein Ausdruck reduzierter Antwortqualität durch eine Vereinfachung der Antwortgabe (z. B.

Tourangeau et al. 2017; Vannette und Krosnick 2014, S. 318). Greifen Befragte bei der Antwortgabe auf Antwortstile oder feste Antwortmuster zurück, steigt der Messfehler durch die Befragten und die Antwort entfernt sich vom wahren Wert des zu messenden Konstrukts, wodurch die Antwortqualität sinkt (Baumgartner und Weijters 2019, S. 321-322). Van Vaerenbergh und Thomas (2013) empfehlen Forschenden, Antwortstile oder Antwortmuster (z. B. Straightlining) aufzudecken und darüber auf die Antwortqualität zu schließen. Die differenzierte Analyse des Antwortverhaltens ist erforderlich, da sich Antwortverzerrungen durch verschiedene Antwortstile auch gegenseitig aufheben können (Stocké 2004).

Der Prozess der gehaltvollen Antwortgabe kann an verschiedenen Stellen gestört werden, wobei sowohl jede Störung exklusiv für sich als auch die Kombination mehrerer Störungen die Antwortqualität reduzieren können (Podsakoff et al. 2003). Insbesondere bei längeren Befragungen nimmt die Antwortqualität im Laufe der Befragungsteilnahme kontinuierlich ab (Galesic und Bosnjak 2009; Schmidt et al. 2020). Gerade dann steigt beispielsweise die Item-Nonresponse und die Befragungen werden häufiger abgebrochen (z. B. Galesic und Bosnjak 2009; Heberlein und Baumgartner 1978; Rolstad, Adler und Rydén 2011; Yammarino et al. 1991). Reduziert sich die Aufmerksamkeit im Verlauf einer Befragung bzw. steigt die Müdigkeit/Erschöpfung der Proband*innen, sinkt dadurch die Qualität der Antworten, die die Befragten geben (Couper et al. 2017, S. 142-145; Lyberg und Stukel 2017, S. 8-9).

3.2.2 Antwortqualität als Prüfkriterium der Smartphoneeignung

Einen wesentlichen Einfluss auf das Antwortverhalten Befragter nimmt der Befragungskontext (z. B. Höhne et al. 2020a; Miller 2008; Mutz und Kämpfer 2011; Pan et al. 2014). Wenn das Ausfüllgerät „Smartphone“ selbst als Kontextfaktor auf die Motivation und somit auch die Aufmerksamkeit im Antwortprozess wirkt, könnte die Antwortqualität systematisch sinken. Kritisch wäre, wenn Befragte am Smartphone per se unaufmerksamer, weniger motiviert oder schneller erschöpft sind und gemäß des TSE der Messfehler durch Befragte sowie die Item-Nonresponse steigen. Reduzierte Aufmerksamkeit kann den Antwortprozess an verschiedenen Stellen stören, sodass sich der Messfehler durch Befragte, der durch den reduzierten kognitiven Aufwand bei der Antwortgabe steigt, divers im Antwortverhalten widerspiegeln kann. Daher müssen zur Bewertung möglicher Moduseffekte des Smartphones hinsichtlich der Antwortqualität die Einflussmöglichkeiten auf den Messfehler durch Befragte besonders fokussiert werden.

Andreadis (2015a, S. 77) empfiehlt für Onlinebefragungen, in denen das Smartphone als Ausfüllgerät genutzt wird, die Ursachen möglicher Antwortqualitätsreduktionen zu betrachten. Für das Smartphone sind dabei verschiedene Einflussmöglichkeiten auf das Antwortverhalten

denkbar. Diese liegen 1) in den Gerätebesonderheiten (siehe Kapitel 2.2.4), 2) in der persönlichen Verbindung zum sowie Erfahrung mit dem Gerät (siehe Kapitel 2.3) und 3) im Nutzungskontext des Smartphones (siehe Kapitel 2.3 und 3.1.1.1). Zunächst könnten die Geräteigenschaften des Smartphones Folgen für die Antwortqualität haben. Schon aufgrund seiner Größe ist das Smartphone anfällig für Verzerrungen (siehe Kapitel 2.2.4). Auch technische Besonderheiten, die sich auf das Antwortverhalten auswirken könnten (z. B. Straightlining aufgrund der geringen Displaygröße), prädestinieren das Smartphone für Reduktionen der Antwortqualität.

Außerdem könnte die psychologische Bedeutung des Smartphones in Onlinebefragungen auf die Antwortqualität wirken. Besonders kritisch sind für das erfolgreiche Durchlaufen des Antwortprozesses nach dem Survey-Response Modell die Motivation und daraus folgende Aufmerksamkeitszustände (siehe Kapitel 3.2.1.2). Smartphones können eine affektive Funktion für ihre Besitzer*innen einnehmen (siehe Kapitel 2.3.2), werden eher als Komfortquelle wahrgenommen und die Smartphonennutzung setzen einige Nutzer*innen als Coping-Strategie bei Stress ein (Melumad und Pham 2020; Rieger et al. 2017). Daher könnte die kognitive Beanspruchung, die Onlinebefragungen benötigen, den Nutzungsgewohnheiten widersprechen und durch diesen Widerspruch am Smartphone die Aufmerksamkeit eher geringer sein. Gerade bei sehr komplexen oder herausfordernden Fragen und Aufgaben sinkt die Qualität der Antworten (Peytchev 2009), wobei am Smartphone in Befragungen genau die wahrgenommene Komplexität der Inhalte höher ist (Dunaway et al. 2018; Ghose et al. 2013; Wenz 2021b).

Zuletzt wirkt der Nutzungskontext des Smartphones auf das Nutzungsempfinden (Yang et al. 2012). Der Ausfüllkontext kann am Smartphone in Onlinebefragungen sehr unterschiedlich sein (siehe Kapitel 3.1.1.1), sodass resultierende Besonderheiten im Antwortverhalten plausibel sind. Ablenkungsreize erschweren in Onlinebefragungen die aufmerksame Teilnahme (z. B. Sendelbah et al. 2016; Wenz 2021b) und könnten, wenn sie am Smartphone vermehrt erlebt oder dominanter wahrgenommen werden, die Antwortqualität im Vergleich zum PC/Laptop stärker schwächen.

Die Bedeutung hoher Antwortqualität für geringe TSE-Ausmaße unterstreicht die Relevanz des Forschungsvorhabens. Über die Antwortqualität könnte das Smartphone, wie dargestellt, vielfältig auf den TSE wirken. Daher legt auch das zu entwickelnde Forschungsprogramm einen besonderen Akzent auf die Prüfung der Antwortqualität, wenn an Onlinebefragungen mit dem Smartphone teilgenommen wird. Sollten für Smartphone-Teilnehmer*innen Unterschiede in der Antwortqualität auftreten, signalisiert dies Moduseffekte, die den Nutzen des Smartphones als Ausfüllgerät begrenzen könnten. Das nachfolgende Kapitel strukturiert daher den bisherigen Forschungsstand zum Einfluss des Smartphones auf die Antwortqualität.

3.2.3 Zusammenfassung des aktuellen Forschungsstands zur Antwortqualität

Struminskaya und Kolleg*innen (2015) konstatieren, dass das Smartphone als Ausfüllgerät die Antwortqualität von Onlinebefragungen gefährdet, und empfehlen daher, den Einfluss des Smartphones auf die Antwortqualität wissenschaftlich zu prüfen. Da die vorliegende Arbeit die Isolierung des reinen Geräteffekts als Grundlage der Prüfung anlegt, werden Einflüsse unterschiedlicher Fragebogendarstellung bei der Prüfung ausgeschlossen (siehe Kapitel 2.4.1). Auch die spezifischen Geräteeffekte und die potenziellen Kontexteffekte sowie deren mögliche Interaktion werden in den Analysen weitestgehend voneinander getrennt untersucht.

Die Arbeit von Antoun und Kolleg*innen (2018) stellt heraus, dass die Befundlage zum Einfluss des Smartphones als Ausfüllgerät auf die Antwortqualität recht breit gefächert, aber nicht auf alle Aspekte bezogen einheitlich ist. Anders als bei Stichprobenkennwerten bei sensiblen Fragen sind Studien, die gar keine oder nur geringe Unterschiede je nach Ausfüllgerät für die Antwortqualität finden (Andreadis 2015b; Buskirk und Andrus 2014; Toepoel und Lugtig 2014), eindeutig in der Minderheit. Zur Schätzung der Antwortqualität werden meist Indikatoren herangezogen, die das Antwortverhalten der Proband*innen charakterisieren (z. B. Bosnjak und Tuten 2001; Deutskens et al. 2004; Díaz de Rada und Domínguez-Álvarez 2014; Göritz 2004; Sánchez-Fernández, Muñoz-Leiva und Montoro-Ríos 2012). Höhne und Schlosser (2020a) zeigen dazu auf, dass die Antwortqualität in Befragungen nur anhand der umfassenden Betrachtung verschiedener Indikatoren beurteilt werden kann. Die häufigsten in der Literatur der Befragungsforschung genutzten Antwortqualitätsindikatoren sind 1) fehlende Werte (sogenannte Missings) (z. B. Bosnjak und Tuten 2001; Göritz 2004; Heerwegh 2006; Sánchez-Fernández et al. 2012), 2) nicht-substanzielle Antworten (z. B. Deutskens et al. 2004; Fricker et al. 2005; Sánchez-Fernández et al. 2012), 3) Befragungsabbrüche (z. B. Bosnjak und Tuten 2001; Göritz 2004; Heerwegh 2006), 4) Länge der Antworten auf offene Fragen (z. B. Deutskens et al. 2004; Díaz de Rada und Domínguez-Álvarez 2014; Sánchez-Fernández et al. 2012) und 5) die Antwortzeit (z. B. Deutskens et al. 2004; Fricker et al. 2005; Heerwegh 2006; Sánchez-Fernández et al. 2012). Des Weiteren werden Maße der Reliabilität und Validität der Konstruktmessung für die Qualitätseinschätzung verwendet (z. B. Cheung et al. 2017; Keusch und Yan 2017; Sommer et al. 2017; Tourangeau et al. 2018). Abseits dieser statistischen Kennwerte kann auch aus Antwortmustern und Antwortstilen, die die Befragten für die Beantwortung einsetzen, auf die Antwortqualität geschlossen werden (z. B. Díaz de Rada und Domínguez-Álvarez 2014; Dolnicar et al. 2009; Fricker et al. 2005; Göritz 2004; Porter und Whitcomb 2003), sofern diese aus reduzierter Aufmerksamkeit bei geringer Motivation während der Teilnahme

resultieren. All diese Indikatoren können auf den Vergleich der Antwortqualität von Smartphone- und PC/Laptop-Teilnehmer*innen übertragen werden (Erens et al. 2019).

Fehlende Werte als eine Variante von Item-Nonresponse stellen einen wesentlichen Unterbereich der Antwortqualität im TSE-Framework dar. Die vorliegende Arbeit verwendet in der Folge den Begriff „Missings“, wenn einzelne Fragen von den Befragten nicht beantwortet werden. Die meisten Studien berichten deckungsgleich, dass Befragte, die mit dem Smartphone teilnehmen, deutlich häufiger Fragen auslassen und keine Antwort geben (z. B. Buskirk et al. 2015; Couper et al. 2017, S. 148; Keusch und Yan 2017; Lugtig und Toepoel 2016; Mavletova und Couper 2014; Struminskaya et al. 2015). Nur selten ist die Anzahl der Missings über alle Geräte hinweg ähnlich (Revilla und Couper 2018; Revilla und Ochoa 2016). Auch wenn dieselben Personen wechselnd mit verschiedenen Ausfüllgeräten teilnehmen, treten intraindividuell mehr Missings am Smartphone auf (Struminskaya et al. 2015). Vor allem offene Fragen werden am Smartphone häufiger ausgelassen (Erens et al. 2019; Lambert und Miller 2015; Lugtig und Toepoel 2016; Struminskaya et al. 2015).

Nicht-substantielle Antworten von Befragten liegen vor, wenn zwar eine Antwort gegeben wird, diese aber keine inhaltliche Aussagekraft hat. Dies betrifft klassische Ausweichkategorien wie „Weiß nicht“ oder „keine Angabe“ und ist im TSE-Framework ebenfalls ein Teil des Unterbereichs der Item-Nonresponse. Ein Einfluss des Smartphones auf das Ausmaß nicht-substantziellen Antwortens findet sich bisher nicht (Revilla et al. 2016a; Revilla und Couper 2018).

Beginnen Proband*innen eine Befragung, beenden sie aber nicht, liegt ein Befragungsabbruch vor (z. B. Steinbrecher et al. 2015). Befragungsabbrüche als Antwortqualitätsmaß unterscheiden sich strukturell von Unit-Nonresponse (Peytchev 2011) und interagieren mit den Rahmenbedingungen der Teilnahme. Ist der Abbruchgrund systematisch, ist die Datengrundlage je nach Themengebiet nicht mehr valide (Steinbrecher et al. 2015). Als Indikator geringer Antwortqualität sollten bei Befragungsabbrüchen nur die Abbruchraten betrachtet werden, da die Antwortqualität von Proband*innen bis zum Abbruch der Befragung nicht zwingend geringer ist (Peytchev 2009). Die Wahrscheinlichkeit für Befragungsabbrüche steigt mit zunehmender Seitenzahl, Fragenlänge, Beanspruchung der Proband*innen und Anzahl offener Fragen (Peytchev 2009). Das Smartphone könnte als Ausfüllgerät besonders anfällig für Befragungsabbrüche sein, da dieselbe Befragung am Smartphone gegenüber dem PC/Laptop in der Wahrnehmung der Teilnehmer*innen herausfordernder erlebt werden kann (Ghose et al. 2013; Wenz 2021b) und teilweise als länger wahrgenommen wird (Bosnjak et al. 2017, S. 61; de Bruijne und Wijnant 2013). Darüber hinaus sind auch technische Probleme am Smartphone häufiger, wenn die Befragung nicht für den Smartphoneeinsatz konzipiert wurde, wodurch Abbrüche

noch wahrscheinlicher werden (Drewes 2014, S. 373). Tatsächlich treten in der Literatur Befragungsabbrüche häufiger bei Personen auf, die das Smartphone als Ausfüllgerät nutzen (z. B. Buskirk und Andrus 2014; Cook 2014; Couper et al. 2017, S. 139; Drewes 2014, S. 376; Keusch und Yan 2017; Lambert und Miller 2015; Mavletova et al. 2018; Mavletova und Couper 2014; Poggio, Bosnjak und Weyandt 2015). Insbesondere, wenn die Darstellung der Befragung am Smartphone als unpassend wahrgenommen wird, steigen die Abbruchraten (Arn et al. 2015). Auch bei adäquater Darstellung aller Inhalte bleiben die Befragungsabbruchraten am Smartphone dennoch höher (Stapleton 2013) und bei offenen Fragen wird am Smartphone besonders häufig abgebrochen (Erens et al. 2019). Unterbrechungen der Befragung, wobei sie später fortgesetzt wird, zeigen sich gerätübergreifend hingegen gleich häufig (Wenz 2021a, S. 113).

Die Länge der Antwort bei offenen Fragen lässt ebenfalls Rückschlüsse auf die Antwortqualität zu. Dabei signalisieren ausführliche Antworten eher höhere Antwortqualität (z. B. Díaz de Rada und Domínguez-Álvarez 2014). Hier zeigt sich: Antworten auf offene Fragen fallen am PC/Laptop ausführlicher aus und die Befragten fassen sich am Smartphone eher kurz (Antoun et al. 2017; Erens et al. 2019; Lambert und Miller 2015; Lugtig und Toepoel 2016; Revilla et al. 2016a; Revilla und Ochoa 2016; Wells et al. 2014). Selbst wenn die Antworten ähnlich lang sind, wird am Smartphone für die Antwortgabe mehr Zeit benötigt (Couper und Peterson 2017; Mavletova 2013).

Daran knüpft die Antwortzeit an, die die Teilnahmedauer an einer Befragung bezeichnet und über die darauf geschlossen wird, wie lange Befragte für die Beantwortung einzelner Items, Seiten oder der gesamten Befragung brauchen. Insbesondere sehr kurze und sehr lange Antwortzeiten können geringe Antwortqualität signalisieren (z. B. Meade und Craig 2012; Read, Wolters und Berinsky 2022). Unter den verschiedenen Antwortqualitätsindikatoren zeigt die Antwortzeit den deutlichsten Unterschied zwischen den beiden in dieser Arbeit untersuchten Ausfüllgeräten: Die Teilnahme an Onlinebefragungen mit dem Smartphone erfordert eindeutig mehr Zeit als mit dem PC/Laptop (Andreadis 2015a, S. 74; Couper et al. 2017, S. 141; Couper und Peterson 2017; Drewes 2014, S. 378; Keusch und Yan 2017; Lambert und Miller 2015; Mavletova 2013; Mavletova et al. 2018; Revilla et al. 2016a, 2017; Revilla und Ochoa 2016; Schlosser und Mays 2018; Sommer et al. 2017; Tourangeau et al. 2018). Auch intraindividuell, wenn im Wechsel verschiedene Geräte für die Teilnahme genutzt werden, ist die Antwortzeit am Smartphone länger (Struminskaya et al. 2015). An dieser Stelle ist ein Interaktionseffekt mit der Bildschirmgröße plausibel, da die Antwortdauer mit sinkender Bildschirmgröße stetig steigt (Liebe et al. 2015; Toninelli und Revilla 2020, S. 361). Die längere Antwortzeit am Smartphone kann dabei nicht vollständig auf die Geräteigenschaften, die Handhabung des

Smartphones und die Internetgeschwindigkeit zurückgeführt werden (Andreadis 2015a, S. 76, 2015b; Schlosser und Silber 2020, S. 278), sodass auch weitere Gerätbesonderheiten zu möglichen Moduseffekten des Smartphones als Ausfüllgerät beitragen könnten. Onlinebefragungen werden selbst bei gleicher Antwortdauer von den Befragten als zeitaufwendiger eingeschätzt, wenn mit dem Smartphone teilgenommen wird (de Bruijne und Wijnant 2013).

Bei der Bewertung der Forschungsbefunde ist es zur Isolation des Effekts von Smartphones wichtig, Darstellungs- und Geräteeffekte voneinander abzugrenzen, da beispielsweise räumliche und zeitliche Nähe von Items in Befragungen in höheren Korrelationen unabhängig vom Inhalt resultieren (Couper, Traugott und Lamias 2001; Weijters, Geuens und Schillewaert 2009). Dies könnte bei größeren Darstellungsunterschieden zur Folge haben, dass Konstruktmessungen je nach Ausfüllgerät verzerrt werden und sich im Extremfall die resultierenden Korrelationsstrukturen verändern. Die Befundlage ist hierfür uneinheitlich: Einerseits deutet sich in einigen Studien bisher eher an, dass über verschiedene Ausfüllgeräte hinweg die Korrelationsstrukturen von Konstrukten gleich sind und Messäquivalenz grundsätzlich gegeben ist (z. B. Mavletova et al. 2018; Revilla et al. 2017). Die Reliabilität der Messungen kann aber mitunter am Smartphone je nach Reliabilitätsmaß geringer ausfallen (Keusch und Yan 2017; Tourangeau et al. 2018). Während Keusch und Yan (2017) beispielsweise geringere interne Konsistenzen gemessen über Cronbachs Alpha (CA) am Smartphone finden, zeigt sich dieser Unterschied wiederum in anderen Studien nicht (Sommer et al. 2017; Tourangeau et al. 2018). Tourangeau und Kolleg*innen (2018) berichten jedoch geringere Reliabilität am Smartphone, wenn die Reliabilitätsmessung über den Ansatz der Item-Response Theorie⁴ erfolgt. Die Validität der Messungen unterscheidet sich zwischen den Ausfüllgeräten hingegen eher nicht (Sommer et al. 2017).

Neben den genannten statistischen Indikatoren (z. B. Missings, Antwortzeit) manifestiert sich die Antwortqualität auch in Antworttendenzen und Antwortstilen (siehe Kapitel 3.2.1.3). Die Strukturierung der Forschungsliteratur offenbart, dass bisher nur eine kleine Auswahl von Antwortstilen unmittelbar auf den Effekt des Ausfüllgeräts überprüft wurde. Die Befundlage bestätigt zunächst das Sichtbarkeitsprinzip für das Smartphone als Ausfüllgerät. Befragte wählen eher Antwortoptionen, die sie unmittelbar sehen, was zu einem Primacy-Effekt am Smartphone führt (Höhne und Schlosser 2019; Stapleton 2013). Werden eigentlich horizontale Skalen im Rahmen responsiver Designanpassungen am Smartphone als Drop-Down Auswahlmenü präsentiert, verstärkt dies Primacy-Effekte zusätzlich (Antoun et al. 2018), was erneut den Ausschluss responsiver Designanpassungen für die vorliegende Arbeit unterstützt (siehe auch Kapitel 2.4.1). Am Smartphone können Teilnehmer*innen je nach Studieneigenschaften anfälliger

⁴ Siehe Embretson und Reise (2013) für eine Einführung in die Item-Response Theorie.

für Non-Differentiation sein (Revilla et al. 2017), wobei dieser Befund nicht immer auftritt (Schlosser und Mays 2018). Ebenfalls kontrovers präsentiert sich die Befundlage zu Straightlining am Smartphone. Mal findet sich am Smartphone vergleichsweise mehr (Struminskaya et al. 2015), mal weniger Straightlining (Keusch und Yan 2017) und zum Teil zeigen sich auch gar keine Unterschiede (Tourangeau et al. 2018). Akquieszenz-Tendenzen scheint das Smartphone sogar abzubauen (Liebe et al. 2015), ohne dass es bisher gegenteilige Befunde gibt.

Tabelle 2 gibt einen Überblick über den Forschungsstand zum differenziellen Einfluss des Smartphones auf die Antwortqualität, der überwiegend negativ ausfällt. Es werden lediglich die Studien aufgelistet, die Antwortqualitätsunterschiede am Smartphone finden, da diese die überwiegende Mehrheit bilden und dadurch klar ersichtlich ist, wo Unterschiede besonders häufig

Unterschiede in Maßen der Antwortqualität bei Teilnahme mit dem Smartphone

Missings	Häufiger am Smartphone: Buskirk et al. 2015; Couper et al. 2017; Erens et al. 2019; Keusch und Yan 2017; Lambert und Miller 2015; Lugtig und Toepoel 2016; Mavletova und Couper 2014; Struminskaya et al. 2015
Befragungsabbrüche	Häufiger am Smartphone: Buskirk und Andrus 2014; Cook 2014; Couper et al. 2017; Drewes 2014; Erens et al. 2019; Keusch und Yan 2017; Lambert und Miller 2015; Mavletova et al. 2018; Mavletova und Couper 2014; Poggio et al. 2015
Antwortlänge bei offenen Fragen	Geringer am Smartphone: Antoun et al. 2017; Couper und Peterson 2017; Erens et al. 2019; Lambert und Miller 2015; Lugtig und Toepoel 2016; Mavletova 2013; Revilla et al. 2016a; Revilla und Ochoa 2016; Wells et al. 2014
Antwortzeit	Länger am Smartphone: Andreadis 2015a; Couper et al. 2017; Couper und Peterson 2017; de Bruijne & Wijnant 2013; Drewes 2014; Keusch und Yan 2017; Lambert und Miller 2015; Mavletova 2013; Mavletova et al. 2018; Revilla et al. 2016a, 2017; Revilla und Ochoa 2016; Schlosser und Mays 2018; Schlosser und Silber 2020; Sommer et al. 2017; Struminskaya et al. 2015; Tourangeau et al. 2018
Reliabilität	Geringer am Smartphone: Keusch und Yan 2017; Tourangeau et al. 2018
Antwortstile	Häufiger am Smartphone: Primacy-Effekte: Antoun et al. 2018; Höhne und Schlosser 2019; Stapleton 2013 Non-Differentiation: Revilla et al. 2017 Straightlining: Struminskaya et al. 2015 Seltener am Smartphone: Straightlining: Keusch und Yan 2017 Akquieszenz: Liebe et al. 2015

Tabelle 2. Antwortqualitätsunterschiede bei der Teilnahme mit dem Smartphone

Die Tabelle gibt einen Überblick über den Forschungsstand zur Antwortqualität, wenn an Onlinebefragungen mit dem Smartphone teilgenommen wird, und begrenzt sich bewusst auf die reichhaltigen gefundenen Unterschiede. Auch wenn die Befundlage nicht komplett einheitlich ist, unterstreicht der dargestellte Umfang derjenigen Studien, die Unterschiede finden, dass das Smartphone den Total Survey Error insbesondere über die Antwortqualität erhöhen könnte. Für die Prüfung von Moduseffekten des Smartphones als Ausfüllgerät bildet Antwortqualität daher in den Hauptuntersuchungen der vorliegenden Arbeit den Forschungsschwerpunkt.

auftreten. Die Vielzahl der Unterschiede unterstreicht die Relevanz der Prüfung der Antwortqualität für den Einsatz von Smartphones in Onlinebefragungen. Der Fokus liegt in der Forschungsliteratur bisher lediglich auf einzelnen Qualitätsindikatoren, während die Hintergründe der Antwortqualität und das generelle Antwortverhalten kaum untersucht werden. Von daher offenbart sich hier eine wesentliche Forschungslücke. Statt lediglich ausgewählte einzelne Maße der Antwortqualität zu betrachten, sollten die Untersuchungen in der vorliegenden Arbeit umfassender angelegt werden.

In der Forschungsliteratur wird diesbezüglich vorgeschlagen, die Prüfung von Moduseffekten, die auf die Antwortqualität im TSE-Framework wirken könnten, über die Analyse von Satisficing vorzunehmen (Faulbaum 2022, S. 579; Jäckle et al. 2010; Weisberg 2009, S. 286). Je größer das Ausmaß des Satisficings als Folge geringer Motivation und Aufmerksamkeit ausfällt, desto problematischer sind die Konsequenzen für die Antwortqualität (z. B. Berinsky, Margolis und Sances 2014; Hong und Cheng 2019; Krosnick 1999; Nichols, Greene und Schmolck 1989). Ist die kognitive Verarbeitung am Smartphone herausfordernder, könnte Satisficing eine willkommene Antwortstrategie darstellen (Bethlehem 2015, S. 127; Schlosser und Mays 2018; Tourangeau et al. 2017). Bisherige Studien finden tendenziell keine Unterschiede im Satisficing zwischen Smartphones und PCs/Laptops (Antoun et al. 2017; Tourangeau et al. 2018), differenzieren aber auch das Ausmaß von Satisficing nicht adäquat und versäumen, den Grad der Antwortqualitätsreduktion und die Konsequenzen für den TSE aufzudecken. Eine Ausnahme bildet die Studie von Struminskaya und Kolleg*innen (2015), die tatsächlich eher Satisficing am Smartphone feststellt. Während für Satisficing schon erste Erkenntnisse bezüglich möglicher Moduseffekte des Smartphones als Ausfüllgerät vorliegen, wird Careless Responding (CR) hingegen in die Forschung zum Geräteeffekt des Smartphones in Onlinebefragungen noch kaum einbezogen. Gerade aber CR als extrem unaufmerksames Antworten verringert die Antwortqualität durch die sehr starke Verkürzung des Antwortprozesses massiv.

Da Aufmerksamkeit als Ausdruck der Motivation in Befragungen eine wesentliche Determinante der Antwortqualität darstellt (siehe Kapitel 3.2.1, insbesondere 3.2.1.2), legt die vorliegende Arbeit die Aufmerksamkeit während der Befragungsteilnahme als Prüfgrundlage möglicher Moduseffekte des Smartphones hinsichtlich der Antwortqualität an. Dies ermöglicht es, dem Anspruch gerecht zu werden, die Hintergründe möglicher hemmender Einflüsse des Smartphones einzubeziehen, und füllt die aufgedeckte Forschungslücke in der wissenschaftlichen Befragungsforschung: Aufgrund der persönlichen Bedeutung und der Kontextbesonderheiten der Nutzung könnte am Smartphone die Antwortqualität möglicherweise sinken, wenn

die Teilnahmemotivation an diesem systematisch beeinflusst wird und somit unaufmerksames Antwortverhalten wahrscheinlicher wird.

Das folgende Kapitel stellt das Themenfeld Careless Responding ausführlich vor. Anhand von CR kann auf Aufmerksamkeitszustände der Befragten geschlossen werden, wobei der Akzent auf Unaufmerksamkeit liegt. Unaufmerksamkeit als wesentlicher Hintergrundfaktor der Antwortqualität und somit auch als Ursache von Satisficing-Tendenzen und Antwortmustern, die aus reduziertem Aufwand bei der Antwortgabe folgen, erweitert mit Blick auf das Smartphone die Erkenntnislage zu möglichen Moduseffekten deutlich. Daher bildet Unaufmerksamkeit bei der Befragungsteilnahme den Forschungsschwerpunkt für die Hauptuntersuchungen der vorliegenden Arbeit. In Vorbereitung der Hauptuntersuchungen wird das CR-Konzept umfänglich vorgestellt, um geeignete Methoden für die Aufmerksamkeitsprüfung zu identifizieren, die möglichst eindeutige Schlüsse hinsichtlich der Antwortqualität zulassen. Dafür werden im folgenden Kapitel die Hintergründe, die Auswirkungen und die Relevanz von Careless Responding in Onlinebefragungen mit Blick auf die Signalwirkung für die Antwortqualität ausgeführt. Außerdem werden die Aufdeckungsmöglichkeiten unaufmerksamen Antwortens vorgestellt, diskutiert und auf die Anwendung als Antwortqualitätsmaß in der vorliegenden Arbeit verdichtet. Als eine direkte Methode der Aufmerksamkeitsmessung könnte Eye-Tracking Hinweise auf gerätspezifischen Aufmerksamkeitsstrukturen liefern. In Onlinebefragungen im Feld kann auf die Aufmerksamkeit jedoch nur indirekt geschlossen werden, indem CR gerätweise verglichen wird, wofür adäquate Aufdeckungsmöglichkeiten konkret von CR nötig sind.

3.3 Careless Responding

Careless Responding (CR) tritt in allen Befragungssituationen auf, in denen die Proband*innen nicht bereit sind, aufmerksam, durchdacht und gewissenhaft zu antworten (z. B. Arias et al. 2020; Arthur, Hagen und George 2021; Curran 2016; Huang et al. 2012; Meade und Craig 2012). Somit beschreibt CR Extremfälle des Messfehlers durch Befragte im TSE-Framework und ist ein fundamentales Problem der Antwortqualität in Befragungen. Allerdings wurde CR in der wissenschaftlichen Forschung sehr lange vernachlässigt (McGonagle, Huang und Walsh 2016). Erst in der jüngeren Vergangenheit wurde es als notwendige Erweiterung des Satisficing-Ansatzes eingeführt und aufgrund der hohen Relevanz seitdem insbesondere in der vergangenen Dekade intensiv beforscht (vgl. Arias et al. 2020; Hong et al. 2020; Huang et al. 2012; Meade und Craig 2012). In den folgenden Unterkapiteln wird zunächst der Begriff Careless Responding definiert (Kapitel 3.3.1) und die Relevanz für die Onlinebefragungsforschung dargestellt (Kapitel 3.3.2). Anschließend wird erarbeitet, wie Aufmerksamkeit und insbesondere

CR in Onlinebefragungen ermittelt werden können (Kapitel 3.3.3). Es werden verschiedene Möglichkeiten der CR-Aufdeckung vorgestellt und der Forschungsstand zu den eingeführten CR-Indikatoren strukturiert. Für die Prüfung von Moduseffekten des Smartphones werden geeignete CR-Verfahren/Maße festgelegt, die in der vorliegenden Arbeit zur Anwendung kommen. Das Kapitel arbeitet abschließend heraus, dass das Smartphone als Ausfüllgerät CR fördern könnte, und leitet Hypothesen ab, die die Entwicklung des Forschungsprogramms zu möglichen Moduseinflüssen des Smartphones auf die Antwortqualität lenken (Kapitel 3.3.4).

3.3.1 Konzeptualisierung von Careless Responding

In der Forschungsliteratur wird CR auch als „Insufficient Effort Responding“ bezeichnet (z. B. Bowling et al. 2016; DeSimone et al. 2020; DeSimone, Harms und DeSimone 2015; Hong et al. 2020; Huang et al. 2015a). Die vorliegende Arbeit verwendet im weiteren Verlauf einheitlich den Terminus Careless Responding (CR), da sich CR und Insufficient Effort Responding inhaltlich nicht unterscheiden. Zudem wird dabei neben der Abkürzung CR verschiedentlich die deutsche Übersetzung „unaufmerksames Antworten“ aus Gründen der Lesbarkeit verwendet.

3.3.1.1 Eingrenzung des Begriffs Careless Responding

Anhand des Survey-Response Modells (siehe Kapitel 3.2.1.1) wird deutlich, dass für hohe Antwortqualität alle Schritte des Antwortprozesses gewissenhaft durchlaufen werden sollten (Krosnick 1991). CR bzw. unaufmerksames Antworten beschreibt jedoch Antwortverhalten Befragter, bei dem diese aufgrund mangelnder Motivation eine Antwort geben, die sehr stark vom Inhalt der zu beantwortenden Frage losgelöst ist (z. B. Curran 2016; DeSimone und Harms 2018; Huang et al. 2012, S. 100; Meade und Craig 2012; Weijters et al. 2013). Bei CR wird der kognitive Aufwand der Antwortgabe so weit wie möglich minimiert, ohne die Befragung abzubrechen, sodass CR den Gegenpol des optimalen Antwortverhaltens (Optimizing) bildet und konträr zum Optimizing zu sehr geringer Antwortqualität führt (Terzi 2017, S. 23-25). Obwohl Befragungsteilnehmer*innen die Fragen beantworten, investieren sie nicht ausreichend Aufwand für idealtypisches Antworten und überspringen sogar – anders als bei den meisten Formen des Satisficing – den initialen Schritt des Frageverständnisses im Survey-Response Modell (Voss 2023). CR ist somit eine eindeutige Ursache geringer Antwortqualität.

Einige Forscher*innen schlagen vor, CR und starkes Satisficing gleichzusetzen und synonym zu verwenden (Anduiza und Galais 2017; Baumgartner und Weijters 2019, S. 295). Die komplette Gleichsetzung von CR und starkem Satisficing kann jedoch irreführend sein. *Abbildung 5* skizziert, dass der Übergang von CR zu starkem Satisficing bezogen auf die jeweils

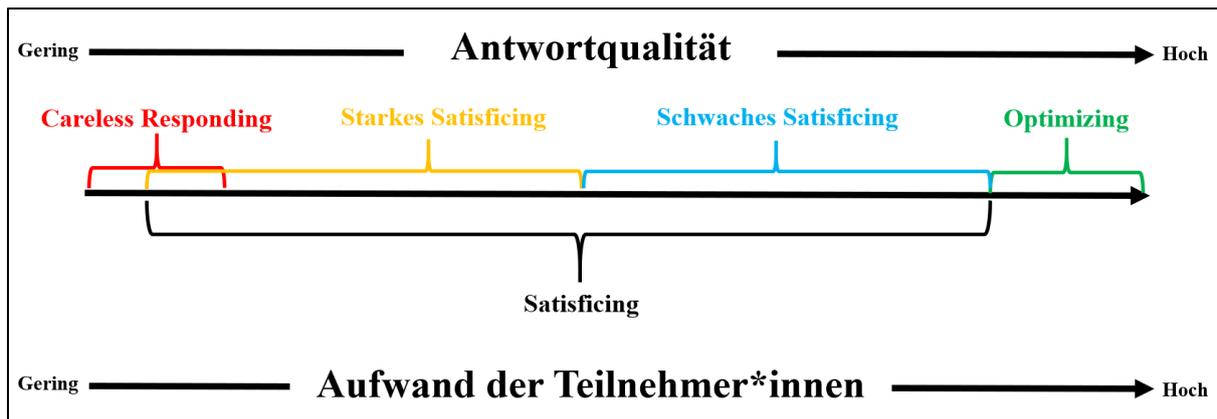


Abbildung 5. Unterscheidung von Careless Responding und Satisficing

Zusammenhang der Antwortqualität und des Aufwands der Befragten während der Teilnahme an Onlinebefragungen als Grundlage der Abgrenzung von Careless Responding (CR) und Satisficing.

reflektierte Antwortqualität zwar fließend ist, CR und starkes Satisficing aber nicht vollständig deckungsgleich sind. Die Antwortqualität als Resultat starken Satisficings variiert, wobei der Antwortprozess zuweilen stärker, aber auch teilweise schwächer vom idealtypischen Antwortprozess abweicht. Starkes Satisficing umfasst sowohl Antwortverhalten, bei dem die Frage sehr wohl verstanden wird (Schritt 1 des Survey-Response Modells), aber die Schritte zwei und drei vor der Antwortgabe übersprungen oder wesentlich abgekürzt werden, als auch das lediglich flüchtige Überfliegen der Items, das in inhaltsunabhängiger Antwortgabe münden kann (Krosnick 1991, 1999). Im ersten Fall verringert das Antwortverhalten die Antwortqualität deutlich weniger stark als CR, während die Verringerung der Antwortqualität bei CR und bei starkem Satisficing durch bloßes Überfliegen der Inhalte annähernd gleich ist (Terzi 2017, S. 23-25; Voss 2023). CR führt zu einer strukturellen Verzerrung der Antwortgabe, die die Antwortqualität immer und wesentlich reduziert, sodass CR das Spektrum des Antwortverhaltens punktuell am unteren Ende der Antwortqualität erweitert. Die sorgfältige Unterscheidung von CR und Satisficing sowie die Berücksichtigung unterschiedlicher Ausmaße von Satisficing bei dieser Abgrenzung erlauben, die Relevanz möglicher Moduseffekte des Smartphones für die Antwortqualität adäquat bewerten zu können.

Die differenzierte Auseinandersetzung mit unaufmerksamem Antworten in Befragungen ist notwendig, da je nach Setting sehr unterschiedliche Antwortmuster auftreten können, die aber jeweils als Folge von CR entstehen (Kam, Meyer und Sun 2021). Manifestiert sich das Antwortverhalten beispielsweise in Antwortstilen, die mit geringerem Aufwandaufwand verbunden sind, können diese eine Konsequenz von CR sein (Alarcon und Lee 2022; Kam und Meyer 2015). In der Vergangenheit wurden klassische Antwortstile schon mehrfach untersucht, allerdings ohne adäquaten Einbezug von CR (vgl. Baumgartner und Weijters 2019; Greenleaf 1992a; van Vaerenbergh und Thomas 2013; Weijters, Schillewaert und Geuens 2008). Ist die

Antwortqualität reduziert, da die Antworten der Befragten von den wahren Werten abweichen, sollte aber geprüft werden, ob dies tatsächlich eine Folge von Unaufmerksamkeit bei der Antwortgabe sein kann und somit auf CR hinweist (Alarcon und Lee 2022). Dies ist bei strategischen Verzerrungen der Antwortgabe, die Bezug zum Iteminhalt aufweisen, wie beispielsweise sozialer Erwünschtheit (z. B. Krumpal 2013; Paulhus 2002) und Faking (z. B. Mueller-Hanson, Heggestad und Thornton III 2006; Steger, Schroeders und Wilhelm 2021), nicht der Fall, da diese eine bewusste Auseinandersetzung mit dem Iteminhalt voraussetzen.

CR definiert sich jedoch genau dadurch, dass die Auseinandersetzung mit den Inhalten fehlt (z. B. Brower 2020, S. 21-22; Denison und Wiernik 2022; Leiner 2019; Maniaci und Rogge 2014; Steger, Schroeders und Gnams 2020). Bei der Tendenz zur Mitte oder den Extremen erfolgt die Antwortgabe losgelöst vom Iteminhalt. In beiden Fällen muss nur grob die genutzte Skala verstanden werden, nicht aber der Iteminhalt, sodass diese Tendenzen eine Folge von CR sein können (Alarcon und Lee 2022). Anders ist es bei Akquieszenz, die zwar den Aufwand der Antwortabgabe reduziert, aber ein gewisses Verständnis des Inhaltes erfordert, um überhaupt zustimmen zu können (Kam et al. 2021; Weijters et al. 2013). Bei CR wird jedoch bereits das Frageverständnis übersprungen. Somit ist Akquieszenz eher ein Ausdruck schwachen Satisficings (z. B. Vannette und Krosnick 2014, S. 317), wobei die Beeinträchtigung der Antwortqualität dann weniger gravierend ist als bei CR. Nutzen Befragte allerdings (wie bei der Tendenz zur Mitte oder den Extremen auch) nur die hinterlegte Skala, um eher zuzustimmen, oder persistiert Akquieszenz über längere Teile der Befragung (insbesondere, wenn sich die Skalierung nicht ändert), löst sich die Antwortgabe stärker vom Frageinhalt (Krosnick 1991, 1999).

3.3.1.2 Formen von Careless Responding

Die jüngere Forschung erweitert die Betrachtung problematischen Antwortverhaltens um CR und hat eine differenzierte Auseinandersetzung mit unaufmerksamem Antworten angestoßen. CR ist ein komplexes Phänomen, das sich situativ verschiedenartig in den Antworten der Befragten ausdrückt (DeSimone et al. 2020; Kam et al. 2021). In der Literatur werden mindestens drei verschiedene Formen diskutiert, in denen sich CR manifestiert (z. B. Denison und Wiernik 2022; Kam und Meyer 2015; Meade und Craig 2012; Schroeders, Schmidt und Gnams 2022; Ward und Meade 2023). Die Unterscheidung dieser Formen ist hilfreich, um in der vorliegenden Arbeit die Evaluation von CR möglichst fundiert vorzunehmen.

Zunächst ist es wichtig, zwischen zwei distinkten Antwortmustern zu unterscheiden, die durch CR entstehen können und sich jeweils unterschiedlich in den Daten ausdrücken (z. B. Alarcon und Lee 2022; Denison und Wiernik 2022; Schroeders et al. 2022). Besonders intuitiv

zeigt sich Random Responding (RR), das Antwortverhalten beschreibt, bei dem die Proband*innen zufällige Antworten geben (z. B. Clark et al. 2003; Credé 2010; DeSimone et al. 2018; Kam und Meyer 2015; Ward und Meade 2023). Statt die Antwort inhaltlich abzuwägen, wird willkürlich eine beliebige Antwortoption ausgewählt. Zufälliges Antwortverhalten führt zu inkonsistenten Antwortmustern, sodass die Aufdeckung über die Analyse der Antwortkonsistenz gelingen kann (z. B. Huang et al. 2012; Huang et al. 2015a; Meade und Craig 2012; Pineseault 2007). Frühe Studien verwenden die Begriffe RR und CR synonym, da die Heterogenität unaufmerksamen Antwortverhaltens noch nicht erkannt wurde (vgl. Beach 1989)⁵.

Die zweite Form, in der sich CR manifestiert, bilden monotone Antworten (z. B. Dunn et al. 2018; Huang et al. 2012). Dabei verwenden Befragte bestimmte Antwortmuster, indem sie zum Beispiel wiederholt stupide Abfolgen (z. B. „2,5,3,4, - 2,5,3,4-...“) bei aufeinanderfolgenden Items reproduzieren (Arthur et al. 2021). Lediglich die initiale Erstellung einer solchen Sequenz erfordert kurzzeitig etwas Aufwand. Die Sequenz wird dann ohne Abruf, Urteilsbildung und ohne Verständnis der Frage fortlaufend angewendet, statt jeweils den Prozess des Survey-Response Modells zu durchlaufen. Bei monotonem Antworten kann selbst der schon geringe Aufwand der Mustererstellung noch weiter reduziert werden, wenn sehr starrere Antwortsequenzen mit weniger Variation verwendet werden (z. B. „2, 3, 2, 3, - 2, 3, 2, 3“). Im Extremfall mündet monotonen Antworten in Straightlining, wobei die Antworten dann sogar ohne Bezug zur Antwortskala generiert werden können (z. B. Zhang und Conrad 2014). Insbesondere wenn das gewählte Antwortmuster sehr monoton ist, können Konsistenzanalysen CR nur bedingt indizieren. Sehr starre Antwortmuster können im Gegensatz zu RR, wenn alle Items gleichgerichtet sind, die Antwortkonsistenz sogar erhöhen, ohne dass dies als positives Zeichen für die Antwortqualität gewertet werden darf (DeSimone et al. 2018; Dunn et al. 2018).

Die dritte Form, in der sich CR in Onlinebefragungen zeigt, korrespondiert mit der Antwortzeit oder Response Time (RT). Ist die RT extrem kurz, kommt es zur CR-Form „Speeding“, bei der die Antworten so schnell gegeben werden, dass eine adäquate (oder auch nur rudimentäre) inhaltliche Verarbeitung der Fragen nicht möglich ist (z. B. Conrad et al. 2017a; Huang et al. 2015a; Ward und Meade 2023). Die Antwortzeit kann dabei sowohl seitenweise als auch als Gesamtzeit der Bearbeitung signalisieren, ob Befragte in bestimmten Befragungsteilen oder insgesamt unaufmerksam antworten (Bowling et al. 2016; Bowling et al. 2021b; Ward und Meade 2023). Sehr schnelles Antworten ist nicht scharf von RR und monotonem Antworten zu trennen, da es mit einem oder auch einer Mischung der beiden Antwortmuster einhergehen

⁵ Bis heute ist die Konzeptualisierung nicht konsistent, was der Diskurs zwischen Goldammer & Kolleg*innen (2020) und Denison und Wiernik (2022) offenbart. Auch dort werden CR und RR teilweise gleichgesetzt.

kann. Nichtsdestotrotz fundiert die Empirie die Unterscheidung der drei CR-Typen, da sie jeweils Mehrwert generieren, um CR in Befragungen möglichst komplett zu charakterisieren (z. B. Hong et al. 2020; Schroeders et al. 2022; Ward und Meade 2023). Werden neben RR und monotonem Antworten beispielsweise Mischformen der beiden, die sich in Speeding zeigen können, nicht berücksichtigt, gleichen sich die verschiedene CR-Formen womöglich gegenseitig aus und das tatsächliche CR-Ausmaß bleibt unerkannt (Curran 2016; DeSimone et al. 2018).

3.3.2 Careless Responding in (Online-)Befragungen

Insbesondere im Onlineformat ohne persönlichen, kontrollierenden Einfluss von Interviewenden repräsentiert CR ein dringliches Problem in Befragungen (vgl. Arias et al. 2020; Bowling und Huang 2018). Auch darum legt die vorliegende Arbeit CR als Evaluationsgrundlage möglicher Moduseffekte des Smartphones als Ausfüllgerät hinsichtlich der Antwortqualität an. Für CR als extreme Form des Antwortverhaltens ist der stärkste negative Einfluss auf die Antwortqualität zu erwarten. Andere moderatere Verzerrungen der Antwortqualität werden in dieser Arbeit aber nicht gänzlich ausgeklammert, sondern erweitern die Betrachtungen zielorientiert.

3.3.2.1 Ursachen von Careless Responding

CR lässt sich auf verschiedene Ursachen zurückführen, die sowohl auf der Ebene der Befragten als auch auf der Ebene der Befragung angesiedelt sind und dazu führen können, dass der idealtypische Antwortprozess des Survey-Response Modells verlassen wird. Auf individueller Ebene bilden 1) Motivation, 2) Persönlichkeit und 3) kognitive Kapazitäten wesentliche Einflussgrößen von CR. Auf Befragungsebene wirken 4) die Befragungseigenschaften und 5) der Kontext der Teilnahme auf CR.

Ward und Meade (2023) führen aus, dass motivationalen Zuständen bei der Untersuchung von CR die größte Bedeutung zukommt. Je motivierter an der Befragung teilgenommen wird, desto aufmerksamer sind Befragte und desto seltener zeigen sie CR (z. B. Arthur et al. 2021; Bowling et al. 2021a; Ward und Meade 2023). Sind Personen etwa an der Befragung oder den Befragungsinhalten interessiert, wird weniger CR beobachtet (z. B. Anduiza und Galais 2017).

Zudem existieren für CR auch Zusammenhänge zu Persönlichkeitseigenschaften. Gewissenhaftere und verträglichere Proband*innen zeigen seltener CR, während die Befunde für Offenheit, Extraversion und emotionale Stabilität gemischt sind (z. B. Berry et al. 2019; Bowling et al. 2016; DeSimone et al. 2020; Maniaci und Rogge 2014; Ward et al. 2017). Einige Befragte offenbaren gar über verschiedene Themen und Studien hinweg stabile CR-Neigungen (Bowling et al. 2016), was die individuelle Komponente unterstreicht. Die Befundlage indiziert darüber

hinaus, dass CR mit den kognitiven Fähigkeiten und Kapazitäten der Versuchspersonen zusammenhängt, da mehr CR auftritt, wenn der kognitive Aufwand während der Fragebeantwortung steigt (z. B. Ansolabehere und Schaffner 2015; Berinsky et al. 2014; Kam und Meyer 2015). Entsprechend sind komplexe Befragungsisems und -inhalte anfälliger für CR.

Damit wird deutlich, dass die Übergänge zu Ursachen von CR, die den Befragungseigenschaften entstammen, fließend sind. So wirkt beispielsweise die Befragungslänge auf die Antwortqualität (Galesic 2006; Galesic und Bosnjak 2009). Für längere Befragungen lässt sich unaufmerksames Antworten empirisch häufiger beobachten (Gibson und Bowling 2020), wobei CR besonders in späteren Teilen langer Befragungen steigt (z. B. Berry et al. 1992; Bowling et al. 2021a; Herzog und Bachman 1981; Kam und Chan 2018). Die Aufmerksamkeit Befragter während der Teilnahme kann stark variieren, sodass manche Befragungsteile durch CR verzerrt werden, während andere frei von CR sind und sorgfältig und gewissenhaft ausgefüllt werden (Brower 2020, S. 87; Clark et al. 2003; Ward et al. 2017).

Zuletzt wirken Kontexteffekte der Teilnahme unmittelbar auf CR (z. B. Gummer, Roßmann und Silber 2021; Ward und Meade 2023). Insbesondere ablenkende Umgebungsreize gefährden das Durchlaufen des idealtypischen Antwortprozesses und führen dazu, dass die Antworten nicht sorgfältig gegeben werden (z. B. Höhne und Schlosser 2018; Podsakoff et al. 2003; Sendelbah et al. 2016; Wenz 2021b; Zwarun und Hall 2014).

3.3.2.2 Auswirkungen von Careless Responding

Die Diversität der Entstehungsbedingungen von CR verdeutlicht, dass unaufmerksames Antworten in (Online-)Befragungen eine Grundgefahr darstellt und in jeder wissenschaftlichen Befragung auftreten kann (Abbey und Meloy 2017; DeSimone und Harms 2018; Hong et al. 2020; Ran et al. 2015). Selbst weniger stark betroffene Untersuchungen quantifizieren den Anteil von unaufmerksam Antwortenden („Careless Responder“) auf mindestens 5-15% (Curran 2016; Hauser und Schwarz 2015; Kurtz und Parrish 2001; Meade und Craig 2012). Andere Studien finden deutlich höhere Anteile von Befragten, die unaufmerksam antworten, von bis zu einem Drittel (Burns et al. 2014; Hong et al. 2020; Kam und Meyer 2015). Abhängig vom Aufdeckungsaufwand, von der Differenziertheit der berücksichtigten CR-Formen und der Breite der eingesetzten Aufdeckungsmaße können mitunter sogar CR-Raten von bis zu 50% oder mehr auftreten (Abbey und Meloy 2017; Berinsky et al. 2014; Oppenheimer, Meyvis und Davidenko 2009). Werden bei der Aufdeckung verschiedene CR-Formen gezielt berücksichtigt, ist das erkannte CR-Ausmaß besonders hoch (Brühlmann et al. 2020).

Die Antwortqualität wird durch CR klar eingeschränkt (z. B. Edwards 2019; Hough et al. 1990; Maniaci und Rogge 2014). Die Befundlage zeigt eindeutig, dass CR zu Verzerrungen in Befragungen führt und dass strukturelle Unterschiede zwischen den Daten von Befragten, die unaufmerksam antworten, und den Daten aufmerksamer Teilnehmer*innen bestehen (z. B. Berinsky et al. 2014; Hong et al. 2020; Huang et al. 2012; Huang und DeSimone 2021). Aufgrund von Unaufmerksamkeit antworten die Befragten häufiger inkonsistent oder übermäßig gleichförmig, sodass die Konstruktmessung (vor allem bei Skalen, die aus mehreren Items bestehen (Multi-Item-Skalen)) weniger akkurat ist und Reliabilität sowie Konstruktvalidität unter CR leiden können (z. B. Arias et al. 2020; Cheung et al. 2017; Maniaci und Rogge 2014; Ward und Meade 2023; Wood et al. 2017). In der Folge kann CR zu einer Veränderung korrelativer Zusammenhänge in den Daten führen, sodass die tatsächlichen Beziehungen zwischen Konstrukten nicht mehr valide geschätzt werden (Huang et al. 2015b; Kam und Meyer 2015; McGonagle et al. 2016; McGrath et al. 2010; Ward et al. 2017; Ward und Meade 2023). CR kann so beispielsweise die Ergebnisse faktoranalytischer Untersuchungen verzerren (Arias et al. 2020; Kam 2019; Reise et al. 2016; Schmitt und Stuits 1985; Voss 2021, S. 81; Ward und Meade 2023). Systematische Über- und Unterschätzungen der Parameter in Korrelations- und Regressionsanalysen sowie Verzerrungen der geschätzten Faktorstrukturen in Faktoranalysen treten dabei schon bei geringen Anteilen von CR (z. B. 5-15%) auf (Arias et al. 2020; Credé 2010; DeSimone und Harms 2018; Huang et al. 2015b; Reise et al. 2016; Schmitt und Stuits 1985).

CR schwächt die Aussagekraft statistischer Hypothesentest und der resultierenden Schlussfolgerungen: Durch CR können sowohl Fehler erster Art (Typ1-Fehler) als auch Fehler zweiter Art (Typ2-Fehler) häufiger auftreten, wodurch sich Sensitivität und Spezifität eingesetzter statistischer Testverfahren reduzieren (z. B. Clark et al. 2003; Huang et al. 2015b; Ophir et al. 2020; Schroeders et al. 2022). Steigt der Anteil von CR, wird diese Verzerrung ebenfalls stärker (Huang und DeSimone 2021; Woods 2006), da durch CR die Abweichung von den wahren Werten einer substanziellen Messung größer wird. Die generellen Empfehlungen von Wilkinson (1999) zur Datenbereinigung im Vorfeld der statistischen Analysen sollten, da Befragungsstudien stets mit den Risiken von Careless Responding konfrontiert sind, in der modernen Befragungsforschung auf CR ausgeweitet werden (z. B. DeSimone et al. 2015; DeSimone et al. 2020; DeSimone und Harms 2018; Hong et al. 2020; Meade und Craig 2012).

3.3.2.3 Umgang mit Careless Responding

Die Prävention von CR sollte eine prominente Rolle in der Studienplanung einnehmen, da sie die bestmögliche Umgangsform darstellt und dem ausschließlich statistischen Umgang mit CR

vorzuziehen ist (Arthur et al. 2021; Goldammer et al. 2020; Ward und Meade 2023). Tritt CR jedoch auf, unterscheiden Arthur und Kolleg*innen (2021) sowie Edwards (2019) für den Umgang mit CR drei grundsätzliche Strategien: 1) den Ausschluss betroffener Datensätze, 2) die Kontrolle von CR in der Datenauswertung und 3) die Anwendung von Subgruppenanalysen.

Der einfachste und häufigste Umgang mit CR ist, Befragte, die unaufmerksam antworten, von der Datenanalyse rigoros auszuschließen (z. B. DeSimone et al. 2018; Kennedy et al. 2020; Wood et al. 2017). Dies eliminiert Verzerrungen in den Daten, die durch unaufmerksames Antworten entstehen (Brühlmann et al. 2020; DeSimone und Harms 2018). Gleichzeitig steigt durch die Verringerung der Verzerrungen in den Daten die Aussagekraft der statistischen Schlüsse aus der Hypothesenprüfung (z. B. Abbey und Meloy 2017; Arias et al. 2020; Maniaci und Rogge 2014; Oppenheimer et al. 2009). Klarer Nachteil der Elimination von CR-Daten ist die damit einhergehende Verringerung der Stichprobengröße. Werden von CR betroffene Datensätze ausgeschlossen, ist es ratsam, die Stichprobe schon im Vorfeld größer anzulegen, um antizipierte Verluste durch CR ausgleichen zu können (Arndt et al. 2022; Goldammer et al. 2020; Marjanovic et al. 2015).

Ferner bietet die Behandlung von CR als Kontrollvariable eine Möglichkeit, den Einfluss verringerter Aufmerksamkeit bis zu einem gewissen Grad in der Datenanalyse zu kontrollieren, ohne den Stichprobenumfang zu verringern. Dafür müssen Variablen erstellt werden, die möglichst eindeutig das Auftreten und das Ausmaß von CR signalisieren und dann in die Auswertungsmodelle eingehen. CR oder die Antwortqualität selbst können dadurch beispielsweise als Moderatorvariablen einbezogen werden (z. B. McGrath et al. 2010; Ward und Meade 2018).

Drittens kann CR als Gruppierungsvariable fungieren, wobei basierend auf der CR-Klassifikation Subgruppen gebildet werden (Edwards 2019; Kurtz und Parrish 2001). Die Analysen können dann in der Folge separat jeweils in der Gruppe der unaufmerksamen und aufmerksamen Befragten durchgeführt und die Ergebnisse miteinander verglichen werden.

3.3.3 Aufdeckung von Careless Responding

In Vorbereitung des zu erarbeitenden Forschungsprogramms werden die Möglichkeiten der CR-Aufdeckung, um unaufmerksame Proband*innen in Befragungen zu identifizieren, vorgestellt und hinsichtlich der Eignung für die vorliegende Arbeit strukturiert. In der jüngeren Forschung wurde eine Vielzahl von potenziellen Aufdeckungsmöglichkeiten für CR vorgeschlagen und evaluiert. Die Diskussion ist fluide, wobei die Anzahl möglicher Indikatoren permanent steigt und diese zunehmend komplexer werden⁶.

⁶ Siehe zum Beispiel Schroeders, Schmidt und Gnambs (2022)

Dabei muss zunächst die Aufdeckungsebene differenziert werden. In Laborstudien bieten sich andere Aufmerksamkeitsmaße an als in Onlinebefragungen im Feld. Gabaix (2019, S. 279) führt aus, dass physische Maße, die nur im Labor verwendet werden können, die direkte Aufmerksamkeitsmessung erlauben. So werden Eye-Tracking Verfahren in der Befragungsforschung häufig genutzt, da sie valide Rückschlüsse auf Aufmerksamkeitsprozesse ermöglichen (z. B. Baumgartner, Weijters und Pieters 2018; Galesic et al. 2008). Über das Blickverhalten der Befragten lässt sich auf die kognitive Verarbeitung schließen, was auch Hinweise auf CR liefern kann. Der erste Schritt der Hauptuntersuchungen der vorliegenden Arbeit zu Moduseffekten des Smartphones hinsichtlich der Antwortqualität betreibt daher über die physische Aufmerksamkeitsmessung in Form von Eye-Tracking Grundlagenforschung (Kapitel 3.3.3.1), indem geprüft wird, ob sich die Aufmerksamkeitsstrukturen gerätweise unterscheiden könnten.

Im Feld kann die Aufmerksamkeitsmessung zwar nur indirekt erfolgen, in Onlinebefragungen liefert aber gerade Careless Responding Aufschluss über Aufmerksamkeitszustände der Befragten. Für die Aufdeckung von CR existieren verschiedene Indikatoren, für die die gesammelten Erkenntnisse schon mehrfach zusammengefasst wurden (z. B. Arthur et al. 2021; Curran 2016; Hong et al. 2020; Meade und Craig 2012). Dabei werden zwei Arten von CR-Indikatoren unterschieden (Brühlmann et al. 2020; Curran 2016; Hong et al. 2020; Huang et al. 2012). Zum einen kann die Studie im Vorfeld mit dem Ziel der CR-Aufdeckung aktiv angepasst werden (nachfolgend „geplante Detektionsmethoden“; Kapitel 3.3.3.2). Zum anderen kann die retrospektive Datenaufbereitung CR identifizieren (nachfolgend „post-hoc Detektionsmethoden“; Kapitel 3.3.3.3). Kapitel 3.3.3.4 konzeptualisiert das Prüfkonzept für die vorliegende Arbeit und identifiziert für die Forschungsfrage geeignete Maße der CR-Aufdeckung.

3.3.3.1 Eye-Tracking

Das etablierteste Instrument der physischen Aufmerksamkeitsmessung, insbesondere in der Marketingforschung oder auch in der Psychologie, ist die Aufzeichnung und Analyse des Blickverhaltens in Form von Eye-Tracking (z. B. Gabaix 2019, S. 281; Meißner und Oll 2019; Rosbergen, Pieters und Wedel 1997; Singh, Göritz und Moser 2015, S. 169). Eye-Tracking Daten ermöglichen, auf kognitive Prozesse der Proband*innen zu schließen (Gabaix 2019, S. 279-281; Lai et al. 2013; Pieters und Warlop 1999; Rayner 2009; Wedel und Pieters 2008, S. 144) und Eye-Tracker sind daher auch ein recht mächtiges Werkzeug, um Verarbeitungs- und Aufmerksamkeitsprozesse während der Befragungsteilnahme abzubilden (Baumgartner et al. 2018; Galesic et al. 2008; Lenzner, Kaczmirek und Galesic 2011; Neuert und Lenzner 2016). Da Aufmerksamkeit für gute Antwortqualität wesentlich ist, entsteht eine reichhaltige Datenbasis als

adäquate Grundlage der Analyse möglicher Antwortqualitätsprobleme in Befragungen (z. B. Duchowski 2017, S. 217-218; Reingold 2014). Durch die fortschreitende technische Entwicklung werden Eye-Tracker – ehemals sperrige und unhandliche Apparate – im Forschungsalltag stetig praktikabler, sodass die Belastung der Proband*innen abnimmt. Moderne Eye-Tracker sind zudem weniger umständlich zu bedienen, einfacher einzurichten und fortgeschrittene Softwarelösungen für die Datenauswertung sind unmittelbar mit dem Erhebungsinstrument gekoppelt, während zudem die Auflösung der Blickaufzeichnung steigt (Duchowski 2017, S. 97-98).

Eye-Tracking Systeme erfassen das Blickverhalten der Proband*innen über Videoaufnahmen des Gesichtsfelds. Parallel werden Lichtimpulse auf die Hornhaut des Auges ausgesendet, deren Reflektionen es den Forscher*innen ermöglichen, das Blickverhalten zu klassifizieren. Die Lichtimpulse entstammen Infrarot-Leuchtdioden (LED), die in die Kamera integriert sind und genau wie die Kamera selbst auf die Augen der Proband*innen gerichtet sind. Ein Teil des Lichtimpulses durchdringt die Hornhaut und trifft auf die Netzhaut, aber ein großer Teil wird von der Hornhaut unmittelbar reflektiert. Anhand der zurückgeworfenen Lichtsignale ermittelt die integrierte Software das Zentrum der Pupille und den Ein- und Ausfallwinkel der Lichtimpulse sowie die Intensität der Hornhautreflektionen. Diese Parameter werden über einen trigonometrischen Algorithmus kleinschrittig in Blickbewegungen umgewandelt, wodurch das Blickverhalten rekonstruiert wird (vgl. Meißner und Oll 2019; Reingold 2014). Für die beobachteten Szenen, Bildschirme oder Oberflächen entsteht so ein Abbild des Blickpfades. Damit diese Projektion gelingt, muss der Eye-Tracker auf die Besonderheiten jedes*r einzelnen Teilnehmer*in eingestellt und eine Kalibrierungsphase vorgeschaltet werden (Duchowski 2017, S. 121-123). Ziel der Blickaufzeichnung ist, die Orientierung der Fovea (Sehgrube) als Zentrum der Blickorientierung abzubilden, die den Bereich des schärfsten Sehens repräsentiert (Honsel 2012, S. 31). Durch die Festlegung von „Areas of Interest“ (AOI) werden die Bereiche definiert, die im Forschungsinteresse liegen und für die das Blickverhalten ermittelt werden soll (z. B. Jacob und Karn 2003, S. 581-584).

Blickbewegungen können in zwei wesentliche Komponenten unterteilt werden: in Fixationen und in Sakkaden⁷ (z. B. Rayner 1998, 2009; Shi, Wedel und Pieters 2013). Fixationen beschreiben stabiles Blickverhalten, bei dem der Fokus für längere Zeit kontinuierlich (mindestens 100-150ms) auf einen räumlich begrenzten Ort gerichtet ist (z. B. Duchowski 2017, S. 44; Jacob und Karn 2003, S. 581-584; Rayner 2009; Wedel und Pieters 2008, S. 124). Sakkaden repräsentieren rasante (20-100ms), ballistische Sprünge des Blickes, die das foveale Sehen auf

⁷ Das Blickverhalten kann noch deutlich kleinschrittiger untergliedert werden. Die vorliegende Arbeit beschränkt sich auf Fixationen und Sakkaden, da grundlegende Verarbeitungsstrukturen im Fokus stehen. Für eine ausführlichere Klassifikation der Augenbewegungen siehe Rayner (1998) und Duchowski (2017, S. 39-45).

einen anderen Ort der visuellen Umgebung ausrichten (z. B. Duchowski 2017, S. 40; Rayner 2009; Wedel und Pieters 2008, S. 124). Im Vorfeld einer Sakkade benötigt das visuelle System eine kurze Programmierungszeit („Sakkadenlatenz“) im Rahmen von etwa 150-175ms, um den finalen Zielort der Sakkade zu bestimmen und die Ausführung der Sakkade vorzubereiten (Rayner 1998). Eye-Tracking Systeme zeichnen das Blickverhalten permanent auf und unterteilen nach einer Datenbereinigung die Blickbewegungen in Sakkaden und Fixationen (z. B. Honsel 2012, S. 35). Aus der Klassifikation in Fixationen und Sakkaden resultieren verschiedene Maße für den Schluss auf kognitive Verarbeitungsprozesse. Klassische Eye-Tracking Maße der kognitiven Verarbeitung sind, basierend auf dieser Klassifikation: 1) die totale Blickzeit innerhalb der definierten AOIs (auch als Anteil an der gesamten Blickzeit), 2) die kumulierte Dauer (ebenfalls auch als Anteil) und 3) Häufigkeit der Fixationen innerhalb der AOIs, 4) die durchschnittliche Fixationsdauer, 5) die verstrichene Zeit bis zur ersten Fixation, 6) die Sakkadenamplitude und 7) -länge, 8) die Nachzeichnung des Blickpfads und 9) die Anzahl an Refixationen in der AOI (wiederholte Fixation nach Verlassen der AOI) (z. B. Jacob und Karn 2003, S. 581-584; Meißner und Oll 2019; Radach und Kennedy 2004; Rayner 2009).

Indikatoren des Blickverhaltens und insbesondere die Blickdauer haben sich als robustes Aufmerksamkeitsmaß in diversen Kontexten mehrfach bewährt (z. B. Jacob und Karn 2003, S. 585; Meißner und Oll 2019; Strick et al. 2010). In der Marketingforschung beispielsweise wird das Blickverhalten (etwa in der Werbebannerforschung) als wesentlicher Aufmerksamkeitsindikator genutzt (vgl. Huddleston et al. 2015; Lee und Ahn 2012; Wedel und Pieters 2008). Um vom Blickverhalten qualifiziert auf Aufmerksamkeitsprozesse schließen zu können, muss der Zusammenhang der Blickindikatoren und der kognitiven Prozesse strukturiert werden. Hintergrund der Eye-Tracking Forschung ist die „Eye-Mind-Hypothese“ (Just und Carpenter 1980), die davon ausgeht, dass das Blickverhalten kognitive Prozesse repräsentiert. Entsprechend indizieren Augenbewegungen, was wahrgenommen wird und wo der Aufmerksamkeitsfokus liegt (z. B. Duchowski 2017, S. 5; Orquin und Mueller Loose 2013). Das Blickverhalten stützt etwa die Präferenzbildung bei Auswahlprozessen und beeinflusst die Entscheidungsfindung (z. B. Guerreiro, Rita und Trigueiros 2015; Krajbich und Rangel 2011; Laeng, Suegami und Aminihajibashi 2016; Meißner und Oll 2019; Orquin und Mueller Loose 2013; Reutskaja et al. 2011). Fixationen signalisieren entsprechend der Eye-Mind-Hypothese den aktuellen Aufmerksamkeitsfokus (Just und Carpenter 1976) und stellen das beste bekannte physische Aufmerksamkeitsmaß dar (z. B. Duchowski 2017, S. 250-251; Honsel 2012, S. 35-36; Lai et al. 2013; Meißner und Oll 2019; Orquin und Mueller Loose 2013; Rayner 2009; Wedel und Pieters 2008, S. 141; Zhang 2020, S. 130). Elemente oder AOIs, die fixiert werden, werden besser erinnert

(Krugman et al. 1994; Loftus 1972; Pashler 1988; Tatler, Gilchrist und Land 2005), effizienter verarbeitet (Jacob und Karn 2003, S. 585) und als interessanter wahrgenommen (Honsel 2012, S. 36; Jacob und Karn 2003, S. 585). Dies äußert sich gleichermaßen in kumulierter Fixationsdauer, durchschnittlicher Fixationsdauer und Fixationshäufigkeit, wobei längere und häufigere Fixationen eine größere Aufmerksamkeit und Verarbeitungstiefe anzeigen (Meißner und Oll 2019; Rayner 2009; Zhang 2020, S. 130). Auch die Anteile der Fixationszeit und -häufigkeit an der Gesamtdauer müssen berücksichtigt und Refixationen explizit einbezogen werden, um ein möglichst vollständiges Bild der Aufmerksamkeit und der Kontinuität der Verarbeitung zu zeichnen (z. B. Guerreiro et al. 2015; Rayner 2009). Für extrem lange Fixationen existiert eine zweite kognitive Signalwirkung: Diese treten etwa auch bei komplexen Inhalten auf und können Probleme der Informationsverarbeitung signalisieren (Jacob und Karn 2003, S. 580).

Über Fixationen hinaus verdeutlicht das Konzept der verdeckten visuellen Aufmerksamkeit, dass Menschen auch außerhalb des fovealen Sehens Informationen rudimentär aufnehmen und verarbeiten können (Meißner und Oll 2019). Obwohl die Aufmerksamkeit in der absoluten Mehrheit der Fälle auf dem Fixationsort liegt, kann sie punktuell in die Peripherie verlagert werden (Duchowski 2017, S. 13; Meißner und Oll 2019; Posner 1980). Daher sind Sakkaden für Aufmerksamkeits- und Verarbeitungsprozesse ebenfalls von Bedeutung. Sakkaden indizieren Aufmerksamkeitsverschiebungen, wobei der Blick von einem Fixationsort zum nächsten springt oder – bei aufeinanderfolgenden Sakkaden – die visuelle Umgebung durchwandert (z. B. Kowler et al. 1995; Mayfrank et al. 1986). In der Regel bildet der Zielort einer Sakkade einen neuen Fixationspunkt, auf dem nun der Aufmerksamkeitsfokus liegt, wodurch die Aufmerksamkeit meist schon vor Sakkadenbeginn teilweise verlagert werden muss, um den Zielort der neuen Fixation zu bestimmen (z. B. Kowler et al. 1995; Rayner 2009). Während einer Sakkade selbst können keine Informationen aufgenommen werden (Meißner und Oll 2019; Rayner 1998). Sehr unstetes Blickverhalten mit häufigen oder sehr langen Sakkadenphasen kann daher Unaufmerksamkeit und ineffiziente Informationsverarbeitung signalisieren (Rayner 1998).

Da die Informationsaufnahme bei der Sakkadenausführung unterdrückt wird und nur während Fixationen stattfindet, stellt sich die Frage nach der Bedeutung der Blickrückkehr – sogenannten Refixationen. Shen, McIntosh und Ryan (2014) demonstrieren einen klaren Zusammenhang zwischen der Kapazität des Arbeitsgedächtnisses und Refixationen. Bei jeder Refixation werden im Arbeitsgedächtnis Informationen zu den bereits vorhandenen Inhalten addiert (Wedel und Pieters 2000). Bleibt die Repräsentation im Arbeitsgedächtnis aus der vorherigen Fixation intakt, sind Refixationen wichtig, wenn sie neue Informationen beisteuern (Shen et al. 2014). Im Arbeitsgedächtnis als visuellem Kurzzeitgedächtnis zerfallen Informationen jedoch

sehr schnell, sodass Refixationen häufig eine Update-Funktion einnehmen (Tatler et al. 2005; Zelinsky, Loschky und Dickinson 2011). Ist die Zeit zwischen Fixation und Refixation zu lang, unterstützen Refixationen also nicht die Speicherung und den Gedächtnisabruf, da sie keine neuen Informationen addieren, sondern die Informationsverarbeitung erst wieder neu anstoßen (Tatler et al. 2005; Zhang, Anderson und Miller 2021).

Eine besondere Bedeutung der Aufmerksamkeitsverteilung kommt in der vorliegenden Arbeit textbasierten Inhalten zu. In (Online-)Befragungen müssen Befragte, um gehaltvoll antworten zu können, die Befragungsinhalte lesen, verstehen und interpretieren. Gerade bei Leseprozessen erlauben Eye-Tracking Ansätze die differenzierte Analyse der Aufmerksamkeitsstrukturen (vgl. Jarodzka und Brand-Gruwel 2017; Rayner 2009) und bieten sich somit für Onlinebefragungen an, um die Aufmerksamkeit Befragter bei der Teilnahme zu charakterisieren. Ein umfassendes Modell zur Erklärung der Blickaktivität während des Lesens bildet das „E-Z Reader Modell“ von Reichle und Kolleg*innen (1998). Dieses wurde in der Folge sukzessive weiterentwickelt, sodass sich aus der Literatursynthese (vgl. Rayner 1998, 2009) das Blickverhalten beim Lesen wie folgt charakterisieren lässt: Während des Lesens ist ein stetes und kontrolliertes Blickverhalten wichtig. Der Blick folgt dem Textverlauf und je synchroner Blickverlauf und Textfluss sind, desto besser ist das Textverständnis. Fixierte Wörter werden unmittelbar verarbeitet und sobald die Verarbeitung abgeschlossen ist, wird eine Sakkade zum nächsten noch nicht verarbeiteten Wort programmiert. Sind Wörter bekannt, benötigen sie kürzere Blickzeit – komplexe und unbekannte Inhalte resultieren in längerer Verarbeitungszeit, die sich in einer ebenfalls längeren Fixationsdauer widerspiegelt. Neue Informationen werden beim Lesen ausschließlich während der Fixation aufgenommen (Rayner 1998) und periphere Informationen werden ausgeblendet (Rayner 2009). Bei konzentriertem Lesen erfolgt die Sakkadenprogrammierung immer inhaltsbezogen auf nachfolgende Textpassagen; Sakkaden, die hingegen aus dem Textbereich hinausgehen, hindern den Lesefluss (Rayner 1998, 2009). Wird der Lesefluss unterbrochen, sind Refixationen notwendig, um den Leseprozess ohne Beeinträchtigung fortsetzen zu können (Rayner 1998, 2009).

Eine wesentliche Gefahr für die Antwortqualität in Onlinebefragungen liegt im Phänomen des „Mind-Wandering“ (Zhang 2020, S. 142; Zhang et al. 2021). Mind-Wandering beschreibt einen Aufmerksamkeitszustand, bei dem die Gedanken der Proband*innen von gerade bearbeiteten Aufgaben abschweifen, und der zu einer schlechteren Leseleistung und Verarbeitungstiefe führt (Steindorf und Rummel 2020; Zhang et al. 2021). Werden dann trotzdem in Onlinebefragungen Antworten gegeben, ist die Gefahr von CR durch die Loslösung von den Befragungsinhalten groß. Je länger eine Aufgabe oder Befragung dauert, desto größer wird die Gefahr für

Mind-Wandering und somit auch für CR (Steindorf und Rummel 2020). Dies manifestiert sich in einer geringeren Anzahl von Fixationen im Textbereich und ebenso in überlangen Fixationen ohne Informationsaufnahme (Steindorf und Rummel 2020; Zhang 2020, S. 128). Zugleich werden häufiger Refixationen benötigt, um den Lesefluss wieder aufnehmen zu können und die Gedächtnisinhalte zu aktualisieren (Tatler et al. 2005). In Onlinebefragungen kann über Eye-Tracking Untersuchungen geprüft werden, inwiefern die effiziente Informationsverarbeitung gelingt und die Aufmerksamkeit ausreichend auf die Befragungsinhalte gerichtet wird. Eye-Tracking wird dabei als zuverlässiger Ansatz für die Aufdeckung von CR beschrieben (z. B. Brower 2020, S. 87; Ward und Meade 2023). Auch auf die Motivation während der Befragungsteilnahme als möglichen Hintergrundfaktor von CR lässt sich über Eye-Tracking Analysen schließen, da längere Fixationen und gewissenhaftere Betrachtungen relevanter Inhalte motivationale Zustände repräsentieren können (z. B. Orquin und Mueller Loose 2013; Pieters und Warlop 1999). Für das Forschungsprogramm der vorliegenden Arbeit qualifiziert sich daher der Einsatz von Eye-Tracking als objektives, physisches Aufmerksamkeitsmaß, um denkbare Einflüsse des Smartphones auf die Antwortqualität abschätzen zu können, die auf mögliche Moduseffekte im Sinne des TSE hinweisen könnten.

3.3.3.2 Geplante Detektionsmethoden

Realistische Settings der Befragungsteilnahme können Eye-Tracking Studien nicht abbilden, da die Erhebungssituation nicht im Alltag der Proband*innen angesiedelt ist, sondern artifiziell als Laborstudie angelegt werden muss. Der Schluss auf Aufmerksamkeitszustände in Onlinebefragungen erfolgt im Feld daher indirekt über die Aufdeckung von Careless Responding anhand spezifischer CR-Indikatoren. Diese werden in zwei Gruppen aufgeteilt und in der Folge erläutert. *Abbildung 6* stellt die verschiedenen Indikatoren kompakt dar, die für die Aufdeckung von CR in Onlinebefragungen in der Literatur diskutiert werden. Die erste Gruppe nicht-apparativer CR-Indikatoren bilden Befragungsinhalte (z. B. Items oder Skalen), die im Vorfeld gezielt für die Aufdeckung von unaufmerksamem Antwortverhalten in die Befragung eingebaut werden (z. B. Arndt et al. 2022; Curran 2016; Hong et al. 2020; Meade und Craig 2012). In der vorliegenden Arbeit werden diese als „geplante Detektionsmethoden“ bezeichnet.

Geplante Detektionsmethoden umfassen unter anderem spezifische Items, deren Beantwortung unmittelbar auf den Aufmerksamkeitsgrad der Antwortenden schließen lässt. Ein erster Ansatz der geplanten CR-Aufdeckung – der „Infrequency“-Ansatz – geht auf die Forschung von Beach (1989) zurück und ist in seiner Idee simpel und zugleich hochgradig augenscheinlich valide. Der Infrequency-Ansatz wird so bezeichnet, da er CR aufdeckt, indem nach nicht

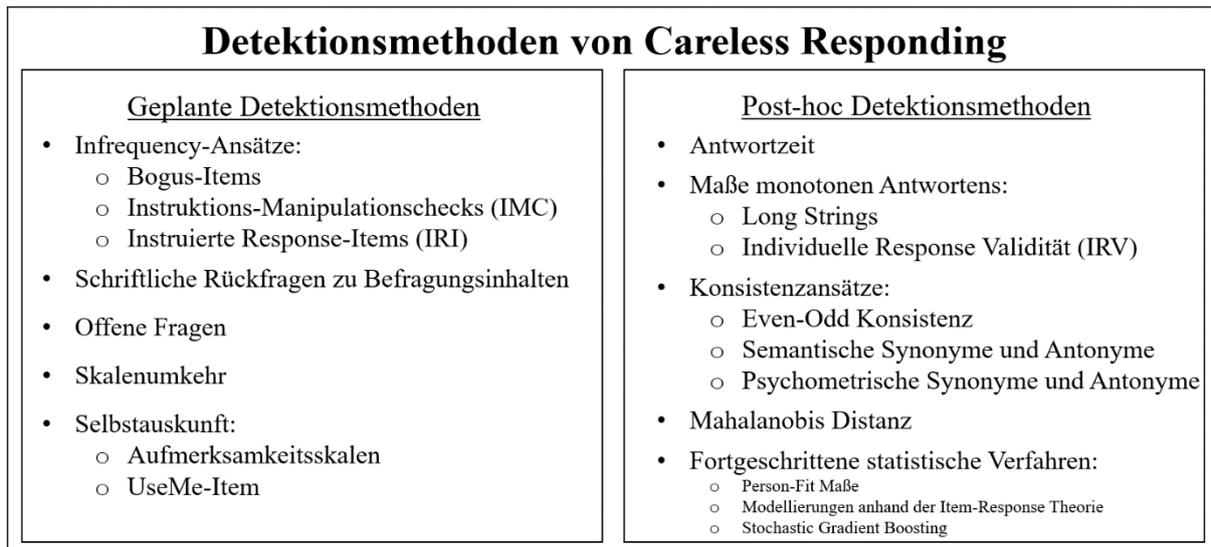


Abbildung 6. Übersicht über verschiedene Detektionsmethoden für Careless Responding

plausiblen Antworten gesucht wird, die bei aufmerksamer Teilnahme nicht entstehen dürften (z. B. Abbey und Meloy 2017; Huang et al. 2012; Huang et al. 2015a). Die intuitivste Form stellen logische Statements oder auch „Bogus-Items“ dar (z. B. Abbey und Meloy 2017; Arthur et al. 2021; Curran 2016; Dunn et al. 2018). Bogus-Items sind so formuliert, dass sie nur eine richtige Antwort zulassen bzw. abweichende Antworten nicht plausibel sind (Beach 1989; Huang et al. 2015a). Sie können beispielsweise sehr seltene oder praktisch unmögliche Verhaltensweisen adressieren („Ich bin schon 23-mal um die Welt gereist“ (Dunn et al. 2018, S. 120)) oder als unlogische Aussagen, zu denen Stellung genommen werden soll („Ich verstehe nicht ein Wort der deutschen Sprache“ (Meade und Craig 2012, S. 441)), formuliert sein⁸. Vermehrt wird empfohlen, statt der ursprünglichen dichotomen Zustimmung oder Ablehnung der Items (Beach 1989; Huang et al. 2015a), mehrstufige Zustimmungsskalen zu hinterlegen, um zufälliges, scheinbar aufmerksames Antworten zu minimieren. Werden hier einer oder mehrere Extremwerte ausgewählt (z. B. „stimme voll und ganz zu“ bzw. „stimme zu“ bei einem Item wie „Ich nehme regelmäßig Arsen ein“ oder „stimme überhaupt nicht zu“ bzw. „stimme nicht zu“ bei einem Item wie „Ich kann lesen“), wird die Antwort als CR klassifiziert (z. B. Berry et al. 2019; Dunn et al. 2018; Olatunji et al. 2007). Die Formulierung der in der Literatur verbreiteten Bogus-Items ist vielfältig und variiert stark. Es existieren ganze Sets, die sich teilweise oder gar komplett aus Bogus-Items zusammensetzen (Berry et al. 2019; Curran und Hauser 2019; Dunn et al. 2018; Huang et al. 2015a; Maniaci und Rogge 2014; Meade und Craig 2012). Mitunter können vermeintliche Bogus-Items wenig geeignet zur CR-Aufdeckung sein, wenn Befragte sie nicht wortwörtlich interpretieren: Beispielsweise könnten sich Personen, die wenig Schlaf

⁸ Diese und weitere zitierte Items des Kapitels wurden aus dem Englischen mit der kollaborativen, iterativen Übersetzungsmethode (Douglas und Craig 2007) ins Deutsche übertragen.

benötigen, schwertun, ein Item wie „Ich schlafe nie“ vollkommen abzulehnen, obwohl dies faktisch die korrekte Antwort wäre. Für Bogus-Items sollte daher vorab gesichert werden, dass abweichende Antworten auf die jeweiligen Items auch unzweifelhaft falsch oder unlogisch sind (Curran und Hauser 2019; Hargittai 2009). Darüber hinaus zeigen einige Literaturbeispiele, dass sich extreme Bogus-Items nur schwerlich thematisch in eine Onlinebefragung einordnen lassen: z. B. „Alle meine Freunde sind Aliens“, „Ich werde zweiwöchentlich von Kobolden bezahlt“ (Meade und Craig 2012, S. 441). Solche Items können irritierend für die Befragten wirken und Reaktanz auslösen (z. B. Baumgartner und Weijters 2012; Breitsohl und Steidelmüller 2018; Curran und Hauser 2019; Hauser et al. 2017). Bogus-Items sollten daher in Befragungen diskret und mit größtmöglicher inhaltlicher Nähe zu den Befragungsthemen integriert werden (Curran 2016; Curran und Hauser 2019; Huang et al. 2015a). Es wird auch empfohlen, Bogus-Items nicht in zu großer Zahl, vorrangig gegen Ende längerer Befragungen und mit Abstand zu weiteren Bogus-Items einzusetzen (Dunn et al. 2018; Meade und Craig 2012). Da Bogus-Items auf Antworten abstellen, die nicht plausibel sind, wenn die Frage aufmerksam gelesen wurde, können prinzipiell alle CR-Formen (RR, monotones Antworten und Speeding) zu unlogischen Antworten führen. Die Aufdeckung ist aber nicht immer verlässlich, da auch zufällig gemäß der gewählten Antwortstrategie vermeintlich korrekt geantwortet werden kann.

Eine weitere etablierte Komponente des Infrequency-Ansatzes der geplanten CR-Aufdeckung ist der „Instruktions-Manipulationscheck“ (IMC), den Oppenheimer und Kolleg*innen (2009) einführen. Ein IMC enthält eine klare Instruktion für die Befragten, wie eine bestimmte folgende Frage beantwortet werden soll, und setzt sich zusammen aus einem längeren Textblock und einem darauffolgenden Item. Das Item vermittelt auf den ersten Blick den Eindruck, dass ein bestimmtes Thema adressiert wird. Statt das Item inhaltlich beantworten zu sollen, wird den Proband*innen im Textblock aber eine ganz bestimmte Antwort oder ein spezielles Verhalten (z. B. Auslassen der Frage) vorgegeben. Der IMC von Oppenheimer und Kolleg*innen (2009) präsentiert im Rahmen einer Onlinebefragung ein Item, das nach der bevorzugten Freizeitaktivität fragt, während der lange Instruktionstext zuvor die Proband*innen auffordert, die Frage nicht zu beantworten, sondern stattdessen auf den Seitentitel zu klicken. Teilnehmer*innen, die die Instruktionen nicht aufmerksam lesen, folgen der Ausfüllanweisung nicht und beantworten stattdessen die Frage. In der Forschungsliteratur finden sich verschiedene IMCs zur Unterscheidung aufmerksamer und unaufmerksamer Proband*innen (z. B. Anduiza und Galais 2017; Goodman, Cryder und Cheema 2013; Hauser und Schwarz 2015). Gerade lange IMCs sollten sich möglichst inhaltlich und formal in Befragungen einbetten, um Irritation und Reaktanz zu vermeiden (z. B. Hauser et al. 2017; Hauser und Schwarz 2015; Oppenheimer

et al. 2009). Je ausführlicher IMCs angelegt sind (z. B. längere Ausfüllanweisungen, wie das Antwortverhalten genau aussehen soll), desto unwahrscheinlicher wird die zufällig richtige Antwort bei Unaufmerksamkeit. Längere und kompliziertere Instruktionsanweisungen werden tatsächlich auch seltener bestanden (Anduiza und Galais 2017; Babakhani, Paas und Dolnicar 2022; Kam und Meyer 2015; Oppenheimer et al. 2009); mitunter befolgen die Hälfte der Proband*innen oder sogar mehr die Ausfüllanweisung nicht korrekt (Anduiza und Galais 2017; Leiner 2019; Oppenheimer et al. 2009). Eine klare Trennung zwischen unaufmerksamem Antworten und sprachlichen bzw. kognitiven Fähigkeiten ist bei IMCs aber nicht immer möglich, sodass auch ein vereinfachtes instruktionsbasiertes CR-Maß existiert, das die Belastung und den Zeitaufwand der Proband*innen reduziert. Instruierte Response-Items (IRI) nutzen die Grundidee des IMC und übertragen sie auf die Itemebene (z. B. Gummer et al. 2021; Kam und Meyer 2015; Maniaci und Rogge 2014). IRIs werden vorrangig in Matrizen integriert, die mehrere Items beinhalten, und weisen die Proband*innen an, auf eine bestimmte Weise zu antworten, statt ein inhaltlich relevantes Konstrukt zu messen: z. B. „Dies ist eine Qualitätscheck. Bitte beantworten Sie diese Frage nicht“ (Paas und Morren 2018, S. 14) oder „Bitte wählen Sie ‚Ich stimme voll und ganz zu‘ für dieses Items aus“ (Kam und Chan 2018, S. 84). IRIs besitzen neben der weniger herausfordernden Instruktion und der besseren Integration in die eigenen Befragungselemente, ohne das Antwortformat ändern zu müssen, den Vorteil, dass sie eine hohe Augenscheinvalidität für die Aufdeckung von CR auszeichnet und dass sie Unaufmerksamkeit recht verlässlich erkennen können (Gummer et al. 2021; Kam und Chan 2018).

Gezielte schriftliche Rückfragen zu Inhalten einer Befragung bilden eine weitere Möglichkeit, unaufmerksames Antwortverhalten zu identifizieren. Die Teilnehmer*innen werden aufgefordert, sich bestimmte Inhalte einzuprägen und diese später zu reproduzieren (z. B. Abbey und Meloy 2017; Arndt et al. 2022). Ist die Anzahl der korrekt beantworteten Fragen hoch, wird die Befragung aufmerksam bearbeitet. Allerdings ist die Verwendung inhaltlicher Rückfragen zur Identifikation nachlässigen Antwortverhaltens problematisch, da bei Leseaufgaben, wie schon beim IMC, die kognitive Fähigkeit der Befragten wesentlich ist (vgl. Britton 1980; Holland und Rabbitt 1990). Da CR und kognitive Fähigkeiten interferieren, das heißt, sehr fähige Befragte auch bei reduzierter Aufmerksamkeit häufiger korrekt antworten können als weniger fähige, aber eigentlich aufmerksamere Teilnehmer*innen, sind gezielte schriftliche Rückfragen zu Befragungsinhalten als alleiniger CR-Indikator nicht ratsam.

In Ergänzung zu schriftlichen Rückfragen fungieren auch offene Fragen als CR-Maß. Offene Fragen können sich als inhaltliche Rückfragen (wie im vorherigen Abschnitt beschrieben) auf zuvor präsentierte Befragungsinhalte beziehen, aber auch davon losgelöst nach der Meinung

oder Sichtweise der Teilnehmer*innen fragen. Die Befragten antworten auf offene Fragen in eigenen Worten in Textform und bei offenen Fragen wird häufig die Textlänge der Antwort als Aufmerksamkeitsmaß genutzt (z. B. Abbey und Meloy 2017; Brühlmann et al. 2020; Huang et al. 2015a). Brühlmann und Kolleg*innen (2020) führen ergänzend aus, dass zur Beurteilung der Antwortqualität bei offenen Fragen insbesondere die Anzahl der Wörter bzw. Buchstaben aber auch der Antwortgehalt (z. B. bei schriftlichen Rückfragen, die als offene Fragen angelegt sind) herangezogen werden können⁹. Bei hoher Aufmerksamkeit antworten Befragte einerseits ausführlicher und andererseits ist die Güte der Antwort höher.

Auch konkrete Fragebogenelemente können zur CR-Aufdeckung genutzt werden. In selbst-administrierten Befragungen wirkt die Richtung der Antwortskala direkt auf das Antwortverhalten (Belson 1966; Friedman und Amoo 1999; Krämer 2017, S. 98). Dies gilt vor allem für Antwortskalen, die mehrstufige aufgebaut sind (Kulas, Klahr und Knights 2019; Yan, Keusch und He 2018). Insbesondere die Skalenumkehr erlaubt dabei, Antwortstile und -muster präzise zu unterscheiden (Baumgartner und Weijters 2019, S. 299-300). Wird die Skalenumkehr von Proband*innen nicht bemerkt, kann dies CR signalisieren (z. B. Abbey und Meloy 2017; Berinsky et al. 2014; Weijters et al. 2013). Bei ähnlichen oder gar identischen Items sollten aufmerksame Befragte unabhängig von der Skalenrichtung eine Antwortkategorie auswählen, die der Antwort auf das vorherige Item (ggf. nach Reskalierung) zumindest sehr ähnlich ist.

Die bisherigen Ansätze zielen darauf ab, CR zwar geplant, aber nicht direkt erfragt zu erfassen. Jedoch können auch Selbstauskünfte zum Aufmerksamkeitsstatus beim Antworten als geplante Detektionsmethoden zur Aufdeckung von CR beitragen (z. B. Aust et al. 2013; Brühlmann et al. 2020; Curran 2016). Eine Möglichkeit der direkten Aufmerksamkeitsabfrage liegt in eigens für die Messung von CR konzipierten Multi-Item-Skalen (z. B. Huang et al. 2012; Maniaci und Rogge 2014; Meade und Craig 2012). Es resultiert ein Skalenwert, der das Ausmaß von CR während der Befragung quantifiziert. Eine subtile Abwandlung komplexer Aufmerksamkeitsskalen stellen Ehrlichkeitschecks dar. Typischerweise sollen die Ausfüllenden bei diesen selbst einschätzen, ob ihre gegebenen Antworten in den Analysen der Studie genutzt werden können (Abbey und Meloy 2017; Aust et al. 2013; Meade und Craig 2012). Konkret werden die Versuchspersonen beispielsweise gefragt: *„Ihrer ehrlichen Meinung nach: Sollten wir Ihre Daten aus dieser Studie in unseren Analysen nutzen?“* (Meade und Craig 2012, S. 442). Darauf wird mit „Ja“ oder „Nein“ geantwortet. Dieser Itemtyp wird in der Literatur als

⁹ Liegt der Studienfokus primär auf der Antwortqualität speziell von offenen Fragen, kann aus der grammatikalischen Korrektheit, ob Sätze komplett ausformuliert werden und der Frequenz der korrekten Antwort auf offene Fragen ein Globalindikator der Antwortqualität auf offene Fragen berechnet werden (Brühlmann et al. 2020).

„UseMe-Item“ bezeichnet und ist sehr etabliert (z. B. Brühlmann et al. 2020; Meade und Craig 2012). Brühlmann und Kolleg*innen (2020) versehen das UseMe-Item, als Voraussetzung für zutreffende Selbsteinschätzungen der Befragten, mit dem Hinweis, keinerlei negative Konsequenzen (z. B. Einbehaltung zugesagter Incentives) befürchten zu müssen, falls die Frage ehrlicherweise verneint wird.

3.3.3.3 Post-hoc Detektionsmethoden

Statt wie bei geplanten Detektionsmethoden Modifikationen des Fragebogens vorzunehmen, decken post-hoc Detektionsmethoden CR durch retrospektive, statistische Datenanalysen auf (z. B. Arndt et al. 2022; Brühlmann et al. 2020; Meade und Craig 2012). Post-hoc Detektionsmethoden suchen dafür nach Besonderheiten in Datensätzen, die im Zusammenhang mit CR auftreten. Einige dieser Indikatoren (z. B. die Antwortzeit oder Long Strings) korrespondieren unmittelbar mit klassischen Antwortqualitätsindikatoren und indizieren somit nicht ausschließlich CR, sondern verringerte Antwortqualität allgemein. Daher muss bei der Analyse der post-hoc Indikatoren immer berücksichtigt werden, welche CR-Form, die sich in den Datenmustern zeigen könnte (Speeding, monotones Antwortverhalten oder RR), jeweils adressiert wird.

Wird ein Fragebogen sehr schnell beantwortet, ist dies ein Indiz für besonders achtloses Antwortverhalten. Die Antwortzeit oder Response Time (RT) spiegelt daher wesentlich den Aufwand der Befragungsteilnehmer*innen bei der Beantwortung der Fragen wider (Wise und Kong 2005). Ein eindeutiger Rückschluss von der RT auf die Unaufmerksamkeit erfordert einen plausiblen und möglichst einheitlichen Cut-Off-Wert (Wise und Kong 2005). Huang und Kolleg*innen (2012) schlagen einen Cut-Off-Wert von zwei Sekunden pro beantwortetem Item (2s/Item) vor, der hohe Augenscheinvalidität besitzt, da es für Antwortzeiten, die unter diesen Schwellenwert fallen, schlicht nicht möglich ist, alle Items so schnell zu lesen und gewissenhaft zu beantworten. Der 2s/Item-Cut-Off-Wert wurde für die CR-Aufdeckung bereits empirisch validiert (z. B. Bowling et al. 2016; Bowling et al. 2021b; Curran 2016; Goldammer et al. 2020). Er sollte dabei nicht itemweise, sondern auf die aufsummierte Antwortzeit über alle Items einer Seite oder auf den gesamten Fragebogen angewendet werden (Bowling et al. 2021b; Ward und Meade 2023). RTs, die aufsummiert geringer als 2s/Item ausfallen, signalisieren dabei dann CR. Der Forschungsstand zeigt, dass sehr kurze Antwortzeiten recht verlässlich auf CR zurückgeführt werden können (z. B. Arthur et al. 2021; Huang et al. 2012; Huang et al. 2015a; Leiner 2019; Wang et al. 2018). Die RT lässt als CR-Maß dabei den eindeutigen Schluss auf Speeding als CR-Form zu.

Jüngere Erkenntnisse erweitern die Bedeutung der Antwortzeit als CR-Maß, da neben der sehr schnellen Beantwortung auch extrem lange Antwortzeiten für CR sprechen können, wenn die Proband*innen stark abgelenkt sind oder parallel anderen Aufgaben nachgehen (Read et al. 2022). Sehr lange RTs können aber nicht so eindeutig wie sehr kurze RTs auf CR attribuiert werden, da sie sowohl Ablenkung und Bearbeitungsunterbrechungen (CR) als auch äußerst intensive Auseinandersetzung mit den Inhalten (kein CR) signalisieren können (Read et al. 2022). Für lange RTs muss daher jeweils im Einzelfall entweder durch den Einbezug weiterer CR-Maße (z. B. kurze Antworten auf offene Fragen trotz langer RTs) oder durch die tiefere Auseinandersetzung mit der Struktur der längeren Antwortzeit die Signalwirkung für unaufmerksames Antworten beurteilt werden (z. B. kontinuierlich längere RTs für alle Fragen und Fragebogenseiten als Indiz für gewissenhafte längere Antwortzeit oder stark fluktuierende RTs in verschiedenen Befragungsteilen, die auf temporäre Ablenkung hinweisen könnten).

Andere post-hoc Detektionsmethoden ermitteln Unaufmerksamkeit über statistische Analysen des Datenmaterials und zielen eher auf die CR-Formen RR oder monotonen Antworten. Straightlining signalisiert beispielsweise mögliche Verringerungen der Aufmerksamkeit in Datensätzen (Lugtig und Toepoel 2016; Struminskaya et al. 2015) und ist ein Ausdruck der CR-Form monotonen Antwortens. Straightlining resultiert in „Long Strings“, bei denen mehrfach hintereinander dieselbe Antwortkategorie ausgewählt wird (z. B. Hong et al. 2020; Huang et al. 2012; Johnson 2005). Für jede Person kann die maximale Anzahl gleicher aufeinanderfolgender Antworten als maximale Stringlänge (Long String) erfasst werden. Je länger die Abfolge gleicher gewählter Antwortkategorien ist, desto höher ist auch die Wahrscheinlichkeit, dass unaufmerksames Antwortverhalten vorliegt. Die maximale Stringlänge kann in einem Diagramm auf der x -Achse gegen die Anzahl derjenigen Personen auf der y -Achse abgetragen werden, die im Datensatz diese maximale Stringlänge aufweisen. Um hier eine Kategorisierung der Unaufmerksamkeit vornehmen zu können, werden die entstehenden Punkte miteinander verbunden, und über ein der Faktorenanalyse entlehntes Scree-Verfahren (vgl. Cattell 1966) wird ein Knick ermittelt, der einen Cut-Off-Wert für die maximale Stringlänge liefert (Johnson 2005). Ist der maximale Long String einer Person größer als dieser Cut-Off-Wert, wird sie als unaufmerksam klassifiziert (Hong et al. 2020; Meade und Craig 2012). Für einzelne unidimensionale Skalen mit gleichgerichteten Items können Long Strings CR nur bedingt signalisieren, da aufmerksame Antworten hier korrekterweise zu einem monotonen Antwortmuster führen. Long Strings sollten daher möglichst skalenübergreifend über die Erfassung verschiedener Konstrukte hinweg ermittelt werden (Gibson und Bowling 2020). Da Straightlining vorrangig bei zunehmender Befragungsdauer auftritt (Herzog und Bachman 1981; Schonlau und Toepoel 2015; Zhang und

Conrad 2014), eignen sich Long Strings gerade gegen Ende längerer Fragebogen für die Aufdeckung von CR.

Der Nutzen von Long Strings ist auf streng monotone Antworten begrenzt, das heißt Proband*innen wählen dieselbe Antwortoption ohne Unterbrechung immer wieder aus. Ist das Antwortverhalten der Befragten zwar monoton, aber nicht absolut gleichförmig, mündet CR nicht zwingend in Long Strings. Monotones Antworten offenbart sich generell in der verringerten Varianz der individuellen Antworten von Proband*innen, wobei ein geringes Ausmaß an Differenziertheit in den Antworten geringeren Antwortaufwand repräsentiert (Kim et al. 2019). Ein flexiblerer Ansatz als Long Strings zur Aufdeckung monotonen Antwortens geht auf Marjanovic und Kolleg*innen (2015) zurück, die die „Inter-Item Standardabweichung“ (ISD) vorstellen. Die ISD kann als individueller Wert für jede Skala und jede*n Proband*in berechnet werden. Die Formel für den individuellen ISD-Wert lautet analog zur regulären Standardabweichung: $\sqrt{\frac{\sum_{j=1}^k (x_j - \bar{x}_i)^2}{(k-1)}}$. Dabei bezeichnet k die Anzahl der Items, auf die sich die Berechnung bezieht, x_j den Antwortwert, der bei den verschiedenen Items jeweils von der betrachteten Versuchsperson angegeben wird und \bar{x}_i den Mittelwert, den diese Versuchsperson über die betrachteten k Items erzielt. Für jedes Item wird die Differenz aus individuellem Itemwert x_j und individuellem Mittelwert \bar{x}_i berechnet, quadriert und über alle Items aufsummiert. Diese Summe wird durch die Anzahl der Items minus eins ($k-1$) geteilt und abschließend die Wurzel gezogen. Die ISD ist in ihrem Ursprungsgedanken auf positiv miteinander korrelierende Items begrenzt (Grable, Heo und Rabbani 2021), sodass hohe ISD-Werte lediglich RR anzeigen (Marjanovic et al. 2015). Die ISD-Berechnung findet dabei statt, nachdem etwaige negativ formulierte oder umgekehrte Items umkodiert werden (Marjanovic et al. 2015). Im Kontext von CR wird die ISD bei gleichbleibender Formel als „Individuelle Response Validität“ (IRV) bezeichnet und das Einsatzgebiet erweitert (z. B. Dunn et al. 2018; Hong et al. 2020). Sind Items nicht gleichförmig formuliert und ist dementsprechend monotoneres Antwortverhalten unplausibel, signalisieren geringe IRV-Werte CR. Andererseits deuten hohe IRV-Werte auf RR hin, wenn alle Items, auf die sich die Berechnung bezieht, in die gleiche Richtung formuliert sind und dasselbe Konstrukt erfassen, da hier starke Antwortvariationen unplausibel sind. Die IRV-Eignung variiert also situativ und kann abhängig von den Rahmenbedingungen jeweils eine bestimmte CR-Form (RR oder monotoneres Antworten) aufdecken (Dunn et al. 2018). Nur im speziellen Fall, dass positive Korrelationen zwischen Items erwartet werden, aber diese weder kongruent noch sehr ähnlich formuliert sind, können gleichzeitig RR (hohe IRV-Werte) und monotoneres Antworten (niedrige IRV-Werte) erkannt werden (Grable et al. 2021). Einheitliche Schwellenwerte für die CR-Klassifikation anhand des IRV existieren bisher nicht, die Randbereiche sehr

hoher bzw. sehr niedriger IRV-Werte gelten jedoch als starke Indizien für CR (z. B. Dunn et al. 2018; Grable et al. 2021; Marjanovic et al. 2015). Die IRV kann für jede einzelne Skala, wenn der Hintergrund adäquat einbezogen wird, aber auch skalenübergreifend oder sogar als Mittelwert aller IRV-Werte einer Person bei verschiedenen Skalen ermittelt werden (Marjanovic et al. 2015). Die Anwendung der IRV wird dabei für die letzten Items einer längeren Befragung empfohlen (Dunn et al. 2018). Dunn und Kolleg*innen (2018) fassen zusammen, dass die IRV vorteilhaft für die Aufdeckung von CR ist, da 1) sich auch Straightlining in sehr kleinen IRV-Werten offenbart, 2) die IRV nicht nur auf die Aufdeckung einer CR-Form begrenzt ist, 3) die Berechnung recht einfach vollzogen werden kann und 4) die IRV flexibel auf verschiedene Fragebogenbereiche sowie auch auf den gesamten Fragebogen angewendet werden kann.

Die retrospektive Datenaufbereitung kann neben der Suche nach statistischen Besonderheiten auch über die Antwortkonsistenz der Proband*innen auf CR schließen. Die Konsistenzprüfung der Antworten von Befragten auf individueller Ebene repräsentiert daher ein wesentliches Feld der post-hoc Detektionsmethoden von CR (z. B. Curran 2016; Huang et al. 2012; Meade und Craig 2012). Gerade längere Befragungen, die multiple Messungen einer Reihe von Konstrukten verwenden, sind für die Konsistenzprüfung des Antwortverhaltens prädestiniert (Schroeders et al. 2022). Bei aufmerksamem Antworten sollten Korrelationen zwischen bestimmten Items sehr hoch und in eine bestimmte Richtung gerichtet sein. Andernfalls signalisieren Antwortreihen Inkonsistenz und deuten auf CR hin. Der einfachste Fall der Konsistenzprüfung liegt vor, wenn alle Items einer Skala dasselbe Konstrukt erfassen und gleichgerichtet sind oder bereits umkodiert wurden. Solche unidimensionalen Skalen können in zwei Hälften aufgeteilt werden, wobei alle geraden Items die eine und alle ungeraden die andere Hälfte bilden (vgl. Jackson 1977; Johnson 2005). Im Anschluss wird die „Even-Odd-Konsistenz“ berechnet, indem auf individueller Ebene die Korrelation zwischen den Antworten der Skalenhälfte der geraden Items und den Antworten der Skalenhälfte ungerader Items als CR-Indikator abgeleitet wird (Brühlmann et al. 2020; Curran 2016; Johnson 2005; Maniaci und Rogge 2014; Meade und Craig 2012). Der Einsatz ist auf unidimensionale Skalen begrenzt und kann nur RR als CR-Form aufdecken. Da mittlerweile robustere Konsistenzansätze existieren, wird die Even-Odd-Konsistenz eher nicht als CR-Indikator empfohlen (Ward und Meade 2023).

Aufgrund ihres Inhalts sind für bestimmte Itempaare höhere Korrelationen bei aufmerksamem Antworten zu erwarten als für andere Itempaare. Unabhängig von der genauen Formulierung und Skalenverankerung sollten Items mit ähnlichem Inhalt (z. B. „Ich spiele gerne Basketball“ und „Basketball gehört zu den Sportarten, die ich am liebsten ausübe“) höher korrelieren als Items, die gegensätzliche oder grundverschiedene Aspekte (z. B. „Ich liebe Hunde“ und

„Ich esse gerne Eis“) adressieren. Für die CR-Aufdeckung lassen sich in diesem Kontext zwei Ansätze unterscheiden: der semantische und der psychometrische Ansatz (Curran 2016; DeSimone et al. 2015). In beiden Fällen sollten nicht ausschließlich gleichförmige Formulierungen, sondern auch umgekehrte Items genutzt und nicht umkodiert werden. Inhaltlich begründete Konsistenzerwartungen werden dann als semantische Synonyme und Antonyme bezeichnet, wobei aufgrund des Inhalts der Items sehr hohe positive („semantische Synonyme“) bzw. negative Korrelationen („semantische Antonyme“) erwartet werden (z. B. Curran 2016; DeSimone et al. 2015). Über verschiedene Itempaare hinweg wird konkret jeweils ein Element der jeweiligen Paare einem Vektor zugeordnet und das jeweils andere einem anderen, bevor die beiden Vektoren auf Personenebene miteinander korreliert werden (DeSimone et al. 2015). In vielen etablierten und validierten Skalen ist es gut möglich, entsprechende Synonyme und Antonyme zu finden, wenn die Konstrukte reflektiv erfasst werden. Für den Fall, dass der Fragebogenaufbau solche Itempaare nicht oder nur vereinzelt aufweist, ist auch der Einsatz eigenständiger Konsistenzskalen möglich (z. B. Maniaci und Rogge 2014; Pineseault 2007), wofür dann gestalterische Anpassungen der Befragung im Vorfeld der Studiendurchführung (wie bei den geplanten Detektionsmethoden) notwendig sind.

Die Eignung des semantischen Ansatzes zur CR-Aufdeckung ist jedoch begrenzt, da Antwortinkonsistenz bei der Mischung regulär und umgekehrt formulierter Items nicht zwingend unaufmerksames Antworten widerspiegelt (Kam et al. 2021) und beispielsweise auch die Darstellungsform der Items die Aussagekraft der Konsistenzprüfung determiniert (Edwards 2019). Alternativ können Antonyme und Synonyme datengetrieben abgeleitet werden. Dafür werden zunächst die Inter-Item-Korrelationen zwischen allen Items eines Fragebogens berechnet und diejenigen Itempaare identifiziert, die im Betrag die höchsten Korrelationen miteinander aufweisen. Als Schwellenwert der Korrelation zur Identifikation von geeigneten Itempaaren zur CR-Aufdeckung wird in der Literatur $r \geq |0,6|$ vorgeschlagen (z. B. Curran 2016; DeSimone et al. 2015; Maniaci und Rogge 2014; Meade und Craig 2012). Paare mit hohen positiven Korrelationen werden als „psychometrische Synonyme“ und Paare mit hohen negativen Korrelationen als „psychometrische Antonyme“ bezeichnet (z. B. Curran 2016; DeSimone et al. 2015; Hong et al. 2020; Huang et al. 2012; Meade und Craig 2012). Nach der Paarbildung erfolgt die Konsistenzprüfung auf Personenebene, indem für jede Versuchsperson jeweils alle Paare der Synonyme und der Antonyme auf jeweils zwei Vektoren aufgeteilt und diese miteinander korreliert werden. Hohe Korrelationen zwischen psychometrischen Synonymen und Antonymen sind bei unaufmerksamem Antworten unwahrscheinlich, sodass bei Personen mit geringen Korrelationswerten eher CR auftritt. Besonders offensichtliche Fälle von CR liegen vor, wenn die

individuellen Korrelationskoeffizienten für psychometrische Synonyme negativ und für Antonyme positiv ausfallen (z. B. Curran 2016; DeSimone et al. 2015; Hong et al. 2020). Sowohl RR als auch monotones Antworten können erkannt werden, da die Itempaare auch skalen- und seitenübergreifend gebildet werden und sich im Skalenformat unterscheiden können¹⁰.

Unaufmerksames Antwortverhalten in Befragungen entspricht optimalerweise nicht der Regel. Wenn die CR-Raten in einer Studie eher gering sind, können auch Ausreißeranalysen daher bei der CR-Aufdeckung helfen (Meade und Craig 2012). Ein komplexeres Ausreißermaß ist die Mahalanobis-Distanz, die die Abweichung eines Punktes vom Zentrum einer Verteilung im multivariaten Fall prüft (vgl. de Maesschalck, Jouan-Rimbaud und Massart 2000; Mahalanobis 1936). Die Mahalanobis-Distanz ermittelt für die Antwortreihe jeder Person im Verhältnis zur durchschnittlichen Antwortverteilung für alle Befragten einen Datenpunkt, sodass die Punktwolke der Punkte aller Personen die Verteilung der Antwortreihen in der Stichprobe darstellt. Konkret wird für jede Versuchsperson die quadrierte Differenz des Vektors ihrer individuellen Antworten vom Vektor der zugehörigen Mittelwerte über alle Befragten berechnet und über die Multiplikation mit der inversen Kovarianzmatrix standardisiert. Der letztliche Mahalanobis-Distanzwert gibt dann an, wie groß die Distanz des ermittelten Datenpunktes jeder einzelnen Person vom Zentroiden der Verteilung der Antwortreihen aller Personen ist. Eine ausführlichere Beschreibung des mathematischen Hintergrunds findet sich bei de Maesschalck und Kolleg*innen (2000). Höhere Mahalanobis-Werte markieren ungewöhnliche Antwortreihen und können ggf. CR signalisieren (z. B. Curran 2016; Hong et al. 2020). Die theoretische Fundierung, ob multivariate Ausreißer tatsächlich ein Ausdruck von CR sind, steht aber noch aus (Arias et al. 2020). Mitunter zeigt die Mahalanobis-Distanz vielversprechende Ansätze als CR-Indikator (Brower 2020, S. 45; Curran 2016; Meade und Craig 2012), während andere Studien keine klaren Vorteile für die CR-Aufdeckung feststellen (Baumgartner und Weijters 2022, S. 131-132). Da ungewöhnliches Verhalten auch abseits von CR entstehen kann und CR beispielsweise bei hohen Anteilen von Befragten, die unaufmerksam antworten, gar nicht ungewöhnlich ist, sollte die Mahalanobis-Distanz nicht exklusiv als CR-Maß verwendet werden.

Die post-hoc Detektionsmethoden von CR werden stetig weiterentwickelt und dabei nimmt die Komplexität der Maße zu. Diese fortgeschrittenen Methoden gehen über die Grenzen der vorliegenden Arbeit hinaus, da hier die Validierung teilweise noch nicht abgeschlossen und die Umsetzung so aufwendig ist, dass der Fokus der eigentlichen Fragestellung schnell verloren gehen kann. Beispiele fortgeschrittener statistischer CR-Maße sind *Person-Fit* Maße (vgl.

¹⁰ Ein letztes Konsistenzmaß, das grundsätzlich als CR-Indikator genutzt werden kann, sind Guttman-Skalen. Reviews zur Nutzung von Guttman-Skalen als CR-Maß finden sich bei Meijer, Niessen und Tendeiro (2016) und Niessen, Meijer und Tendeiro (2016).

Hong et al. 2020; Meijer, Niessen und Tendeiro 2016; Niessen, Meijer und Tendeiro 2016), Modellierungen anhand der Item-Response Theorie (vgl. Ulitzsch et al. 2022) oder der Machine Learning Ansatz des *Stochastic Gradient Boosting* (vgl. Schroeders et al. 2022).

3.3.3.4 Übertragung der CR-Aufdeckung auf die vorliegende Arbeit

Die verlässliche Analyse von möglichen Moduseffekten des Smartphones, die in Onlinebefragungen zu unaufmerksamem Antworten und dadurch zu ungenügender Antwortqualität führen könnten, erfordert die möglichst umfassende Aufdeckung von CR, da bisher kein valider Globalindikator existiert (z. B. Abbey und Meloy 2017; DeSimone et al. 2015; DeSimone und Harms 2018; Hong et al. 2020; Ward und Meade 2023). Der erste Schritt der Annäherung an mögliche Moduseinflüsse des Smartphones auf die Aufmerksamkeit liegt in der vorliegenden Arbeit in der physischen Betrachtung des Blickverhaltens über Eye-Tracking, wodurch die Aufmerksamkeit direkt gemessen wird. Losgelöst von etwaigen Kontexteinflüssen kann dies mögliche Aufmerksamkeitsunterschiede zwischen den Ausfüllgeräten PC/Laptop und Smartphone offenlegen. Eye-Tracking Daten lassen Rückschlüsse darauf zu, ob Befragungsinhalte gelesen und wahrgenommen werden und wie der Aufmerksamkeitsfokus der Proband*innen verteilt ist. Dies ermöglicht somit auch wichtige Erkenntnisse für CR (vgl. Brower 2020), allerdings ohne dass unmittelbare Rückschlüsse auf die Antwortqualität möglich sind. In der vorliegenden Arbeit wird das Blickverhalten daher zunächst geräteweise verglichen, um potenzielle gerätgebundene Ursachen für Antwortqualitätseinschränkungen betrachten zu können.

Da die Teilnahme an Onlinebefragungen in der Regel nicht unter Laborbedingungen, sondern in der Alltagssituation der Befragten erfolgt, stellt sich die Frage, ob die Eye-Tracking Befunde zu möglichen gerätspezifischen Unterschieden hinsichtlich der Aufmerksamkeitsprozesse auch auf das Feld übertragbar sind. In Alltagssituationen sind apparative Untersuchungen zur direkten Aufmerksamkeitsmessung nicht möglich, weshalb dort geplante und post-hoc Detektionsmethoden für die CR-Aufdeckung verwendet werden. Die Beurteilung der Aufmerksamkeit erfolgt über CR-Indikatoren zwar indirekt, gleichwohl ermöglichen diese aber klare Rückschlüsse auf die Antwortqualität. Dabei ist eine möglichst breite Auswahl verschiedener Methoden bei der Evaluation von CR sinnvoll (z. B. Abbey und Meloy 2017; Arthur et al. 2021; Berinsky et al. 2014; Hong et al. 2020; Schroeders et al. 2022), da einzelne Indikatoren teilweise auch nur einzelne CR-Formen verlässlich aufdecken können (z. B. Credé 2010; Curran 2016; Dunn et al. 2018; Schroeders et al. 2022). Werden verschiedene CR-Indikatoren gemeinsam verwendet, erhöht dies die Akkuratessse der CR-Aufdeckung (z. B. Abbey und Meloy 2017; Arndt et al. 2022; Berinsky et al. 2014; Curran 2016; Höhne und Schlosser 2018; Meade und

Craig 2012). Selbst wenn CR-Indikatoren stark miteinander zusammenhängen, fördern sie doch stets auch jeweils inkrementell die Konvergenzvalidität der CR-Aufdeckung (Brühlmann et al. 2020; Huang et al. 2012). Insbesondere wird empfohlen, Methoden der geplanten Detektion mit denen der post-hoc Detektion zu kombinieren (Ward und Meade 2023). Zusätzlich sind einige Indikatoren auch für anderes problematisches Antwortverhalten über CR hinaus (z. B. Satisficing-Tendenzen verschiedenen Ausmaßes oder Antwortstile) sensibel (Arias et al. 2020).

Allerdings ist es nicht praktikabel und vor allem ineffizient, alle möglichen CR-Indikatoren zur Anwendung zu bringen, da dies den Aufwand unnütz erhöht (Hong et al. 2020; Huang et al. 2015a). Mit der Anzahl an CR-Indikatoren steigt zudem die Gefahr falsch positiver CR-Klassifikationen (DeSimone und Harms 2018; Goldammer et al. 2020). Auch für die vorliegende Arbeit ist daher eine Vorauswahl der Indikatoren für die Feldstudie angezeigt. Die Vorauswahl sollte, wie ausgeführt, mehrere Kriterien erfüllen, um die zugrundeliegende Fragestellung adäquat zu untersuchen: 1) Es sollte nicht nur ein einzelner CR-Indikatoren einbezogen werden; 2) Die Indikatoren sollten CR in seiner Breite (über verschiedene CR-Formen hinweg) aufdecken (siehe Kapitel 3.3.1.2); 3) Die CR-Indikatoren sollten aussagekräftig sein; 4) Die Indikatoren sollten möglichst einfach implementierbar sein und die Befragung nicht überfrachten und 5) Die Überprüfung sollte für die Befragten nicht zu offensichtlich angelegt sein.

Mit Blick auf die Befundlage und in Einklang mit den definierten Auswahlkriterien eignen sich als CR-Maße im Feld insbesondere: CR-Selbstberichte in Form von 1) Aufmerksamkeitskalen (Arthur et al. 2021; Edwards 2019) oder 2) UseMe-Items (Aust et al. 2013; Baumgartner und Weijters 2019, S. 301; Brühlmann et al. 2020; Meade und Craig 2012), 3) Ausfüllinstruktionen wie IRIs oder IMCs (Baumgartner und Weijters 2019, S. 301; Kam und Chan 2018; Paas und Morren 2018), 4) Bogus-Items (Brühlmann et al. 2020; Credé 2010; Huang et al. 2015a), 5) offene Fragen und 6) gezielte Rückfragen (Brühlmann et al. 2020), 7) Long Strings (Curran 2016; DeSimone und Harms 2018; Niessen et al. 2016) und 8) die IRV (Curran 2016; Dunn et al. 2018; Huang et al. 2012), 9) Konsistenzprüfungen über Synonyme und Antonyme (Curran 2016; Huang et al. 2012; Meade und Craig 2012), 10) sehr kurze Antwortzeiten (Brower 2020, S. 67; Huang et al. 2012; Leiner 2019; Meade und Craig 2012; Niessen et al. 2016; Wang und Xu 2015) und 11) die Mahalanobis-Distanz (Curran 2016; Meade und Craig 2012).

Die vorliegende Arbeit folgt diesen Literaturempfehlungen für die CR-Aufdeckung in der Feldstudie der Hauptuntersuchungen größtenteils und nutzt die genannten CR-Indikatoren. Anhand der definierten Auswahlkriterien wird geprüft, ob gewisse Eingrenzungen möglich bzw. Erweiterungen notwendig sind. Dies führt dazu, dass nicht alle von der Literatur vorgeschlagenen Indikatoren zur Anwendung gebracht werden, sondern die Vorauswahl punktuell weiter

eingegrenzt wird. Da CR den klaren Forschungsfokus darstellt, werden aber gewisse Überschneidungen in den Methoden der CR-Aufdeckung bewusst toleriert und bei der Auswertung berücksichtigt. Bei den Ausfüllinstruktionen versprechen IMCs aufgrund der erhöhten Komplexität sensitivere Aufdeckungsraten als IRIs (Clifford und Jerit 2015; Morren und Paas 2020). Der zusätzliche Erkenntnisgewinn durch IRIs über den IMC und andere Indikatoren hinaus ist hingegen sehr gering (Kam und Chan 2018; Paas und Morren 2018), sodass die vorliegende Arbeit lediglich den IMC integriert. Die verwendeten Indikatoren des Infrequency-Ansatzes (der IMC und Bogus-Items) werden entsprechend der Literaturempfehlungen möglichst diskret und thematisch passend in die Befragung einbezogen, um Reaktanz der Befragten zu vermeiden (siehe Kapitel 3.3.3.2). Als Konsistenzmaß sind psychometrische Synonyme und Antonyme für die vorliegende Arbeit zweckmäßig, da sie weniger planerischen Zusatzaufwand als ihre semantischen Pendanten erfordern und sowohl skalenübergreifend als auch skalenspezifisch unabhängig von der Skalenpolarität geeignet sind (Ward und Meade 2023). Neben den Indikatoren aus den Literaturempfehlungen kann gerade bei längeren Befragungen die unangekündigte Skalenumkehr (siehe Kapitel 3.3.3.2) zusätzliche Hinweise für CR liefern und wird ebenfalls integriert. *Tabelle 3* listet die Indikatoren auf, die für die CR-Aufdeckung im Feld in der vorliegenden Arbeit letztlich verwendet werden. Die eingesetzten Indikatoren umspannen sowohl dichotome als auch kontinuierliche CR-Maße für die CR-Klassifikation, was eine möglichst ganzheitliche CR-Aufdeckung sowie die kleinschrittige statistische Prüfung ermöglicht (z. B. Arthur et al. 2021; Edwards 2019; Goldammer et al. 2020; Ward und Meade 2023). Für den Großteil dieser Aufmerksamkeitsmaße (insbesondere Bogus-Items, IMC, IRV, Long Strings, Psychometrische Synonyme und Antonyme) blieb bisher noch nahezu ungeprüft, ob CR häufiger

Aufdeckung von Careless Responding im Feld im Forschungsprogramm

Geplante Detektionsmethoden	Post-hoc Detektionsmethoden
Infrequency: <u>Bogus-Items</u> Instruktions- Manipulationscheck (IMC)	Konsistenz: <u>Psychometrische Synonyme und Antonyme</u> Umgekehrte Skalierung
Gezielte Rückfragen zu Inhalten	Antwortzeit (RT)
Offene Fragen	Long Strings
Selbstbericht: <u>Use-Me Item</u> Aufmerksamkeitsskala	Individuelle Response Validität (IRV) Mahalanobis-Distanz

Tabelle 3. Aufdeckung von Careless Responding in der Studie 4 der vorliegenden Arbeit

auftritt, wenn Smartphones als Ausfüllgeräte genutzt werden. Die vorliegende Arbeit erweitert damit die Erkenntnislage (siehe Kapitel 3.2.3 und nachfolgend 3.3.4) zum potenziellen Einfluss des Smartphones auf die Antwortqualität und somit auf den TSE wesentlich.

Das nachfolgende Unterkapitel betrachtet in Vorbereitung des Forschungsprogramms der Hauptuntersuchungen die mögliche Rolle, die das Smartphone als Ausfüllgerät in Onlinebefragungen für CR einnehmen könnte. Dafür wird die Forschungsliteratur zum Arbeitsgedächtnis, das die Aufmerksamkeit im kognitiven System steuert, aufbereitet und herausgearbeitet, ob und inwiefern mögliche aufmerksamkeitsrelevante Einflüsse des Smartphones existieren könnten. Zudem wird der bisherige Forschungsstand zu CR mit Blick auf das Smartphone als Ausfüllgerät in Onlinebefragungen strukturiert. Die Herleitung der resultierenden Forschungshypothesen, an deren Prüfung in der vorliegenden Arbeit die empirischen Untersuchungen möglicher Moduseffekte des Smartphones als Ausfüllgerät hinsichtlich der Antwortqualität ausgerichtet werden, vervollständigt das Kapitel.

3.3.4 Das Smartphone als Hintergrundfaktor von Careless Responding

Wie in Kapitel 3.2.1 ausgeführt, verringert insbesondere Unaufmerksamkeit als Folge fehlender Motivation – Careless Responding – die Antwortqualität in Onlinebefragungen. Das folgende Kapitel erkundet, ob CR womöglich am Smartphone häufiger auftreten könnte, wodurch Moduseffekte hinsichtlich der Antwortqualität entstehen würden. Dafür wird zunächst (Kapitel 3.3.4.1) das menschliche Arbeitsgedächtnis vorgestellt, das die Grundlage der menschlichen Aufmerksamkeit im kognitiven System darstellt (z. B. Oberauer 2019). Im Anschluss werden die Hypothesen für die Hauptuntersuchungen der vorliegenden Arbeit hergeleitet, indem die Besonderheiten des Smartphones als Ausfüllgerät in Onlinebefragungen und die Erkenntnisse zu CR in Onlinebefragungen aufeinander bezogen werden (Kapitel 3.3.4.2).

3.3.4.1 Das Arbeitsgedächtnis als Grundlage der Aufmerksamkeit

Das menschliche Arbeitsgedächtnis ist der Teil des Gedächtnisses, in dem aufgenommene Informationen initial verarbeitet und kurzzeitig gespeichert werden. Es besteht aus vier wesentlichen Substrukturen: 1) der Phonologischen Schleife, 2) dem Visuell-Räumlichen Notizblock, 3) dem Episodischen Puffer und 4) der Zentralen Exekutive (vgl. Baddeley 1992, 2003; Baddeley und Hitch 1974). Diese Substrukturen steuern alle Prozesse, durch die im kognitiven System mentale Repräsentationen gebildet werden, und die die Verarbeitung und letztlich den Abruf von Informationen vorbereiten (Oberauer 2019). Die Phonologische Schleife speichert und verarbeitet akustische sprachliche Reize und legt aufgenommene Informationen als Phoneme

ab (Baddeley 1992, 2003). Visuelle Informationen gelangen in den Visuell-Räumlichen Notizblock (Baddeley 1992, 2003). Informationen, die die Phonologische Schleife oder der Visuell-Räumliche Notizblock aufnehmen, müssen aktiv angesteuert werden, damit das kognitive System sie verarbeiten und verorten kann. Die Zentrale Exekutive steuert diese Informationsverarbeitung und entscheidet, welchen Reizen Aufmerksamkeit zugeteilt wird und wie intensiv diese verarbeitet werden (Baddeley 1992, 2003). Die aufgenommenen akustischen und visuellen Informationen zerfallen dabei jeweils sehr schnell, wenn sie nicht von der Zentralen Exekutive – durch sogenanntes Rehearsal – aktiv aufrechterhalten werden (Baddeley 2003). Die Informationsaufnahme visueller und akustischer Reize ist voneinander größtenteils unabhängig und beeinflusst sich gegenseitig nur schwach (z. B. Baddeley 2003; Logie, Zucco und Baddeley 1990). Da in bestimmten Fällen auch multisensorische Informationen in Form von Episoden gespeichert werden, bildet der Episodische Puffer explizit für multisensorische Inhalte eine zusätzliche Komponente des Arbeitsgedächtnisses (Baddeley 2000, 2003). In Onlinebefragungen liegt der unmittelbare Verarbeitungsfokus, solange keine auditiven Inhalte integriert werden, auf visuellen Informationen, sodass das visuelle Arbeitsgedächtnis die Grundlage der Analyse von Aufmerksamkeitsmechanismen in Onlinebefragungen und daraus folgenden Konsequenzen für die Antwortqualität im TSE-Framework darstellt. Die phonologische Schleife und der episodische Puffer sind für die vorliegende Arbeit daher nur von untergeordneter Bedeutung.

Das visuelle System kann initial eine große Anzahl an Informationen aufnehmen und sensorisch abbilden (Phillips 1974). Der visuelle Teil des Arbeitsgedächtnisses strukturiert, welche dieser Informationen verarbeitet und für den Gedächtnisabruf oder zur Bearbeitung laufender Aufgaben genutzt werden können (Luck und Vogel 2013). Aufmerksamkeitsstrukturen werden hier offensichtlich, da nur sensorische Informationen, die nicht zerfallen, sondern durch Rehearsal aufrechterhalten werden, Aufmerksamkeit erhalten und folglich tatsächlich verarbeitet werden (Oberauer 2019). Für die Aufmerksamkeitsallokation entscheidet das Arbeitsgedächtnis situativ, welche Elemente aus der Informationsvielfalt herausgefiltert und verarbeitet werden. Gerade für Textelemente als Hauptbestandteil von Onlinebefragungen sind Aufmerksamkeitsentscheidungen notwendig: Die perzeptuelle Spanne, die den Wahrnehmungsraum für Textinhalte beschreibt, ist räumlich begrenzt und stellt gleichzeitig die Grundlage der Informationsverarbeitung im Arbeitsgedächtnis dar (Daneman und Carpenter 1980). Die generelle Kapazität des Arbeitsgedächtnisses für sensorische Informationen ist ebenfalls limitiert, sodass nur eine begrenzte Anzahl aufgenommener Informationen gleichzeitig verarbeitet werden kann (z. B. Alvarez und Cavanagh 2004; Baddeley 2003). Diese Aufmerksamkeitsbegrenzung könnte im kognitiven System entweder als starrer Selektionsmechanismus oder als begrenzte

flexible Ressource angelegt sein, wobei in der Forschungsliteratur eine klare Einigung auf eines der Modelle noch aussteht (Oberauer 2019). Der Selektionsansatz konzeptualisiert die Aufmerksamkeitsverteilung als diskrete Slots, wobei die Informationsaufnahme nicht mehr möglich ist, sobald alle Slots gefüllt sind (Luck und Vogel 2013; Oberauer 2019). Der Ressourcenansatz beschreibt Aufmerksamkeit hingegen als dynamischen Prozess: Die Verarbeitung und die Speicherung von Informationen sind in diesem voneinander entkoppelt, sodass die Quantität der Arbeitsgedächtnisinhalte situativ variieren kann (Ma, Husain und Bays 2014; Oberauer 2019). Sind die Informationen sehr komplex, ist die Gesamtzahl der gleichzeitig verarbeiteten Inhalte geringer. Andererseits kann die Aufmerksamkeitsressource bei trivialen Informationen eine größere Zahl von Inhalten parallel adressieren. Für die vorliegende Arbeit wird festgehalten, dass der Umfang der visuellen Informationen, die das Arbeitsgedächtnis verarbeiten kann, limitiert wird durch die Anzahl und die Komplexität der Informationen (Alvarez und Cavanagh 2004; Drew, Boettcher und Wolfe 2017).

3.3.4.2 Herleitung der Hypothesen für die Hauptuntersuchungen

Da an Onlinebefragungen mit technischen Geräten teilgenommen wird, muss die Aufmerksamkeitsallokation am Bildschirm stattfinden, damit Teilnehmer*innen die Fragebogenbeantwortung aufmerksam ausführen können. Aufgrund seines relativ kleinen Displays könnte das Smartphone für die Aufmerksamkeitsbindung nachteilig sein. Größere Elemente (wie der Bildschirm eines PCs/Laptops) erhalten grundsätzlich eher mehr Aufmerksamkeit als kleinere Elemente (Orquin und Mueller Loose 2013). Zusätzlich fällt an größeren Bildschirmen denselben Inhalten mehr Aufmerksamkeit zu als an kleineren Bildschirmen (Reeves et al. 1999) und Personen benötigen an kleineren Bildschirmen mehr Zeit für die Verarbeitung derselben Inhalte (Kim et al. 2015). Durch die geringe Screengröße des Smartphones könnten in Onlinebefragungen also Unterschiede in den Aufmerksamkeitsstrukturen zwischen dem Smartphone und dem PC/Laptop entstehen. Ein genereller Nachteil des Smartphones als Ausfüllgerät kann jedoch bisher nicht eindeutig gefolgert werden, da die geringe Bildschirmgröße nicht pauschal die Aufmerksamkeit reduziert (Kim et al. 2016). Die Aufmerksamkeit schweift an kleineren Screens zwar potenziell schneller ab, kann durch Schlüsselreize (z. B. emotionale Reize mit Aufgabenbezug) aber auch nachhaltig gebunden werden (Ravaja et al. 2004).

Abseits der geringen Bildschirmgröße ist das Smartphone als Ausfüllgerät auch aufgrund seiner persönlichen Bedeutung für Aufmerksamkeitseinschränkungen prädestiniert. Wie in Kapitel 3.2.2 beschrieben, liegt in den Nutzungsmustern und der psychologischen Bedeutung eine Ursache für mögliche Unterschiede in der Antwortqualität. Mit dem Smartphone assoziierte

Aufmerksamkeitsprozesse sind sprunghafter, die Nutzungsphasen kürzer (Dunaway et al. 2018; Molyneux 2018) und auch die Aufmerksamkeitsspanne ist demnach potenziell verkürzt. Hedonistische Nutzungsziele haben am Smartphone einen hohen Stellenwert (siehe Kapitel 2.3), so dass hier die Tendenz zur Vermeidung von Aufwand bei der Befragungsteilnahme größer sein könnte. Smartphones sind zudem sehr persönliche und private Gegenstände und besitzen für ihre Nutzer*innen einen hohen emotionalen Wert (siehe Kapitel 2.3.2). Als Ausfüllgeräte könnten sie daher andere Verarbeitungsstrukturen anregen als PCs/Laptops (Tourangeau et al. 2017).

Moderne Eye-Tracking Systeme können die kognitive Verarbeitung von Bildschirmhalten im Rahmen von Onlinebefragungen recht unkompliziert abbilden (siehe Kapitel 3.3.3.1) und somit konkret für die Untersuchung der Aufmerksamkeit während der Smartphonennutzung herangezogen werden (Kim et al. 2016; Lagun et al. 2014, S. 122). Für die Untersuchung möglicher Moduseffekte des Smartphones auf die Antwortqualität wird zunächst Grundlagenforschung betrieben und analysiert, ob sich Aufmerksamkeitsstrukturen in Onlinebefragungen als wesentliche Grundlage der Antwortqualität womöglich gerätspezifisch unterscheiden und ob die Befragten ihre Aufmerksamkeit gerätweise unterschiedlich häufig auf den Bildschirm und die Befragungsinhalte richten. Die Besonderheiten des Smartphones und die Herausforderungen während der Befragungsteilnahme könnten dazu führen, dass sich Befragte weniger stark mit den Befragungsinhalten auseinandersetzen und schneller abgelenkt werden als am PC/Laptop. Diese grundsätzliche Anfälligkeit für geringere Aufmerksamkeit bei der Smartphonennutzung führt zu folgender Hypothese:

H3: *Proband*innen, die an Onlinebefragungen mit dem Smartphone teilnehmen, blicken häufiger von den Befragungsinhalten weg als Proband*innen, die mit dem PC/Laptop teilnehmen.*

Anhand der Hypothese **H3** kann eruiert werden, ob das Smartphone womöglich tatsächlich ein Hintergrundfaktor von Careless Responding sein könnte und somit Moduseffekte verringerter Antwortqualität am Smartphone wahrscheinlicher sind. In der Eye-Tracking Untersuchung werden bei der direkten Aufmerksamkeitsmessung zunächst die Geräteeffekte isoliert und von möglichen Kontexteffekten getrennt. CR-Tendenzen selbst sind hingegen stark kontext- und situationsabhängig (siehe Kapitel 3.3.2.1.) und am Smartphone sind die Kontexteffekte in Onlinebefragungen diverser als am PC/Laptop (z. B. Antoun et al. 2017; de Bruijne und Oudejans 2015, S. 141; Höhne und Schlosser 2019; Sendelbah et al. 2016), sodass Aufmerksamkeitsverluste schon aufgrund des Ausfüllkontexts wahrscheinlicher werden. Die Erkenntnisse aus den Analysen des Blickverhaltens und daraus folgende Konsequenzen für mögliche gerätbezogene Aufmerksamkeitsbesonderheiten müssen daher zusätzlich ins Feld übertragen werden. In der entsprechenden Feldstudie sind dann die Teilnahmebedingungen mitzuerheben, um

Kontexteinflüsse der Smartphoneteilnahme und mögliche Moduseffekte des Smartphones als Ausfüllgerät, die dem Gerät selbst entstammen, abgrenzen zu können. Dadurch können explizit mögliche Einflüsse des Smartphones als Ausfüllgerät auf CR adressiert werden, die eine klarere Signalwirkung für die Antwortqualität hätten als Eye-Tracking Daten.

Neben der grundsätzlichen Anfälligkeit des Smartphones für unaufmerksames Antworten in Onlinebefragungen, lassen sich drei weitere befragungsspezifische Hauptbereiche ausmachen, durch die Moduseffekte des Smartphones entstehen könnten, die CR womöglich begünstigen: 1) Die Wechselwirkungen von Komplexität und kognitiven Herausforderungen am Smartphone, 2) Multi-Tasking und 3) Die Wirkung von Ablenkung am Smartphone. Der Lesefluss ist bei komplexen Inhalten und höherer Belastung grundsätzlich gehemmt (Caplan und Waters 1999). In Befragungen schweifen Proband*innen häufiger ab, wenn die kognitive Beanspruchung größer wird (Ansolabehere und Schaffner 2015). Die Leistung und Aufmerksamkeitskapazität von Smartphone-Besitzer*innen wird, zumindest bei komplexeren Aufgaben, bereits durch die bloße Präsenz des Smartphones reduziert (Thornton et al. 2014), was sich auf die aktive Nutzung als Ausfüllgerät übertragen könnte. Auch dadurch, dass dieselben Inhalte am Smartphone ein höheres Maß an kognitiver Beanspruchung mit sich bringen als am PC/Laptop (Ghose et al. 2013), könnte die Teilnahme an Onlinebefragungen mit dem Smartphone herausfordernder erlebt werden (Wenz 2021b). Gerade größere kognitive Herausforderung und empfundene hohe Komplexität erhöhen zugleich die Gefahr für unaufmerksames Antwortverhalten (Kam und Meyer 2015). Die Nutzung von Smartphones in Onlinebefragungen könnte also für CR kritisch sein, da die Proband*innen hier Inhalte vergleichsweise komplexer wahrnehmen können und die kognitive Beanspruchung höher sein kann.

Multi-Tasking tritt auf, wenn mehrere Aufgaben parallel ausgeführt werden (Carrier et al. 2009; Spink, Cole und Waller 2008). Die fortschreitende technische Entwicklung ermöglicht es Nutzer*innen elektronischer Geräte, vielfältige verschiedene Aufgaben in sehr diversen Settings auszuführen (Spink et al. 2008). Auch in selbst-administrierten Onlinebefragungen kommt es während der Teilnahme durchaus zu Multi-Tasking, wenn Befragte zeitgleich noch weiteren Aktivitäten nachgehen (Schober et al. 2015). Multi-Tasking schmälert die Leistungsfähigkeit (Adler und Benbunan-Fich 2012) und führt in Befragungen zu längeren Antwortzeiten (Antoun und Cernat 2020). Das Smartphone als Ausfüllgerät steigert noch zusätzlich die Anfälligkeit für Multi-Tasking aufgrund seiner Portabilität und der Bandbreite an Tätigkeiten, denen die Befragten parallel zu Befragungsteilnahme nachgehen können. Multi-Tasking kann dabei sowohl am Smartphone selbst (gerätgebunden) als auch abseits des Smartphones durch die Nutzung anderer Geräte oder durch analoge Umgebungseinflüsse (gerätgesondert) entstehen

(z. B. Sendelbah et al. 2016; Spink et al. 2008). Wird die Befragung kurz verlassen, um eine andere Aufgabe zu absolvieren, sinkt die Aufmerksamkeit unmittelbar und die Teilnahmedauer verlängert sich (z. B. Höhne und Schlosser 2018; Zwarun und Hall 2014). Multi-Tasking tritt häufiger bei jüngeren und gebildeteren Personen auf (Höhne et al. 2020b; Zwarun und Hall 2014) und genau diese Personengruppen nehmen an Onlinebefragungen vorrangig mit dem Smartphone teil (siehe Kapitel 3.1.2). Die Befundlage zu Multi-Tasking während der Teilnahme an Onlinebefragungen ist dabei kontrovers: Vereinzelt tritt mehr Multi-Tasking am Smartphone auf, gerade wenn auch Dritte anwesend sind (Antoun et al. 2017). Allerdings finden andere Studien keine Unterschiede im Multi-Tasking je nach Ausfüllgerät (Antoun und Cernat 2020; Revilla und Couper 2018). Die Gefahr für gerätegebundenes Multi-Tasking scheint mitunter sogar am PC/Laptop höher zu sein (Höhne et al. 2020b).

Ablenkung während der Befragungsteilnahme geht häufig mit einer Verringerung des Aufwands im Antwortprozess einher (z. B. Höhne und Schlosser 2018; Sendelbah et al. 2016; Zwarun und Hall 2014). Die Effizienz der Verarbeitung leidet besonders dann, wenn Ablenkungsreize und Zielreize sich ähneln (Rayner 1998). Insbesondere unruhige Umgebungen wirken ablenkend und erschweren die aufmerksame und gewissenhafte Teilnahme (Wenz 2021b). Am Smartphone sind Ablenkungen während der Befragungsteilnahme wahrscheinlicher (siehe Kapitel 3.1.1 und 3.1.2) und treten auch häufiger auf (z. B. Ansolabehere und Schaffner 2015; de Bruijne und Oudejans 2015, S. 141; Höhne und Schlosser 2019; Wenz 2021b; Zwarun und Hall 2014). Solche Umgebungsreize können dabei je nach Erhebungsbedingung unterschiedlich auf die Proband*innen wirken (Schober et al. 2015). Moduseffekte des Smartphones als Ausfüllgerät würden hier entstehen, wenn die Ablenkungsreize am Smartphone in anderer Form als am PC/Laptop auf die Aufmerksamkeit wirken und beispielsweise CR auch bei gleicher Ablenkung am Smartphone häufiger ist. Für eine größere CR-Anfälligkeit des Smartphones bei Ablenkung spricht etwa, dass Smartphone-Teilnehmer*innen weniger fokussiert und konzentriert sind (de Bruijne und Oudejans 2015, S. 138). Ob sich Ablenkungen tatsächlich gerätespezifisch unterschiedlich auswirken und so Moduseffekte entstehen, die auf die Gerätesonderheiten zurückzuführen sind, ist dabei noch weitestgehend unklar (Wenz 2021b). Dies soll im Rahmen der vorliegenden Arbeit gezielt analysiert werden, indem Kontexteffekte der Teilnahme und potenziell vorhandene Geräteeffekte möglichst getrennt werden.

Der Forschungsstand verdeutlicht, dass sich bei der Teilnahme mit dem Smartphone Unterschiede für verschiedene Antwortqualitätsindikatoren zeigen (siehe Kapitel 3.2.3, insbesondere *Tabelle 2*), die auch ein Ausdruck von CR sein können (siehe Kapitel 3.3.3). Allerdings zeigen sich auch größere Forschungslücken: Während Moduseffekte des Smartphones schon in Bezug

auf Satisficing bisher nur rudimentär untersucht wurden (siehe Kapitel 3.2.3), ist der Erkenntnisstand für CR und verschiedene CR-Formen noch vager. Die Untersuchung möglicher Moduseffekte des Smartphones für Careless Responding erweitert den Forschungsstand daher wesentlich. Obwohl mögliche Einflüsse des Smartphones auf CR aufgrund der Gerätebesonderheiten und der Hintergründe des Entstehens von CR durchaus plausibel sind (siehe Kapitel 3.3.4), wurde eine Verbindung zwischen dem Smartphone und unaufmerksamem Antworten bisher nicht hergestellt. Die einzige Ausnahme bildet eine aktuelle Studie von Magraw-Mickelson, Wang und Gollwitzer (2022), die zu dem Ergebnis kommt, dass CR am Smartphone eher nicht häufiger auftritt. Diese nutzt aber relativ kleine Stichproben ($n < 120$) und begrenzt die Auswertungen darauf, dass alle Befragten eindeutig als unaufmerksam klassifiziert werden, ohne etwa verschiedene CR-Formen differenziert zu berücksichtigen. Für die robuste Abschätzung möglicher Moduseffekte hinsichtlich der Antwortqualität müssen die Erkenntnisse weiter vertieft und verschiedene Ausmaße und Konsequenzen der Antwortqualitätsreduktion adäquat differenziert werden. Statt eine Globalklassifikation für CR vorzunehmen, ist es sinnvoll mehrere CR-Indikatoren geräteweise zu vergleichen. Anknüpfend an die **H3** lassen die Hintergründe der Smartphone-nutzung (siehe Kapitel 2) und die Ausführungen des theoretischen Teils der Arbeit (siehe Kapitel 3) für die Hauptuntersuchungen zu möglichen Moduseffekten des Smartphones hinsichtlich der Antwortqualität vermuten, dass Careless Responding gerade am Smartphone vermehrt auftreten könnte. Dies führt zur Formulierung der Hypothese:

H4: *Proband*innen, die mit dem Smartphone teilnehmen, zeigen in Onlinebefragungen häufiger Careless Responding als Proband*innen, die mit dem PC/Laptop teilnehmen.*

CR zielt explizit auf die wesentliche Grundlage der Antwortqualität im TSE-Framework – die Aufmerksamkeit (siehe Kapitel 3.2.1). Die Ausführungen des Kapitels 3.3 demonstrieren, dass sich dafür zur Abschätzung möglicher Moduseffekte des Smartphones CR als Quantifizierungsmaß des Messfehlers durch Befragte anbietet. Die Analysen werden zudem um die Einflüsse des Smartphones als Ausfüllgerät auf die Item-Nonresponse erweitert. Sollten sich trotz seiner Gerätanfälligkeit am Smartphone keine höheren CR-Ausmaße zeigen, würde dies eher gegen Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen hinsichtlich der Antwortqualität sprechen. Entstehen durch das Smartphone jedoch (zumindest teilweise) Moduseffekte dahingehend, dass CR am Smartphone häufiger auftritt, sollte die Untersuchung der Zusammenhänge in künftigen Studien vertieft werden. Wenn Smartphones als Ausfüllgeräte in Onlinebefragungen zu Anwendung kommen, müsste dann darauf umsichtig reagiert werden, um CR weitestmöglich vorzubeugen.

4 Forschungsprogramm

Die vorliegende Arbeit prüft, ob das Smartphone als Ausfüllgerät in Onlinebefragungen Moduseffekte auf relevante Komponenten des TSE ausübt. *Tabelle 4* bündelt die erarbeiteten übergeordneten Forschungshypothesen und listet auf, in welchen Kapiteln die Herleitung jeweils nachvollzogen werden kann. Zudem wird das Unterkapitel des Forschungsprogramms referenziert, das die Hypothesen jeweils ausdifferenziert und die empirische Prüfung vorbereitet.

Das Forschungsprogramm der vorliegenden Arbeit zielt darauf, aussagekräftige Erkenntnisse zu etwaigen Moduseffekten des Smartphones zu sammeln. Dafür ergeben sich einige Vorgaben für die Konzeption der zugehörigen Studien. Zunächst sollten Onlinebefragungen generell vertrauenswürdig sein, den Befragten alle relevanten Informationen bereitstellen und den Aufwand für Proband*innen auf ein vertretbares Maß begrenzen (vgl. Reips 2000). Die Onlinebefragungen in der vorliegenden Arbeit müssen zudem ausdrücklich berücksichtigen, dass auch Smartphones als Ausfüllgeräte genutzt werden. Insbesondere die geringe Bildschirmgröße von Smartphones (siehe auch Kapitel 2.2.4 und 3.3.4) konfrontiert Befragte in Onlinebefragungen mit erheblichen Herausforderungen (z. B. Duchnicky und Kolers 1983; Jones et al. 1999; Krämer 2017, S. 95; Toninelli und Revilla 2020, S. 361). Das Forschungsprogramm zielt daher auf die bestmögliche Gestaltung von Onlinebefragungen für den Smartphoneeinsatz, verbunden

Forschungshypothese	Herleitung	Forschungsprogramm
H1: Personen, die an einer Onlinebefragung mit dem eigenen Smartphone teilnehmen, beantworten sensible Fragen aufrichtiger als Personen, die unter gleichen Bedingungen mit dem eigenen PC/Laptop teilnehmen.	➤ Kapitel 3.1 ➤ Kapitel 3.1.1	Kapitel 4.1
H2: Proband*innen, entscheiden sich eher für die Teilnahme mit dem Smartphone und gegen die Teilnahme mit dem PC/Laptop, wenn ihnen bestimmte Bedürfnisse für die Befragungsteilnahme wichtig sind.	➤ Kapitel 3.1 ➤ Kapitel 3.1.2	Kapitel 4.2
H3: Proband*innen, die an Onlinebefragungen mit dem Smartphone teilnehmen, blicken häufiger von den Befragungsinhalten weg als Proband*innen, die mit dem PC/Laptop teilnehmen.	➤ Kapitel 3.2.1 ➤ Kapitel 3.3.3.1 ➤ Kapitel 3.3.4	Kapitel 4.3
H4: Proband*innen, die mit dem Smartphone teilnehmen, zeigen in Onlinebefragungen häufiger Careless Responding als Proband*innen, die mit dem PC/Laptop teilnehmen.	➤ Kapitel 3.2.1 ➤ Kapitel 3.3 ➤ Kapitel 3.3.4	Kapitel 4.4

Tabelle 4: Überblick über die erarbeiteten Forschungshypothesen

mit der Prüfung möglicher Moduseffekte. Die Darstellung muss zwar an das Smartphone angepasst sein (z. B. Antoun 2015, S. 115; Arn et al. 2015), von der Verwendung responsiver Designs für Smartphones wird aber abgesehen, da diese die gerätübergreifende Vergleichbarkeit untergraben und nicht in tatsächlich optimaler Darstellung münden (z. B. Lütters 2017, S. 72; Mavletova und Couper 2015, S. 89; Revilla und Ochoa 2016; Toepoel 2016, S. 126-127). Für den angestrebten Gerätvergleich ist eine gerätübergreifend weitgehend identische Darstellung zweckmäßig, da diese erlaubt, mögliche Geräteeffekte zu isolieren und von Darstellungseinflüssen zu trennen (de Leeuw und Hox 2015, S. 27; Doty und Glick 1998). In Einklang mit den Empfehlungen der Literatur wird daher die Darstellung der Befragungen des Forschungsprogramms am kleineren Gerät, dem Smartphone, ausgerichtet und eine gerätübergreifend möglichst gleiche Darstellung angestrebt (z. B. Antoun et al. 2018, S. 560; Díaz de Rada und Domínguez-Álvarez 2014; Krämer 2017, S. 99; Lütters 2017, S. 73; Revilla et al. 2016a; Theobald 2017, S. 307). Die Teilnahme mit dem Smartphone soll dafür in allen Studien grundsätzlich im Landscape-Modus erfolgen, der die gesamte Orientierung des Smartphones horizontal anlegt und so eine ähnliche Darstellung wie am PC/Laptop ermöglicht (Höhne und Lenzner 2015; Revilla und Couper 2018). Damit am Smartphone alle Inhalte gut sichtbar und ansteuerbar sind, werden feste Designentscheidungen für die vergleichbare Darstellung (z. B. Pixelbreite, Abstände) getroffen (Schlosser und Silber 2020, S. 275). Vertikales Scrolling muss dabei nicht grundsätzlich vermieden werden (Antoun et al. 2018; de Bruijne 2015, S. 135; Mavletova und Couper 2014, 2016), sollte aber möglichst wenig erforderlich sein (Andreadis 2015b; Couper und Peterson 2017; Jones et al. 1999; Mavletova und Couper 2016). In allen Onlinebefragungen des Forschungsprogramms können demnach mehrere Fragen auf einer Seite platziert werden, die Itemzahl pro Seite wird aber möglichst klein gehalten und die maximale Itemzahl pro Seite auf 15 festgelegt, um Scrolling auf ein geringes Ausmaß zu begrenzen.

Für die Prüfung auf Moduseffekte des Smartphones als Ausfüllgerät werden die Studien so gestaltet, dass mögliche Geräteeffekte des Smartphones in der jeweiligen Befragungssituation wahrscheinlich sind. Zugleich werden stets Personencharakteristika der Befragten erfasst (insbesondere Alter und Geschlecht¹¹), die mit der Gerätenutzung assoziiert sind (z. B. Andreadis 2015a, S. 76; de Bruijne und Wijnant 2013, 2014; Heerwegh und Loosveldt 2008; North et al. 2014; Revilla und Ochoa 2016; Schlosser und Silber 2020, S. 270-271; Toninelli und Revilla 2020, S. 369), und als Kontrollvariablen genutzt. Die inhaltliche Breite der eingesetzten Inventare und Skalen approximiert insgesamt auch die Breite in der Befragungspraxis (Revilla 2017).

¹¹ Da die Fallzahlen diversgeschlechtlicher Personen in der vorliegenden Arbeit für statistische Analysen zu gering sind, wird Geschlecht in den Auswertungen stets als dichotome Variable (weiblich vs. männlich) behandelt

In der Literatur wird empfohlen, vorrangig etablierte Skalen und Inventare einzusetzen, wenn der Fokus nicht auf der Entwicklung neuer Konstruktmessungen liegt (Baumgartner und Weijters 2019, S. 282). Die vorliegende Arbeit folgt so weit als möglich dieser Empfehlung, wobei alle verwendeten Items, sofern sie nicht in deutscher Übersetzung vorliegen, anhand des kollaborativen, iterativen Übersetzungsansatzes in die deutsche Sprache übertragen werden (Douglas und Craig 2007). Die Antwortskalen werden dabei nicht vertikal, sondern horizontal präsentiert, da dies Primacy-Effekte am Smartphone verringert (Couper et al. 2004; Krebs und Höhne 2021). Die Onlinebefragungen des Forschungsprogramms verwenden unter anderem mehrstufige Antwortskalen, da diese besonders kritisch für die Antwortqualität am Smartphone sind (Mavletova et al. 2018), Textfelder als modussensibles Antwortformat (Antoun et al. 2018) sowie heterogene Items und Item-Sets, die diverse Varianten der Itemformulierung enthalten (z. B. Mischung regulärer und umgekehrter Items), anhand derer differenzielle Antworteffekte akkurat trennbar sind (z. B. Baumgartner und Weijters 2019, S. 369-371; Greenleaf 1992b). Für die Onlinebefragungen des Forschungsprogramms wird ganz bewusst auch die Matrixdarstellung für Items genutzt, die auf derselben Antwortskala beantwortet werden. Matrizen sind in der Forschungslandschaft sehr verbreitet und Unterschiede in der Antwortqualität, die durch das Smartphone als Ausfüllgerät entstehen könnten, sind für dieses Frageformat besonders plausibel (z. B. Antoun et al. 2018; Couper und Peterson 2017; Mavletova et al. 2018).

Die Untersuchung möglicher Moduseffekte insbesondere auf die Antwortqualität kann nur verdeckt geschehen, das heißt, ohne die Befragten über den Studienhintergrund aufzuklären. Andernfalls passen diese ihr Antwortverhalten womöglich dem Forschungsziel an. Glaubhafte Instruktionen und Cover-Stories helfen, solche bewussten Anpassungen des Antwortverhaltens zu vermeiden (z. B. Podsakoff et al. 2012), und werden daher in das Forschungsprogramm integriert. Die Aussagekraft empirischer Studien hängt von der Teststärke ab, die durch eine adäquate Stichprobenplanung abgesichert wird. Für die Studien des Forschungsprogramms wird bei der Ermittlung der zu erhebenden Stichprobe jeweils die mindestens relevante Effektstärke für die statistischen Testverfahren angelegt, die bei den Analysen der adressierten abhängigen Variablen verwendet werden (Meyvis und van Osselaer 2018). Zur Realisierung geplanter Stichproben können in Onlinebefragungen Incentives oder Aufwandsentschädigungen bei der Akquise helfen (Reips 2000, S. 110). Gerade Gewinnverlosungen (Deutskens et al. 2004; Heerwegh 2006) und materielle Aufwandsentschädigungen (Manzo und Burke 2012, S. 329) werden von Befragten goutiert und sorgsam in das Forschungsprogramm einbezogen.

Abbildung 7 gibt einen Überblick über das Forschungsprogramm der vorliegenden Arbeit und veranschaulicht, wie die Prüfung auf Moduseffekte des Smartphones aufgebaut ist. Zwei

Voruntersuchungen eröffnen die Forschungsreihe, in denen zunächst – entsprechend der Empfehlungen von Tourangeau (2017, S. 117) – getrennt mögliche Moduseinflüsse 1) auf die Antwortqualität gezielt für sensible Fragen (in Form von Unterschieden der Stichprobenkennwerte und der Antwortbereitschaft) und 2) durch befragungsrelevante Bedürfnisstrukturen der Gerätewahl (ohne festgelegten Themenbereich), die Hinweise auf die Unit-Nonresponse geben könnten, untersucht werden. Die Hauptuntersuchungen des Forschungsprogramms konzentrieren sich in der Folge, gelöst von sensiblen Themen, auf die breitere Betrachtung der Antwortqualität, wofür ein Methodenmix eingesetzt wird. Als Grundlage der Antwortqualität wird dort die Aufmerksamkeit gerätweise verglichen. Zunächst eruiert eine apparative Eye-Tracking Studie, ob das Smartphone für mögliche Aufmerksamkeitsablenkungen anfälliger ist als der PC/Laptop. Abschließend wird in der letzten Studie untersucht, ob etwaige Unterschiede aus der direkten Aufmerksamkeitsmessung im Labor sich auch in einer breit angelegten Onlinebefragung im Feld finden lassen. Es wird geprüft, ob Careless Responding am Smartphone häufiger auftritt, was Antwortqualitätsunterschiede zwischen verschiedenen Ausfüllgeräten anzeigen und für Moduseffekte des Smartphones hinsichtlich der Antwortqualität sprechen würde. CR signalisiert insbesondere sehr starke Messfehler durch Befragte. Zusätzlich wird die Item-Nonresponse als Teilbereich der Antwortqualität einbezogen und weitere Antwortqualitätsreduktionen, die durch das Smartphone entstehen könnten, finden ebenfalls Berücksichtigung.

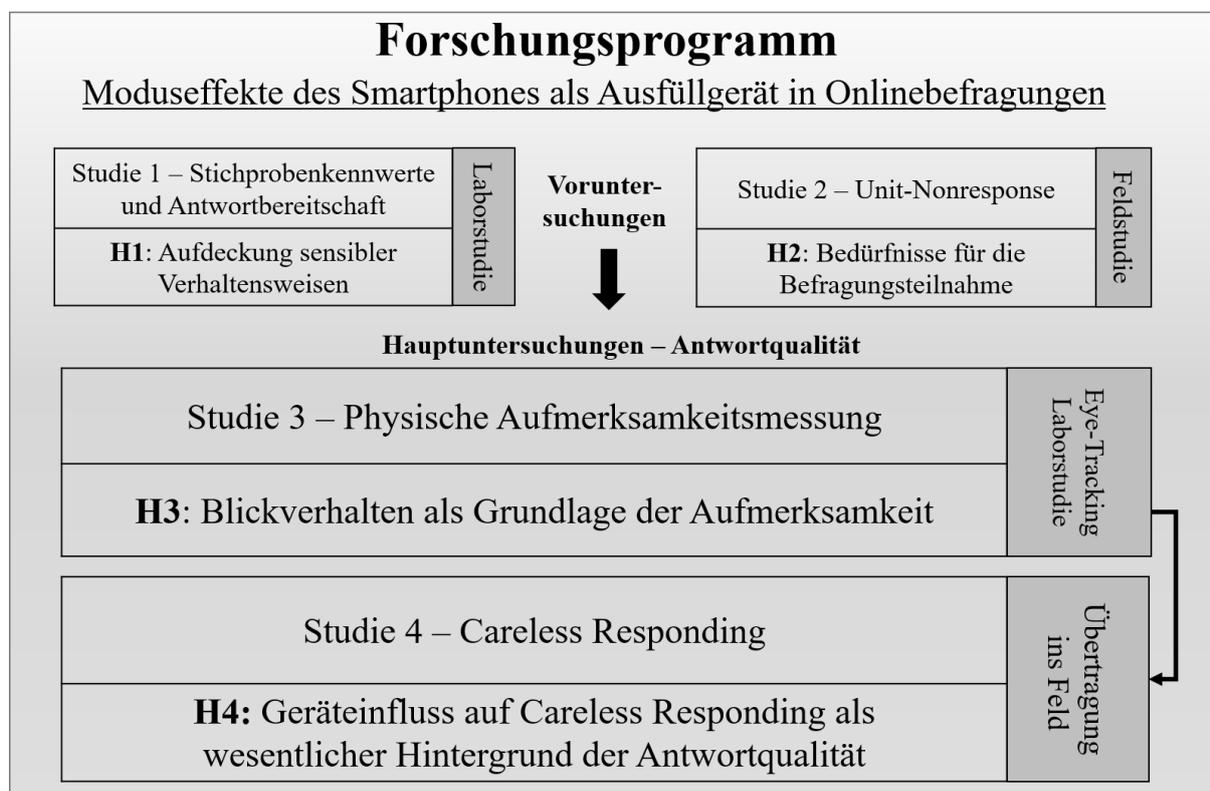


Abbildung 7. Das Forschungsprogramm der vorliegenden Arbeit
Auf die Hypothesenkürzel (**H1-H4**) folgen Kurzbeschreibungen der Studien, die diese jeweils prüfen.

4.1 Prüfung möglicher Moduseffekte des Smartphones bei sensiblen Fragen

Sensible Themenbereiche (siehe Kapitel 2.1 und 3.1.1) bieten sich als erster Ansatzpunkt zur Überprüfung von Moduseffekten des Smartphones an (vgl. Mavletova und Couper 2013; Toninelli und Revilla 2016b). Die erste Voruntersuchung der vorliegenden Arbeit prüft daher mögliche Moduseffekte des Smartphones für sensible Themenbereiche. Als Teilbereich der Antwortqualität im TSE-Framework werden zunächst mögliche Einflüsse auf Stichprobenkennwerte (als besonderes deutlicher Ausdruck des Messfehlers durch Befragte) und auf die Antwortbereitschaft (als bewusste Item-Nonresponse bei sensiblen Fragen) untersucht.

Bisher zeigen sich in der Forschungsliteratur für das Smartphone keine Moduseffekte bei der Erhebung sensibler Informationen (Antoun et al. 2017; Mavletova 2013; Mavletova und Couper 2013; Toninelli und Revilla 2016b), wobei allerdings davon ausgegangen wird, dass am Smartphone weniger aufrichtig geantwortet wird, da der Teilnahmekontext diverser und öffentlicher sein könnte (vgl. Mavletova und Couper 2013; Toninelli und Revilla 2016b). Die adäquate Überprüfung erfordert aber die isolierte Analyse des Geräteinflusses des Smartphones. Vorteile der Selbst-Administration und des Onlineformats für den Selbstbericht bei sensiblen Fragen (siehe Kapitel 2.1 und 3.1.1) gelten sowohl für PCs/Laptops als auch für Smartphones, etwaige weiterführende Unterschiede zwischen den Geräten werden jedoch bisher nicht angemessen adressiert. Die psychologische Bedeutung des Smartphones (siehe Kapitel 2.3.2) demonstriert, dass Smartphones für ihre Besitzer*innen sehr persönliche und im Alltag sehr präsente Objekte sind, mit denen vertrauensvoll interagiert wird (z. B. Fullwood et al. 2017; Keusch et al. 2022; Melumad und Pham 2020; Walsh et al. 2011). Die Privatheit einer Situation fördert in der Befragungsforschung die Aufdeckung sensibler Verhaltensweisen (de Leeuw 2005; Ong und Weiss 2000), demzufolge könnten sich Moduseffekte des Smartphones bei sensiblen Themen darin zeigen, dass am Smartphone ehrlichere Antworten gegeben werden. Am Smartphone berichten Nutzer*innen zudem höheres Kontrollerleben (Brandimarte et al. 2013; Sutanto et al. 2013), was Selbstoffenbarungen weiter fördern könnte.

In Laborstudien können Smartphoneeinflüsse auf die Nutzer*innen mit hoher interner Validität isoliert und von Kontexteinflüssen entkoppelt werden (z. B. Dufau et al. 2011; Schouten et al. 2013; Tourangeau et al. 2017). Zudem sind Selektionseffekte ausgeschlossen, wenn das Ausfüllgerät zufällig zugewiesen wird (Tourangeau 2017, S. 124). Daher bildet eine Laborstudie, in der Personen in identischen Situationen mit unterschiedlichen, zugewiesenen Ausfüllgeräten an derselben Onlinebefragung mit sensiblen Inhalten teilnehmen, die erste Studie des Forschungsprogramms als Voruntersuchung zur Antwortqualität im TSE-Framework. Die Teilnehmer*innen nutzen in dieser Studie ihr eigenes Ausfüllgerät (entweder eigener Laptop oder

eigenes Smartphone, die sie beide jeweils mitführen), da der Besitzfaktor wesentlich für die vermuteten Vorteile des Smartphones ist. Zu eigenen Geräten und Objekten besteht eine engere Bindung und sie werden positiver bewertet (Beggan 1992; Cunningham et al. 2008; Huang, Wang und Shi 2009; Strahilevitz und Loewenstein 1998). Gerade für Smartphones ist der Besitzaspekt relevant, da die Verbindung zum eigenen Gerät emotional gefärbt und deutlich intensiver ist als zu fremden Geräten (z. B. Brasel und Gips 2014; Konok et al. 2017; Melumad und Pham 2020; Thorsteinsson und Page 2014). Für die erste Voruntersuchung wird eine Onlinebefragung konzipiert, an der im Labor teilgenommen wird, und die klassische sensible Themen wie Substanzkonsum, Delinquenz sowie Sexualverhalten (vgl. Kleck und Roberts 2012; Krumpal und Näher 2012; Langhaug et al. 2010; McCallum und Peterson 2012; Tourangeau und Yan 2007; Wolter 2012) integriert. Um sowohl sehr starke als auch eher schwache positive Moduseffekte erkennen zu können, wird der Grad der Sensibilität bei den verwendeten sensiblen Fragen bewusst variiert. Die übergeordnete Hypothese der **H1**

H1: *Personen, die an einer Onlinebefragung mit dem eigenen Smartphone teilnehmen, beantworten sensible Fragen aufrichtiger als Personen, die unter gleichen Bedingungen mit dem eigenen PC/Laptop teilnehmen.*

wird für die Hypothesenprüfung ausdifferenziert und konkret auf den Vergleich von Smartphones und Laptops bezogen (PCs können nicht mitgeführt werden). Sollten Befragte am Smartphone aufrichtiger antworten, könnte dies dazu führen, dass sensible Verhaltensweisen in größerem Ausmaß berichtet werden (geringerer Messfehler durch Befragte), höhere Antwortbereitschaft vorliegt (weniger Item-Nonresponse) und eine größere Antwortbreite gefunden wird, da ein größerer Ausschnitt der Antwortoptionen bei der Antwort auf sensible Fragen genutzt wird (höhere Antwortvarianz). Entsprechend lauten die Unterhypothesen:

H1.1a: *Personen, die die Befragung mit dem Smartphone ausfüllen, berichten in höherem Ausmaß sensibles Verhalten als Personen, die die Befragung mit dem Laptop ausfüllen.*

H1.1b: *Personen, die die Befragung mit dem Smartphone ausfüllen, antworten häufiger auf sensible Fragen als Personen, die die Befragung mit dem Laptop ausfüllen.*

H1.1c: *Personen, die die Befragung mit dem Smartphone ausfüllen, antworten mit größerer Varianz auf sensible Fragen als Personen, die die Befragung mit dem Laptop ausfüllen.*

Zudem wird sozial erwünschtes Verhalten geprüft, da dieses ebenfalls die Aufrichtigkeit im Selbstbericht der Proband*innen widerspiegelt (vgl. Jann 2015; Lee 1993). Im privaten und geschützten Rahmen des eigenen Smartphones könnte das Ausmaß sozialer Erwünschtheit geringer sein. Dies führt zu der letzten Unterhypothese der ersten Voruntersuchung:

H1.2: *Personen, die die Befragung mit dem Smartphone ausfüllen, antworten weniger sozial erwünscht als Personen, die die Befragung mit dem Laptop ausfüllen.*

4.2 Gerätspezifische Bedürfnisstrukturen in Onlinebefragungen

Die zweite Voruntersuchung der vorliegenden Arbeit zu möglichen Moduseffekten des Smartphones als Ausfüllgerät nähert sich über die Untersuchung der befragungsrelevanten Bedürfnisse von Befragten der Unit-Nonresponse als Teilbereich des TSE an. Da Unit-Nonresponse nicht direkt betrachtet werden kann, müssen über Vergleiche der Gruppen, die verschiedene Ausfüllgeräte nutzen, Hinweise auf mögliche Variablen gesammelt werden, die mit der Teilnahmeentscheidung zusammenhängen. Diese könnten ggf. auf gerätbedingte Unterschiede in den Raten der Unit-Nonresponse hindeuten. Teilnahmeverweigerungen könnten problematisch sein, wenn sie mit den Gerätkarakteristika assoziiert sind und dadurch gerätspezifische Unterschiede in der Unit-Nonresponse entstehen. Da weltweit die Antwortraten in der wissenschaftlichen Befragungsforschung zurückgehen (Brick und Williams 2013; Stedman et al. 2019), ist es aus Forschungssicht ratsam, dass die Bedürfnisse Befragter für die Befragungsteilnahme befriedigt werden, damit sie sich für die Teilnahme entscheiden (Biemer 2010). Daraus folgen auch Implikationen für das Smartphone als Ausfüllgerät. So wird die prinzipielle Wahlmöglichkeit des Ausfüllgeräts explizit gewünscht (z. B. Axinn, Gatny und Wagner 2015; de Bruijne und Wijnant 2013) und zugleich steigt in Onlinebefragungen die Präferenz für das Smartphone (z. B. Krämer 2017, S. 100; Revilla et al. 2016b).

Für die Teilnahme an Onlinebefragungen könnten Smartphones aufgrund ihrer Geräteeigenschaften und ihrer psychologischen Bedeutung womöglich befragungsrelevante Bedürfnisse unterschiedlich gut saturieren (siehe Kapitel 2.2 und 2.3), da die Passung von Erhebungssituation und Bedürfnissen wichtig ist (z. B. Kim et al. 2017; Park 2015; Wenz 2021a, S. 102). Die Teilnahmeentscheidungen könnten sich gerätweise unterscheiden, je nachdem, welche Geräte welche Bedürfnisse besser befriedigen, wodurch Unit-Nonresponse gerätspezifisch gesteigert oder verringert werden könnte. Sollten jeweils verschiedene Geräte für die Befriedigung bestimmter Bedürfnisse für die Befragungsteilnahme von allen Befragten als geeigneter eingeschätzt werden, könnten systematische Moduseffekte entstehen, da sich Befragte, denen diese Bedürfnisse wichtig sind, bevorzugt oder sogar ausschließlich für die Teilnahme mit dem überlegenen Gerät entscheiden. In Kapitel 3.1.2 wurde die übergeordnete Hypothese hergeleitet:

H2: *Proband*innen, entscheiden sich eher für die Teilnahme mit dem Smartphone und gegen die Teilnahme mit dem PC/Laptop, wenn ihnen bestimmte Bedürfnisse für die Befragungsteilnahme wichtig sind.*

Die Einschätzung des Gerätnutzens basiert dafür auf konditionalen Erwartungen. Die Gerätwahl wird von Einstellungen, Erfahrungen und Bedürfnissen gesteuert und durch den Nutzungskontext charakterisiert, wobei situationsspezifisch diejenigen Geräte gewählt werden, die jeweils als die geeignetsten erscheinen (de Reuver et al. 2016; Huang et al. 2009; Park und Han 2013; Venkatesh et al. 2003; Yang et al. 2012). Das TAM verdeutlicht, dass WN, WEdN und Selbstwirksamkeit auf solche Gerätwahrnehmungen wirken können (Davis 1993; Joo und Sang 2013; Park und Chen 2007). Zunächst muss geprüft werden, ob verschiedene Ausfüllgeräte in der Wahrnehmung aller Befragten konsistent mit bestimmten Rahmenbedingungen der Teilnahme assoziiert sind. Da die Teilnahme mit dem Smartphone mehr Zeit beansprucht (siehe Kapitel 3.2.3), wird vermutet, dass auch Befragte selber den Zeitaufwand am Smartphone als größer wahrnehmen könnten als am PC/Laptop. Die Entscheidung zur Teilnahme basiert zudem auf der empfundenen Privatheit des Settings (Buchanan und Hvizdak 2009; Hartmann 1995; Taddicken 2009, S. 88-89). Aufgrund der diversen Nutzungsmuster und der besonderen psychologischen Bedeutung des Smartphones (siehe Kapitel 2.2 und 2.3) könnte auch die Privatheit der Teilnahme je nach Gerät unterschiedlich bewertet werden. Da noch unklar ist, welches Ausfüllgerät (Smartphone vs. PC/Laptop) für Privatheit pauschal als besser geeignet erachtet werden könnte, untersucht Studie 2 initial, in welche Richtung ein möglicher Effekt wirkt. Sollten mit Zeitaufwand und Privatheit verknüpfte Wahrnehmungen und Bedürfnisse verhaltensrelevant sein, schätzen Befragte, die ein für diese Aspekte eigentlich objektiv schlechter bewertetes Gerät für die Teilnahme nutzen, dieses vermutlich weniger schlecht ein als Personen, die das vermeintlich überlegene Gerät verwenden. Dies führt zu folgenden Unterhypothesen:

H2.1a: *Proband*innen schätzen die Teilnahme mit dem Smartphone als zeitaufwendiger ein als die Teilnahme mit dem PC/Laptop.*

H2.1b: *Proband*innen, die mit dem Smartphone teilnehmen, schätzen das Smartphone im Vergleich zu Proband*innen, die mit dem PC/Laptop teilnehmen, hinsichtlich des Zeitaufwands weniger schlecht ein.*

H2.1c: *Bei der Einschätzung der Privatheit der Befragungsteilnahme unterscheiden sich das Smartphone und der PC/Laptop als Ausfüllgeräte.*

H2.1d: *Für das eigene gewählte Ausfüllgerät wird die Privatheit selbst besser eingeschätzt als durch Personen, die das jeweils andere Ausfüllgerät nutzen.*

De Bruijne und Wijnant (2014) untersuchen den prädiktiven Charakter soziodemografischer Variablen für die Gerätwahl in Onlinebefragungen über eine logistische Regressionsanalyse. Dieser Ansatz wird in Studie 2 auf die Untersuchung der **H2** übertragen, um sich der Unit-

Nonresponse anzunähern, indem psychosoziale Variablen und Bedürfnisstrukturen als Prädiktoren der Gerätwahl ergänzt werden. Smartphones ermöglichen orts- und zeitflexible Teilnahme und maximieren Flexibilität und Selbstbestimmtheit des Teilnahmekontexts (siehe Kapitel 2.2.2 und 2.2.4). Andererseits dominieren am Smartphone hedonistische Aktivitäten und das Nutzungserleben ist eher komfortzentriert (z. B. Kim et al. 2005; Melumad und Pham 2020). Befragungen erfordern am Smartphone aber mehr Zeit (siehe Kapitel 3.2.3), was hingegen den Komfort senkt und eher für PCs/Laptops als Ausfüllgeräte sprechen könnte. Wenn für Befragte eine oder mehrere dieser Dimensionen sehr wichtig sind, sollten sie bevorzugt mit dem Gerät teilnehmen, das für diese Dimension oder Dimensionen von allen Befragten besser (Smartphone besser für örtliche und zeitliche Flexibilität und PC/Laptop besser für kurze Befragungszeit) eingeschätzt wird. Als Approximation der Teilnahmeentscheidung und Hinweis auf mögliche gerätspezifische Unit-Nonresponseraten agiert die Prognosekraft der Bedeutung dieser Bedürfnisse für die Befragten bei der Gerätwahl. Die konditionalen Bedürfnisse für den Teilnehmerahmen einer Onlinebefragung führen zu den Unterhypothesen:

H2.2a: *Die Gerätwahl in Onlinebefragungen hängt von dem Bedürfnis nach zeitlicher Flexibilität ab, wobei Personen, die zeitliche Flexibilität schätzen, eher das Smartphone nutzen.*

H2.2b: *Die Gerätwahl in Onlinebefragungen hängt von dem Bedürfnis nach örtlicher Flexibilität ab, wobei Personen, die örtliche Flexibilität schätzen, eher das Smartphone nutzen.*

H2.2c: *Die Gerätwahl in Onlinebefragungen hängt von der Bedeutung der Teilnahmedauer ab, wobei Personen, die kürzere Befragungszeiten schätzen, eher den PC/Laptop nutzen.*

Für die Erfassung der Privatheit werden Vertraulichkeit und Anonymität konzeptionell getrennt (vgl. Ong und Weiss 2000; Whelan 2007). Zusätzlich determiniert im Internetzeitalter auch Datensicherheit die Privatheit. Da die Hypothese **H2.1c** für Privatheit zunächst noch die mögliche Effektrichtung untersucht, wird die zugehörige Hypothese des möglichen Prognosewerts des Bedürfnisses nach Privatheit in Onlinebefragungen ebenfalls ungerichtet formuliert:

H2.2d: *Die Gerätwahl in Onlinebefragungen hängt von dem Bedürfnis nach Privatheit ab.*

Zusätzlich wichtig für die Gerätwahl in Onlinebefragungen kann die persönliche Bedeutung des Ausfüllgeräts sein. Nutzer*innen können starke Bindungen an technische Geräte entwickeln, welche sowohl die Interaktionsfrequenz als auch die Nutzungsintensität beeinflussen und daher auch die Gerätwahl determinieren (z. B. Carter und Grover 2015; Raptis et al. 2013, S. 132). Wird ein bestimmtes Gerät als Teil der eigenen Identität wahrgenommen, kann dies also auch die Gerätwahl beeinflussen. Die Intensität und Verbundenheit zu technischen Geräten variiert interindividuell, sodass verschiedene Proband*innen divergierende Bindungsstrukturen

zum Smartphone und zu PCs/Laptops aufweisen. Daraus ergeben sich die finalen Unterthesen:

H2.3a: *Proband*innen, die eine hohe Bindung zum Smartphone berichten, wählen dieses auch eher für die Teilnahme an Onlinebefragungen.*

H2.3b: *Proband*innen, die eine hohe Bindung zum PC/Laptop berichten, wählen diesen auch eher für die Teilnahme an Onlinebefragungen.*

In die regressionsanalytischen Analysen der Gerätwahl werden außerdem relevante Kontrollvariablen integriert. Dies umfasst wahlrelevante Besonderheiten der Befragtengruppen wie Alter, Geschlecht und Gerätefahrung (z. B. de Bruijne und Wijnant 2014; Park 2015; Revilla und Ochoa 2016; Venkatesh et al. 2003). Zeigen sich hier Unterschiede, sollten Studien, wenn sich die Anteile der Smartphone- und PC/Laptop-Teilnahme unterscheiden und eine dieser Variablen mit Zielvariablen verknüpft sein könnte, dies in den Analysen berücksichtigen. Das Analysemodell wird abgerundet durch die Variable Technologieaffinität, da Smartphone-Teilnehmer*innen in Onlinebefragungen sich selbst als progressiver und offener beschreiben (de Bruijne und Wijnant 2014). Das TAM zeigt zudem, dass die wahrgenommene Nützlichkeit (WN) große Relevanz für die Gerätnutzung besitzt (Davis et al. 1989). Da die WN aber höchst individuell ist, wird WN nicht mit ins Regressionsmodell aufgenommen, sondern die Analysen werden durch Zusatzbetrachtungen zur WN des letztlich gewählten Ausfüllgeräts vertieft.

Die zweite Voruntersuchung ist als Grundlagenforschung angelegt, um für wissenschaftliche Onlinebefragungen zu ergründen, ob und, falls ja, in welcher Form Personen ihre Entscheidung für die Teilnahme an Onlinebefragungen mit dem Smartphone von ihren befragungsrelevanten Bedürfnisstrukturen abhängig machen und inwiefern es dabei Unterschiede zum PC/Laptop gibt. Zeigen sich gerätspezifische Unterschiede liefert das Hinweise darauf, dass die Unit-Non-responseraten gerätweise variieren könnten. Das Wissen um etwaige Modusunterschiede ermöglicht Forscher*innen, die Gerätbesonderheiten adäquat zu berücksichtigen, was insbesondere dann angezeigt ist, wenn mögliche gerätspezifische Unterschiede in befragungsrelevanten Bedürfnissen mit Zielvariablen der empirischen Untersuchung assoziiert sind (z. B. könnte das Bedürfnis nach ortsflexibler Teilnahme mit der Variable „variety seeking“ (z. B. McAlister und Pessemer 1982) in Studien zum Konsumverhalten zusammenhängen).

4.3 Anfälligkeit des Smartphones für Aufmerksamkeitsablenkung

Die beiden Voruntersuchungen prüfen mögliche Moduseffekte des Smartphones getrennt für die beiden Teilbereiche des TSE, auf die Geräteinflüsse wirken könnten (Antwortqualität und

Unit-Nonresponse, siehe Kapitel 2.4.2). In der ersten Voruntersuchung werden insbesondere mögliche positive Moduseffekte des Smartphones betrachtet, die sich in Unterschieden in den Stichprobenkennwerten und in der Antwortbereitschaft bei sensiblen Fragen ausdrücken könnten. Die Hauptuntersuchungen des Forschungsprogramms legen die Beurteilung der Antwortqualität breiter an und lösen sich von der Eingrenzung auf sensible Themen. Mögliche Einflüsse des Smartphones auf die Antwortqualität sind vielschichtig, da diese sich in mehreren Facetten ausdrücken kann. Die Forschungsliteratur konstatiert, dass sich für verschiedene Indikatoren der Antwortqualität Unterschiede im Vergleich zum PC/Laptop zeigen (z. B. Antwortzeit, Missings), wenn das Smartphone als Ausfüllgerät genutzt wird. Dies spricht somit eher für negative Moduseffekte des Smartphones hinsichtlich der Antwortqualität, die den TSE erhöhen würden (siehe Kapitel 3.2 bzw. 3.2.3 und insbesondere *Tabelle 2*).

Die Gerätesonderheiten und der Nutzungskontext prädestinieren das Smartphone als Hintergrundfaktor von Careless Responding (siehe Kapitel 3.3.4). Tatsächlich unterscheiden sich die Aufmerksamkeitsstrukturen, insbesondere beim Leseverhalten, zwischen Bildschirmen unterschiedlicher Größe (z. B. Chae und Kim 2004; Dillon et al. 1990; Jones et al. 1999). Die Antwortzeit für gleiche Inhalte ist bei kleineren Bildschirmen länger (Kim et al. 2015) und durch Scrolling wird die Interaktion herausfordernder (Kim et al. 2016). In der dritten Studie des Forschungsprogramms steht zunächst die interne Validität im Vordergrund und es wird Grundlagenforschung betrieben. Es soll geprüft werden, ob gerätspezifische Unterschiede in den Aufmerksamkeitsstrukturen während der Teilnahme an Onlinebefragungen existieren und dadurch Antwortqualitätsdefizite am Smartphone tatsächlich wahrscheinlicher sein könnten. Eye-Tracking Maße bieten sich hierfür als direktes Aufmerksamkeitsmaß an (siehe Kapitel 3.3.3.1). Um Aufmerksamkeitsprozesse als Hintergrund der Antwortqualität (siehe Kapitel 3.2.1) möglichst objektiv zu erfassen, wird zunächst das Blickverhalten der Proband*innen während der Befragungsteilnahme analysiert. Die in der Herleitung aufgezeigten Besonderheiten des Smartphones als Ausfüllgerät (siehe Kapitel 3.3.4) führen zu folgender Hypothese:

H3: *Proband*innen, die an Onlinebefragungen mit dem Smartphone teilnehmen, blicken häufiger von den Befragungsinhalten weg als Proband*innen, die mit dem PC/Laptop teilnehmen.*

Am Smartphone können moderne Eye-Tracking Systeme Aufmerksamkeitsstrukturen sehr robust aufdecken (z. B. Chen et al. 2003; Kim et al. 2015, 2016). Durch die Aufzeichnung des Blickverhaltens im Labor werden in Studie 3 dabei aufmerksamkeitswirksame Kontexteffekte konstant gehalten, sodass Unterschiede im Blickverhalten primär auf das Ausfüllgerät zurückzuführen sind. In dieser dritten Studie steht also die Fundierung der mutmaßlichen Anfälligkeit des Smartphones für geringere Aufmerksamkeit als Grundlage der Antwortqualität im TSE-

Framework im Vordergrund. Die reine Gerätgröße des Smartphones macht es wahrscheinlich, dass sich der Blick häufiger vom Gerät und der Befragung entfernt. Es wird geprüft, ob Personen am Smartphone während der Befragung häufiger von den Befragungsinhalten und dem Gerät abschweifen. Dadurch wird das potenzielle Risiko für Aufmerksamkeitsablenkungen und somit CR abgebildet. Dies ist vor allem für Fixationen relevant, die direkt den Aufmerksamkeitsfokus anzeigen (siehe Kapitel 3.3.3.1). Aus diesen Überlegungen ergeben sich die Unterhypothesen:

H3.1: *Der Blick der Proband*innen, die an Befragungen mit dem Smartphone teilnehmen, bewegt sich häufiger außerhalb des Bereichs der Befragung als bei Proband*innen, die mit dem PC/Laptop teilnehmen.*

H3.2: *Die Fixationsdauer auf Befragungsinhalte ist bei Proband*innen, die mit dem Smartphone teilnehmen, geringer als bei Proband*innen, die mit dem PC/Laptop teilnehmen.*

Sollte häufiger von der Befragung weggeschaut, trotzdem aber stets geantwortet werden, sind mehr Refixationen nötig und auch mehr Fixationen insgesamt. Unsteteres Blickverhalten im Laborsetting bei gleichen Außeneinflüssen kann dabei zudem über die durchschnittliche Fixationsdauer abgebildet werden. Da am Smartphone konstant mehr Zeit für die Bearbeitung der gleichen Inhalte benötigt wird (siehe Kapitel 3.2.3), sind kürzere, aber dafür häufigere Fixationen am Smartphone wahrscheinlich. Daraus ergeben sich die Hypothesen:

H3.3: *Proband*innen, die am Smartphone an der Befragung teilnehmen, zeigen häufiger Refixationen in die Befragung als Personen, die mit dem PC/Laptop teilnehmen.*

H3.4: *Für Proband*innen, die am Smartphone an der Befragung teilnehmen, ist die Anzahl der Fixationen im Bereich der Befragungsinhalte größer als für Proband*innen, die mit dem PC/Laptop teilnehmen.*

H3.5: *Die durchschnittliche Fixationsdauer ist bei Smartphone-Teilnehmer*innen geringer als bei PC/Laptop-Teilnehmer*innen.*

4.4 Das Smartphone als Quelle von Careless Responding?

Die beiden Hauptuntersuchungen des Forschungsprogramms betrachten mögliche Moduseffekte des Smartphones als Ausfüllgerät im TSE-Framework hinsichtlich der Antwortqualität. Dafür werden in der vierten Studie die Unterbereiche der Antwortqualität sorgfältig analysiert, für die Moduseffekte des Smartphones denkbar sind (Messfehler durch Befragte und Item-Nonresponse, siehe Kapitel 2.4 und 2.4.2.2). Die Prüfung baut auf dem Forschungsstand auf, der das Smartphone als anfälliger für CR charakterisiert (siehe Kapitel 3.3.4), und erweitert die

Befunde zum Blickverhalten während der Befragungsteilnahme (siehe Kapitel 4.3). Potenzielle gerätspezifische Aufmerksamkeitsunterschiede werden im Unterschied zur vorangegangenen Studie mit ihrer künstlichen Laborsituation nun im realen Kontext untersucht (Gosling und Mason 2015).

Gerade Onlinebefragungen finden im Feld in diversen und unkontrollierten Kontexten statt (z. B. Wharton et al. 2003), was Ablenkungen und Aufmerksamkeitsverluste während der Befragung generell wahrscheinlicher macht (z. B. Sendelbah et al. 2016; Wenz 2021b). Die mobile Internetnutzung während der Befragung potenziert die Kontextvariabilität zusätzlich. Aufgrund der starken Kontextabhängigkeit der Antwortqualität müssen die mit dem für die Befragung verwendeten Ausfüllgerät verbundenen Besonderheiten, die schon in Kapitel 3.1.1.1 ausgeführt wurden, erneut für die Beurteilung möglicher Moduseffekte berücksichtigt werden. Smartphones werden häufiger unterwegs, in Bewegung, in Anwesenheit Dritter und in geräuschvoller/ablenkender Umgebung genutzt (z. B. Höhne und Schlosser 2019; Kim et al. 2005; Lee et al. 2005). Auch an Onlinebefragungen nehmen Smartphone-Teilnehmer*innen daher häufiger an anderen Orten als dem eigenen Zuhause und in Anwesenheit Dritter teil, als dies am PC/Laptop vorkommt (z. B. de Bruijne und Oudejans 2015, S. 136; Mavletova und Couper 2013; Revilla et al. 2016a; Toninelli und Revilla 2016a, S. 4). Um den Einfluss des Smartphones als Ausfüllgerät auf CR von möglichen Kontexteffekten getrennt zu untersuchen, werden in Studie 4 demnach der Teilnahmeort, die Teilnahmesituation und mögliche anwesende Bystander kontrolliert (vgl. Lee et al. 2005; Verkasalo 2009). Da es am Smartphone häufiger zu Ablenkung während der Teilnahme kommt (siehe Kapitel 3.3.4), wird zusätzlich die subjektiv empfundene Ablenkung erfragt und ebenfalls in die Analysen einbezogen. In Onlinebefragungen können Befragte parallel anderen Aktivitäten nachgehen, anstatt sich ausschließlich auf die Befragung zu konzentrieren (Couper und Peterson 2017; Höhne et al. 2020b; Zwarun und Hall 2014). Daher zielt die Kontrolle zusätzlich darauf, ob Multi-Tasking während der Teilnahme stattfindet. Ebenfalls kontrolliert werden entsprechend der Literaturempfehlungen die soziodemografischen Variablen Alter und Geschlecht (z. B. Heerwegh und Loosveldt 2008; Revilla und Ochoa 2016; Schlosser und Silber 2020, S. 270-271). Durch den Einbezug der Kontextbedingungen und der personenbezogenen Charakteristika kann die Hypothese **H4** den isolierten Geräteinfluss des Smartphones betrachten:

H4: *Proband*innen, die mit dem Smartphone teilnehmen, zeigen in Onlinebefragungen häufiger Careless Responding als Proband*innen, die mit dem PC/Laptop teilnehmen.*

Die **H4** wird in der Folge weiter untergliedert. Sollte das Smartphone als Ausfüllgerät CR fördern und somit, da die Messfehler durch Befragte steigen, den TSE systematisch erhöhen,

würde dies adäquate Anpassungen der Befragungspraxis erfordern. Daher ist die gewissenhafte Prüfung der **H4** wichtig. Die CR-Prävalenz sollte in der durchgeführten Feldstudie nicht zu gering sein, damit Bodeneffekte für CR vermieden werden. Bodeneffekte könnten vorliegen, wenn in der Befragung kaum oder nur sehr wenig CR auftritt. Dies könnte etwa der Fall sein, wenn Befragte an den Befragungsinhalten interessiert sind und daher seltener unaufmerksames Antwortverhalten zeigen (Arthur et al. 2021; Gummer und Kunz 2022). Themen, die eine besonders hohe Aktivierung der Teilnehmer*innen begünstigen, werden im verwendeten Fragebogen daher vermieden. Zudem sinken in längeren Befragungen im Verlauf Gewissenhaftigkeit und Aufmerksamkeit der Teilnehmer*innen (z. B. Duffy et al. 2005; Galesic und Bosnjak 2009), da Befragte stetig weniger Zeit in die Antwortgabe investieren (Bäckström und Björklund 2019). Werden in späteren Befragungsteilen trotz geringerer Aufmerksamkeit Antworten gegeben, zeigen sich auch ganz konkret höhere CR-Raten (z. B. Bowling et al. 2021a; Gibson und Bowling 2020). Die Befragung in der vierten Studie wird daher so angelegt, dass sie eine gewisse Mindestlänge nicht unterschreitet. Für das vorliegende Forschungsprogramm wird eine Onlinebefragung konzipiert, die bei gewissenhafter Teilnahme mehr als 20 Minuten Zeit beansprucht. Diese 20 Minuten gelten für Befragungen (insbesondere Onlinebefragungen) als Grenzwert, ab dem Einschränkungen der Antwortqualität wahrscheinlich werden (z. B. Herzing 2019, S. 8; Revilla und Ochoa 2017). Die Antwortqualität in Befragungen kann auch aktiv beeinflusst werden. CR-Tendenzen können verringert werden, wenn 1) vor der Aufdeckung unaufmerksamen Antwortens gewarnt wird (Bowling et al. 2021a; Huang et al. 2012; Huang et al. 2015a; Ward und Pond 2015), 2) Befragte negative Konsequenzen bei geringer Antwortqualität befürchten (Gibson und Bowling 2020), 3) Instruktionen Befragte zu Beginn gezielt auffordern, mit hoher Qualität zu antworten (Meade und Craig 2012; Ward und Meade 2018) und 4) dynamisches Feedback zur Antwortqualität während der Teilnahme dargeboten wird (Conrad et al. 2017a; Ward und Pond 2015). Um von den CR-Ausmaßen möglichst verlässlich auf potenzielle Moduseffekte des Smartphones schließen zu können, werden diese Aspekte, die von sich aus aufmerksamkeitsaktivierend wirken, ebenfalls nicht integriert.

Bei der Stichprobenakquise für die Studie 4 wird auch auf eine Studienbörse zurückgegriffen, da für Studienbörsen bekannt ist, dass dort häufig Antwortqualitätsdefizite auftreten (z. B. Cheung et al. 2017; Fleischer, Mead und Huang 2015; Kennedy et al. 2020). Teilnehmer*innen aus Studienbörsen sind sehr befragungserfahren und zugleich auf die Minimierung des persönlichen Aufwands bedacht, sodass Motivationseinschränkungen durchaus plausibel sind.

Die für das Forschungsprogramm gewählten Methoden der CR-Detektion (siehe *Tabelle 3*) diktieren außerdem einige Rahmenbedingungen für das zu wählende Befragungsdesign. Damit

verschiedene post-hoc Detektionsmethoden greifen können, werden sie auf heterogene Item-Sets (Mischung regulärer und umgekehrter Items) angewendet (z. B. Grable et al. 2021; Hong et al. 2020; Marjanovic et al. 2015; Schroeders et al. 2022). Für die skalenübergreifende Anwendung von post-hoc Detektionsmethoden bieten sich Multi-Item-Skalen an (Gibson und Bowling 2020; Goldammer et al. 2020; Greenleaf 1992a; Marjanovic et al. 2015), die in den Fragebogen integriert werden und für die auch bewusst die Matrixdarstellung verwendet wird. Da komplexe und herausfordernde Elemente (z. B. offene Fragen, lange Texte) besonders anfällig für Antwortqualitätseinschränkungen sind (Alvarez und Cavanagh 2004; Caplan und Waters 1999; Thornton et al. 2014), werden auch diese in den Fragebogen einbezogen.

Die Besonderheit der konzipierten Studie liegt darin, dass sie nicht auf ein inhaltliches Themengebiet festgelegt ist, da das Untersuchungsziel auf strukturelle und statistische Effekte ausgerichtet ist, die sich themenunabhängig untersuchen lassen. Vor diesem Hintergrund wird ein Themengebiet gewählt, das breit genug ist, um eine längere Befragung zu rechtfertigen, und die Erfassung mehrerer ähnlicher Konstrukte zur Konsistenzprüfung ermöglicht. Einige der erfragten Konstrukte sollten zumindest teilweise miteinander korrelieren, um konsistenzbezogene CR-Indikatoren auch skalenübergreifend anwenden zu können. Daher werden in der Befragung Redundanzen geschaffen und mehrfach ähnliche Konstrukte abgefragt. Konkret wird Umweltbewusstsein, als eines der Leitthemen moderner Gesellschaften und insbesondere der letzten Jahre, als Befragungsinhalt gewählt, da dieses auch stark mit persönlichen Werten und Eigenschaften verbunden ist (vgl. Schultz und Kaiser 2012; Schultz und Zelezny 1999; Steg et al. 2014). Dies eröffnet breite Einsatzmöglichkeiten für etablierte und verwandte Multi-Item-Skalen. Thematisch wird die Abfrage des Umweltbewusstseins in die – zum Zeitpunkt der Erhebung (2020) omnipräsente und wegweisende – Covid-19 Pandemie eingebettet (vgl. Ciotti et al. 2020; Velavan und Meyer 2020; World Health Organization 2020).

Zusammenfassend wird in Studie 4 zur Prüfung der Bedeutung des Smartphones für CR eine umfassende Onlinebefragung (mehr als 150 Items) konzipiert, die mehrere thematisch heterogene Multi-Item-Skalen integriert. Diese birgt keine negativen Konsequenzen für die Proband*innen, wenn sie unaufmerksam antworten und wird thematisch nicht zu interessant und aktivierend gestaltet (z. B. Kim et al. 2019; Ward und Meade 2023). Für die verwendeten CR-Indikatoren in der vorliegenden Arbeit (siehe *Tabelle 3*) lässt sich für die Onlinebefragung im Feld die **H4** in folgende Unterhypothesen untergliedern, wobei verschiedene CR-Formen berücksichtigt werden:

H4.1: *Bei Proband*innen, die mit dem Smartphone teilnehmen, signalisieren Infrequency-Indikatoren häufiger CR als bei Teilnehmer*innen mit dem PC/Laptop.*

H4.1.1: *Am Smartphone werden Bogus-Items seltener korrekt beantwortet als am PC/Laptop.*

H4.1.2: *Am Smartphone bestehen Proband*innen einen Instruktions-Manipulationscheck (IMC) seltener als am PC/Laptop.*

H4.2: *Proband*innen, die mit dem Smartphone teilnehmen, berichten selbst häufiger unaufmerksames Antwortverhalten als Proband*innen, die mit dem PC/Laptop teilnehmen.*

H4.2.1: *Bei Proband*innen, die mit dem Smartphone teilnehmen, ist die selbstberichtete Aufmerksamkeit geringer als bei Proband*innen, die mit dem PC/Laptop teilnehmen.*

H4.2.2: *Proband*innen, die mit dem Smartphone teilnehmen, geben häufiger an, dass ihre Daten für die Auswertung nicht genutzt werden sollten.*

Beim monotonen Antworten ist es für Long Strings zusätzlich möglich, für die CR-Aufdeckung die durchschnittliche Stringlänge zu ermitteln, die angibt, wie oft Befragte im Schnitt auf aufeinanderfolgende Items identisch antworten. Ist die Stringlänge hoch, deutet auch dies auf CR hin.

H4.3: *Proband*innen, die mit dem Smartphone teilnehmen, antworten häufiger monoton als Proband*innen, die mit dem PC/Laptop teilnehmen.*

H4.3.1: *Am Smartphone treten in den Antwortmustern der Proband*innen häufiger sehr lange Long Strings auf als am PC/Laptop.*

H4.3.2: *Am Smartphone sind Long Strings durchschnittlich länger als am PC/Laptop.*

H4.3.3: *Am Smartphone ist die individuelle Response Validität (IRV) geringer als am PC/Laptop.*

H4.4: *Am Smartphone ist die Antwortkonsistenz der Teilnehmer*innen geringer als am PC/Laptop.*

H4.4.1: *Am Smartphone antworten die Teilnehmer*innen weniger konsistent auf psychometrische Synonyme als am PC/Laptop.*

H4.4.2: *Am Smartphone antworten die Teilnehmer*innen weniger konsistent auf psychometrische Antonyme als am PC/Laptop.*

H4.4.3: *Am Smartphone erkennen die Proband*innen seltener als am PC/Laptop, dass Skalenumkehr vorliegt, und antworten dadurch inkonsistenter.*

H4.5: *Proband*innen, die mit dem Smartphone teilnehmen, setzen sich weniger bewusst mit den Befragungsinhalten auseinander als Proband*innen, die mit dem PC/Laptop teilnehmen.*

H4.5.1: *Am Smartphone werden offene Fragen weniger ausführlich beantwortet als am PC/Laptop.*

H4.5.2: *Am Smartphone antworten Proband*innen auf inhaltliche Rückfragen seltener korrekt als am PC/Laptop.*

H4.5.3: *Proband*innen, die mit dem Smartphone teilnehmen, werden häufiger aufgrund sehr schneller Antwortzeiten als Careless Responder klassifiziert als Proband*innen, die mit dem PC/Laptop teilnehmen.*

Sehr kurze Antwortzeiten werden anhand des Cut-Off-Werts 2s/Item als CR klassifiziert (siehe Kapitel 3.3.3.3). Allgemein zeigen sich in Onlinebefragungen am Smartphone jedoch stets längere Antwortzeiten, die ebenfalls die Gefahr von CR erhöhen können. Somit erweitert sich die **H.4.5** um eine weitere Hypothese:

H4.5.4: *Am Smartphone benötigen Teilnehmer*innen generell mehr Zeit als am PC/Laptop, um die Befragung durchzuführen.*

Die Prüfung auf Moduseffekte des Smartphones wird um weitere Antwortqualitätsmaße (vgl. Erens et al. 2019) und Ausreißeranalysen ergänzt, um über CR hinaus mögliche Antwortqualitätseinschränkungen durch das Smartphone zu berücksichtigen. Über Missings werden auch mögliche Einflüsse des Smartphones auf die Item-Nonresponse als Teilbereich der Antwortqualität im TSE-Framework explizit einbezogen.

H4.6a: *Proband*innen, die mit dem Smartphone teilnehmen, weisen höhere Mahalanobis-Distanzwerte auf als Proband*innen, die mit dem PC/Laptop teilnehmen.*

H4.6b: *Für Proband*innen, die mit dem Smartphone teilnehmen, treten mehr Missings in den Antworten auf als für Proband*innen, die mit dem PC/Laptop teilnehmen.*

H4.6c: *Proband*innen, die mit dem Smartphone teilnehmen, brechen die Befragungsteilnahme häufiger ab als Proband*innen, die mit dem PC/Laptop teilnehmen.*

Diese Hypothesen werden im zugehörigen Methodenkapitel (siehe Kapitel 5.4.1) jeweils konkret auf die eingesetzten Skalen und das jeweils betrachtete CR-Maß bezogen. Die breite Anlage der Aufdeckung unaufmerksamen Antwortverhaltens ermöglicht es nicht nur, CR als Ursache extrem verringerter Antwortqualität zu erkennen, sondern gibt auch Einblicke in weniger starke Einschränkungen der Antwortqualität. Sollte sich trotz der demonstrierten Anfälligkeit (siehe Kapitel 3.4.4) des Smartphones an diesem, wenn es als Ausfüllgerät verwendet wird, kein stärkeres Ausmaß von CR zeigen und die Antwortqualität gerätübergreifend gleich sein, stärkt dies den Einsatz des Smartphones als Ausfüllgerät in Onlinebefragungen und spricht gegen Moduseffekte.

5 Empirische Studien

Die Voruntersuchungen der vorliegenden Arbeit bestehen aus zwei empirischen Studien, die die möglichst verlässliche Prüfung etwaiger Moduseffekte des Smartphones im TSE-Framework vorbereiten. Sie betrachten zunächst mögliche gerätweise Unterschiede in den Stichprobenkennwerten und der Antwortbereitschaft bei sensiblen Fragen (Kapitel 5.1) und, getrennt davon, ob es durch den Smartphoneeinsatz womöglich zu Unterschieden in der Unit-Nonresponse kommen könnte (Kapitel 5.2). Anschließend fokussieren die Hauptuntersuchungen Moduseffekte hinsichtlich der Antwortqualität, da hierfür mögliche Beiträge des Smartphones zum TSE besonders plausibel sind. Die Forschungsreihe legt dazu zwei aufeinander aufbauende empirische Studien an, die sukzessive Einblicke in Aufmerksamkeitsstrukturen und die Antwortqualität bei der Teilnahme mit dem Smartphone geben. Diese beiden Studien werden in den Unterkapiteln 5.3 und 5.4 des fünften Kapitels vorgestellt. In der Hypothesenprüfung aller Studien wird stets eine Irrtumswahrscheinlichkeit von $\alpha = 5\%$ angelegt und für die Berechnung der erforderlichen Stichprobengröße eine Power von mindestens 80% vorausgesetzt.

5.1 Studie 1 – Laborexperiment zur Aufrichtigkeit bei sensiblen Themenbereichen

Für sensible Themenbereiche sind Moduseffekte des Smartphones besonders wahrscheinlich (siehe Kapitel 2.1 und 3.1.1). Entsprechend der Ausführungen des Forschungsprogramms in Kapitel 4.1 wird in einem Laborexperiment geprüft, ob durch das Smartphone Unterschiede in den Stichprobenkennwerten (Messfehler durch Befragte) und der Antwortbereitschaft (Item-Nonresponse) bei sensiblen Fragen entstehen, die besonders offensichtliche Moduseffekte für die Antwortqualität signalisieren würden. Nach Kenntnisstand des Autors wird erstmals gezielt die psychologische Bedeutung des Smartphones bei der Untersuchung berücksichtigt und der Geräteinfluss in einer Laborstudie isoliert. Dies erweitert die Erkenntnislage, da bislang mögliche Moduseffekte des Smartphones für sensible Themenbereiche nur im Feld untersucht werden und dort die Kontexteinflüsse der Erhebungssituation im Fokus stehen (vgl. Mavletova und Couper 2013; Toninelli und Revilla 2016b). Diese Kontexteinflüsse werden in Studie 1 eliminiert bzw. konstant gehalten. Zudem ist der Besitzfaktor von Objekten bei technischen Geräten relevant, mit denen die Interaktion zumindest teilweise haptisch erfolgt (z. B. Beggan 1992; Brasel und Gips 2014; Cunningham et al. 2008; Huang et al. 2009; Strahilevitz und Loewenstein 1998). Der Besitz wird kontrolliert, indem Proband*innen eines ihrer eigenen Ausfüllgeräte (Smartphone vs. Laptop) in der Befragungssituation für die Teilnahme nutzen. Zusammenfassend nehmen die Versuchspersonen in Studie 1 also jeweils mit ihrem eigenen Ausfüllgerät in einer kontrollierten Umgebung an einer Befragung zu sensiblen Themenbereichen teil.

5.1.1 Methode

Die Stichprobenplanung fand im Vorfeld der Erhebung mithilfe der Software G*Power (Faul et al. 2007; Faul et al. 2009) statt. Durch den innovativen Ansatz der vorliegenden Studie konnte eine antizipierte Effektstärke nicht der Literatur entnommen werden. Stattdessen wird eine Mindestgröße der Effektstärke angelegt. Schon ab einer niedrigen mittleren Effektstärke (Cohen's $d = 0,5$) würden Befundabweichungen – beispielsweise in Form von Mittelwertdifferenzen (Unterschiede im Messfehler durch Befragte) im Zwei-Gruppen-Vergleich (Laptop vs. Smartphone) oder höherer Antwortbereitschaft (weniger Item-Nonresponse) am Smartphone – bedeutsame Moduseffekte signalisieren, die für exklusive Geräteinflüsse sprechen. Um einen Effekt mittlerer Stärke verlässlich aufdecken zu können, sollten mindestens $N = 102$ Personen teilnehmen und sich möglichst gleich auf die Gerätgruppen (Laptop und Smartphone) verteilen. Entsprechend werden mögliche Moduseffekte im Between-Subjects-Design geprüft.

5.1.1.1 Versuchsablauf

Die Erhebung fand im Zeitraum vom 25.06. bis zum 19.07.2018 an der Bergischen Universität Wuppertal (BUW) statt. Die Proband*innen wurden über persönliche Ansprache am universitären Hauptcampus rekrutiert. Es wurden vornehmlich Personen zur Teilnahme eingeladen, die sowohl einen Laptop als auch ein Smartphone mit sich führten, um die Zuteilung zu den Untersuchungsgruppen (Laptop vs. Smartphone) möglichst zufällig vornehmen zu können. In den wenigen Fällen ($n = 7$), in denen Teilnehmer*innen nur über eines der Geräte verfügten (jeweils nur ein Smartphone), wurden sie Teil der Gruppe des mitgeführten Geräts, sodass die letztliche Gruppenzuweisung eingeschränkt randomisiert erfolgte, da eine vollständige Randomisierung nicht möglich war (Campbell und Stanley 1963, S. 34-37; Shadish, Cook und Campbell 2002, S. 294-303). Die Versuchspersonen wurden nach ihrer Zustimmung zur Teilnahme in den reizarmen Untersuchungsraum geleitet. Dort empfing sie die Versuchsleitung, die ihnen einen Arbeitsplatz für die Befragungsteilnahme zuwies. Die Teilnehmer*innen nahmen auf einem höhenverstellbaren Stuhl Platz, richteten sich initial am Arbeitsplatz ein und wählten eine für sie bequeme Sitzhaltung. Ein Sichtschutz schirmte sie von externen Störreizen ab.

Im Untersuchungsraum wurde kabelloser Internetzugang über das Netzwerk der BUW sichergestellt. Die Interaktion der Versuchsleitung und der Teilnehmer*innen war auf ein Mindestmaß reduziert. Die Teilnehmer*innen öffneten mit ihrem eigenen Ausfüllgerät – Smartphone oder Laptop je nach Gruppenzuteilung – den Link zu einer Onlinebefragung. Die Datenerhebung der Onlinebefragung erfolgte über die Befragungsplattform SoSci Survey (Leiner 2023). Die Proband*innen wurden instruiert, im zugewiesenen Setting an der gerade geöffneten

Onlinebefragung teilzunehmen, in der alle weiteren Informationen im Einleitungstext dargestellt wurden. Mögliche weitere Funktionen oder Programme des Ausfüllgeräts sollten für die Dauer der Untersuchung nicht parallel genutzt werden. Für Teilnehmer*innen mit dem Smartphone wurde sichergestellt, dass sie im horizontalen Landscape-Modus an der Befragung teilnahmen, um die Fragebogendarstellung möglichst gleich zu halten. Nachdem die Onlinebefragung aufgerufen wurde, erfolgte der Hinweis, dass diese nur den ersten Befragungsteil darstelle und durch eine zweite kurze und schriftliche Befragung im Anschluss im Nebenraum ergänzt werde. Sofern keine weiteren Fragen auftraten, verließ die Versuchsleitung den Laborraum und wartete im Nebenraum auf die Versuchspersonen, die zum Abschluss des Onlinebefragungsteils gebeten wurden, sich dort zu melden.

Die Proband*innen waren während der Onlinebefragung isoliert, da sensible Themenbereiche adressiert wurden und die Anwesenheit Dritter die Bereitschaft zur ehrlichen Selbstausskunft bei sensiblen Fragen hemmt (siehe Kapitel 3.1.1). Die Onlinebefragung beinhaltete 37 Items und die Bearbeitung nahm im Schnitt 7,23 Minuten in Anspruch. Nach Beendigung der Onlinebefragung empfing die Versuchsleitung die Proband*innen im Nebenraum, wo eine schriftliche Nachbefragung, bestehend aus 31 Items, angeschlossen wurde. Diese wurde separat durchgeführt und fragte affektive Komponenten der Befragungsteilnahme und technische Erfahrungen ab. Um hier mögliche Einflüsse des Ausfüllgeräts bei der Beantwortung auszuschließen, wurde das Ausfüllformat gewechselt und beide Gruppen füllten einen ausgedruckten Fragebogen schriftlich aus. Zum Abschluss wurden alle Teilnehmer*innen in Form eines Debriefings ausführlich über den Studienhintergrund und die untersuchte Fragestellung aufgeklärt. Unter allen Versuchspersonen wurden drei Amazon-Gutscheine im Wert von jeweils 20 Euro verlost. Die Teilnahmedauer für die gesamte Untersuchung lag zwischen 15 und 20 Minuten.

5.1.1.2 Stichprobe

Die Stichprobe setzte sich ausnahmslos aus Studierenden der BUW zusammen. Insgesamt konnten 115 Versuchspersonen gewonnen werden. Nach Ausschlüssen (je einmal: nicht ausreichende Deutschkenntnisse und Mitführen eines geborgten statt eines eigenen Laptops; zweimal: sehr häufige Unterbrechungen der Onlinebefragung) verblieben insgesamt $N = 111$ Versuchspersonen im Datensatz, die in die Analysen eingehen. Die Altersspanne lag zwischen 18 und 35 Jahren ($\bar{x} = 22,88$, $SD = 2,68$). 59,46% der Stichprobe waren weiblich ($n = 66$), womit die Stichprobe sich aus signifikant mehr Frauen als Männern zusammensetzte ($\chi^2(1) = 3,970$, $p = 0,046$). 54 Proband*innen gehörten der Laptop-Gruppe an und die Gruppe des Smartphones

umfasste 57 Teilnehmer*innen. Bezogen auf die Untersuchungsgruppen zeigte sich für die Geschlechterverteilung je nach Ausfüllgerät kein Unterschied ($\chi^2(1) = 0,535, p = 0,464$).

5.1.1.3 Fragebogenbeschreibung

Der Befragung wurde ein Pretest mit ausgewählten Mitarbeiter*innen des methodischen Fachbereichs der Fakultät der Wirtschaftswissenschaft an der BUW vorgeschaltet. Im Folgenden wird die finale Version der Befragung vorgestellt. Die Bezeichnungen der Variablen sind kursiv dargestellt. Die Einleitung (S. 1) stellte als Cover-Thema der Befragung „Work-Life Balance der Studierenden der Universität Wuppertal“ vor, mit dem vermeintlichen Ziel, die Bedürfnisse, Einstellungen und Verhaltensweisen der Studierenden besser abzubilden. Unter diesem Vorwand konnten private und sensible Themenbereiche in der Befragung adressiert werden. Es wurde die Zustimmung der Befragten eingeholt, dass die erhobenen Daten im Rahmen der europäischen Datenschutzgrundverordnung (DSGVO) zu rein wissenschaftlichen Zwecken verarbeitet und genutzt werden können. Im Anschluss generierten die Teilnehmer*innen einen individuellen, anonymisierten sechsstelligen Code (S. 2), über den die Datenreihen der Onlinebefragung und der anschließenden schriftlichen Befragung einander zugeordnet wurden. In der Folge (S. 3) wurden der höchste bisher erreichte *Bildungsabschluss* und die Anzahl der bisherigen *Hochschulsemester* inklusive des aktuellen erfragt.

Der inhaltliche Teil der Befragung (S. 4-10) mischte bewusst verschiedene Frageformate, um unabhängig von etwaigen Darstellungsbesonderheiten am kleinen Smartphonebildschirm eine Einschätzung möglicher Moduseffekte zu erlauben. Den inhaltlichen Einstieg markierten vier Items zum Studium an der BUW, wobei das Ausmaß der Zufriedenheit mit dem Angebot der *Mensa*, der wissenschaftlichen *Betreuung* und dem *Lehrangebot* im eigenen Fachbereich sowie dem *Hochschulsportangebot* abgefragt wurde (S. 4). Hinterlegt war eine validierte siebenstufige, vollverbalisierte Zufriedenheitsskala von 1 = „sehr unzufrieden“ bis 7 = „sehr zufrieden“ (z. B. Fischer, Hüttermann und Werther 2021, S. 3). Zudem standen zwei Ausweichkategorien („keine Angabe“ und „kann ich nicht beurteilen“) zur Verfügung.

Die folgenden sechs Fragen (S. 5+6) adressierten sensible Verhaltensweisen im Studium und in der Freizeit. Die Items zu sensiblem bzw. privatem Verhalten wurden der Literatur entnommen und gegebenenfalls auf den Kontext des Studiums angepasst. Die insgesamt sechs Items umfassten *Klausurbetrug* und *Plagiarismus* (Ong und Weiss 2000, S. 1696) sowie Umgang mit Alkohol, wobei Konsumfrequenz (*Alkoholkonsum*) und möglicher Kontrollverlust (*Kontrollverlust Alkoholkonsum*) beim Konsum (Mavletova und Couper 2013, S. 204; Toninelli und Revilla 2016b, S. 166) entsprechend der Literaturempfehlungen von Johnston und O'Malley

(1997) getrennt abgefragt wurden. Des Weiteren erfragten zwei nach dem C-OAR-SE Ansatz der Skalenentwicklung selbst erstellte, inhaltssvalide Einzelitems (vgl. Rossiter 2002) die *Prokrastination* im Studium und die *Häufigkeit des Treffens* mit Kommiliton*innen. Die Befragten schätzten die Häufigkeit dieser Verhaltensweisen zumeist auf einer siebenstufigen (1 = „nie“ bis 7 = „(fast) immer“) validierten, vollverbalisierten Antwortskala ein (Mohr und Müller 2004, S. 2) mit der Ausweichoption „keine Angabe“. *Alkoholkonsum* wurde, auf eine durchschnittliche Woche bezogen, auf einer fünfstufigen, vollverbalisierten Skala abgefragt (1 = „nie“ bis 5 = „jeden Tag“). Für die Variable *Kontrollverlust Alkoholkonsum* wurde in Folge des Pretests die genutzte Häufigkeitsskala auf vier Antwortoptionen gekürzt (1 = „nie“, 2 = „selten“, 3 = „oft“, 4 = „(fast) immer“). Die sechs Fragen wurden jeweils als Einzelitems dargestellt.

Sieben weitere aus der Literatur extrahierte, dichotome Items (1 = „Ja“, 2 = „Nein“) erfragten, ob bestimmte sensible Verhaltensweisen im eigenen Leben schon vorkamen, und wurden in der Reihenfolge rotiert als Matrix präsentiert (S. 7). Die Items umfassten *Stehlen* in Geschäften (Ong und Weiss 2000, S. 1696), *Untreue* in der Partnerschaft (Toninelli und Revilla 2016b, S. 166) – umgekehrt formuliert als stete Treue –, *Scham nach Alkoholkonsum* (*Scham Alkohol*) (Mavletova und Couper 2013, S. 204), *Konsum harter Drogen* abseits von Alkohol und Marihuana (Gnambs und Kaspar 2015, S. 1242; Tourangeau, Smith und Rasinski 1997b, S. 215), *Zahlung oder Annahme von Geld für sexuelle Handlungen* (*Geld für Sex*) (Gnambs und Kaspar 2015, S. 1242), *alkoholisierte Fahrzeugführung* (*Alkoholisiertes Fahren*) (Gnambs und Kaspar 2015, S. 1242) und *Erschleichen einer Krankschreibung* durch Krankheitssimulation (Mavletova und Couper 2013, S. 204; Toninelli und Revilla 2016b, S. 166). Erneut konnte die Antwort umgangen werden („keine Angabe“).

Die nachfolgende Seite 8 präsentierte eine Kurzskala zur sozialen Erwünschtheit bestehend aus sechs Items (Kemper et al. 2012, S. 25), die um vier weitere Items der Aufdeckung sozial erwünschter Verhaltensweisen (Stöber 1999, S. 175) ergänzt wurde. Die insgesamt zehn Items wurden rotiert und auf einer siebenstufigen, vollverbalisierten Likert-Skala (1 = „trifft überhaupt nicht zu“ bis 7 = „trifft vollständig zu“) beantwortet (z. B. Janke und Glöckner-Rist 2012, S. 2). Die interne Konsistenz der zehn Items ($\alpha = 0,647$) charakterisiert den Skalenmittelwert (*Soziale Erwünschtheit*) nur bedingt als reliabel. Gute interne Konsistenz erfordert Werte von Cronbachs Alpha (CA) von $\alpha > 0,7$ (Taber 2018), sodass die Beurteilung der sozialen Erwünschtheit sowohl über den Mittelwert als auch separat jeweils für die einzelnen Items erfolgt.

Auf Seite 9 teilten die Befragten ihren *Familienstand* mit und bewerteten auf derselben siebenstufigen Zufriedenheitsskala wie zuvor (Fischer et al. 2021, S. 3) ihre *Lebenszufriedenheit* (Mavletova 2013, S. 739). Als letzte sensible Themenbereiche gaben die Versuchspersonen per

Freitextantwort die durchschnittliche Häufigkeit des *Sport*-Treibens pro Woche (gilt ebenfalls als sensible Information, wobei geringere Werte aufrichtigeres Antworten signalisieren (Tourangeau et al. 1997b)), und die Häufigkeit des *Geschlechtsverkehrs* im letzten Monat an. Zum Befragungsende (S. 10) wurde nach Whelan (2007, S. 6) ebenfalls auf siebenstufiger, vollverbalisierter Skala des Zutreffens erfragt (Janke und Glöckner-Rist 2012, S. 2), inwiefern die Befragung als vertraulich (*Vertraulichkeit*) und anonym (*Anonymität*) erlebt wurde.

Der schriftliche Fragebogenteil nach Wechsel des Raums wurde in Anwesenheit der Versuchsleitung ausgefüllt. Zunächst wurde erneut der individuelle Befragungscode generiert (S. 1). Im Anschluss beantworteten die Befragten, ob sie schon an wissenschaftlichen Studien (*Teilnahmeerfahrung*) teilgenommen haben, und falls ja, an wie vielen im letzten Jahr (*Teilnahme im letzten Jahr*). Ein inhaltsvalides, nach dem C-OAR-SE Ansatz (vgl. Rossiter 2002) erstelltes Einzelitem erhob über eine endpunktverbalisierte, siebenstufige Skala (1 = „sehr niedrig“ bis 7 = „sehr hoch“) das generelle *Vertrauen in wissenschaftliche Untersuchungen*.

Im nächsten Abschnitt (S. 1+2) wurde die Nutzungserfahrung für die beiden potenziellen Ausfüllgerätkategorien PC/Laptop und Smartphone adressiert. Die Proband*innen gaben an, seit wann sie die (Internet-)Funktionen eines PCs/Laptops bzw. eines Smartphones verwenden (*PC-Erfahrung* und *SP-Erfahrung*) und beurteilten auf einer siebenstufigen, vollverbalisierten Skala (1 = „nie“ bis 7 = „mindestens alle 10 Minuten“), wie häufig sie diese an einem durchschnittlichen Tag nutzen (*PC-Frequenz* und *SP-Frequenz*) (Vorderer et al. 2016, S. 699). Zudem wurde bei allen Teilnehmer*innen die Technologieakzeptanz explizit für das Smartphone ermittelt. Dafür wurden sechs reflektive Items, jeweils drei für *WN* und drei für *WE_dN*, aus der Literatur identifiziert (Davis et al. 1989, S. 991) und auf den Kontext des Studiums angepasst. Die sehr hohen internen Konsistenzwerte ($\alpha > 0,95$) der Ursprungsskala (Davis 1993) deuten auf unnütze Redundanz hin, sodass die Skalen in verschiedenen Studien bereits gekürzt auf jeweils unterschiedliche technische Systeme angewendet wurden (Jockisch 2009, S. 246; Kim und Sundar 2014; Venkatesh und Bala 2008; Venkatesh, Thong und Xu 2012). Auch für Smartphones existieren angepasste TAM-Skalen, die als direkte Vorlage für die Umsetzung in der vorliegenden Studie herangezogen werden konnten (Joo und Sang 2013; Park und Chen 2007). Die sechs Items für die Technologieakzeptanz wurden umsichtig im Abgleich mit den zitierten Literaturquellen formuliert. Sie resultierten in drei reflektiven Items der *WE_dN*: „*Den Umgang mit meinem Smartphone zu lernen, fiel mir leicht*“, „*Ich finde es einfach, mein Smartphone zu bedienen*“ und „*Der Umgang mit meinem Smartphone ist für mich klar und verständlich*“ sowie in drei Items für *WN*: „*Die Nutzung meines Smartphones erlaubt es mir, Aufgaben schneller zu erledigen*“, „*Die Nutzung meines Smartphones erhöht meine Effektivität im Studium*“ und „*Die*

Nutzung meines Smartphones erleichtert das Studium“ und wurden ebenfalls anhand der siebenstufigen, vollverbalisierten Likert-Skala des Zutreffens beantwortet (Janke und Glöckner-Rist 2012, S. 2). Als Maß der *WN* ($\alpha = 0,762$) und der *WE_{dN}* ($\alpha = 0,839$) agierte jeweils der Skalenmittelwert. Die interne Konsistenz war für *WN* und *WE_{dN}* jeweils angemessen ($\alpha > 0,7$).

Neben der Technologieakzeptanz wurde auch die Smartphonebindung (S. 3) zweistufig erfasst: Zum einen als reflektive Skala, bestehend aus acht Items, die auf einer siebenstufigen, vollverbalisierten Likert-Skala (1 = „trifft überhaupt nicht zu“ bis 7 = „trifft vollständig zu“) beantwortet wurden (Walsh et al. 2010, S. 199) und deren Skalenmittelwert ($\alpha = 0,750$) die Bindung (*Attachment1*) repräsentierte. Zum anderen wurde die Smartphonebindung als gradu-elles Einzelitem in Analogie zur Erfassung der Markenbindung konzeptualisiert (Park et al. 2010, S. 6). Die zugehörige ursprüngliche Antwortskala wurde in ein siebenstufiges semantisches Differential umgewandelt (Jiménez und Voss 2014), sodass die Proband*innen das Ausmaß ihrer persönlichen Bindung zum Smartphone (*Attachment2*) auf einem numerisch verankerten Kontinuum zwischen 0 = „Ich empfinde keine persönliche Beziehung zu meinem Smartphone“ und 6 = „Ich empfinde eine starke persönliche Beziehung zu meinem Smartphone“ verorten konnten. Zusätzlich wurde – analog zur Erfassung der Markenbedeutung im Alltag (Park et al. 2010, S. 6) – auf der zuvor verwendeten siebenstufigen, vollverbalisierten Häufigkeits-skala (1 = „nie“ bis 7 = „(fast) immer“) erfragt, wie häufig die eigenen Gedanken im Alltag zum Smartphone wandern (*Gedankenfluss*). In Ergänzung (S. 4) beurteilten die Befragten auf einer numerisch verankerten, siebenstufigen, endpunktverbalisierten Skala (1 = „gar nicht wahrscheinlich“ bis 7 = „sehr wahrscheinlich“) die Wahrscheinlichkeit der *zukünftigen Smartphone-nutzung* (Park und Chen 2007, S. 1363).

Abschließend (S. 4+5) gaben die Proband*innen an, mit welchem Gerät (*Wunschgerät*) sie den Onlineteil der Studie am liebsten bearbeitet hätten. Außerdem wurden jeweils auf fünfstufigen, numerisch verankerten, endpunktverbalisierten Skalen die *wahrgenommene Befragungsdauer* (1 = „sehr kurz“ bis 5 = „sehr lang“), das *Interesse* an der Befragung (1 = „gar nicht interessant“ bis 5 = „sehr interessant“) und die *Gesamtbewertung* der Befragung (1 = „sehr schlecht“ bis 5 = „sehr gut“) erfasst (de Bruijne und Wijnant 2013, S. 490-492). Zuletzt gaben die Proband*innen, ihr *Geschlecht* und ihr *Alter* an. Die gesamten Fragebogen des Onlineteils und des schriftlichen Teils der ersten Studie können in **Anhang A** eingesehen werden.

5.1.1.4 Operationalisierung der Hypothesen

Die Fragebogen der Studie 1 wurden aufbauend auf dem Forschungsmodell (siehe Kapitel 4.1) zur möglichst eindeutigen Überprüfung der zugehörigen Hypothesen entwickelt. *Tabelle 5* gibt

eine Übersicht über die Analysen und Variablenzuordnungen. Als unabhängige Variable (UV) fungiert in allen Analysen die Gruppenzugehörigkeit (*Gerät*) als dichotome nominale Variable, wobei der Laptop die Referenzkategorie (0) bildet und das Smartphone mit 1 kodiert wird. Für die Prüfung der Hypothese **H1.1a** (*Personen, die die Befragung mit dem Smartphone ausfüllen, berichten in höherem Ausmaß sensibles Verhalten als Personen, die die Befragung mit dem Laptop ausfüllen*) werden die Stichprobenkennwerte (z. B. Mittelwerte, Häufigkeiten) aller sensiblen Fragen zwischen den Ausfüllgeräten verglichen. Dies betrifft einerseits die Variablen *Prokrastination*, *Plagiarismus*, *Klausurbetrug*, *Alkoholkonsum* und *Kontrollverlust Alkoholkonsum*, die die Befragten jeweils auf vier- bis siebenstufigen Antwortskalen beantworteten. Ab fünf Stufen oder wenn mehrere Items zusammengefasst werden, können die Antworten der

Hypothese	UV	Kontrollvariablen	Analysemethode	AV
H1.1a und H1.1c	<i>Gerät</i>	Modell 1: <i>Geschlecht</i> <i>Alter</i> <i>WEEdN</i> <i>WN</i> <i>Attachment1</i> <i>Attachment2</i>	Kovarianzanalyse	<i>Prokrastination</i>
				<i>Plagiarismus</i>
				<i>Klausurbetrug</i>
				<i>Alkoholkonsum</i>
				<i>Kontrollverlust Alkoholkonsum (nur H1.1a)</i>
				<i>Geschlechtsverkehr</i>
H1.1a		Modell 2: zusätzlich: <i>SP-Erfahrung</i> <i>SP-Frequenz</i> <i>Familienstand</i>	Logistische Regression	<i>Stehlen</i>
				<i>Untreue</i>
				<i>Scham Alkohol</i>
				<i>Drogen</i>
				<i>Geld für Sex</i>
				<i>Krankschreibung</i>
H1.1b	Kovarianzanalyse	<i>Summe Handlungen</i>		
		<i>Missings</i>		
H1.2	Logistische Regression	<i>Missings_dichotom</i>		
		<i>Soziale Erwünschtheit</i>		
				10 Einzelitems separat

Tabelle 5: Übersicht über das Auswertungsmodell von Studie 1

Befragten als quasimetrisch behandelt werden (Harpe 2015; Urban und Mayerl 2018, S. 301-302). Andernfalls werden nicht-parametrische Auswertungsmethoden herangezogen. Außerdem werden die beiden metrischen Variablen *Geschlechtsverkehr* und *Sport* einbezogen, bei denen die Häufigkeit des Geschlechtsverkehrs im letzten Monat und des wöchentlichen Sport-Treibens von den Versuchspersonen jeweils als absolute Zahl angegeben wurde. Zuletzt schließen die dichotomen Variablen *Stehlen*, *Untreue*, *Scham Alkohol*, *Drogen*, *Geld für Sex*, *Krankschreibung* und *Alkoholisiertes Fahren* die Hypothesenprüfung ab. Für diese wird noch ein Summenscore (*Summe Handlungen*) erstellt und ergänzend analysiert, der angibt, bei wie vielen der Items die Befragten sensibles Verhalten berichten. In alle Analysen werden zudem relevante Kontrollvariablen berücksichtigt. Aus der Literatur resultiert die Empfehlung, *Alter* und *Geschlecht* bei sensiblen Themen zu kontrollieren (z. B. Toninelli und Revilla 2016b). Auch die besondere psychologische Bedeutung des Smartphones muss berücksichtigt werden, was über die Variablen *WEdN*, *WN*, *Attachment1* und *Attachment2* erfolgt¹². Die genannten Kontrollvariablen formen gemeinsam mit der UV *Gerät* das Grundmodell der Analysen (*Modell 1*). Ein zweites Modell (*Modell 2*) erweitert die Betrachtung um mögliche weitere Drittvariablen der Nutzungserfahrung des Smartphones (*SP-Erfahrung* und *SP-Frequenz*). Aufgrund inhaltlicher Überlegungen wird zusätzlich der *Familienstand* (dichotom: 0 = Single, 1 = liiert) integriert, da dieser mit einigen der betrachteten sensiblen Themenbereiche (z. B. *Geschlechtsverkehr*) zusammenhängen könnte.

Für die Analyse der Hypothese **H1.1b** (*Personen, die die Befragung mit dem Smartphone ausfüllen, antworten häufiger auf sensible Fragen als Personen, die die Befragung mit dem Laptop ausfüllen*) wird die Anzahl fehlender Werte (*Missings*) für die 14 sensiblen Einzelitems als Indikator der Antwortbereitschaft ermittelt. Diese wird zusätzlich auch dichotomisiert (0 = keine Missings, 1 = ein Missing oder mehr) ausgewertet (*Missings_dichotom*). Als Missings gelten in Studie 1 sowohl fehlende Werte als auch nicht-substanzielle Antworten, wenn Ausweichkategorien (z. B. „keine Angabe“) gewählt werden.

Die Hypothese **H1.1c** (*Personen, die die Befragung mit dem Smartphone ausfüllen, antworten mit größerer Varianz auf sensible Fragen als Personen, die die Befragung mit dem Laptop ausfüllen*) betrachtet die Variation in den Antworten der Befragten auf sensible Fragen. Dafür werden die Antwortvarianzen aller mindestens quasimetrischen sensiblen Variablen (*Geschlechtsverkehr*, *Sport*, *Prokrastination*, *Plagiarismus*, *Klausurbetrug* und *Alkoholkonsum*) gerätweise verglichen, wobei höhere Varianz eine breitere Nutzung der Antwortskalen anzeigt.

¹² Zusätzlich hätte auch die Variable *Gedankenfluss* integriert werden können. In Kapitel 5.1.2.1 wird die Begrenzung auf die beiden Attachment-Variablen ausgeführt.

Für die Prüfung der Hypothese **H1.2** (*Personen, die die Befragung mit dem Smartphone ausfüllen, antworten weniger sozial erwünscht als Personen, die die Befragung mit dem Laptop ausfüllen*) wird primär die Variable *Soziale Erwünschtheit* als Mittelwert über alle zehn Einzelitems analysiert. Zudem werden die Antworten auf alle zehn Items einzeln ausgewertet, um mögliche sozial erwünschte Antworttendenzen engmaschig aufdecken zu können.

5.1.2 Ergebnisse

Die Datenauswertung erfolgte situationsadäquat mit den Auswertungsprogrammen SPSS (IBM Corp. 2021), JASP (JASP Team 2023) und R (R Core Team 2023). *Tabelle 5* ordnet die Analysemethoden den untersuchten abhängigen Variablen (AVn) zu. Für mindestens quasimetrische AVn werden einfaktorielle Kovarianzanalysen (ANCOVAs) in Form allgemeiner linearer Modelle (ALM) durchgeführt (z. B. Bortz und Schuster 2016, S. 305-319). Die dichotomen AVn werden über multiple logistische Regressionen analysiert (z. B. Backhaus et al. 2021, S. 289-382). Da die Gruppenzugehörigkeit dichotom ist, werden in der Ergebnisdarstellung der logistischen Regression die nicht-standardisierten Regressionsgewichte (*b*) berichtet. Bei der ANCOVA kann bei Verletzung der Voraussetzungen relativ flexibel auf nicht-parametrische Prüfkriterien des UV-Einflusses ausgewichen werden, sodass die erste Voruntersuchung die ANCOVA nutzt, um gegebenenfalls situationsadäquat reagieren zu können.

5.1.2.1 Voranalysen und Voraussetzungsprüfung

Da Studie 1 auf Unterschiede in Datenstrukturen zielt und ein kontrolliertes Laborsetting vorliegt, wird keine herkömmliche Ausreißeranalyse vorgenommen. Gerade besondere Datenmuster und auch fehlende Werte liegen im Untersuchungsfokus. Statt Proband*innen pauschal auszuschließen, variiert die Stichprobengröße der einzelnen Analysen, da Datenreihen situativ von diesen ausgeschlossen werden, wenn sie für eine der jeweils beteiligten Modellvariablen fehlende Werte oder extreme Ausreißer aufweisen.

Die logistische Regression erfordert unabhängige Beobachtungen, Fallzahlen von mindestens 25 pro Gruppe und dass die Daten möglichst frei von Multikollinearität sind, wobei die Analyse recht robust gegenüber Voraussetzungsverletzungen ist (Backhaus et al. 2021, S. 379). Die bivariaten Pearson-Produkt-Moment-Korrelationen (Bortz und Schuster 2016, S. 156-169) aller Kontrollvariablen der Analysemodelle sind teilweise signifikant: Die Länge des Smartphonebesitzes (*SP-Erfahrung*) geht mit höherer *WN* einher ($r = 0,266, p = 0,005$) und je häufiger ein Smartphone täglich genutzt wird (*SP-Frequenz*), desto eher ist auch die *WE_{dN}* höher ($r = 0,248, p = 0,009$). Zudem korrelieren *Alter* und *SP-Erfahrung* signifikant ($r = 0,525, p <$

0,001). Weibliche Personen (*Geschlecht*) sind eher liiert (*Familienstand*) ($r = 0,240, p = 0,015$), berichten höhere *SP-Frequenz* ($r = 0,242, p = 0,011$) und zeigen graduell abgefragt (*Attachment2*) eine stärkere Bindung zum Smartphone ($r = 0,331, p < 0,001$). *WN* und *WEdN* korrelieren aber nicht signifikant mit Maßen der Smartphonebindung. Die korrelativen Beobachtungen stützen die Aufteilung in zwei Analysemodelle mit unterschiedlicher Anzahl an Kontrollvariablen, um Multikollinearität aufzufangen. Erst im zweiten Modell werden die Variablen *SP-Erfahrung*, *SP-Frequenz* und *Familienstand* ergänzt, die wiederholt mit anderen Variablen des ersten Modells korrelieren. Zudem zeigen die drei Variablen *Attachment1*, *Attachment2* und *Gedankenfluss* hohe korrelative Zusammenhänge, wobei die Variable *Gedankenfluss* signifikant mit beiden Attachment-Variablen korreliert (*Attachment1*: $r = 0,611, p < 0,001$; *Attachment2*: $r = 0,417, p < 0,001$), die wiederum selbst miteinander zusammenhängen ($r = 0,433, p < 0,001$). Beide Attachment-Variablen werden aufgrund der konzeptionellen Unterschiede und des divergierenden Abfrageformats beibehalten. Die Variable *Gedankenfluss* verspricht hingegen eher keinen zusätzlichen Erkenntnisgewinn und geht nicht in die Analysen ein.

Die Voraussetzungen für ANCOVAs in Form von ALMs bestehen aus Linearität, Normalverteilung der AVn, Homoskedastizität als Varianzhomogenität der Residuen, geringer Multikollinearität und Homogenität der Regressionssteigungen innerhalb der verglichenen Gruppen, wobei die Analysen recht robust bei Verletzungen sind (Bortz und Schuster 2016, S. 311-348). Die Kennwerte der AVn nähern sich gemäß des zentralen Grenzwertsatzes ab einer Gruppengröße von $n > 30$, wie vorliegend, einer Normalverteilung an, wodurch die Voraussetzung erfüllt ist (Bortz und Schuster 2016, S. 586). Im Rahmen der Analysen wird bei Verletzung der Homoskedastizität und auch der Varianzhomogenität der AVn der nicht-parametrische Kruskal-Wallis-Test zur Anwendung gebracht (z. B. Leonhart 2022, S. 274). Die Homogenität der Regressionssteigungen wurde durch die Bildung aller möglichen Interaktionsterme der UV *Gerät* mit mindestens quasimetrische Kontrollvariablen für alle AVn überprüft und war jeweils gegeben ($p > 0,05$)¹³. Alle Voraussetzung werden stets geprüft und etwaige Verletzungen dokumentiert. Für die dichotomen Kontrollvariablen (*Geschlecht*, *Familienstand*) wird kein expliziter Interaktionseffekt mit der nominalen UV (*Gerät*) ergänzt, da aufgrund des Laborsettings und der Gruppenzuweisung kein inhaltlich begründeter Zusammenhang für die nominalen Kontrollvariablen mit dem genutzten Ausfüllgerät vermutet wird und auch die Voranalysen keine systematischen Zusammenhänge nahelegen¹⁴.

¹³ Die einzigen Ausnahmen (bei den insgesamt über 60 berechneten Interaktionseffekten) zeigten sich bei der AV *Geschlechtsverkehr* für die Kontrollvariable *Attachment2* ($p = 0,004$) und bei der AV *Missings* für die Kontrollvariable *SP-Erfahrung* ($p = 0,032$). In beiden Fällen verändern sich die Befundmuster der zugehörigen ANCOVAs nicht, wenn die jeweils betroffene Kontrollvariable aus den Analysen ausgeschlossen wird.

¹⁴ Dies wurde in allen Analysen überprüft. Der Einbezug der Interaktionseffekte ändert die Befundmuster nicht.

Die weiteren Voranalysen über unabhängige t -Tests und χ^2 -Tests vergleichen die Gruppen der Laptop- und Smartphone-Teilnehmer*innen. Die Voraussetzungen des χ^2 -Tests, dass die Beobachtungen unabhängig sind, jede Versuchsperson einer Zelle eindeutig zugeordnet werden kann und jede Zelle mindestens eine Beobachtung aufweist (Bühner und Ziegler 2017, S. 353), sind in allen Fällen erfüllt. t -Tests erfordern unabhängige Beobachtungen, Normalverteilung der AV in den Gruppen, Intervallskalenniveau der AV und Varianzhomogenität (Bühner und Ziegler 2017, S. 301). Die Intervallskalierung wird über die Quasimetrik mindestens fünfstufiger Skalen gesichert (Harpe 2015; Urban und Mayerl 2018, S. 301-302) und es greift der zentrale Grenzwertsatz für die Normalverteilung. In allen Analysen wird Varianzhomogenität geprüft und liegt bei den berichteten Ergebnissen vor, wenn nicht anders angegeben. Die Voranalysen zeigen, dass keine signifikanten demografischen Unterschiede zwischen Laptop- und Smartphone-Teilnehmer*innen bestehen und sich auch die relevanten Kontrollvariablen nicht zwischen den Gruppen unterscheiden. Zusätzlich sichert die Betrachtung der Geräteigenschaften die Gruppenzuteilung ab. Die Studierenden nutzten zwar ihr eigenes Ausfüllgerät mit individuell differierenden Eigenschaften, die Smartphones der Smartphone-Gruppe sind dabei aber unzweifelhaft kleiner als die Laptops der anderen Gruppe und besitzen im Mittel eine signifikant geringere *Bildschirmbreite* ($t(98,55) = 39,278, p < 0,001$), *Bildschirmhöhe* ($t(109) = 6,265, p < 0,001$) und auch eine geringere Breite der Fragebogendarstellung (*Fragebogenbreite*) ($t(56) = 15,550, p < 0,001$). Für beide Breitenmaße liegt keine Varianzhomogenität vor (*Bildschirmbreite*: $F(1, 109) = 7,109, p = 0,009$; *Fragebogenbreite*: $F(1, 109) = 42,460, p < 0,001$), weshalb hier die Ergebnisse des Welch- t -Tests als Alternative für ungleiche Varianzen berichtet werden.

5.1.2.2 Hypothesenprüfung

Der Hypothese **H1.1a** liegt die Vermutung zu Grunde, dass am Smartphone aufrichtiger auf sensible Fragen geantwortet wird. Insgesamt werden 15 AVn für die Hypothesenprüfung herangezogen und jeweils beide Analysemodelle (*Modell 1* und *Modell 2*) angelegt. Die Analysen entsprechen der Darstellung in *Tabelle 5*. Bei der Analyse der AV *Geschlechtsverkehr* wird die Variable *Familienstand* einmalig schon in das erste Analysemodell (*Modell 1*) integriert. Berichtet werden stets die Ergebnisse des Grundmodells (*Modell 1*) und zusätzlich die Ergebnisse von *Modell 2*, falls sich Abweichungen von den Ergebnissen des Grundmodells zeigen. *Modell 1* liefert unter Umständen etwas robustere Daten, da die Datenbasis (im Schnitt $\bar{n} = 98,50$ Datenreihen) aufgrund fehlender Werte der zusätzlichen Kontrollvariablen des *Modells 2* ($\bar{n} = 87,75$) reichhaltiger ist.

Bei den Kovarianzanalysen sollten die Mittelwerte sensibler Items am Smartphone höher ausfallen und dadurch mehr sensibles Verhalten offenbaren als am Laptop. Für die ordinal skalierte, vierstufige Variable *Kontrollverlust Alkoholkonsum* erfolgt die Prüfung über den nicht-parametrischen Kruskal-Wallis-Test. Da für die AVn *Prokrastination* ($F(1, 102) = 4,863$, $p = 0,030$) und *Klausurbetrug* ($F(1, 101) = 7,730$, $p = 0,006$) keine Varianzhomogenität vorliegt, kommt auch hier der Kruskal-Wallis-Test zu Anwendung. Bei den logistischen Regressionen vermutet die **H1.1a** signifikante positive Regressionsgewichte der Variable *Gerät*, sodass am Smartphone häufiger sensibles Verhalten zugegeben wird. Die Variable der steten Treue wird vorab umkodiert in *Untreue*, damit auch hier die Antwortkategorie 1 geäußerte *Untreue* indiziert. Für die 15 Analysen zur Prüfung der **H1.1a** entsteht kein einheitliches Befundbild. Die wenigen signifikanten Unterschiede sind meist kontraintuitiv und offenbaren entgegen der Hypothese eher mehr berichtete sensible Verhaltensweisen am Laptop. Die logistischen Regressionen zeigen, dass am Laptop die Befragten häufiger angeben, nach Alkoholkonsum am nächsten Tag Scham empfunden (*Scham Alkohol*) zu haben ($b = -1,291$, $W(1) = 6,779$, $p = 0,009$) und schon alkoholisiert ein Fahrzeug (*Alkoholisiertes Fahren*) geführt zu haben ($b = -1,083$, $W(1) = 3,966$, $p = 0,046$). In allen Analysen spricht der Varianzinflationsfaktor (VIF) gegen kritische Multikollinearität, da auch konservative Grenzwerte deutlich nicht erreicht werden und die Werte jeweils knapp (alle $VIF < 1,8$) über dem optimalen Wert $VIF = 1$ liegen (Backhaus et al. 2021, S. 100). Die Kovarianzanalysen finden ebenfalls kaum signifikante Gruppenunterschiede. Für die Variable *Summe Handlungen* als Summenscore der dichotomen sensiblen Verhaltensweisen fällt die Selbstoffenbarung am Laptop entgegen der Vermutungen höher aus als am Smartphone ($F(1, 101) = 6,735$, $p = 0,011$). Lediglich bei der Variable *Sport* ist der signifikante Unterschied hypothesenkonform: Proband*innen, die mit dem Smartphone teilnehmen, berichten weniger wöchentliche sportliche Aktivität ($F(1, 102) = 4,821$, $p = 0,030$). Höhere Werte könnten hier eine Überschätzung indizieren, sodass niedrigere Angaben eher als aufrichtige Antwort gelten. Für alle Analysen sind die Ergebnisse für *Modell 1* und *Modell 2* hinsichtlich signifikanter Einflüsse und der Größe der Koeffizienten nahezu gleich.

Für die sechs mindestens quasimetrischen AVn (*Prokrastination*, *Plagiarismus*, *Klausurbetrug*, *Alkoholkonsum*, *Sport*, *Geschlechtsverkehr*) wird im Rahmen der Kovarianzanalysen auch die Hypothese **H1.1c** untersucht, indem die Antwortvarianz der Untersuchungsgruppen verglichen wird. Dies geschieht konkret über die Prüfung der Varianzhomogenität der AVn. Sollte sich das Antwortverhalten auf sensible Fragen unterscheiden und je nach Ausfüllgerät mehr Aufdeckung ermöglichen, wäre der Wertebereich der Antworten in der Smartphone-Gruppe breiter. Erneut sind die Befundmuster uneinheitlich. Varianzheterogenität liegt lediglich bei den

Variablen *Prokrastination* ($F(1, 102) = 4,863, p = 0,031$) und *Klausurbetrug* ($F(1, 101) = 7,730, p = 0,006$) vor. Dabei ist die Standardabweichung am Smartphone nur bei *Prokrastination*, wie erwartet, signifikant größer als am Laptop ($SD_{Smartphone} = 1,55 > SD_{Laptop} = 1,31$) – bei *Klausurbetrug* hingegen geringer ($SD_{Smartphone} = 0,90 < SD_{Laptop} = 1,42$).

Die Prüfung der Hypothese **H1.1b**, ob Befragte am Smartphone häufiger auf sensible Fragen antworten, rundet die Analysen zu sensiblen Fragen ab. Die Gesamtzahl der *Missings* wird über eine ANCOVA analysiert und offenbart entgegen der Hypothesenformulierung tendenziell mehr *Missings* am Smartphone ($F(1, 108) = 3,238, p = 0,075$). Dieser Effekt verschwindet in *Modell 2* ($F(1, 89) = 0,608, p = 0,437$). Hier könnte ein Bodeneffekt vorliegen, da der Mittelwert der *Missings* in beiden Untersuchungsgruppen recht gering ist ($\bar{x}_{Smartphone} = 0,59 > \bar{x}_{Laptop} = 0,33$). Die dichotomisierte Missing-Variable (*Missings_dichotom*) liefert einen ähnlichen Befund. Am Smartphone wird signifikant häufiger mindestens eine sensible Frage nicht beantwortet ($b = 1,178, W(1) = 5,640, p = 0,018$). Erneut zeigt sich der Einfluss in *Modell 2* nicht ($b = 0,931, W(1) = 2,613, p = 0,106$). Insgesamt deuten die Ergebnisse, anders als vermutet, auf eine Tendenz zu mehr nicht-substanziellen Antworten bzw. zu mehr fehlenden Werten und somit geringerer Antwortbereitschaft am Smartphone hin.

Die finale Hypothese **H1.2** vermutet Unterschiede im Ausmaß sozial erwünschten Antwortens. Per ANCOVA werden die Gesamtvariable *Soziale Erwünschtheit* und alle Einzelitems als quasimetrische Variablen analysiert. Für *Soziale Erwünschtheit* liegt kein signifikanter Unterschied vor ($F(1, 102) = 0,803, p = 0,372$) und für die einzelnen Variablen entsteht erneut kein einheitliches Bild. Bei acht der zehn Variablen unterscheiden sich die Antworten nicht. Die Gruppe der Smartphone-Teilnehmer*innen gibt an, anderen Menschen weniger gut zuzuhören ($F(1, 102) = 4,567, p = 0,035$) und sich weniger gesund zu ernähren ($F(1, 102) = 4,919, p = 0,029$). Damit wird in diesen Fällen am Smartphone weniger sozial erwünscht geantwortet. In beiden Fällen ist der Unterschied in *Modell 2* nicht signifikant (*Zuhören*: $F(1, 89) = 3,042, p = 0,085$; *Ernährung*: $F(1, 89) = 2,466, p = 0,120$). Aufgrund der multiplen Testung kumuliert hier der alpha-Fehler (vgl. Holm 1979), sodass die wenigen signifikanten Unterschiede auch auf zufällige Fehler in der Hypothesenprüfung zurückzuführen sein könnten, was gemeinsam mit der Uneinheitlichkeit der Befunde gegen einen bedeutsamen Effekt spricht. Auch über die berichteten Analysen hinaus sind die Befunde inkonsistent. Rein deskriptiv weisen Smartphone-Teilnehmer*innen bei vier der zehn Fragen höhere Mittelwerte auf als die Proband*innen der Laptop-Gruppe und antworten somit bei diesen Items deskriptiv sozial erwünschter. In allen Analysen zur sozialen Erwünschtheit zeigen die AVn Varianzhomogenität mit Blick auf die Hypothese **H1.1c**. *Tabelle 6* fasst die Befunde der Hypothesenprüfungen zusammen und

konstatiert, dass für alle vier Hypothesen die vermuteten Effekte nicht auftreten. Die wenigen Unterschiede deuten an, dass das Antwortverhalten zwar nicht an beiden Ausfüllgeräten komplett identisch ist, aber eher keine Moduseffekte auftreten, die sich in Unterschieden in den Stichprobenkennwerten bei sensiblen Fragen äußern. Die Antwortbereitschaft als Ausdruck der Item-Nonresponse scheint hingegen tendenziell am Smartphone höher zu sein, wobei die geringe Zahl der Missings eindeutige Rückschlüsse nicht zulässt.

Hypothese	Vermutung	Empirischer Befund	Entscheidung
H1.1a <i>Aufrichtigkeit bei sensiblen Fragen</i>	Smartphone > Laptop	Kein Unterschied	abgelehnt
H1.1b <i>Missings</i>	Smartphone < Laptop	Tendenz: Smartphone > Laptop	abgelehnt
H1.1c <i>Antwortvarianz</i>	Smartphone > Laptop	Kein Unterschied (aber variierende Antwortmuster)	abgelehnt
H1.2 <i>Soziale Erwünschtheit</i>	Smartphone < Laptop	Kein Unterschied	abgelehnt

Tabelle 6: Übersicht über die Befunde der ersten Studie

5.1.2.3 Zusatzbetrachtungen

Für sensible Themenbereiche ist die subjektive Bewertung der *Anonymität* und *Vertraulichkeit* wichtig. Sollte sich diese zwischen den Untersuchungsgruppen unterscheiden, würde dies eine Anpassung der Analysemodelle erfordern. Kovarianzanalysen mit allen Kontrollvariablen der *Modelle 1* und *2* finden jedoch keine Unterschiede zwischen den Untersuchungsgruppen bei der wahrgenommenen *Anonymität* ($F(1, 102) = 0,349, p = 0,556$) und der *Vertraulichkeit* ($F(1, 102) = 0,134, p = 0,716$). Auch wenn aus *Anonymität* und *Vertraulichkeit* ein Mittelwert gebildet wird, tritt für diesen kein Unterschied auf ($F(1, 102) = 0,246, p = 0,621$). Für die Abschlussbewertungen hinsichtlich des *Interesses* an den Studieninhalten ($F(1, 102) = 0,076, p = 0,784$), der subjektiven *Gesamtbewertung* der Studie ($F(1, 102) = 0,368, p = 0,546$) und der *wahrgenommenen Befragungsdauer* ($F(1, 102) = 0,001, p = 0,987$) zeigen sich ebenfalls keine Gerätunterschiede. Der letzte Befund ist dabei interessant, da die *wahrgenommene Befragungsdauer* offenbar nur bedingt mit der *Antwortzeit* korrespondiert. Der Vergleich der *Antwortzeit* (anhand eines unabhängigen *t*-Tests) ergibt, dass am Smartphone signifikant mehr Zeit für die Beantwortung des Fragebogens benötigt wird ($t(109) = 2,547, p = 0,012$). Auch wenn der lange Einführungstext ausgeklammert und nur die *reine Antwortzeit* verglichen wird, dauert die Bearbeitung mit dem Smartphone signifikant länger ($t(109) = 2,007, p = 0,047$). In beiden Fällen sind die Voraussetzungen für die Anwendung eines *t*-Tests erfüllt. Diese Befunde persistieren, wenn der Vergleich unter Einbezug der Kontrollvariablen als ANCOVA vorgenommen wird.

5.1.3 Diskussion

In Onlinebefragungen werden Smartphones vermehrt als Ausfüllgeräte verwendet (siehe Kapitel 2.2.2), weshalb die wissenschaftliche Forschung prüfen sollte, inwiefern durch das Smartphone womöglich Moduseffekte entstehen. Als Auftakt des vorliegenden Forschungsprogramms nähert sich die erste Studie über Stichprobenkennwerte und Antwortbereitschaft bei sensiblen Themen der Prüfung etwaiger Moduseffekte des Smartphones hinsichtlich der Antwortqualität an. Für die angestrebte Prüfung sind sensible Themenbereiche besonders geeignet, da Aufdeckungsbereitschaft und ehrliches Antworten individuell stark variieren können (Barnett 1998; Jann 2015, S. 102; Tourangeau und Yan 2007) und sensible Themen besonders modussensibel sind (Barnett 1998; Gnambs und Kaspar 2015; Jann 2015, S. 104). In der vorliegenden Studie werden dabei nach Kenntnisstand des Autors erstmals Kontexteinflüsse bei der Teilnahme mit dem Smartphone, die bisher meist die Grundlage der empirischen Untersuchungen bilden (vgl. Mavletova und Couper 2013; Toninelli und Revilla 2016b), durch das Laborsetting explizit ausgeschlossen. Zudem wird der Besitzfaktor konstant gehalten. Ausdrücklich berücksichtigt wird in Studie 1 hingegen die besondere emotionale und psychologische Bedeutung des Smartphones (z. B. Fullwood et al. 2017; Melumad und Pham 2020; Park und Kaye 2019; Walsh et al. 2011), die Kapitel 2.3 dezidiert ausführt. Smartphones sind sehr vertraute Geräte, die mit der eigenen Person verknüpft sind. Daher könnte die Bereitschaft, private und persönliche Informationen weiterzugeben, gerätspezifisch variieren und gerade am Smartphone höher sein und somit Moduseffekte bereithalten, die den TSE womöglich sogar senken.

Die Befunde der Studie 1 sprechen jedoch gegen Moduseffekte des Smartphones bei der Aufdeckung sensibler Verhaltensweisen. Personen, die mit unterschiedlichen Geräten an Befragungen teilnehmen, unterscheiden sich nicht systematisch in ihrer Bereitschaft, ehrlich auf sensible Fragen zu antworten. Da auf besonders sensible Fragen generell weniger bereitwillig geantwortet wird (Gnambs und Kaspar 2015), wurde der Grad der Sensibilität der Inhalte in der Befragung variiert. Entgegen den Hypothesen zeigt sich kein Vorteil des Smartphones als Ausfüllgerät für den aufrichtigen Selbstbericht über das gesamte Spektrum dieser Variation. Die erste Hypothese der Untersuchung (**H1.1a**) muss damit abgelehnt werden. Das Smartphone führt trotz der starken affektiven Funktion, die es für seine Besitzer*innen einnehmen kann (siehe Kapitel 2.3.2), nicht dazu, dass authentischer und ehrlicher auf sensible Fragen geantwortet wird, und es können keine Modusunterschiede erkannt werden. Ein Nulleffekt tritt auch für soziale Erwünschtheit auf. Das Ausmaß sozial erwünschten Antwortverhaltens wird eher nicht durch das Ausfüllgerät verändert, sodass auch die Hypothese **H1.2** abgelehnt werden muss. Die Isolierung des Geräteeffekts der vorliegenden Studie vervollständigt die Befundlage

zum Smartphone als Ausfüllgerät für sensible Themenbereiche: Am Smartphone werden sensible Fragen ähnlich aufrichtig beantwortet (Revilla et al. 2016a; Toninelli und Revilla 2016b) und das Gerät selbst übt keinen isolierten, kontextunabhängigen Einfluss aus. Die vorliegende Studie schließt somit Moduseffekte des Smartphones aus, die sich auf die Stichprobenkennwerte bei sensiblen Fragen auswirken. Daraus kann geschlossen werden, dass der Messfehler durch Befragte (**H1.1a** und **H1.2**) am Smartphone bei sensiblen Fragen eher nicht reduziert ist.

Aus den Ergebnissen der Prüfungen der Hypothesen **H1.1b** und **H1.1c** folgen aber Implikationen für das weitere Forschungsprogramm. Zunächst müssen auch diese grundsätzlich abgelehnt werden, da am Smartphone weder größere Antwortvarianzen bei sensiblen Fragen (**H1.1c**) noch weniger Missings (**H1.1b**) auftreten. Die Verteilungen der Antworten (und auch die Antwortvarianzen) zeigen in Studie 1 mitunter jedoch heterogene Strukturen, die sich zwar nicht in Unterschieden in den Stichprobenkennwerten manifestieren, aber die Prüfung auf weitere Moduseffekte des Smartphones als Ausfüllgerät hinsichtlich der Antwortqualität unterstreichen. Gerade die Antwortqualität hängt erheblich von solchen Verteilungsformen und auch der Antwortkonsistenz ab (z. B. Bautista 2012, S. 38; Peytchev 2013; Schlosser und Silber 2020, S. 278). Es bleibt also fraglich, ob mit verschiedenen Ausfüllgeräten erhobenes Datenmaterial holistisch vergleichbar ist. Auch Missings können als Antwortqualitätsmaß verwendet werden, wobei höhere Ausmaße reduzierte Antwortqualität signalisieren (z. B. Bosnjak und Tuten 2001; Göritz 2004; Heerwegh 2006; Sánchez-Fernández et al. 2012). In den vorliegenden Ergebnissen zeigt sich eine Tendenz, dass am Smartphone (anders als **H1.1b** vermutet) mehr Missings auftreten. Die Antwortbereitschaft ist am Smartphone also eher geringer und die Item-Nonresponse potenziell erhöht. Somit liegen eindeutig keine positiven Moduseffekte des Smartphones bei sensiblen Fragen vor. Da Missings hier sowohl fehlende als auch nicht-substantielle Antworten einschließen, werden nicht rein fehlende Werte untersucht. Die Anzahl der Missings ist zusätzlich äußerst gering und der gefundene Unterschied lediglich marginalsignifikant, so dass keine ganz eindeutigen Schlussfolgerungen für die Antwortqualität möglich sind. Die Studienbefunde stützen damit die Herleitung des theoretischen Teils der Arbeit in den Kapiteln 2 und 3: In den Hauptuntersuchungen sollte die Antwortqualität als Teilbereich des TSE fokussiert werden, da hierbei negative Moduseffekte des Smartphones besonders plausibel sind.

In den Zusatzbetrachtungen sind die Ergebnisse der Analysen der objektiv gemessenen Antwortzeit konform zu den bisherigen Befunden der Forschungsliteratur: Für die Teilnahme mit dem Smartphone wird mehr Zeit benötigt (z. B. Antoun und Cernat 2020; Couper und Peterson 2017; Keusch und Yan 2017; Revilla et al. 2016a). Dies ist auch in der hier durchgeführten ersten Studie der Fall, trotz vergleichsweise kurzer Gesamtdauer. Außerdem fällt auf, dass sich

die persönliche und emotionale Bedeutung des Geräts eher nicht auf die Ausfüllsituation in Onlinebefragungen überträgt. Werden Kontexteinflüsse (wie z. B. Teilnahmeort und Ablenkungsreize) konstant gehalten, verändert das Ausfüllgerät nicht die Einschätzung der Anonymität und Vertraulichkeit der Situation. Für die Analyse möglicher Moduseffekte des Smartphones als Ausfüllgerät sollte daher den Empfehlungen gefolgt werden, Kontexteffekte bei der Evaluation des Datenmaterials stets einzubeziehen (z. B. Gummer et al. 2021; Höhne et al. 2020a; Miller 2008; Pan et al. 2014). Entsprechend wird der Kontext der Teilnahme in den Hauptuntersuchungen der vorliegenden Arbeit gezielt berücksichtigt und ggf. kontrolliert.

Die Anlage der Studie als Laborexperiment maximiert die interne Validität der Analysen des reinen Geräteeffekts. Zugleich liegt in den fehlenden Außeneinflüssen aber eine klare Limitation. Der Charakter der Untersuchungssituation ist stets präsent und könnte die Aufdeckungsbereitschaft der Befragten systematisch beeinflussen. Die Antwortraten und die berichteten Ausmaße sensibler Verhaltensweisen sind zwar eher hoch, sodass vermutlich keine Bodeneffekte vorliegen, doch können mögliche Einflüsse der artifiziellen Laborsituation nicht ausgeschlossen werden. Zudem setzt sich die Stichprobe ausschließlich aus Studierenden zusammen. Reine Studierendensamples sind eher homogen und bilden die Zielpopulation nur bedingt ab, wenn über Studierende hinaus Schlussfolgerungen abgeleitet werden sollen (z. B. Henrich, Heine und Norenzayan 2010; Peterson 2001). Für die Studie 1 könnten beispielsweise die Technologieaffinität und die Vertrautheit mit verschiedenen Ausfüllgeräten vergleichsweise junger Studierender die Aussagekraft der Befunde einschränken. Außerdem kann die Stichprobe aufgrund der Freiwilligkeit der Teilnahme und der Möglichkeit der Verweigerung nicht als Zufallsauswahl bezeichnet werden. Ggf. setzt sich die Stichprobe daher durch Selbstselektionseffekte aus eher offenen und interessierten Personen zusammen, deren Antwortverhalten auch eher wohlwollend für die Studienleitung ausfällt. Möglicherweise erfolgen die Antworten daher insgesamt aufrichtiger und werden weniger vom Ausfüllgerät beeinflusst. Dies unterstreicht die Bedeutung der separaten personenzentrierten Analyse der Bedürfnisstrukturen der Befragten bei der Teilnahme an Onlinebefragungen in der anschließenden Studie 2. Diese löst sich von der Eingrenzung auf Studierende, verringert die Interaktion zwischen Versuchsleitung und Versuchsperson und legt den Betrachtungsschwerpunkt auf die Rolle des Ausfüllgeräts bei der Befriedigung verschiedener Bedürfnisse während der Befragungsteilnahme.

5.2 Studie 2 – Gerätspezifische Bedürfnisstrukturen der Befragungsteilnahme

Moduseffekte des Smartphones könnten im TSE-Framework auch für Unit-Nonresponse entstehen (siehe Kapitel 3.1.2). Da in Onlinebefragungen nicht direkt auf die Unit-Nonresponse

geschlossen werden kann, ist die zweite Studie als Voruntersuchung der Unit-Nonresponse angelegt. Sollte das Smartphone vorrangig von bestimmten Personengruppen für die Teilnahme genutzt werden, könnte dies Hinweise auf gerätspezifische Unit-Nonresponseraten liefern. Soziodemografische Unterschiede zwischen Personen, die verschiedene Ausfüllgeräte nutzen, werden stetig geringer und verlieren mit zunehmender Verbreitung des Smartphones an Bedeutung (siehe Kapitel 3.1.2.1). Stattdessen betrachtet die zweite Voruntersuchung des Forschungsprogramms die Bedürfnisstrukturen der Teilnahme. Womöglich gewichten Personen, die mit dem Smartphone teilnehmen, ihre Bedürfnisse für die Befragungsteilnahme anders als Personen, die PCs/Laptops einsetzen und dies drückt sich in der Gerätwahl aus (siehe Kapitel 4.2). Die Studie 2 der vorliegenden Arbeit orientiert sich an der Untersuchung von de Bruijne und Wijnant (2014). Statt die Gerätwahl lediglich auf Soziodemografika zu regressieren, wird der Untersuchungsansatz auf Teilnahmebedürfnisse übertragen und der prädiktive Wert dieser Bedürfnisse für die Gerätwahl geprüft. Die Befragten nehmen im Feld mit einem selbstgewählten Gerät an einer Onlinebefragung teil und beantworten Fragen zu ihren Einstellungen, Erfahrungen und Bedürfnissen bei der Befragungsteilnahme.

5.2.1 Methode

Auch in Studie 2 wurde die erforderliche Stichprobengröße im Vorfeld mit G*Power ermittelt (Faul et al. 2007; Faul et al. 2009). Angelegt wurde eine kleine Effektstärke, um potenzielle systematische Einflüsse der Bedürfnisse auf die Gerätwahl möglichst sicher aufzudecken. Als Untersuchungsmodell wird gemäß der Studie von de Bruijne und Wijnant (2014) die logistische Regression eingesetzt, für die äquivalent zur kleinen Effektstärke nach Cohen (Cohen's $d = 0,2$) ein Odds-Ratio-Wert von 1,68 angelegt wird (Chen, Cohen und Chen 2010). Zudem wird festgelegt, dass die Varianzaufklärung der in der Hypothesenprüfung jeweils betrachteten Prädiktoren durch die anderen Modellprädiktoren maximal 50% beträgt. Dies führt zu einer Stichprobe von mindestens $N = 266$. Die Studie ist als Between-Subjects-Design angelegt.

5.2.1.1 Versuchsablauf

Die Onlinebefragung wurde auf der Befragungsplattform SoSci Survey (Leiner 2023) angelegt, über die die Datenerhebung im Zeitraum vom 11.06. bis zum 02.07.2019 stattfand. Die Onlinebefragung beinhaltete 50 Items und die Befragten benötigten im Durchschnitt 7,08 Minuten für die Beantwortung. Die Einladung zur Teilnahme wurde am Hauptcampus der BUW und im Umfeld projektbeteiligter Studierender breit gestreut. Die Befragung sollte nicht mit einem bestimmten Ausfüllgerät einfacher aufgerufen werden können. Dafür wurde der Einladungstext

mit dem Aufruf zur Befragungsteilnahme, der auch den gekürzten Befragungslink enthielt, ausgedruckt. Dies ermöglichte den Versuchspersonen, selbstbestimmt teilzunehmen und auch das Ausfüllgerät unabhängig vom Format der Einladung zu wählen.

5.2.1.2 Stichprobe

Die Befragung wurde insgesamt von 448 Personen aufgerufen. 44 Personen nahmen die Befragung nach dem Aufruf der Webseite gar nicht erst auf. Als notwendiges Kriterium der Teilnahme mussten die Befragten sowohl über ein eigenes Smartphone als auch über einen eigenen PC/Laptop verfügen, da andernfalls die freie Gerätwahl nicht möglich war. 36 potenzielle Teilnehmer*innen erfüllten diese Grundbedingung nicht (6 ohne Smartphone, 28 ohne PC/Laptop, 2 ohne Angabe). Zudem mussten 26 weitere Proband*innen ausgeschlossen werden, die die Bearbeitung kurz nach Beginn abbrachen oder zu viele Antworten ausließen und somit nicht in die Schätzung des Regressionsmodells einbezogen werden konnten. Von den verbliebenen 342 Versuchspersonen nahmen 39 (11,4%) per Tablet und eine Person mit einem Convertible teil und wurden ebenfalls ausgeschlossen. Abschließend wurden elf sehr junge Personen (jünger als 16 Jahre) aus dem Datensatz entfernt, wodurch in der Gesamtstichprobe $N = 291$ Datenreihen verbleiben. Der Altersschnitt der finalen Stichprobe lag bei 36,8 Jahren ($SD = 15,90$) – die jüngste Person war 16 und die älteste 78 Jahre alt. 45,1% ($n = 128$) waren männlich und 54,9% ($n = 156$) weiblich. 154 Teilnehmer*innen nutzten den PC/Laptop als Ausfüllgerät und 137 das Smartphone. Sowohl die Häufigkeiten des Geschlechts ($\chi^2(1) = 2,761, p = 0,097$) als auch des Ausfüllgeräts ($\chi^2(1) = 0,993, p = 0,319$) unterschieden sich in der Stichprobe nicht signifikant.

5.2.1.3 Fragebogenbeschreibung und Operationalisierung der Hypothesen

Der Feldphase der Erhebung wurde ein Pretest mit ausgewählten Teilnehmer*innen aus dem methodischen Fachbereich der Fakultät der Wirtschaftswissenschaft an der BUW sowie mit neun Befragten, die die letztliche Zielstichprobe repräsentierten (mindestens 16 Jahre alt, Besitz eines Smartphones und PCs/Laptops) vorgeschaltet. Mit dem Ziel, mögliche Verständnisschwierigkeiten und Unklarheiten zu reduzieren, wurden die ursprünglichen Befragungsinhalte dann ggf. modifiziert. Im Folgenden wird die finale Version der Onlinebefragung beschrieben. Die Bezeichnungen der Variablen werden kursiv angelegt.

Zu Beginn der Befragung (S.1) wurde den Teilnehmer*innen das Untersuchungsziel verkürzt vorgestellt. Den Proband*innen wurde als Befragungsthema die optimale Gestaltung moderner Onlinebefragungen genannt, die sich an den individuellen Bedürfnissen und Wünschen für die Befragungsteilnahme ausrichten wolle. Da die freie Wahl des Ausfüllgeräts zugelassen

werden solle, werde auch die persönliche Wahrnehmung verschiedener Ausfüllgeräte integriert. Lediglich das eigentliche Untersuchungsziel, die Bedürfnisse für die Teilnahme auf die Gerätewahl zu beziehen, wurde zunächst nicht offengelegt, um mögliche Übertragungseffekte der tatsächlichen Gerätewahl auf die Antworten zu minimieren. Alle Befragten stimmten der Verarbeitung und Nutzung ihrer Daten zu rein wissenschaftlichen Zwecken im Rahmen der europäischen DSGVO zu. Die folgende Seite (S. 2) bat darum, bei der Nutzung eines Smartphones im horizontalen Landscape-Modus an der Befragung teilzunehmen. Dies sicherte die übersichtliche und gerätübergreifend annähernd gleiche Darstellung des Fragebogens. Anschließend (S. 3) wurde geprüft, ob die Versuchspersonen die Teilnahmevoraussetzungen der Studie erfüllten und mindestens über ein eigenes Smartphone sowie einen eigenen PC/Laptop verfügten. Diese Abfrage wurde ergänzt um Besitzfragen zu Tablets und einem Smart TV, um den Fokus auf den Vergleich von Smartphones und PCs/Laptops zu verdecken. Für alle vier Geräte wurde der Besitz dichotom erfragt (ja/nein). Eine Filterführung leitete lediglich die Personen zur nächsten Seite, die sowohl einen PC/Laptop als auch ein Smartphone besaßen – andernfalls sprang die Befragung direkt zum Befragungsende.

Der erste inhaltliche Teil der Befragung (S. 4+5) adressierte die persönliche Wahrnehmung der Ausfüllgeräte Smartphone und PC/Laptop sowie die individuellen Bedürfnisse der Befragten bei der Teilnahme an Onlinebefragungen. Um mögliche Reihenfolgeeffekte auszuschließen, wurde die Abfolge dieser Seiten interindividuell randomisiert. Auf Seite 4 (bzw. 5) wurde für die Ausfüllgeräte abgefragt, inwiefern sie hinsichtlich verschiedener Aspekte der Befragungsteilnahme nach Wahrnehmung der Befragten zweckmäßiger sind. Auf einem zwanzigstufigen Kontinuum beurteilten die Proband*innen den Unterschied zwischen Smartphones und PCs/Laptops als Ausfüllgeräte hinsichtlich des Zeitaufwands (*Zeit*), der *Datensicherheit*, der *Vertraulichkeit* und der *Anonymität* (siehe *Abbildung 8*). Entsprechend der Vorgaben von Whelan (2007, S. 6) wurden *Vertraulichkeit* und *Anonymität* separat abgefragt und bilden zusammen mit der *Datensicherheit* die Grundlage der Privatheit, die während der Befragungsteilnahme erlebt wird. Die vier Items wurden als inhaltssvalide Einzelitems im Sinne des C-OAR-SE Ansatzes der Skalenentwicklung erstellt (vgl. Rossiter 2002) und im Rahmen des Pretests abgesichert. Für die Antwort konnte trotz möglicher Probleme bei Smartphones (Toepoel 2016, S. 126) ein Schieberegler verwendet werden, um die feinstufige Unterscheidung der Einschätzung hinsichtlich der Ausfüllgeräte anzuregen, da in Studie 2 nicht der Geräteinfluss auf das Antwortverhalten im Vordergrund steht. Die Geräte (PC/Laptop linksstehend und Smartphone rechtsstehend) stellten die Extrempunkte des hinterlegten Differenzials dar, sodass für alle vier Dimensionen eine klare, individuelle Einschätzung der wahrgenommenen Gerätbesonderheiten

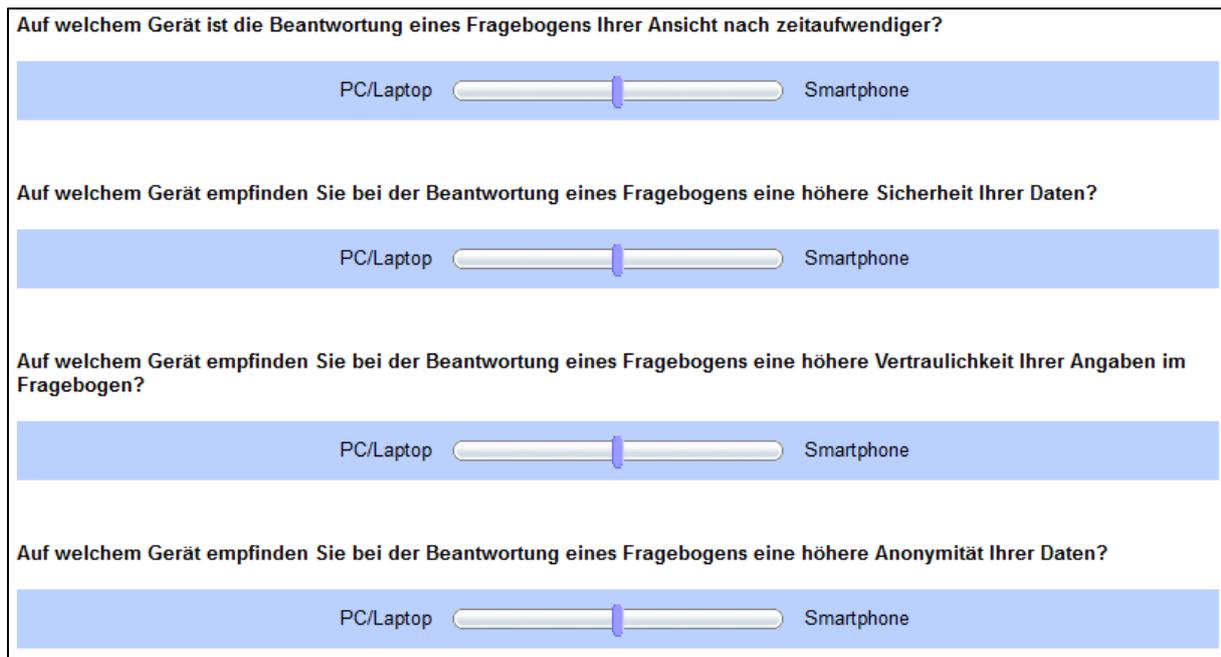


Abbildung 8. Gerätevergleich hinsichtlich der Dimensionen Zeitaufwand und Privatheit

Dargestellt ist ein Auszug der Seite 4 (bzw. 5) der Onlinebefragung aus Studie 2. Die Proband*innen beurteilen mittels eines Schiebereglers, der sich zwischen den Geräten PC/Laptop und Smartphone positionieren ließ, welches Gerät ihrer persönlichen Wahrnehmung nach als Ausfüllgerät einen höheren Zeitaufwand verlangt sowie höhere Datensicherheit, höhere Vertraulichkeit und höhere Anonymität ermöglicht.

Quelle: Eigener Fragebogen, SoSci Survey (Leiner 2023)

resultierte. Voreingestellt war jeweils die mittlere Position der Skala, die keinen Unterschied zwischen den Geräten signalisierte. Je weiter rechts der Schieberegler von den Befragten platziert wurde, desto größer ausgeprägt wurde die Besonderheit des Smartphones für diese Dimension wahrgenommen. Mit Verschiebungen des Reglers vom Mittelpunkt nach links indizierten die Befragten ein in ihrer Wahrnehmung stärkeres Gewicht des PCs/Laptops bezüglich der zu beurteilenden Dimension. Während für die Proband*innen keine Zahlen erkennbar waren, verlief die hinterlegte Skala aufsteigend von links (PC/Laptop = 1) nach rechts (Smartphone = 20). Der Startpunkt der mittleren Position wurde mit 10,5 kodiert. Die Items *Zeit*, *Datensicherheit*, *Vertraulichkeit* und *Anonymität* bilden die AVn der Prüfung der Hypothesen **H2.1a** (Proband*innen schätzen die Teilnahme mit dem Smartphone als zeitaufwendiger ein als die Teilnahme mit dem PC/Laptop), **H2.1b** (Proband*innen, die mit dem Smartphone teilnehmen, schätzen das Smartphone im Vergleich zu Proband*innen, die mit dem PC/Laptop teilnehmen, hinsichtlich des Zeitaufwands weniger schlecht ein), **H2.1c** (Bei der Einschätzung der Privatheit der Befragungsteilnahme unterscheiden sich das Smartphone und der PC/Laptop als Ausfüllgeräte) und **H2.1d** (Für das eigene gewählte Ausfüllgerät wird die Privatheit selbst besser eingeschätzt als durch Personen, die das jeweils andere Ausfüllgerät nutzen). Für alle Dimensionen wurden die Mittelwerte über alle Personen berechnet.

Seite 5 (bzw. 4) fragte für insgesamt sechs Bedürfnisse für die Befragungsteilnahme die individuelle Bedeutung ab, die Befragte diesen beimessen: Der Bedarf nach kurzer Bearbeitungszeit (*Zeitbedürfnis*), das Bedürfnis nach örtlicher Flexibilität bei der Teilnahme (*Ortsflexibilität*) und das Bedürfnis nach zeitlich flexibler Teilnahme (*Zeitflexibilität*) wurden als konkrete, inhaltsvalide Einzelitems konzipiert (vgl. Rossiter 2002). Das Bedürfnis nach *Datensicherheit* wurde angelehnt an Featherman und Pavlou (2003) erfragt und die Bedürfnisse nach Vertraulichkeit (*Vertraulichkeitsbedürfnis*) und nach Anonymität (*Anonymitätsbedürfnis*) nach Whelan (2007, S. 6) erhoben. Aus den drei letztgenannten Items wird der Skalenmittelwert zur Messung des Konzepts *Privatheitsbedürfnis* ($\alpha = 0,703$)¹⁵ gebildet, dessen interne Konsistenz den kritischen CA-Wert von 0,7 knapp übersteigt (vgl. Taber 2018). Die sechs Items der Bedürfnisse wurden in ihrer Abfolge rotiert und auf einer fünfstufigen, vollverbalisierten und numerisch verankerten Wichtigkeitsskala (1 = „überhaupt nicht wichtig“ bis 5 = „sehr wichtig“) beantwortet (z. B. Prüfer, Vazansky und Wystup 2003, S. 28). Die Variable *Zeitflexibilität* ist die Grundlage der Prüfung der Hypothese **H2.2a** (*Die Gerätwahl in Onlinebefragungen hängt von dem Bedürfnis nach zeitlicher Flexibilität ab, wobei Personen, die zeitliche Flexibilität schätzen, eher das Smartphone nutzen*); über *Ortsflexibilität* wird die **H2.2b** (*Die Gerätwahl in Onlinebefragungen hängt von dem Bedürfnis nach örtlicher Flexibilität ab, wobei Personen, die örtliche Flexibilität schätzen, eher das Smartphone nutzen*) geprüft; die Hypothesenprüfung der **H2.2c** (*Die Gerätwahl in Onlinebefragungen hängt von der Bedeutung der Teilnahmedauer ab, wobei Personen, die kürzere Befragungszeiten schätzen, eher den PC/Laptop nutzen*) basiert auf dem *Zeitbedürfnis* und für die Hypothese **H2.2d** (*Die Gerätwahl in Onlinebefragungen hängt von dem Bedürfnis nach Privatheit ab*) wird das *Privatheitsbedürfnis* genutzt. Für alle vier Hypothesen wird die dichotome AV *Gerät* (PC/Laptop = 0, Smartphone = 1) angelegt.

Auf Seite 6 wurde die *Technologieaffinität* erfragt, die in die Hypothesenprüfung als Kontrollvariable eingeht. Die Befragten beantworteten die Items „*Ich fühle mich in der Lage, über neue Technologien auf dem Laufenden zu bleiben*“ und „*Ich bin mir sicher, dass ich den Umgang mit neuen Technologien lernen kann*“ (Stüber 2013, S. 173) auf einer – angelehnt an eine Skala der ALLBUS-Bevölkerungsumfrage (Prüfer et al. 2003, S. 44) entwickelten und im Pretest abgesicherten – siebenstufigen, vollverbalisierten und numerisch verankerten Zustimmungsskala (1 = „stimme überhaupt nicht zu“ bis 7 = „stimme vollständig zu“). Der Skalenmittelwert der beiden Items bildet die *Technologieaffinität* ab. Für Zwei-Item-Skalen eignet sich die Spearman-Brown-Reliabilität als Konsistenzmaß eher als Cronbachs Alpha. Der Wert von $\rho = 0,819$ deutet hier auf gute interne Konsistenz hin (Eisinga, Grotenhuis und Pelzer 2013).

¹⁵ Die Eindimensionalität der drei Indikatoren wurde anhand einer exploratorischen Faktoranalyse überprüft.

Die nächsten Seiten (S. 7+8) erfassten die Gerätbindung der Befragten, wobei die Seitenabfolge erneut interindividuell randomisiert wurde, um Reihenfolgeeffekte auszuschließen. Eine Seite erfragte die Smartphonebindung (*Smartphone-Attachment*) und die andere Seite die PC-/Laptopbindung (*PC-Attachment*). *Smartphone-Attachment* wurde über eine reflektive Multi-Item-Skala der Bindung an das Smartphone erfasst (Melumad 2017, S. 211; Melumad und Pham 2020), die übereinstimmend mit der Originalskala ebenfalls auf der siebenstufigen, vollverbalisierten und numerisch verankerten Zustimmungsskala (s. o.) beantwortet wurde. Über alle sechs Items wurde der Skalenmittelwert gebildet ($\alpha = 0,861$) und für die Prüfung der Hypothese **H2.3a** (*Proband*innen, die eine hohe Bindung zum Smartphone berichten, wählen dieses auch eher für die Teilnahme an Onlinebefragungen*) verwendet. Die ursprünglich für die Smartphonebindung konzipierte Skala wurde auch auf den PC/Laptop angepasst und analog ebenfalls mit sechs Items gemessen. Das *PC-Attachment* ergab sich als Mittelwert dieser sechs Items ($\alpha = 0,790$) und wird als Prüfvariable der Hypothese **H2.3b** (*Proband*innen, die eine hohe Bindung zum PC/Laptop berichten, wählen diesen auch eher für die Teilnahme an Onlinebefragungen*) eingesetzt. Auf beiden Seiten wurden die Items in ihrer Abfolge rotiert.

Als Variable für Zusatzbetrachtungen im Anschluss an die Hypothesenprüfung wurde die *WN* desjenigen Geräts integriert, mit dem tatsächlich an der Befragung teilgenommen wurde (S. 9). Die Erfassung der *WN* basierte auf der Zehn-Item-Skala von Brooke (1996), die sorgsam auf den Studienkontext angepasst und im Rahmen des Pretests auf fünf Items gekürzt wurde. Erneut wurden die fünf Items auf derselben siebenstufigen Zustimmungsskala (s. o.) beantwortet und die *WN* ergab sich als Skalenmittelwert ($\alpha = 0,738$).

Im weiteren Verlauf wurden das *Alter* und das *Geschlecht* (S. 10), welche als zusätzliche Kontrollvariablen in die Analysen einbezogen wurden, der höchste bisher erreichte *Schulabschluss* und die aktuelle *Berufssituation* (S. 11), sowie die *Haushaltsgröße* und das eigene *Nettoeinkommen* (S. 12) als soziodemografische Variablen erfragt. Der Abschluss der Befragung bezog sich auf den Teilnahmehintergrund. Das Ausfüllgerät, mit dem die Studie *erstmalig geöffnet* wurde, und das *letztliche Ausfüllgerät* wurden separat erfragt (S. 13). Angegeben wurden zudem der *Ausfüllort*, ob *Dritte anwesend* waren, die *Körperhaltung* bei der Teilnahme und der Grad der *Ablenkung* (siebenstufig, endpunktverbalisiert und numerisch verankert: 1 = „sehr geringe Ablenkung“, 7 = „sehr starke Ablenkung“) während der Teilnahme (S. 14). Abschließend (S. 15) bewerteten die Befragten auf der siebenstufigen Zustimmungsskala (s. o.) die *wahrgenommene Befragungsdauer* und, analog zu den Bedürfnisitems (s. o.), drei Items zur empfundenen Datensicherheit, Vertraulichkeit und Anonymität während der aktuellen Teilnahme, aus denen sich die *empfundene Privatheit* als Skalenmittelwert ergab ($\alpha = 0,903$). Die

Verabschiedung beinhaltete ein vollständiges Debriefing zum Studienziel. Der Fragebogen der zweiten Studie ist in *Anhang B* dargestellt.

5.2.2 Ergebnisse

In den Analysen werden die Statistikprogramme JASP (JASP Team 2023), SPSS (IBM Corp. 2021) und situativ zur Ergänzung R (R Core Team 2023) verwendet. Für die Prüfung möglicher Gruppenunterschiede wird bei mindestens quasimetrischen AVn der t -Test für unabhängige Stichproben eingesetzt (Bortz und Schuster 2016, S. 120-124). Wenn alle Proband*innen zusammengefasst werden, erfolgt die Hypothesenprüfung mit dem Ein-Stichproben- t -Test (Bortz und Schuster 2016, S. 118-120). Für die Analyse der Hintergründe der Gerätwahl wird die multiple logistische Regression verwendet, wobei das gewählte Ausfüllgerät (*Gerät*: PC/Laptop = 0 vs. Smartphone = 1) die dichotome AV darstellt (Backhaus et al. 2021, S. 289-382). *Tabelle 7* gibt einen Überblick, welche Verfahren zur Prüfung welcher Hypothesen herangezogen werden. Da es sich um eine Feldstudie handelt, fand im Vorfeld eine Ausreißeranalyse bezüglich fehlender Werte und Abbrüche statt (siehe Kapitel 5.2.1.2). Extremwerte bei den Modellvariablen traten nicht auf, sodass eine weitere Bereinigung des Datensatzes nicht erforderlich war. Die Stichprobengröße der einzelnen Analysen variiert leicht, da Datenreihen von den Analysen ausgeschlossen wurden, wenn für jeweils einbezogene Variablen fehlende Werte auftraten.

	Angewendet für Hypothese...					
Ein-Stichproben- t -Test	H2.1a	H2.1c				
t -Test unabhängige Stichproben	H2.1b	H2.1d	H2.3a	H2.3b		
Logistische Regression	H2.2a	H2.2b	H2.2c	H2.2d	H2.3a	H2.3b

Tabelle 7. Aufbau der Hypothesenprüfung in Studie 2

5.2.2.1 Voranalysen und Voraussetzungsprüfung

Die statistische Hypothesenprüfung wird durch einige Vorbetrachtungen vorbereitet. Zunächst werden die beiden Untersuchungsgruppen (Smartphone vs. PC/Laptop) verglichen. Für die Unterschiedsprüfung der Voranalysen werden χ^2 -Tests und t -Tests für unabhängige Stichproben herangezogen. Die Voraussetzungen des χ^2 -Tests, dass die Beobachtungen unabhängig sind, jede Versuchsperson einer Zelle eindeutig zugeordnet werden kann und jede Zelle mindestens eine Beobachtung aufweist (Bühner und Ziegler 2017, S. 353), waren dabei für alle ausgewerteten Variablen erfüllt. Die verwendeten t -Tests erfordern unabhängige Beobachtungen, Normalverteilung der AV in den Gruppen, Intervallskalenniveau der AV und Varianzhomogenität (Bühner und Ziegler 2017, S. 301). Mindestens fünfstufige Skalen können als quasimetrisch

angesehen werden (Harpe 2015; Urban und Mayerl 2018, S. 301-302) und für alle betrachteten Variablen greift der zentrale Grenzwertsatz, nach dem sich ab einer Gruppengröße von $n > 30$ die Stichprobenkennwerte einer Normalverteilung annähern (Bortz und Schuster 2016, S. 586). Dies gilt auch für die unabhängigen t -Tests der letztlichen Hypothesenprüfung. Für alle berichteten Analysen wird Varianzhomogenität geprüft und ist erfüllt, wenn nicht anders angegeben.

Die nachfolgenden Analysen beziehen sich auf die Stichprobe ($N = 291$) nach der grundlegenden Datenbereinigung (siehe Kapitel 5.2.1.2). Auch wenn sich weder für das Verhältnis der Ausfüllgeräte (PC/Laptop = 154, Smartphone = 137) noch im Geschlechtsverhältnis (weiblich = 156, männlich = 128, 7 fehlende Werte) signifikante Unterschiede zeigen, sind die Variablen *Gerät* und *Geschlecht* nicht unabhängig voneinander. An der Befragung nahmen Frauen häufiger mit dem Smartphone und Männer häufiger mit dem PC/Laptop teil ($\chi^2(1) = 11,825$, $p < 0,001$). *Geschlecht* fungiert in Konsequenz als Kontrollvariable in der Hypothesenprüfung. Dasselbe gilt für das *Alter* der Proband*innen, da die PC/Laptop-Teilnehmer*innen ($\bar{x} = 40,88$ Jahre, $SD = 15,44$) signifikant älter als die Smartphone-Teilnehmer*innen ($\bar{x} = 32,21$ Jahre, $SD = 15,20$) waren ($t(287) = 4,800$, $p < 0,001$). In Vorbereitung der Hypothesenprüfung werden die Charakteristika der Ausfüllgeräte der Teilnehmer*innen der beiden Gerätgruppen verglichen, um die Gruppenzuteilung zu fundieren. Für *Bildschirmbreite* ($F(1, 289) = 177,221$, $p < 0,001$), *Bildschirmhöhe* ($F(1, 289) = 45,945$, $p < 0,001$) und *Fragebogenbreite* ($F(1, 289) = 29,975$, $p < 0,001$), die SoSci Survey für jedes Ausfüllgerät automatisch ermittelt, liegt keine Varianzhomogenität vor, sodass der Welch- t -Test verwendet wird, der für ungleiche Varianzen eher geeignet ist. Die Geräte der Smartphone-Teilnehmer*innen besitzen eine geringere *Bildschirmbreite* ($t(196,92) = 46,224$, $p < 0,001$), eine geringere *Bildschirmhöhe* ($t(248,25) = 17,952$, $p < 0,001$) sowie auch eine geringere Breite der Fragebogendarstellung (*Fragebogenbreite*) ($t(144,56) = 91,106$, $p < 0,001$) und sind somit deutlich kleiner.

Neben dem t -Test für unabhängige Stichproben kommen bei der Hypothesenprüfung auch der Ein-Stichproben- t -Test und die multiple logistische Regression zum Einsatz. Die Voraussetzungen des Ein-Stichproben- t -Tests sind, dass die Beobachtungen einer Zufallsstichprobe entnommen werden und die betrachtete Variable normalverteilt ist (Bortz und Schuster 2016, S. 119-120). Reine einfache Zufallsstichproben sind in der empirischen Wirtschafts- und Sozialforschung nur sehr aufwendig zu realisieren. Im vorliegenden Fall wurde zumindest keine systematische Eingrenzung der Auswahlpopulation vorgenommen. Für die Normalverteilung greift jeweils der zentrale Grenzwertsatz bei den zur Analyse herangezogenen Mittelwerten. Im Zuge der Hypothesenprüfung wird zusätzlich der nicht-parametrische Wilcoxon-Test berichtet (Bortz und Schuster 2016, S. 125), um die Robustheit der Prüfung zu erhöhen.

Für die logistische Regression sind die Voraussetzungen der Unabhängigkeit der Beobachtungen, von Fallzahlen von mindestens $n = 25$ pro Untersuchungsgruppe und von geringer Multikollinearität der Modellvariablen erfüllt, wobei die Analyse recht robust gegenüber Verletzungen ist (Backhaus et al. 2021, S. 379). In der Hypothesenprüfung wird die AV *Gerät* auf die verschiedenen Bedürfnisse der Teilnehmer*innen regressiert (*Zeitbedürfnis*, *Privatheitsbedürfnis*, *Zeitflexibilität*, *Ortsflexibilität*). Als weitere Prädiktoren werden *Smartphone-Attachment*, *PC-Attachment* und *Technologieaffinität* ergänzt und auch die soziodemografischen Hintergrundvariablen *Alter* und *Geschlecht* ins Regressionsmodell einbezogen. Für alle beteiligten Variablen kann Multikollinearität ausgeschlossen werden, da VIF (stets $< 1,55$) und Toleranz (stets $> 0,65$) eng um den Wert 1 schwanken (Backhaus et al. 2021, S. 100).

5.2.2.2 Hypothesenprüfung

Die **H2.1a** unterstellt, dass die Teilnahme am Smartphone im Vergleich zum PC/Laptop konsistent von allen Befragten als zeitaufwendiger eingeschätzt wird. Die Variable *Zeit* bildet die Einschätzung der Proband*innen auf einem 20-stufigen Differenzial ab, wobei hohe Zahlen einen höheren wahrgenommenen Zeitaufwand des Smartphones signalisieren und niedrige Zahlen eher für höheren Zeitaufwand am PC/Laptop sprechen (siehe *Abbildung 8*). Wird der Zeitaufwand an beiden Ausfüllgeräte als gleich wahrgenommen, müsste die Einschätzung in der Mitte bei 10,5 liegen. Der einseitige Ein-Stichproben-*t*-Test bestätigt die Hypothese und demonstriert, dass im Mittel ($\bar{x} = 12,13$, $SD = 6,39$, $\tilde{x} = 14,0$) über alle Proband*innen der Studie hinweg die Teilnahme am Smartphone zeitaufwendiger erlebt wird ($t(290) = 4,343$, $p < 0,001$). Dies zeigt sich auch in der nicht-parametrischen Prüfung über den Wilcoxon-Test ($p < 0,001$).

Hypothese **H2.1c** vermutet, dass auch die Privatheit der Befragungsteilnahme in der Wahrnehmung der Befragten gerätübergreifend variiert. Für alle drei Subvariablen der Privatheit – *Datensicherheit* ($\bar{x} = 6,66$, $SD = 4,61$, $\tilde{x} = 6,0$; $t(290) = -14,191$, $p < 0,001$), *Vertraulichkeit* ($\bar{x} = 7,75$, $SD = 4,46$, $\tilde{x} = 9,0$; $t(290) = -10,512$, $p < 0,001$) und *Anonymität* ($\bar{x} = 7,77$, $SD = 4,35$, $\tilde{x} = 9,0$; $t(290) = -10,697$, $p < 0,001$) – wird über alle Befragten hinweg der PC/Laptop bei zweiseitiger Testung, da die Hypothese keine Richtung des Effekts vermutet, als privater eingeschätzt und die Mittelwerte unterscheiden sich signifikant vom neutralen Wert 10,5. Die nicht-parametrischen Wilcoxon-Prüfungen bestätigen die Befunde (jeweils $p < 0,001$).

Anknüpfend an die konsistenten Wahrnehmungsunterschiede für die untersuchten Dimensionen der Geräteignung werden zur Untersuchung der Hypothesen **H2.1b** und **H2.1d** die Gerätebewertungen getrennt nach Gerätgruppe (Smartphone vs. PC/Laptop), der die Befragten angehören, verglichen. Es wird vermutet, dass das tatsächlich verwendete Ausfüllgerät selbst

bezüglich der Dimensionen besser eingeschätzt wird als von Personen, die das jeweils andere Gerät verwenden (z. B., dass der Zeitaufwand am Smartphone weniger schlecht von Smartphone-Teilnehmer*innen bewertet wird als von PC/Laptop-Teilnehmer*innen). Hierfür werden die Gruppenmittelwerte mithilfe des t -Tests für unabhängige Stichproben verglichen und führen zu hypothesenkonformen Befunden. Für die Variable *Zeit* liegt Varianzheterogenität vor ($F(1, 289) = 13,873, p < 0,001$), weshalb der Welch- t -Wert berichtet wird. Die Smartphone-Gruppe ($\bar{x} = 8,44, SD = 5,88$) schätzt die Teilnahme mit dem Smartphone weniger zeitintensiv ein, als dies die PC/Laptop-Gruppe ($\bar{x} = 15,41, SD = 4,85$) für das Smartphone tut ($t(264,43) = 10,946, p < 0,001$). Die Smartphone-Teilnehmer*innen halten – entgegen der Effektrichtung in der Gesamtstichprobe – die Teilnahme mit dem PC/Laptop sogar für zeitaufwendiger. Unter Ausschluss aller Befragten, die den PC/Laptop nutzen, liefert der zweiseitige Ein-Stichproben- t -Test nun das gegenteilige Ergebnis, wobei der Mittelwert der Smartphone-Gruppe signifikant nach unten vom neutralen Mittel (10,5) abweicht ($t(136) = -4,105, p < 0,001$). Für die drei Variablen der Privatheit zeigt sich ein ähnliches Muster. Der Effekt dreht sich hier in der Smartphone-Gruppe zwar nicht um und auch diese bewertet weiterhin den PC/Laptop als privater. Für *Datensicherheit* ($t(289) = -3,051, p = 0,012$), *Vertraulichkeit* ($t(289) = -3,287, p < 0,001$) und *Anonymität* ($t(289) = -3,488, p < 0,001$) schätzen Smartphone-Teilnehmer*innen die Vorteile des PCs/Laptops aber signifikant weniger stark ein als die Gruppe der PC/Laptop-Nutzer*innen. *Tabelle 8* veranschaulicht die Befunde durch die Darstellung der Mittelwerte im Gesamten und aufgeteilt auf die beiden Gerätgruppen.

Die Prüfung der weiteren Hypothesen **H2.2a-d** sowie **H2.3a** und **H2.3b** erfolgt gesammelt in der logistischen Regressionsanalyse. Für alle Prädiktoren wird ermittelt, ob sie das *Gerät* (PC/Laptop = 0, Smartphone = 1) vorhersagen. Die Prüfung ist kleinschrittig angelegt und umfasst neben einem Basismodell (*Modell 1*), das nur die vier Bedürfnisvariablen (*Zeitbedürfnis*, *Ortsflexibilität*, *Zeitflexibilität* und *Privatheitsbedürfnis*) beinhaltet, zwei weitere Modelle, die

	Gesamt	PC/Laptop	Smartphone
<i>Zeit</i>	12,13	15,41	8,44
<i>Datensicherheit</i>	6,66	5,90	7,53
<i>Vertraulichkeit</i>	7,75	6,98	8,65
<i>Anonymität</i>	7,77	6,95	8,70

Tabelle 8. Mittelwerte für die Prüfungen der Hypothesen **H2.1a-H2.1d**

Dargestellt sind die Mittelwerte der vier Variablen *Zeit*, *Datensicherheit*, *Vertraulichkeit* und *Anonymität* für die gesamte Stichprobe sowie für die beiden Gerätegruppen. Der neutrale Mittelpunkt liegt bei 10,5. *Blaue Werte* signalisieren den relativen Vorteil des PCs/Laptops bei der jeweiligen Dimension. *Rote Werte* entsprechen eher einem Vorteil des Smartphones.

jeweils um relevante Variablen erweitert werden. *Modell 2* ergänzt die beiden Variablen des Attachments (*Smartphone-Attachment* und *PC-Attachment*) sowie *Technologieaffinität*. Das dritte Modell (*Modell 3*) erweitert das Regressionsmodell zusätzlich um die soziodemografischen Variablen *Alter* und *Geschlecht* und stellt somit das Gesamtmodell der Prüfung dar. *Tabelle 9* fasst die Befunde modellspezifisch zusammen. Das Gesamtmodell (*Modell 3*) erzielt dabei verglichen mit den *Modellen 1* und *2* deutlich die höchste Modellgüte (*McFaddens* $R^2 = 0,106$, *Nagelkerke* $R^2 = 0,182$, $AIC = 364,581$, $BIC = 400,893$, siehe *Tabelle 9*), weshalb sich die Ergebnisdarstellung in der Folge auf *Modell 3* bezieht und punktuell bei abweichenden Befundmustern der anderen Modelle um deren Ergebnisse ergänzt wird. Für *Zeitbedürfnis* (**H2.2c**) finden die Analysen modellübergreifend keinen signifikanten Einfluss auf die Gerätwahl ($b = -0,031$, $W(1) = 0,032$, $p = 0,858$)¹⁶. Sowohl das Bedürfnis nach *Zeitflexibilität* (**H2.2a**) ($b = -0,090$, $W(1) = 0,330$, $p = 0,566$) als auch das *Privatheitsbedürfnis*¹⁷ (**H2.2d**) ($b = 0,022$, $W(1) = 0,012$, $p = 0,913$) beeinflussen die Gerätwahl ebenfalls nicht. Für *Ortsflexibilität* (**H2.2b**) findet das Gesamtmodell keinen signifikanten, aber einen marginalsignifikanten Einfluss auf die Gerätwahl ($b = 0,203$, $W(1) = 2,766$, $p = 0,096$), wobei Personen mit größerem

Prädiktor	Modell 1			Modell 2			Modell 3		
	<i>b</i>	<i>Wald</i>	<i>p</i>	<i>b</i>	<i>Wald</i>	<i>p</i>	<i>b</i>	<i>Wald</i>	<i>p</i>
<i>Zeitbedürfnis</i>	-0,095	0,381	0,537	-0,072	0,206	0,650	-0,031	0,032	0,858
<i>Privatheitsbedürfnis</i>	-0,135	0,554	0,457	-0,090	0,230	0,631	0,022	0,012	0,913
<i>Zeitflexibilität</i>	-0,035	0,066	0,797	-0,114	0,626	0,429	-0,090	0,330	0,566
<i>Ortsflexibilität</i>	0,212	3,452	0,063°	0,189	2,604	0,107	0,203	2,766	0,096°
<i>Technologieaffinität</i>				-0,127	1,034	0,309	-0,210	2,294	0,130
<i>Smartphone-Attachment</i>				0,459	13,924	<0,001*	0,244	3,316	0,069°
<i>PC-Attachment</i>				-0,368	5,711	0,017*	-0,228	1,958	0,162
<i>Alter</i>							-0,038	14,053	<0,001*
<i>Geschlecht</i>							0,549	4,087	0,043*
Modell-Fit	<i>McFaddens R²</i>	0,012		<i>McFaddens R²</i>	0,051		<i>McFaddens R²</i>	0,106	
	<i>Nagelkerke R²</i>	0,021		<i>Nagelkerke R²</i>	0,091		<i>Nagelkerke R²</i>	0,182	
	<i>AIC</i>	404,787		<i>AIC</i>	392,377		<i>AIC</i>	364,581	
	<i>BIC</i>	423,119		<i>BIC</i>	421,653		<i>BIC</i>	400,893	

Tabelle 9. Ausgewählte Ergebnisse der logistischen Regression der Studie 2

* = signifikant ($p < 0,05$); ° = marginalsignifikant ($p < 0,1$) (beides **fett** gedruckt)

¹⁶ Berichtet werden die nicht-standardisierten Regressionskoeffizienten (*b*). Dabei wird lediglich das Vorzeichen interpretiert, wenn ein signifikanter Effekt vorliegt, um die Richtung des Effekts einschätzen zu können.

¹⁷ Da sich das Konstrukt *Privatheitsbedürfnis* aus drei Einzelitems zusammensetzt und Cronbachs Alpha den Schwellenwert von 0,7 nur knapp übersteigt, wurde zusätzlich geprüft, ob sich die Befundmuster verändern, wenn die drei Items einzeln betrachtet werden. Dies war für alle Analysen über alle Modelle hinweg nicht der Fall.

Wunsch nach *Ortsflexibilität* tendenziell eher zum Smartphone greifen. Werden in *Modell 1* nur die vier Bedürfnisvariablen betrachtet, nähert sich der Einfluss der *Ortsflexibilität* auf die Gerätwahl einem signifikanten Befund stärker an ($b = 0,212$, $W(1) = 3,452$, $p = 0,063$). Der Modellunterschied demonstriert, dass die Bedürfnisse mit anderen Hintergrundfaktoren assoziiert sind und die Geräteignung für die Bedürfnisbefriedigung interindividuell variieren kann. Damit wird die Bedeutung des Einbezugs der gewählten Kontrollvariablen unterstrichen.

Die Befunde der Hypothesenprüfung der **H2.3a** und **H2.3b** der Rolle der Gerätbindung festigen diesen Eindruck. Ohne Einbezug der soziodemografischen Hintergrundvariablen zeigen sich in *Modell 2* signifikante Effekte dahingehend, dass Personen mit höheren Werten beim *PC-Attachment* (**H2.3b**) eher mit dem PC/Laptop teilnehmen ($b = -0,368$, $W(1) = 5,711$, $p = 0,017$) und Personen mit höherer Ausprägung beim *Smartphone-Attachment* (**H2.3a**) eher mit dem Smartphone ($b = 0,459$, $W(1) = 13,924$, $p < 0,001$). Dieser Effekt verschwindet beim *PC-Attachment* jedoch im Gesamtmodell (*Modell 3*) vollständig ($b = -0,228$, $W(1) = 1,958$, $p = 0,162$) und ist für das *Smartphone-Attachment*, ebenfalls deutlich abgeschwächt, lediglich marginalsignifikant ($b = 0,244$, $W(1) = 3,316$, $p = 0,069$). Konform zu diesen Ergebnissen erzielen Befragte am Smartphone höhere Werte (**H2.3a**) für die Variable *Smartphone-Attachment* ($t(289) = -2,909$, $p = 0,004$). Für die Bindung an den PC/Laptop (*PC-Attachment*) treten hingegen keine Gruppenunterschiede ($t(287) = 0,777$, $p = 0,438$) auf (**H2.3b**). *Abbildung 9* gibt einen Überblick über die gesamten Ergebnisse der Hypothesenprüfungen der Studie 2.

Gerätwahrnehmung	Bedürfnisbefriedigung	Bindung an das Gerät
H2.1a: Teilnahme mit dem Smartphone wird zeitaufwendiger eingeschätzt. ✓	H2.2a: Ein hohes Bedürfnis nach zeitlicher Flexibilität führt eher zur Teilnahme mit dem Smartphone. ✗	H2.3a: Personen, die eine höhere Bindung an das Smartphone berichten, nehmen eher mit diesem teil. *** ✓
H2.1b: Teilnehmer*innen mit dem Smartphone schätzen den höheren Zeitaufwand der Smartphones selbst geringer ein. ✓	H2.2b: Ein hohes Bedürfnis nach örtlicher Flexibilität führt eher zur Teilnahme mit dem Smartphone. ✗	H2.3b: Personen, die eine höhere Bindung an den PC/Laptop berichten, nehmen eher mit diesem teil. ✗
H2.1c: Ausfüllgeräte unterscheiden sich hinsichtlich der wahrgenommenen gewährleisteten Privatheit bei der Teilnahme. ✓	H2.2c: Ein hohes Bedürfnis nach kurzer Bearbeitungszeit führt eher zur Teilnahme mit dem PC/Laptop. ✗	
H2.1d: Die Privatheit des selbst genutzten Geräts wird selbst besser eingeschätzt als von Nutzer*innen des anderen Gerätes. ✓	H2.2d: Die Gerätwahl wird beeinflusst vom Bedürfnis nach privater Teilnahme. ✗	
		✓ Hypothese angenommen ✗ Hypothese abgelehnt *** uneinheitliche Befunde

Abbildung 9. Überblick über die Ergebnisse der Hypothesenprüfung der Studie 2

5.2.2.3 Zusatzbetrachtungen

Gestützt auf die Befunde der Hypothesenprüfung vertiefen Zusatzanalysen die Bedeutung der beteiligten Kontrollvariablen und knüpfen an die dargestellten Ergebnisse aus der *Tabelle 9* an.

Für *Technologieaffinität* ($b = -0,210$, $W(1) = 2,294$, $p = 0,130$) treten keine signifikanten Effekte auf die Gerätwahl auf. Eine besondere Bedeutung offenbart sich in den Zusatzbetrachtungen aber für die soziodemografischen Kontrollvariablen. Die Ungleichheit bei der Verteilung des *Geschlechts* in den Untersuchungsgruppen ist bereits in den Vorbetrachtungen aufgefallen. Weibliche Personen nutzen eher das Smartphone als den PC/Laptop zur Teilnahme und der Einfluss des *Geschlechts* ist im Regressionsmodell signifikant ($b = 0,549$, $W(1) = 4,087$, $p = 0,043$). Noch eindeutiger stellt sich der Einfluss des *Alters* auf die Gerätwahl dar ($b = -0,038$, $W(1) = 14,053$, $p < 0,001$): Ältere Menschen greifen eher zum PC/Laptop. Die Bedeutung der soziodemografischen Hintergrundvariablen wird durch den Vergleich der Ergebnisse der verschiedenen Analysemodelle verstärkt. Für alle Bedürfnisvariablen und die gerätbezogenen Einstellungen verschwinden signifikante Einflüsse, wenn *Geschlecht* und *Alter* als Kontrollvariablen einbezogen werden und die Modellgüte steigt durch den Einbezug deutlich an.

Zusätzlich werden die Gerätgruppen noch auf weitere möglicherweise für die Gerätwahl bedeutsame Unterschiede geprüft. Der erste Vergleich bezieht sich darauf, ob sich die *WN* des genutzten Geräts zwischen den beiden Untersuchungsgruppen unterscheidet. In beiden Gruppen schätzen die Proband*innen ihr jeweiliges Ausfüllgerät als ähnlich nützlich ein ($t(289) = -1,386$, $p = 0,167$). Damit bestätigt auch der Gruppenvergleich der *WN*, dass das jeweils genutzte Gerät (spätestens nach der Wahl) stets für die Teilnahme an Onlinebefragungen als geeignet wahrgenommen wird. Beim Teilnahmekontext der Onlinebefragung zeigen sich gemischte Muster im Vergleich der Gerätgruppen. Proband*innen sind während der Teilnahme im Mittel gleich stark abgelenkt (*Ablenkung*) ($t(285) = 0,530$, $p = 0,596$) und schätzen die *empfundene Privatheit* in der tatsächlichen Teilnahme-situation ähnlich ein ($t(286) = 1,050$, $p = 0,295$). Auch die *wahrgenommene Befragungsdauer* unterscheidet sich nicht signifikant ($t(283) = -1,874$, $p = 0,062$). Die Signifikanz wird allerdings nur knapp verfehlt, wobei am Smartphone die Befragungsdauer marginalsignifikant länger wahrgenommen wird. Dies korrespondiert auch mit einer höheren *Antwortzeit* am Smartphone ($t(289) = -2,873$, $p = 0,004$), da die Versuchspersonen am Smartphone signifikant länger für die Befragungsteilnahme ($\bar{x} = 7,42$ Minuten, $SD = 1,98$) benötigen als am PC/Laptop ($\bar{x} = 6,77$ Minuten, $SD = 1,89$).

5.2.3 Diskussion und Zwischenfazit der Voruntersuchungen

Das Smartphone wird als Ausfüllgerät in Onlinebefragungen zunehmend präferiert (Krämer 2017, S. 100; Revilla et al. 2016b), wobei noch fraglich ist, ob dadurch Moduseffekte hinsichtlich der Unit-Nonresponse entstehen. Da die Unit-Nonresponse nicht direkt beurteilt werden kann (siehe Kapitel 2.4), nähert sich Studie 2 der Unit-Nonresponse über die Untersuchung der

Gruppenunterschiede zwischen Smartphone- und PC/Laptop-Teilnehmer*innen in Onlinebefragungen an. Versuchspersonen, die das Smartphone als Ausfüllgerät nutzen, unterscheiden sich hinsichtlich soziodemografischer Merkmale von Personen, die vorrangig den PC/Laptop verwenden. Kritische Moduseffekte der Unit-Nonresponse werden hierfür aber eher ausgeschlossen (siehe Kapitel 3.1.2). Die zweite Studie der vorliegenden Forschungsreihe erweitert die Überlegungen zur Unit-Nonresponse auf die Gerätesonderheiten und prüft, ob die verschiedenen Ausfüllgeräte unterschiedlich gut geeignet sind, befragungsrelevante Bedürfnisse zu befriedigen. Gerade aufgrund seiner psychologischen Bedeutung könnte das Smartphone als Ausfüllgerät in der Einschätzung der Befragten für die Bedürfnisbefriedigung in Onlinebefragungen geeigneter oder weniger geeignet sein (siehe Kapitel 2.2 und 2.3). Daraus könnten ggf. gerätweise Unterschiede in den Unit-Nonresponseraten resultieren, die den TSE erhöhen.

Studie 2 zeigt, dass Befragte Smartphones und PCs/Laptops bezogen auf befragungsrelevante Bedürfnisdimensionen unterschiedlich bewerten. Insgesamt ordnen die Befragten dem PC/Laptop ein größeres Vermögen beim Schutz der Privatheit in Form von Datensicherheit, Vertraulichkeit und Anonymität zu und die Teilnahme am Smartphone wird über alle Befragten hinweg als insgesamt zeitaufwendiger eingeschätzt. Dies bestätigt die Hypothesen **H2.1a** und **H2.1c**: Verschiedene Ausfüllgeräte können in Onlinebefragungen also als unterschiedlich geeignet für die Befriedigung nutzungsrelevanter Bedürfnisse wahrgenommen werden. Dadurch wird der Annäherungsansatz an die Unit-Nonresponse fundiert, den Studie 2 wählt. Dieser besagt, dass Befragte sich für oder gegen die Teilnahme mit einem bestimmten Gerät entscheiden könnten, wenn Geräte bezogen auf spezifische Bedürfnisdimensionen konsistent als unterschiedlich gut eingeschätzt werden und den Befragten diese Dimensionen sehr wichtig sind. Für Personen, die geringe Teilnahmezeiten oder größere Privatheit (jeweils durch PC/Laptop besser gewahrt) bei der Teilnahme schätzen, könnten etwa gerätspezifische Unit-Nonresponseraten entstehen. Dies ist auch für Orts- und Zeitflexibilität denkbar, für die Vorteile des Smartphones erwartbar sind (siehe Kapitel 2.2.2 und 2.2.4). Allerdings bewähren sich ebenfalls die Hypothesen **H2.1b** und **H2.1d**, die postulieren, dass das jeweils genutzte Gerät vergleichsweise besser in Bezug auf die Bedürfnisdimensionen eingeschätzt wird. Befragte, die mit dem Smartphone teilnehmen, sehen zwar auch die Privatheit am PC/Laptop eher gewahrt, nehmen den Vorteil gegenüber dem Smartphone aber geringer wahr. Für den Zeitaufwand kehrt sich der Effekt sogar um und Smartphone-Teilnehmer*innen schätzen den Zeitbedarf des Smartphones als Ausfüllgerät geringer ein als den des PCs/Laptops. Neben den als konsistent wahrgenommenen Unterschieden zwischen den Ausfüllgeräten für die Bedürfnisdimensionen gibt es bei

der Eignungseinschätzung der Geräte also noch eine individuelle Komponente in der Wahrnehmung der Befragten. Dabei bewerten Personen das Gerät, das sie zur Teilnahme wählen, besser als Personen, die das jeweils andere Gerät wählen. Dies kann darauf zurückzuführen sein, dass die Interaktion mit technischen Geräten sehr individuell und mit vielfältigen Motivations- und Bedürfnisstrukturen verknüpft ist (z. B. Keusch et al. 2022; Park 2015). Daher sind nicht immer stabile Einschätzungen plausibel, die die Gerätwahl für alle Befragten gleich beeinflussen.

Nichtsdestotrotz könnten die Dimensionen, in denen ein Gerät konsistent besser wahrgenommen wird, wahrelevant sein, wenn sie für Befragte wichtig sind. In der zweiten Studie wird analog zum Vorgehen von de Bruijne und Wijnant (2014) die Gerätwahl auf die Bedürfnisse der Befragten bei der Teilnahme an Onlinebefragungen regressiert. Die Regressionsergebnisse unterstreichen aber, dass die Gerätwahl nicht eindeutig auf die Erwartungen zurückführbar ist, die Befragte an die Befragungsteilnahme stellen. Keins der Bedürfnisse nach zeitlicher Flexibilität (**H2.2a**), örtlicher Flexibilität (**H2.2b**), kurzer Teilnahmedauer (**H2.2c**) und nach Schutz der Privatheit (**H2.2d**) besitzt einen signifikanten prädiktiven Wert für die Gerätwahl, wodurch alle Hypothesen abgelehnt werden müssen. Dieser Befund ist rückblickend auf den vorherigen Absatz schlüssig, da die interindividuell variierende Eignungseinschätzung der Geräte dazu führt, dass verschiedene Personen für dieselbe Tätigkeit individuell eher das Smartphone oder eher den PC/Laptop als bedürfnisbefriedigend einschätzen. Nur wenn diese Einordnung interindividuell komplett stabil wäre, könnte sie die Gerätwahl potent prädizieren und hätte Konsequenzen für die Unit-Nonresponse. Die Gerätwahrnehmungen sind jedoch letztlich so divers, dass verschiedene Geräte je nach Person dieselben Bedürfnisse unterschiedlich gut befriedigen können. Indes deutet sich in den Ergebnissen eine gewisse emotionale Komponente des Smartphones zumindest an: Personen, die eine höhere Bindung zum Smartphone berichten, nehmen tendenziell auch eher mit diesem teil (**H2.3a**), während die Gerätbindung an den PC/Laptop keinen prädiktiven Wert besitzt (**H2.3b**). Die emotionale und psychologische Bedeutung des Smartphones (siehe Kapitel 2.3.2) im Unterschied zu herkömmlichen Ausfüllgeräten (PC/Laptop) manifestiert sich demnach tendenziell in den Daten der Studie 2, die die handlungs- und gedankenleitenden Funktionen des Smartphones bestätigen. Dabei scheinen grundsätzliche Gerätepräferenzen zu existieren, die aus der wohlwollenden, interindividuell variierenden Einschätzung der Kerngrößen der Geräteignung resultieren und daher wohl gerade nicht die gerätspezifische Unit-Nonresponse fördern. Gleichzeitig entsprechen die gefundenen Muster zur Antwortzeit den Befunden der Literatur: Analog zum Forschungsstand (z. B. Antoun und Cernat 2020; Couper und Peterson 2017; Keusch und Yan 2017; Revilla et al. 2016a) benötigen Teilnehmer*innen am Smartphone länger für die Befragungsteilnahme als am PC/Laptop.

Die Ergebnisse der Studie 2 schließen als zweite Voruntersuchung die Vorbereitung der adäquaten Prüfung von möglichen Moduseffekten des Smartphones als Ausfüllgerät im TSE-Framework ab. Die Untersuchung der Bedürfnisse der Versuchspersonen zeigt, dass die einzelnen Nutzungsgruppen (Smartphone vs. PC/Laptop) sich nicht in psychosozialen Charakteristika unterscheiden. Die Präferenz für ein Ausfüllgerät scheint eher nicht systematisch auf befragungsrelevante Bedürfnisse der Ausfüllenden zurückzuführen zu sein. Allerdings ordnen sich die soziodemografischen Gruppenbesonderheiten nahtlos in die bisherige Erkenntnislage ein: Smartphone-Teilnehmer*innen sind jünger als Teilnehmer*innen mit dem PC/Laptop (z. B. Buskirk et al. 2015; de Bruijne und Wijnant 2014; Mavletova 2013; Sommer et al. 2017; Toepoel und Lugtig 2014) und eher weiblich (z. B. de Bruijne und Wijnant 2014; Erens et al. 2019; Keusch und Yan 2017; Sommer et al. 2017). Die Analyseergebnisse unterstreichen die Relevanz dieser Variablen für die Gerätwahl. Insbesondere das Alter hat einen hohen prädiktiven Wert, aber auch das Geschlecht agiert als relevante Prognosevariable. Die Modellgüte der logistischen Regression steigt deutlich an, wenn Alter und Geschlecht als Prädiktoren ergänzt werden. Dabei reduzieren der Einbezug des Alters und des Geschlechts den Einfluss der untersuchten Bedürfnisvariablen auf die Gerätwahl und stellen die einzigen robusten Komponenten der Gerätwahl dar. Daraus ergibt sich für wissenschaftliche Onlinebefragungen die Implikation, das genutzte Ausfüllgerät in der Datenauswertung zu berücksichtigen, wenn Zielvariablen der Studie mit Alter oder Geschlecht korrelieren könnten (z. B. Einkommen, Familienstand) und nicht gleich viele Befragte mit dem Smartphone und dem PC/Laptop teilnehmen.

Die konditionalen Teilnahmebedürfnisse für den Zeitaufwand, die Privatheit und die Flexibilität (örtlich und zeitlich) bei der Teilnahme variieren hingegen alters- und geschlechtsspezifisch und die psychologische Bedeutung der Geräte sowie die Bedürfnisbefriedigung durch bestimmte Ausfüllgeräte sind sehr individuell. Dementsprechend findet eine natürliche Kontrolle der Teilnahmebedürfnisse dann statt, wenn das Ausfüllgerät selbst gewählt werden kann. Keins der Geräte befriedigt pauschal bestimmte Bedürfnisse besser, sondern die Effekte heben sich über alle Versuchspersonen hinweg auf. Insofern löst der Smartphoneeinsatz mögliche Probleme der Unit-Nonresponse eher, da diese nur zu erwarten sind, wenn die Entscheidung für ein Ausfüllgerät nicht selbstbestimmt getroffen werden kann. Zusammenfassend untersucht die zweite Studie entsprechend der Empfehlungen von Tourangeau (2017, S. 117) Moduseffekte möglichst exklusiv für die Unit-Nonresponse. Der Unit-Nonresponse wird sich basierend auf den Teilnahmebedürfnissen angenähert, die gerätweise verglichen werden. Die Ergebnisse zeigen, dass durch das Smartphone eher keine Moduseffekte entstehen, sondern dass die freie Teilnahmeentscheidung mit verschiedenen Geräten sogar eher Repräsentativität fördern kann.

Die Studie 2 ist mit einigen Limitationen verknüpft. Möglicherweise trifft die ursprüngliche Grundannahme nicht zu. Diese geht davon aus, dass Nutzer*innen situativ diejenigen Geräte wählen, die ihnen jeweils am geeignetsten erscheinen und die eigenen Bedürfnisse besser abbilden (Wang 2016). Basierend auf den Studienergebnissen könnte sich dieser Wirkmechanismus aber auch umkehren oder zumindest wechselseitig wirken. Eventuell beeinflusst das gewählte Ausfüllgerät selbst die persönliche Wahrnehmung der Geräteignung hinsichtlich der Bedürfnisdimensionen. Zur Absicherung des Befundes könnte künftig diese mögliche gegenteilige Wirkrichtung überprüft und die Gerätebewertung vom tatsächlichen Ausfüllgerät in Onlinebefragungen entkoppelt werden. Die untersuchten Bedürfnisse während der Befragungsteilnahme stellen darüber hinaus nur eine Auswahl dar. Weitere teilnahmerelevante Bedürfnisse wie beispielsweise Themeninteresse, Studiendesign oder Aktivierung durch Befragungsinhalte könnten ebenfalls auf die Teilnahmebereitschaft und die Wahl des Ausfüllgeräts wirken.

Im Gegensatz zur ersten Studie der vorliegenden Arbeit löst sich Studie 2 vom artifiziellen Laborkontext und legt auch den Auswahlrahmen der Stichprobe breiter an. Durch die Übertragung ins Feld ist die Anlage realitätsnäher und ermöglicht direkte Rückschlüsse auf die Bedürfnisse der Teilnehmer*innen und die tatsächliche und möglicherweise situativ variierende Gerätwahl. Neben dieser Stärke offenbart sich aber auch eine Schwäche des Forschungsdesigns: Kausale Interpretationen der Wirkmechanismen sind nicht möglich und es werden nicht alle möglichen Hintergrundfaktoren erfasst, die auf die erhobenen Daten wirken könnten. Mögliche weitere Drittvariablen (z. B. Vertrauen in die wissenschaftliche Forschung) könnten das Antwortverhalten und die Gerätwahl ebenfalls systematisch beeinflussen. Entsprechend wird der Geräteeffekt nicht gänzlich isoliert, sondern ist situativ eingebettet. Auch die Situation, in der die Einladung zur Studie erhalten wurde, kann nicht kontrolliert werden. Hier ist ein Bezug zur Entscheidung für das letztlich gewählte Ausfüllgerät denkbar. Besonders problematisch ist, dass in der Studie 2 die Unit-Nonresponse nicht direkt untersucht werden kann. Zukünftige Studien könnten versuchen die Gründe in den Blick zu nehmen, die Personen zur Teilnahmeverweigerung bewegen und die möglicherweise mit einem Ausfüllgerät zusammenhängen. Dafür müssten gezielt Untersuchungen mit Personen konzipiert werden, die an Onlinebefragungen nicht teilnehmen oder nicht mit bestimmten Geräten teilnehmen. Der Zugang zu und die Identifikation von solchen Personen ist praktisch aber nicht möglich ohne die Erhebung personenbezogener Daten, über die Individuen eindeutig identifiziert werden können, die sich gegen die Teilnahme entscheiden. Auch generelle Selektionseffekte, unabhängig vom Ausfüllgerät, sind nicht auszuschließen. Zukünftige Studien könnten daher die Stichprobenziehung breiter anlegen und auf einen größeren Umfang der zu realisierenden Stichprobe zielen.

Zusammengefasst zeigen die Voruntersuchungen der vorliegenden Arbeit weder positive Moduseffekte des Smartphones für sensible Fragen als ersten Ansatzpunkt der Antwortqualität noch Hinweise auf Moduseffekte des Smartphones für die Unit-Nonresponse. Während sich in der ersten Studie keine Unterschiede für Stichprobenkennwerte zeigen, fällt die Item-Nonresponse tendenziell am Smartphone höher aus. Die zweite Studie demonstriert zudem, dass das Smartphone und der PC/Laptop als Ausfüllgeräte in Onlinebefragungen nicht komplett identisch wahrgenommen werden. Die Hauptuntersuchungen der vorliegenden Arbeit erweitern die Prüfung auf Moduseffekte des Smartphones im TSE-Framework gezielt für die Antwortqualität und gerätassoziierte Aufmerksamkeitszustände (siehe Herleitung in den Kapiteln 3.2, 3.3, 4.3 und 4.4). Damit wird ein bisher unterrepräsentiertes und teilweise gänzlich unbeachtetes Forschungsfeld des Smartphoneinflusses auf den TSE erschlossen.

5.3 Studie 3 – Aufmerksamkeitsstrukturen während der Befragungsteilnahme

Die Hauptuntersuchungen der vorliegenden Arbeit legen den Forschungsschwerpunkt zur Abschätzung möglicher Moduseffekte im Framework des TSE auf den Einfluss, den das Smartphone auf die Antwortqualität in Onlinebefragungen ausübt. Smartphones führen zwar bei sensiblen Themen zu vergleichbaren Stichprobenkennwerten (siehe Kapitel 5.1), die gesamte Breite möglicher Einflüsse auf die Antwortqualität wurde aber noch nicht untersucht. Die Studie 3 der vorliegenden Arbeit nähert sich der Antwortqualität zunächst über die Analyse der Aufmerksamkeitsstrukturen an (siehe Kapitel 3.2.1). Anhand der Aufzeichnung des Blickverhaltens als physischer Ansatz der direkten Aufmerksamkeitsmessung wird in Studie 3 geprüft, ob Smartphones potenziell anfälliger für Aufmerksamkeitsablenkung sein könnten. Dafür schließt die dritte Studie (siehe Kapitel 4.3) als Between-Subjects-Laborstudie Kontexteinflüsse bewusst aus und prüft, ob sich das Blickverhalten während der Teilnahme an Onlinebefragungen je nach Ausfüllgerät unterscheidet. Die Versuchspersonen beantworten entweder mit einem handelsüblichen Smartphone oder an einem PC mit großem Bildschirm dieselbe Onlinebefragung. Um auch gerätbezogene Erfahrungseffekte und mögliche psychologische Einflüsse des eigenen Geräts auszuschließen, nutzen alle Versuchspersonen entweder dasselbe Smartphone oder denselben PC.

5.3.1 Methode

Analog zur ersten Studie der Forschungsreihe ist der Ansatz von Studie 3 ebenfalls innovativ. Sie stellt nach Kenntnis des Autors die erste wissenschaftliche Studie dar, die das Blickverhalten während der Befragungsteilnahme am Smartphone aufzeichnet und mit anderen Geräten

vergleicht. Entsprechend konnte für die Ermittlung der erforderlichen Stichprobengröße keine Effektstärke der Literatur entnommen werden, sondern erneut wird eine mindestens niedrige mittlere Effektstärke (Cohen's $d = 0,5$) für den Zwei-Gruppen-Vergleich (Smartphone vs. PC) angelegt. Die erforderliche Stichprobengröße umfasste entsprechend der Berechnung des Programms G*Power (Faul et al. 2007; Faul et al. 2009) $N = 102$ Proband*innen. Diese werden möglichst gleich auf die Untersuchungsgruppen Smartphone und PC aufgeteilt.

5.3.1.1 Versuchsablauf

Die Datenerhebung fand im Zeitraum vom 25.11. bis zum 18.12.2019 am Hauptcampus der BUW statt. Die Proband*innen wurden über persönliche Ansprache rekrutiert und hatten während ihrer Teilnahme die Möglichkeit, ein komprimiertes Persönlichkeitsprofil basierend auf ihren Angaben in der Befragung zu erhalten. Zudem wurde als Aufwandsentschädigung eine Süßwarenauswahl bereitgestellt. Die Erhebung wurde in einem geschlossenen und reizarmen Büroraum durchgeführt. Die Versuchsleitung war während der Teilnahme anwesend, um das Gelingen der Blickaufzeichnung zu kontrollieren. Den Teilnehmer*innen wurde ein Sitzplatz zugewiesen, der sie durch einen Sichtschutz von externen Störeinflüssen und dem direkten Kontakt zur Versuchsleitung, die ebenfalls mit Sichtschutz umgeben im Abstand von zwei Metern gegenüber saß, abschirmte. Pro Versuchstag wurde das Ausfüllgerät konstant gehalten, sodass die Versuchsgruppen täglich alternierten. Alle Versuchspersonen nahmen in der Bedingung teil, die an ihrem Untersuchungstag angesetzt war (Smartphone oder PC), ohne Einsicht in die Unterschiedlichkeit der Gruppen zu haben. Die Zuteilung zu den Versuchsgruppen war daher nur eingeschränkt randomisiert, da eine vollständige Randomisierung aufgrund der freiwilligen Teilnahme und des tageweisen Wechsels der Versuchsbedingung nicht möglich war (Campbell und Stanley 1963, S. 34-37; Shadish et al. 2002, S. 294-303). Die Proband*innen konnten ihre Gruppenzugehörigkeit aber nicht selbst aktiv beeinflussen und auch die Versuchsleitung nahm keinen aktiven Einfluss auf die Einteilung einzelner Personen zu bestimmten Gruppen. Vor Beginn der Untersuchung wurden die Proband*innen transparent über die Rahmenbedingungen der Teilnahme aufgeklärt. Teil der Blickaufzeichnung waren Videoaufnahmen der Proband*innen, denen diese vorab schriftlich zustimmten (siehe **Anhang C**). Es wurde betont, dass sie die Teilnahme jederzeit ohne die Angabe von Gründen abbrechen konnten.

In beiden Versuchsbedingungen (PC und Smartphone) saßen die Proband*innen auf einem höhenverstellbaren Stuhl an einem Schreibtisch, auf dem das Gerät der jeweiligen Versuchsbedingung platziert war. Sowohl die Blickaufzeichnung am PC als auch am Smartphone fand mit

einem modernen „Tobii-X2 @60Hz“-Eye-Tracker statt und wurde über die Forschungssoftware „iMotions 8.1“ umgesetzt (iMotions 2021). Der Eye-Tracker wurde je nach Ausfüllgerät umpositioniert, um optimale Blickerkennung zu garantieren. In der PC-Bedingung arbeiteten die Proband*innen an einem 24"-Bildschirm (Auflösung: 1920x1080 Pixel), wobei der Eye-Tracker unterhalb des Bildschirms angebracht wurde und eine Kamera auf der Oberkante des Monitors die Teilnehmer*innen filmte (siehe *Abbildung 10*). Das Blickverhalten wird dabei direkt auf den Bildschirminhalt übertragen. Als Smartphone fungierte ein Huawei Mate 20 Lite (Bildschirmgröße: 6,3"; Auflösung: 1280x720 Pixel). Für die Blickaufzeichnung am Smartphone wurde ein „Mobile-Device Stand“ der Firma iMotions genutzt. Dieser besteht aus einer Plattform, auf der das Smartphone positioniert wird, und einem Unterbau, in den der Eye-Tracker eingelegt wird (siehe *Abbildung 10*). Der Bildschirm des Smartphones und die direkte Umgebung werden von einer Kamera, die an der Spitze des Stands befestigt wird, abgefilmt. In diese Videoaufzeichnung wird das Blickverhalten der Teilnehmer*innen projiziert. Die Armbewegungen der Nutzer*innen des Smartphones werden durch zwei Stützpfosten geleitet, damit der Eye-Tracker die Lichtreflektionen der Augenbewegungen ungehindert aufzeichnen kann. Rechts vom Mobile-Device Stand wurde auf dem Tisch eine zweite Kamera platziert, die die Versuchspersonen filmte. Die Einstellung der Sitzhöhe des Arbeitsplatzes richtete sich nach den Softwarevorgaben und variierte je nach Größe der Teilnehmer*innen. Der Abstand des

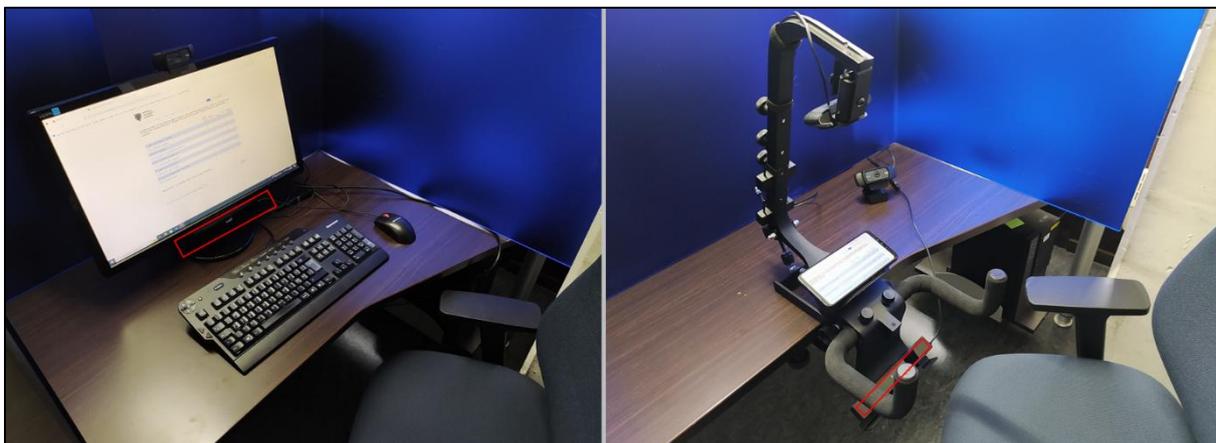


Abbildung 10. Versuchsaufbau der dritten Studie

Links ist der Versuchsaufbau der PC-Bedingung dargestellt. Rot umrandet ist der unter dem Monitor angebrachte Eye-Tracker. Auf dem Monitor ist zusätzlich eine Kamera platziert. Im rechten Teil ist der Versuchsaufbau der Smartphone-Bedingung dargestellt. Der Eye-Tracker (rot umrandet) wird in den Unterbau des Mobile-Device Stands eingelegt. Die Stützpfosten des Mobile-Device Stands leiten die Handbewegungen der Teilnehmer*innen, die um diese herumgreifen, damit die Blickaufzeichnung nicht durch die eigenen Arme oder Hände gestört wird. An der Spitze des Mobile-Device Stands ist eine Kamera platziert, die den Bildschirm und die unmittelbare Umgebung abfilmt, um das Blickverhalten in diese Szenerie übertragen zu können. Zusätzlich filmt eine zweite Kamera auf dem Tisch die Teilnehmer*innen.

Quelle: Eigene Aufnahme

Gesichts zum Bildschirm lag sowohl für den PC als auch für das Smartphone zwischen 60 und 70 cm, wobei 60 cm nicht unterschritten werden durften. Zu Beginn der Blickaufzeichnung wurde der Eye-Tracker jeweils auf die Versuchsperson kalibriert. Die Proband*innen blickten dabei nach Aufforderung auf neun verschiedene Punkte (von oben links nach unten rechts) im Projektionsbereich der Blickaufzeichnung. In der PC-Gruppe wurden die Kalibrierungspunkte direkt am Bildschirm präsentiert. Für die Smartphone-Gruppe wurde eine Kalibrierungsplatte in den Stand eingespannt. Erst nach erfolgreicher Kalibrierung startete die Untersuchung. In beiden Untersuchungsgruppen war dann die Onlinebefragung bereits geöffnet und der Internetzugang permanent gewährleistet. Die Smartphone-Gruppe nahm für die größtmögliche gerätübergreifende Darstellungskongruenz im horizontalen Landscape-Modus teil.

Der Fragebogen der Onlinebefragung wurde über die Befragungsplattform SoSci Survey (Leiner 2023) umgesetzt. Parallel lief die Blickaufzeichnung über die Software iMotions 8.1 (iMotions 2021). Die Teilnehmer*innen benötigten zur Beantwortung der 74 Items im Schnitt 9,71 Minuten. Da die Items der Onlinebefragung nicht inhaltlich für die Hypothesenprüfung ausgewertet werden, sondern diese auf das Blickverhalten während der Teilnahme zielt, wurde auf einen Pretest des Fragebogens verzichtet. Zudem besteht die Befragung hauptsächlich aus einer etablierten, validierten Skala (siehe Kapitel 5.3.1.3). Der Versuchsaufbau hingegen wurde ausführlich getestet und mehrfach angepasst, um eine möglichst angenehme Teilnahme an der Eye-Tracking Untersuchung zu gewährleisten. Moderne Eye-Tracker gewähren den Versuchspersonen relativ viel Bewegungsspielraum und sind recht robust in der Blickaufzeichnung, allerdings ist die hohe räumliche Auflösung bei zu starker Bewegung oder Haltungsänderungen gefährdet. Nach Abschluss der Onlinebefragung wurde die Video- und Blickaufzeichnung beendet und ein umfangreiches Debriefing durchgeführt, das die zugrundeliegende Fragestellung umfassend vorstellte. Für die gesamte Untersuchung mit Begrüßung, Einstellung des Eye-Trackers, Teilnahme an der Onlinebefragung und Verabschiedung betrug der Zeitbedarf pro Versuchsperson zwischen 15 und 25 Minuten.

5.3.1.2 Stichprobe

Die Stichprobe bestand vornehmlich aus Studierenden der BUW und wurde punktuell durch Personen aus dem Umfeld des Forschungsteams ergänzt, die dem Teilnahmeaufruf folgten. Menschliche Blickprozesse als Aufmerksamkeitsmaß sind universale menschliche Variablen, sodass hier auch Studierendensamples und soziodemografisch eher homogene Stichproben als repräsentativ angesehen werden können (Rad, Martingano und Ginges 2018). Die Gesamtzahl

der Teilnehmer*innen betrug initial $N = 124$. Trotz des technisch weit entwickelten Eye-Trackers traten vereinzelt Ausfälle auf, bei denen die Blickaufzeichnung aussetzte. Für das Studienziel ist die permanente Blickaufzeichnung aber grundlegend, sodass eine gerätspezifische Ausreißeranalyse durchgeführt wurde, die Proband*innen markierte, deren Gesamtdauer der Blickaufzeichnung als Anteil an der Studiendauer je nach Ausfüllgerät mehr als zwei Standardabweichungen nach unten vom Mittelwert abwich. Dies kann als Indikator für wiederholte technische Aussetzer gewertet werden, da die Dauer der Blickaufzeichnung die Dauer der Fragebogenbearbeitung hier übermäßig stark unterschreitet. Neben den 13 Datenreihen, die auf diesem Weg ausgeschlossen wurden, schlug die Datenerhebung auch für zwei weitere Proband*innen fehl, bei denen durch zu starke Veränderungen der Sitzposition die Kalibrierungseinstellungen nicht mehr adäquat waren. Die finale Gesamtstichprobe für die Datenanalyse umfasste $N = 109$ Personen, die im Schnitt 24,13 Jahre alt waren ($SD = 4,01$, Altersspanne: 18 bis 51 Jahre). 55 Teilnehmer*innen nahmen mit dem PC teil und 54 Personen mit dem Smartphone. 53,21% der Stichprobe waren männlich ($n = 58$) und 46,79% weiblich ($n = 51$). Die Proband*innen verteilten sich nach Geschlecht gleichermaßen auf die Untersuchungsgruppen ($\chi^2(1) = 1,102$, $p = 0,294$). Auch das Alter unterschied sich zwischen der PC-Gruppe ($\bar{x} = 24,31$, $SD = 3,70$) und der Smartphone-Gruppe ($\bar{x} = 23,94$, $SD = 4,58$) nicht ($t(107) = 0,463$, $p = 0,644$).

5.3.1.3 Fragebogenbeschreibung

Bei der Vorstellung des Fragebogens sind alle Variablenbezeichnungen, die sich auf Items oder Konstrukte beziehen, kursiv dargestellt. Die erste Fragebogenseite ließ die Teilnehmer*innen über das eigentliche Untersuchungsziel im Unklaren und stellte ein Cover-Thema vor. Ziel der Studie sei es, die Verarbeitungsstrukturen während der Befragungsteilnahme anhand der Eye-Tracking Daten hinsichtlich verschiedener Persönlichkeitsdimensionen zu organisieren, um daraus Implikationen für die Gestaltung von Onlinebefragungen abzuleiten. Für Smartphone-Teilnehmer*innen wurde der Hinweis ergänzt, dass das Smartphone als modernes Ausfüllgerät für die Untersuchung verwendet wird. Entsprechend konnten Proband*innen im Vorfeld nicht erkennen, dass das Ausfüllgerät interindividuell variierte. Zudem unterschied sich – basierend auf den Erkenntnissen der Erprobungsphase des Versuchsaufbaus – der Zeithorizont, der jeweils als Teilnahmedauer genannt wurde. PC-Teilnehmer*innen erhielten den Hinweis, dass die durchschnittliche Teilnahmedauer zwischen acht und zehn Minuten liege, während diese für das Smartphone (10 bis 15 Minuten) höher angesetzt wurde. Die Teilnehmer*innen bestätigten, dass ihre Daten im Rahmen der europäischen DSGVO zu rein wissenschaftlichen Zwecken verarbeitet und genutzt werden dürfen. Zu Beginn des Befragungsteils (S. 2) gaben die

Proband*innen ihr *Geschlecht*, ihr *Alter* und ihren höchsten erreichten Bildungsabschluss (*Bildung*) an.

Die nächsten sechs Seiten (S. 3-8) präsentierten den Proband*innen jeweils zehn Items der deutschen Version des Big-Five-Persönlichkeitsinventars BFI-2 (Danner et al. 2016; Danner et al. 2019). Das BFI-2 erfasst mit insgesamt 60 Items die fünf Persönlichkeitsdimensionen *Verträglichkeit*, *Negative Emotionalität*, *Offenheit für Erfahrungen*, *Gewissenhaftigkeit* und *Extraversion* (jeweils zwölf Items). Die Items werden auf einer vollverbalisierten, fünfstufigen Zustimmungsskala (1 = „stimme überhaupt nicht zu“ bis 5 = „stimme voll und ganz zu“) beantwortet. Dabei wurde die Reihenfolge der Items des Originalinventars beibehalten.

Die folgende Seite (S. 9) erfragte, ob schon an wissenschaftlichen Untersuchungen teilgenommen (*Teilnahmeerfahrung*) wurde (ja/nein), wenn ja, an wie vielen im vergangenen Monat (*Teilnahme im letzten Monat*), und ob schon konkret an einer *Eye-Tracking Studie* teilgenommen wurde (ja/nein). Zudem gaben die Versuchspersonen an, welches Gerät sie für die Teilnahme an Onlinebefragungen (*Wunschgerät*) bevorzugen (Smartphone, Tablet, Laptop, PC, anderes Gerät und zwar ...) und konnten mit einem Freitext auf die offene Frage antworten, wie Onlinebefragungen aufgebaut sein sollten, damit sie motiviert sind, teilzunehmen (*Studienaufbau*). Zudem wurde gefragt, ob das eigene, komprimierte Persönlichkeitsprofil angezeigt werden soll (ja/nein). Dieses wurde basierend auf den eigenen Angaben bei den 60 BFI-Items für jede*n Teilnehmer*in im Hintergrund automatisiert erstellt, indem die Werte zunächst dimensionsweise aggregiert und dann nach Geschlecht unterschieden an die Vergleichswerte der Normstichprobe des Inventars angelegt wurden (vgl. Danner et al. 2016). Seite 10 stellte das Persönlichkeitsprofil, wenn gewünscht, einschließlich einer knappen Definition der fünf Dimensionen dar. Die Definitionen der Dimensionen und die Formulierungen der Textbausteine des Persönlichkeitsprofils wurden sorgfältig aus der Big-Five Forschung und den Formulierungen des Originalinventars abgeleitet (vgl. Costa und McCrae 2008; Danner et al. 2019; McCrae und Costa 1987). Abschließend beurteilten die Befragten jeweils auf fünf- bis siebenstufigen, numerisch verankerten, endpunktverbalisierten Skalen die *wahrgenommene Befragungsdauer* (1 = „sehr kurz“ bis 7 = „sehr lang“), das *Interesse* an der Befragung (1 = „gar nicht interessant“ bis 5 = „sehr interessant“) und die *Gesamtbewertung* der Befragung (1 = „sehr schlecht“ bis 5 = „sehr gut“) (de Bruijne und Wijnant 2013, S. 490-492). Der gesamte Fragebogen der Studie 3 ist in *Anhang D* einzusehen.

5.3.1.4 Eye-Tracking Maße und Operationalisierungen der Hypothesen

Grundlage der Hypothesenprüfung ist die **H3** (*Proband*innen, die an Onlinebefragungen mit dem Smartphone teilnehmen, blicken häufiger von den Befragungsinhalten weg als Proband*innen, die mit dem PC/Laptop teilnehmen*). Verglichen werden die Bildschirme der verschiedenen Ausfüllgeräte als Areas of Interest (AOIs) und das Blickverhalten innerhalb der Bildschirmgrenzen. Auch für das vergleichsweise kleinere Smartphone stellt der gesamte Bildschirm eine ausreichend große AOI dar. Die hohe Bildschirmauflösung (1280x720 Pixel) überschreitet den kritischen Grenzwert an kleinen Bildschirmen für Eye-Tracking (25x30) deutlich (Chen et al. 2003). In der Gerätkategorie PC/Laptop wird als Ausfüllgerät ein PC eingesetzt.

Die Aufmerksamkeitsstrukturen der Befragungsteilnahme werden anhand der erhobenen Blickdaten gruppiert. Bei der Blickaufzeichnung werden kontinuierlich die Geschwindigkeit der Augenbewegungen und für räumlich engere Augenbewegungen die Grenzen, in denen sich diese bewegen, ermittelt und abgespeichert. Die Software iMotions 8.1 übersetzt die Rohdaten der Blickaufzeichnung in Fixationen und Sakkaden, die für die Hypothesenprüfung herangezogen werden (iMotions 2021). Die Klassifikation in Sakkaden und Fixationen erfolgt über den „Velocity-Threshold Identification“-Filter (I-VT-Filter) (Holmqvist et al. 2011, S. 147-175), wobei für beide Versuchsbedingungen (PC und Smartphone) einheitliche Grenzwerte genutzt werden. Es werden zunächst 20 Millisekunden (ms) als Zeitfenster eines Datenpunktes festgelegt, um Sequenzen von Augenbewegungen bilden zu können und darauf aufbauend die Geschwindigkeit der Augenbewegungen zu ermitteln. Innerhalb dieses Zeitfensters wird für jeden Datenpunkt die Winkelgeschwindigkeit (in °/Sekunden) ermittelt, die angibt, wie schnell sich der Winkel, den das Auge beschreibt, im Zeitverlauf um eine Achse verändern. Die axiale Abweichung (in °) zwischen den beiden Extremwerten eines Zeitintervalls wird in Abhängigkeit von der Entfernung des Auges vom Bildschirm ermittelt und durch den zeitlichen Abstand (in Sekunden) zwischen diesen Extremwerten dividiert. Für die Winkelgeschwindigkeit wird ein Grenzwert von 30°/Sekunde angelegt. Wird der Grenzwert unterschritten, ist die Winkelgeschwindigkeit sehr gering und das Auge verharrt an einer Stelle. Dies wird als Fixation klassifiziert. Wird der Schwellenwert hingegen überschritten, findet eine Augenbewegung in Form einer Sakkade statt.

Bezogen auf die AOIs (Bildschirmgröße und Befragungsinhalte, siehe nachfolgend Kapitel 5.3.2) der beiden Vergleichsgruppen manifestiert sich das Blickverhalten in verschiedenen Metriken, die iMotions für festgelegte AOIs ausgibt. Die Bezeichnungen aller Variablen, die aus der Blickaufzeichnung für die Auswertung resultieren, werden kursiv dargestellt. Bei allen Hypothesen agiert das Ausfüllgerät (*Gerät*) als dichotome UV mit den beiden Ausprägungen

PC (Referenzkategorie = 0) und Smartphone (mit 1 kodiert). Für die Hypothese **H3.1** (*Der Blick der Proband*innen, die an Befragungen mit dem Smartphone teilnehmen, bewegt sich häufiger außerhalb des Bereichs der Befragung als bei Proband*innen, die mit dem PC/Laptop teilnehmen*) wird geprüft, wie groß der Anteil der Blickbewegung innerhalb der AOI im Vergleich zur Gesamtzeit der Aufzeichnung ist (*BlickzeitAOI*). Die Hypothese **H3.2** (*Die Fixationsdauer auf Befragungsinhalte ist bei Proband*innen, die mit dem Smartphone teilnehmen, geringer als bei Proband*innen, die mit dem PC/Laptop teilnehmen*) nutzt die Fixationen und zieht analog zur **H3.1** den Anteil der Fixationsdauer innerhalb der AOI an der Gesamtzeit der Blickaufzeichnung heran (*FixationenAOI*). Die **H3.3** postuliert: *Proband*innen, die am Smartphone an der Befragung teilnehmen, zeigen häufiger Refixationen in die Befragung als Personen, die mit dem PC/Laptop teilnehmen*. Hierfür wird die Häufigkeit der *Refixationen* in die AOI ermittelt, die angibt, wie oft nach Verlassen der AOI wieder längere Zeit am Stück (mindestens 100ms) in die AOI geblickt und somit dort fixiert wird. Die Analyse der **H3.3** wird ergänzt durch die Variable *Rückkehr*, die ermittelt, wie häufig Proband*innen mit dem Blick in die AOI zurückkehren, ohne dort zu fixieren (< 100ms). Zuletzt konzentrieren sich die Hypothesen **H3.4** (*Für Proband*innen, die am Smartphone an der Befragung teilnehmen, ist die Anzahl der Fixationen im Bereich der Befragungsinhalte größer als für Proband*innen, die mit dem PC/Laptop teilnehmen*) und **H3.5** (*Die durchschnittliche Fixationsdauer ist bei Smartphone-Teilnehmer*innen geringer als bei PC/Laptop-Teilnehmer*innen*) auf die Verarbeitungsqualität der Blickbewegungen. Für die **H3.4** wird die *Fixationszahl* berechnet als Anzahl der Fixationen in der AOI und die **H3.5** legt die durchschnittliche *Fixationsdauer* dieser Fixationen an.

5.3.2 Ergebnisse

Die Datenauswertung erfolgte hauptsächlich mit der Statistiksoftware JASP (JASP Team 2023) und wurde punktuell durch die Programme SPSS (IBM Corp. 2021) und R (R Core Team 2023) ergänzt. Alle AVn basieren auf den Eye-Tracking Daten und sind metrisch. Die Hypothesenprüfung erfolgt einheitlich über unabhängige *t*-Tests (Bortz und Schuster 2016, S. 120-124). In den Zusatzbetrachtungen werden mindestens fünfstufige Ratingskalen als quasimetrisch behandelt (Harpe 2015; Urban und Mayerl 2018, S. 301-302).

Die Analysen der Studie 3 vergleichen das Blickverhalten der Smartphone- und PC-Teilnehmer*innen während der Befragungsteilnahme. Grundlage dieser Analysen ist die Festlegung aussagekräftiger AOIs. Für beide Geräte werden dabei jeweils zwei AOIs definiert. Am PC agiert einerseits der Gesamtbildschirm als räumlich weitere AOI (*PC_weit*) und zusätzlich wird

eine engere AOI festgelegt (*PC_eng*), die lediglich die Befragungsinhalte umgrenzt (siehe *Abbildung 11*). Am Smartphone entsprechen die Gerätgrenzen direkt den Befragungsgrenzen, so dass die Bildschirmgröße hier die engere AOI repräsentiert (*SP_eng*). Zudem wird aber auch der deutlich kleineren Gerätgröße des Smartphones Rechnung getragen und eine räumlich weitere AOI (*SP_weit*) definiert, die um die Gerätgrenzen einen Rahmen mit 1,5 cm Abstand zeichnet, um leichte Abweichungen und mögliche Aufzeichnungsartefakte abzufedern (siehe *Abbildung 11*). Die Hypothesenprüfung erfolgt in drei Analyseschritten. Der erste Schritt legt den Fokus auf die Gerätgrenzen. Hierfür werden die AOIs *SP_eng* und *PC_weit* verglichen, deren Grenzen jeweils den Bildschirmrändern entsprechen. Dafür ist der erwartete Unterschied maximal. Im zweiten Schritt vergleichen die Analysen das Blickverhalten nur für tatsächliche Befragungsinhalte und somit die beiden engen AOIs *SP_eng* und *PC_eng*. Bei Abweichungen verlässt der Blick zwar die Befragungsinhalte, aber am PC nicht zwingend das Gerät. Der letzte Schritt prüft die Robustheit möglicher Unterschiede und vergleicht die räumlich weiteren AOIs *SP_weit* und *PC_weit*, wobei mögliche Nachteile des kleineren Smartphones und geringe Schwankungen in der Blickaufzeichnung teilweise aufgefangen werden.

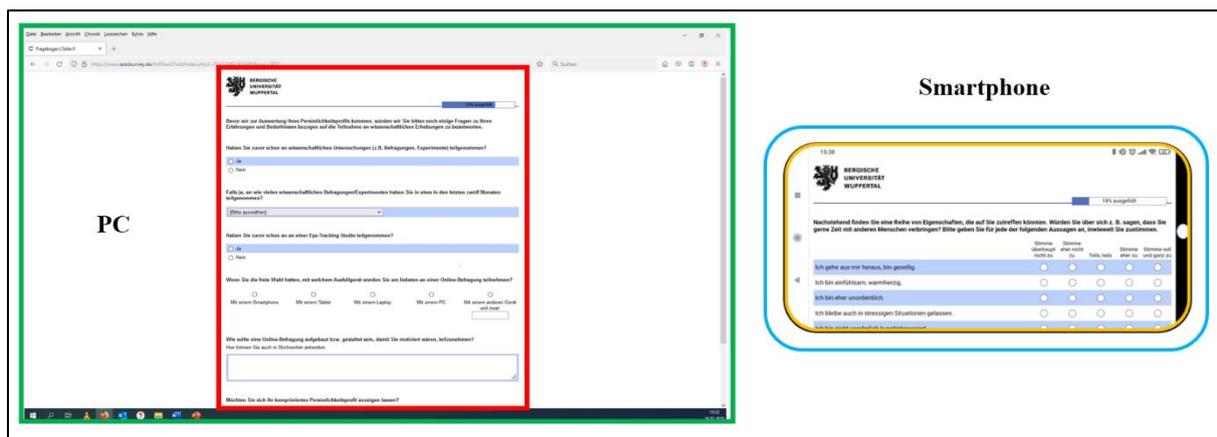


Abbildung 11. Areas of Interest (AOI) in Studie 3

Rot = Befragung PC (*PC_eng*) **Orange** = Befragung/Gerätsgrenzen Smartphone (*SP_eng*)
Grün = Gerätsgrenzen PC (*PC_weit*) **Blau** = Spielraum Smartphone (*SP_weit*)

5.3.2.1 Voranalysen und Voraussetzungsprüfung

Abseits der grundsätzlichen Datenbereinigung (siehe Kapitel 5.3.1.2) können alle verbleibenden Datensätze in die Analysen eingehen. Während die drei AVn *BlickzeitAOI*, *FixationenAOI* und *Fixationsdauer* unabhängig von der Dauer der AOI-Aktivierung sind, muss diese für die kumulierten Daten einbezogen werden. In Vorbereitung der Hypothesenprüfung werden die AVn *Refixationen*, *Rückkehr* und *Fixationszahl* aufgrund der interindividuell variierenden Bearbeitungsdauer an dieser normiert. Die *Fixationszahl* wird dafür durch die Gesamtdauer der AOI-Aktivierung geteilt und der resultierende Wert in die Hypothesenprüfung einbezogen. Die

Variablen *Refixationen* und *Rückkehr* werden in gleicher Weise normiert und zusätzlich mit 10000 multipliziert, da ansonsten sehr kleine Werte entstehen, deren geringe Variation die Interpretation der inferenzstatistischen Hypothesenprüfung erschwert. Durch die Normierung können alle Hypothesen über *t*-Tests für unabhängige Stichproben geprüft werden, für die die Voraussetzungen gelten, dass die Beobachtungen unabhängig sind, die AV in den Gruppen normalverteilt ist, Varianzhomogenität vorliegt und die AV mindestens intervallskaliert ist (Bühner und Ziegler 2017, S. 301). Das experimentelle Forschungsdesign stellt die Unabhängigkeit der Beobachtungen sicher und alle ermittelten AVn sind metrisch. Aufgrund der Gruppengrößen ($n > 30$) greift erneut der zentrale Grenzwertsatz, dass die Stichprobenkennwerte annähernd normalverteilt sind. Für alle berichteten Analysen wird die Varianzhomogenität geprüft und liegt vor, wenn nicht anders angegeben. Bei Varianzheterogenität wird alternativ der Welch-*t*-Wert berichtet. Die Laborsituation qualifiziert *t*-Tests als Prüfverfahren, da keine variierenden Kontextvariablen vorliegen und da demografische Unterschiede das Blickverhalten und das generelle Blicksystem des Menschen nicht beeinflussen¹⁸.

In Vorbereitung der Hypothesenprüfung werden zunächst intraindividuelle Vergleiche innerhalb der beiden Untersuchungsgruppen für alle AVn angestellt. Sowohl für die PC-Teilnehmer*innen als auch für die Smartphone-Gruppe wird geprüft, ob sich die Blickparameter für die enge und die weite AOI jeweils unterscheiden und ob diese Aufteilung sinnvoll ist. Für die Vergleiche werden *t*-Tests für abhängige Stichproben verwendet (Bortz und Schuster 2016, S. 124-125). Für diese gelten die Voraussetzungen, dass die abhängigen Variablen mindestens intervallskaliert sein müssen (erfüllt), und die Differenzen der Beobachtungen sollten zusätzlich normalverteilt sein (Bortz und Schuster 2016, S. 125). Bei mehr als 30 Beobachtungen pro Gruppe greift auch für die Mittelwerte der Differenzen der zentrale Grenzwertsatz (Bortz und Schuster 2016, S. 586), sodass in der Folge die Ergebnisse des *t*-Tests berichtet werden können¹⁹. Für fast alle AVn (*BlickzeitAOI*, *FixationenAOI*, *Refixationen*, *Rückkehr*, *Fixationszahl*, *Fixationsdauer*) unterscheiden sich die Werte der engen AOIs und der weiten AOIs in beiden Gruppen jeweils signifikant ($t_{min}(54) = 6,453, p < 0,001$). Lediglich der Anteil der kumulierten Fixationsdauer an der Gesamtzeit (*FixationenAOI*) ist am PC für die beiden AOIs *PC_eng* und *PC_weit* ähnlich. Die Ergebnisse fundieren die Unterscheidung der weiten und der engen AOI und das mehrstufige Vorgehen der Hypothesenprüfung. Die vollständigen Ergebnisse der abhängigen *t*-Tests finden sich in **Anhang E**.

¹⁸ Alle berichteten Befunde persistieren, wenn die Gruppenvergleiche unter Einbezug möglicher Kontrollvariablen (*Alter*, *Geschlecht*, *Bildung*) als allgemeines lineares Modell (ALM) durchgeführt werden.

¹⁹ Die Prüfung wurde über nicht-parametrische Prüfverfahren abgesichert, die zu denselben Ergebnissen führten.

5.3.2.2 Hypothesenprüfung

Die Blickdatenermittlung für die AOIs bezieht sich auf die gesamte Befragungsdauer mit Ausnahme der beiden Freitextvariablen (Altersangabe und Beschreibung der Bedingungen, die die Befragten zur Teilnahme an Onlinebefragungen motivieren), da sich dort das Format der Texteingabe zwischen den Versuchsgruppen stark unterscheidet. Während am Smartphone im Touchformat auf dem Bildschirm selbst geantwortet wird, erfolgt die Texteingabe am PC über eine externe Tastatur, sodass hier auch bei aufmerksamem Antworten der Blick temporär vom Bildschirm entfernt wird. Die Hauptanalysen schließen diese Zeitabschnitte daher aus.

Die fünf Hypothesen (**H3.1**, **H3.2**, **H3.3**, **H3.4** und **H3.5**) werden, wie beschrieben, in drei Schritten geprüft. Im ersten Schritt werden die Gerätgrenzen verglichen (*PC_weit* vs. *SP_eng*), im zweiten Schritt die Befragungsgrenzen (*PC_eng* vs. *SP_eng*) und zuletzt im dritten Schritt ein robuster Vergleich mit Puffer für die Smartphonegrenzen angestrebt (*PC_weit* vs. *SP_weit*). Da alle Hypothesen gerichtet formuliert sind, werden die Testungen einseitig angelegt. Berichtet werden stets die Ergebnisse des ersten Analyseschrittes und von diesen abweichende Befunde aus den nachfolgenden Schritten. *Tabelle 10* stellt die Ergebnisse aller Analysen gebündelt dar. Für die Variable *BlickzeitAOI* als Grundlage der Hypothese **H3.1** resultiert am PC ein höherer Anteil an Blickbewegung innerhalb der Gerätgrenzen (keine Varianzhomogenität: $F(1, 107) = 4,035, p = 0,047$; $t(90,37) = 3,078, p = 0,001$) und der Befragungsinhalte ($t(107) = 2,554, p = 0,006$) als am Smartphone. Für den robusten Vergleich des dritten Schrittes verschwindet der signifikante Unterschied ($t(107) = 0,853, p = 0,200$). Konform zur Hypothese **H3.2** ist der Anteil der Fixationszeit an der Gesamtzeit (*FixationenAOI*) für alle drei Vergleiche am PC signifikant größer als am Smartphone (keine Varianzhomogenität: $F(1, 107) = 11,451,$

	Schritt 1: <i>PC_weit</i> vs. <i>SP_eng</i>		Schritt 2: <i>PC_eng</i> vs. <i>SP_eng</i>		Schritt 3: <i>PC_weit</i> vs. <i>SP_weit</i>		
	<i>t</i> (df)	<i>p</i>	<i>t</i> (df)	<i>p</i>	<i>t</i> (df)	<i>p</i>	
H3.1: <i>BlickzeitAOI</i>	3,078 _(90,37)	* < 0,001	2,554 ₍₁₀₇₎	* 0,006	0,853 ₍₁₀₇₎	0,200	
H3.2: <i>FixationenAOI</i>	4,914 _(82,88)	* < 0,001	4,845 _(83,56)	* < 0,001	4,681 _(84,40)	* < 0,001	
H3.3:	<i>Rückkehr</i>	-1,829 ₍₁₀₇₎	* 0,035	-1,421 ₍₁₀₇₎	° 0,079	-0,463 ₍₁₀₇₎	0,322
	<i>Refixationen</i>	-1,799 ₍₁₀₇₎	* 0,038	-1,652 ₍₁₀₇₎	° 0,051	-1,260 ₍₁₀₇₎	0,195
H3.4: <i>Fixationszahl</i>	-5,560 _(84,67)	* < 0,001	-5,626 _(84,43)	* < 0,001	-5,736 _(83,74)	* < 0,001	
H3.5: <i>Fixationsdauer</i>	5,905 ₍₁₀₇₎	* < 0,001	5,945 ₍₁₀₇₎	* < 0,001	5,966 ₍₁₀₇₎	* < 0,001	

Tabelle 10. Ergebnisse der Hypothesenprüfung der Studie 3

Für *t*-Werte, deren Freiheitsgradzahl mindestens eine Dezimalstelle besitzt, liegt keine Varianzhomogenität für die AV vor, sodass hier der *t*-Wert des Welch-Tests berichtet wird.

* = signifikant ($p < 0,05$); ° = marginalsignifikant ($p < 0,1$)

$p < 0,001$; $t(82,88) = 4,914$, $p < 0,001$). Die normierte *Fixationszahl* der **H3.4** (keine Varianzhomogenität: $F(1, 107) = 10,613$, $p = 0,002$; $t(84,67) = -5,560$, $p < 0,001$) fällt am Smartphone konstant für alle Vergleiche größer aus, während die durchschnittliche *Fixationsdauer* der **H3.5** ($t(107) = 5,905$, $p < 0,001$) über alle drei Vergleichsschritte ebenfalls hypothesenkonform am PC signifikant größer ist als am Smartphone.

Etwas uneinheitlicher ist das Bild für *Refixationen* und *Rückkehr* des Blickes als Grundlage der Hypothese **H3.3**. *Refixationen* ($t(107) = -1,799$, $p = 0,038$) und *Blick-Rückkehr* ($t(107) = -1,829$, $p = 0,035$) in die AOIs finden beim Vergleich der Gerätgrenzen (Schritt 1) am Smartphone signifikant häufiger statt als am PC. Für die Befragungsgrenzen (Schritt 2) schwächt sich dieser Unterschied ab und verfehlt trotz einseitiger Testung das Signifikanzniveau knapp für die *Rückkehr* ($t(107) = -1,421$, $p = 0,079$) bzw. sehr knapp für die *Refixationen* ($t(107) = -1,652$, $p = 0,051$). Im dritten robusten Vergleich zeigt sich kein signifikanter Unterschied sowohl für *Refixationen* ($t(107) = -1,260$, $p = 0,195$) als auch für *Rückkehr* ($t(107) = -0,463$, $p = 0,322$).

5.3.2.3 Zusatzbetrachtungen

In die Hypothesenprüfung wurde aufgrund der gerätspezifischen Eingabemodalitäten die Freitextantwort der Variable *Studienaufbau* nicht einbezogen. Nicht alle Versuchspersonen beantworteten die Frage, was sie motivieren würde, an Onlinebefragungen teilzunehmen, sodass für die Blickzeitanalyse der Freitextantwort ein Subdatensatz ($n = 85$) verwendet wird. Teilnehmer*innen am PC und am Smartphone antworteten gleich häufig inhaltlich auf die Frage ($\chi^2(1) = 2,068$, $p = 0,151$) und benötigten dafür durchschnittlich unabhängig vom Gerät ähnlich viel Zeit ($t(84) = -1,485$, $p = 0,141$). Auch in der Textlänge unterschieden sich die Antworten der offenen Frage nicht. An beiden Geräten werden annähernd gleich viele Zeichen für die Formulierung der Antwort genutzt ($t(84) = 0,857$, $p = 0,394$).

Die Parameter der Blickaufzeichnung (sprich: die AVn der Hypothesenprüfung) bestätigen die Potenz des Eye-Tracking Ansatzes für die Messung der Aufmerksamkeit und zeigen erwartungskonforme Muster für die Freitextantworten. Hinsichtlich der AVn drehen sich viele der Effekte für alle drei Vergleichsschritte um, wenn diese nur für die Texteingabe betrachtet werden. Für die Variablen *BlickzeitAOI*, *FixationenAOI* und *Fixationszahl* sind die Werte nun am Smartphone signifikant höher als am PC (Angaben für geringsten t -Wert: keine Varianzhomogenität: $F(1, 84) = 12,019$, $p < 0,001$; $t_{min}(62,62) = -7,558$, $p < 0,001$), da Smartphone-Nutzer*innen bei der Texteingabe aufgrund des Touchformats permanent auf den Bildschirm schauen, während am PC der Blick wiederholt auf die separate Tastatur gerichtet werden muss. Gleichwohl bleiben einige gerätassoziierte Blickmuster erhalten. Die Fixationen am PC (*Fixationsdauer*) sind erneut für alle drei Vergleichsschritte durchschnittlich länger ($t_{min}(84) = 4,032$,

$p < 0,001$). Der Blick am Smartphone bleibt zudem sprunghaft und geht häufig über die Gerätgrenzen hinaus. Durch die Tastatureingabe verschwindet der Unterschied in den *Refixationen* zwischen den Untersuchungsgruppen für alle drei Vergleichsschritte ($t_{max}(84) = -0,504, p = 0,616$), da nun auch am PC häufiger weggeschaut wird. Auch wenn nach Wegblicken der Fokus gleich häufig wieder auf die Befragung gelegt wird, schauen trotz der Texteingabe am Gerät Smartphone-Nutzer*innen weiterhin insgesamt noch häufig von der Befragung weg (Mittelwert des Wegschauens über alle Vergleichsschritte: $\bar{x} = 36,26$). Die *Rückkehr* zur AOI ohne Fixation ist am PC nun sogar signifikant häufiger als am Smartphone ($t_{min}(84) = 4,375, p < 0,001$). Die vollständigen Befundmuster der Subanalysen zur Freitextantwort finden sich in **Anhang E**.

Unabhängig vom Ausfüllgerät bewerten die Proband*innen die Studie (*Gesamtbewertung*) insgesamt gleich gut ($t(107) = -0,165, p = 0,869$), gleich interessant (*Interesse*) ($t(107) = -1,548, p = 0,125$) und schätzen auch die *wahrgenommene Befragungsdauer* ähnlich lang ein ($t(107) = -1,237, p = 0,219$). Die persönliche Zeitwahrnehmung unterscheidet sich dabei von der *tatsächlichen Antwortzeit*. Am Smartphone benötigen die Teilnehmer*innen signifikant länger für die Beantwortung der Onlinebefragung als am PC ($t(107) = -2,480, p = 0,015$).

Abschließend zeigen die Zusatzbetrachtungen, dass das Ausfüllgerät auf die Wahrnehmung der Proband*innen wirkt. Die Teilnehmer*innen gaben abschließend das *Wunschgerät* für Onlinebefragungen an. Es findet sich, dass der PC/Laptop ($n = 56$) insgesamt signifikant häufiger als *Wunschgerät* genannt wird als das Smartphone ($n = 28; \chi^2(1) = 9,333, p = 0,002$). Diese klare Präferenz verschwindet bei der separaten Betrachtung der beiden Untersuchungsgruppen. Am Smartphone (48,57%) deklarieren im Vergleich zum PC (22,45%) signifikant mehr Personen das Smartphone auch als präferiertes Ausfüllgerät ($\chi^2(1) = 6,269, p = 0,012$).

5.3.3 Diskussion

Die Unterschiede zwischen klassischen Ausfüllgeräten wie dem PC/Laptop und dem Smartphone als vergleichsweise neuerem Gerät und die Eigenheiten des Smartphones werfen die Frage auf, inwiefern das Antwortverhalten in Onlinebefragungen durch das Ausfüllgerät selbst beeinflusst wird. Gerätspezifische Aufmerksamkeitsstrukturen bei der Teilnahme an Onlinebefragungen werden in der Forschungsliteratur bisher kaum untersucht. Der Theorieteil (siehe Kapitel 3.2 und 3.3) und das Forschungsprogramm (siehe Kapitel 4.3 und 4.4) der vorliegenden Arbeit führen indessen aus, dass gerade Aufmerksamkeitsstrukturen die Grundlage der Antwortqualität im TSE-Framework sind und daher in die Beurteilung möglicher Moduseffekte des Smartphones in Onlinebefragungen einbezogen werden sollten. Diese Aufmerksamkeitsstrukturen werden in der hier zu diskutierenden dritten Studie über die physische Messung des

Blickverhaltens explizit adressiert. Apparative Eye-Tracker erheben Blickindikatoren, die Aufmerksamkeitsprozesse in Onlinebefragungen direkt und robust abbilden (z. B. Baumgartner et al. 2018; Lenzner et al. 2011; Reingold 2014; Wedel und Pieters 2008, S. 142).

Die Ergebnisse der Studie 3 zeigen, dass sich das Blickverhalten am Smartphone während der Befragungsteilnahme systematisch vom Blickverhalten am PC unterscheidet. Am PC orientieren sich Proband*innen mit ihrem Blick insgesamt häufiger auf dem Bildschirm und im Sektor der Befragungsinhalte und fixieren auch länger innerhalb dieser Bereiche, was die Hypothesen **H3.1** und **H3.2** bestätigt. Somit richten Proband*innen am PC ihren Blick häufiger auf den Bildschirm des Ausfüllgeräts und auf den Befragungsinhalt als Befragte, die das Smartphone nutzen. Das Blickverhalten bildet eine Vielzahl der kognitiven Verarbeitungsstrukturen ab (siehe Kapitel 3.3.3.1), wobei die längere Blickdauer auf bestimmte Inhalte indiziert, dass diese auch besser verarbeitet werden (z. B. Jacob und Karn 2003, S. 585; Orquin und Mueller Loose 2013). Zudem entspricht der Ort der Fixation zumeist dem Aufmerksamkeitsfokus (z. B. Duchowski 2017, S. 250-251; Mayfrank et al. 1986), der somit am PC häufiger auf den Befragungsbereich und auch insgesamt auf das Gerät gerichtet ist. Darüber hinaus sind die Fixationen innerhalb der Gerät- und Befragungsgrenzen am PC in der vorgestellten Studie durchschnittlich länger, was Hypothese **H3.5** bestätigt. Längere Fixationsdauern können anzeigen, dass die fixierten Inhalte kognitiv effizienter verarbeitet werden (z. B. Honsel 2012, S. 36; Meißner und Oll 2019). Der Anteil der Blickbewegungen und der Anteil der Fixationen innerhalb der Gerätgrenzen sowie die durchschnittliche Fixationslänge auf Befragungsinhalte bewegen sich jeweils in der vorliegenden Studie bei PC-Teilnehmer*innen auf einem signifikant höheren Niveau als bei Befragten, die das Smartphone nutzen. Dies deutet an, dass am Smartphone die Aufmerksamkeit weniger nachhaltig auf dem Ausfüllgerät und somit auf der Befragung liegt und Smartphones potenziell anfälliger für Aufmerksamkeitsablenkungen sein könnten.

Neben der Fixationsdauer kann die Aufmerksamkeit auch anhand der Fixationshäufigkeit beschrieben werden. Häufige Fixationen kürzerer Dauer signalisieren dabei eher ineffiziente Verarbeitungsmechanismen (Jacob und Karn 2003, S. 585; Meißner und Oll 2019). Die Eye-Tracking Daten unterstützen die Hypothese **H3.4** und dokumentieren eine signifikant größere Anzahl von Fixationen am Smartphone als am PC bei zugleich kürzerer durchschnittlicher Fixationsdauer (durch **H3.5** bestätigt). Das Blickverhalten ist während der Befragungsteilnahme am Smartphone also deutlich unsteter. Während sprunghafter Augenbewegungen (Sakkaden) werden keine Informationen aufgenommen (z. B. Rayner 1998, 2009) und Sakkaden signalisieren Aufmerksamkeitsverlagerungen (Kowler et al. 1995). Mit dem Blick verlagert sich demnach auch die Aufmerksamkeit am Smartphone häufiger sprunghaft und ist weniger stabil.

Abseits der Blickmuster bezogen auf die Befragungsinhalte wird der Blick am Smartphone auch häufiger komplett vom Bildschirm und somit von der Befragung abgewendet. Die Hypothese **H3.3** wird durch eine größere Anzahl an Refixationen am Smartphone und zudem eine größere Anzahl an Blickabweichungen von der Befragung, die sich in der Häufigkeit der Blickrückkehr ohne Fixation ausdrückt, überwiegend bestätigt. Nur wenn die Vergleiche weniger streng sind und am Smartphone nicht die tatsächlichen Befragungsgrenzen als AOIs angelegt werden, zeigen sich weniger Unterschiede. Beide Befundmuster erweitern gleichwohl die Erkenntnisse. Es kommt häufiger zu Refixationen, die zwar potenziell auch dazu befähigen können, Aufmerksamkeitsverschiebung wieder aufzufangen, indem sie die Gedächtnisinhalte aktualisieren und so die Verarbeitung fördern (Tatler et al. 2005; Wedel und Pieters 2000; Zelinsky et al. 2011). Dies ist allerdings nur der Fall, wenn das Blickverhalten konstant und zielgerichtet ist (Gwizdka und Zhang 2015, S. 814). Die Ergebnisse der Studie 3 zeichnen aber ein gegenteiliges Bild. Die kürzeren Fixationsdauern (**H3.5**) bei zugleich häufigeren Fixationen untermauern die potenzielle Anfälligkeit des Smartphones für ineffiziente kognitive Verarbeitung und Aufmerksamkeitsdefizite während der Teilnahme. Ergänzend charakterisiert auch die häufigere Blickrückkehr zum Smartphone ohne folgende Fixation die Aufmerksamkeitsstrukturen ebenso als unstet und sprunghaft. Der Blick orientiert sich zwar häufiger zum Gerät zurück, allerdings ohne, dass dort dann kognitive Verarbeitung stattfindet.

Die Ergebnisse der Studie 3 bestätigen eindeutig, dass die Aufmerksamkeit der Nutzer*innen nicht unabhängig vom Ausfüllgerät ist (Dunaway et al. 2018). Die Aufmerksamkeit ist in Onlinebefragungen bei der Teilnahme mit dem Smartphone potenziell eingeschränkt, da das Blickverhalten unsteter und sprunghafter ist. Aufmerksamkeitsdefizite reduzieren die Antwortqualität und somit die Validität der Befragungsdaten (z. B. Bautista 2012, S. 38; Groves und Lyberg 2010; Wiersma 2013). Grundsätzlich binden größere Bildschirme häufiger die Aufmerksamkeit und die Leistung der Nutzer*innen ist insbesondere bei komplexen Aufgaben an größeren Bildschirmen besser (z. B. Chae und Kim 2004; Jones et al. 1999; Orquin und Mueller Loose 2013). Schon die geringere Bildschirmgröße des Smartphones könnte die aufmerksame Teilnahme also erschweren. Zudem sind die Nutzungsphasen am Smartphone weniger nachhaltig und flüchtiger (z. B. Kim et al. 2017; Lee et al. 2005; Molyneux 2018), sodass auch die mit dem Gerät verknüpften Gewohnheiten dazu führen könnten, dass Aufmerksamkeitsverluste am Smartphone häufiger sind. Die Interaktion mit dem Smartphone kann vielfältiger und herausfordernder erlebt werden (Ghose et al. 2013; Schlosser und Mays 2018; Tourangeau et al. 2017), was ebenfalls eher zu Aufmerksamkeitsverlagerungen führen könnte. Auch die fast immer längere Teilnahmedauer an Onlinebefragungen mit dem Smartphone (z. B. Andreadis

2015a, S. 74; Mavletova 2013; Struminskaya et al. 2015; Tourangeau et al. 2018) zeigt sich in Studie 3, wobei die Blickdaten hierfür einen neuen möglichen Erklärungsansatz eröffnen. Da Proband*innen am Smartphone häufiger von der Befragung wegschauen, ihr Blickziel kleiner und ihr Blickverhalten weniger beständig ist, benötigen sie auch mehr Zeit für Refixationen und die Aktualisierung der Gedächtnisinhalte, sodass eine gewissenhafte und aufmerksame Teilnahme nur mit erhöhtem Zeitaufwand möglich ist.

Neben der Hypothesenprüfung unterstreichen die Analysen der dritten Studie erneut die Notwendigkeit der Smartphoneteilnahme an Onlinebefragungen. Analog zur Forschungsliteratur zeigt sich für einen wesentlichen Teil der Proband*innen eine Präferenz für die Teilnahme mit dem Smartphone (Revilla et al. 2016b). Zudem untermauern die Ergebnisse der Studie 3 auch die Erkenntnis aus der zweiten Studie der Forschungsreihe, dass das genutzte Ausfüllgerät die Einschätzung von Nutzer*innen insofern klar beeinflusst, dass sie ihr verwendetes Gerät selbst eher als geeignet wahrnehmen. Die Versuchspersonen gestalten den Einsatz verschiedener Ausfüllgeräte also interindividuell selbst so aus, dass ihre sozialen und psychologischen Bedürfnisse saturiert werden. Diese Bedürfnisbefriedigung ist elementar für die motivierte Teilnahme an Befragungen (Groves et al. 1992). Allerdings wurden in Studie 3 PCs und Laptops nicht als gemeinsame Antwortkategorie präsentiert, was die Präferenzhäufigkeiten systematisch verzerren könnte und in künftigen Studien durch die Kopplung beider Geräte als gemeinsame Antwortoption bei der Abfrage abgesichert werden sollte.

Die artifizielle Laborumgebung – zusätzlich gesteigert durch den apparativen Aufbau des Eye-Trackers – maximiert zwar die interne Validität des Gerätvergleichs, unterscheidet sich aber klar von realen Teilnahmebedingungen. Die Antwortqualität in Befragungen ist aber kontextsensibel (z. B. Huang et al. 2012; Johnson 2005; Wiersma 2013) und Umgebungseinflüsse wirken unmittelbar auf das Antwortverhalten (Höhne et al. 2020a; Miller 2008). Das Smartphone könnte die Antwortqualität daher systematisch gefährden, da Onlinebefragungen für Störungen durch äußere Einflüsse im Erhebungsprozess besonders anfällig sind (Edwards 2019; Struminskaya et al. 2015) und der Teilnahmekontext am Smartphone im Feld nicht kontrollierbar und diverser ist als bei klassischen Ausfüllgeräten wie dem PC/Laptop (z. B. Antoun et al. 2017; Höhne und Schlosser 2019; Mavletova und Couper 2013; Revilla et al. 2016a). Am Smartphone sind entsprechend multiple Ablenkungsreize während der Teilnahme denkbar und wahrscheinlich (de Bruijne und Oudejans 2015, S. 141; Revilla et al. 2016a; Toninelli und Revilla 2016b; Zwarun und Hall 2014). Das Ausfüllverhalten im Feld könnte dabei einerseits von diesen Kontexteinflüssen direkt beeinflusst werden und andererseits könnten die Einflüsse am Smartphone anders wirken als am PC/Laptop und dadurch die Antwortqualität verringern.

Für klassische Antwortqualitätsindikatoren (siehe Kapitel 3.2.3) finden sich im Feld bereits wiederholt Unterschiede, wenn das Smartphone als Ausfüllgerät genutzt wird. Während in der Forschungsliteratur die Antworten auf offene Fragen am Smartphone kürzer ausfallen oder diese häufiger ausgelassen werden (z. B. Erens et al. 2019; Lambert und Miller 2015; Mavletova 2013; Wells et al. 2014), tritt das in Studie 3 nicht auf. Als relative kurze Laborstudie könnte der soziale Druck der Anwesenheit der Versuchsleitung das Antwortverhalten der Proband*innen beeinflussen, weswegen die Analysen auf die objektiven Blickmuster begrenzt werden. Erst ohne persönliche Interaktion und Reaktanz der Experimentalsituation werden etwa kürzere Antworten auf offene Fragen, Item-Nonresponse oder Straightlining wahrscheinlicher. Zudem verwenden die Versuchspersonen in Studie 3 nicht ihre eigenen Geräte, zu denen aber eine stärkere Bindung besteht und die mit distinkten Nutzungsgewohnheiten verknüpft sein können (z. B. Brasel und Gips 2014; Melumad und Pham 2020; Thorsteinsson und Page 2014).

Im nächsten Schritt muss die vorläufig konstatierte Anfälligkeit des Smartphones für unaufmerksames Antwortverhalten und verringerte Antwortqualität daher ins Feld übertragen werden. Die Erkenntnisse zu Aufmerksamkeitsmechanismen, die durch das Ausfüllgerät beeinflusst werden könnten, werden in der vierten Studie konkret für Careless Responding (CR) als indirektes Aufmerksamkeitsmaß in Onlinebefragungen getestet (siehe Kapitel 3.3.1 und 3.3.2). Eye-Tracking Maße liefern wichtige Hinweise auf CR in Onlinebefragungen (siehe Kapitel 3.3.3.1), da sie Strukturen kognitiver Verarbeitung reflektieren (Brower 2020, S. 87; Shen et al. 2014). Gerade Leseprozesse als Grundelement von Befragungen werden durch unstetes Blickverhalten beeinträchtigt (Jarodzka und Brand-Gruwel 2017; Rayner 1998), sodass CR aufgrund des unsteten Blickverhaltens am Smartphone durchaus wahrscheinlicher ist. Abseits der genannten praktischen Limitationen der Blickaufzeichnung im Feld erweitert die nachfolgende Feldstudie (Studie 4) die Messung der Aufmerksamkeit über die rein physische Betrachtung hinaus, da in Ausnahmefällen Aufmerksamkeitsallokation auch außerhalb des fovealen Sehens und somit abseits des Fixationszentrums stattfinden kann (Duchowski 2017, S. 10; Jacob und Karn 2003, S. 580; Meißner und Oll 2019). Dies konnte in Studie 3 nicht abgebildet werden und limitiert die Aussagekraft der Befunde bezüglich der Folgen für die Antwortqualität.

In einem breit angelegten Vergleich der Aufmerksamkeit und Antwortqualität im Feld beschließt die vierte Studie der Forschungsreihe die Prüfung möglicher Moduseffekte des Smartphones als Ausfüllgerät in Onlinebefragungen. Die finale Studie der Hauptuntersuchungen analysiert konkret, ob sich die potenziell größere Anfälligkeit des Smartphones für Unaufmerksamkeit auch im Feld zeigt und Smartphone-Teilnehmer*innen in Befragungssituationen, die in ihrem Alltag angesiedelt sind, häufiger unaufmerksam antworten. Ist das der Fall, übt das

Smartphone durch die Reduzierung der Antwortqualität womöglich Moduseffekte aus, die den TSE erhöhen.

5.4 Studie 4 – Aufmerksamkeit und Careless Responding am Smartphone

Die Ergebnisse der Studie 3 fundieren die theoretische Herleitung (siehe Kapitel 3.2 und 3.3, insbesondere 3.3.4), die das Smartphone als möglicherweise anfälliger für Aufmerksamkeitsablenkungen kennzeichnet. Das Blickverhalten während der Befragungsteilnahme ist bei Onlinebefragungen am Smartphone tatsächlich unsteter und sprunghafter. Entsprechend ist die Aufmerksamkeit während der Befragung potenziell häufiger gehemmt oder entfernt sich von der Befragung. Die potenziellen Aufmerksamkeitseinschränkungen am Smartphone könnten dazu führen, dass in Onlinebefragungen am Smartphone weniger aufmerksam teilgenommen wird und Moduseffekte entstehen. Als wesentlicher Hintergrundfaktor für die Antwortqualität (siehe Kapitel 3.2.1) sollte Aufmerksamkeit aber gerätübergreifend möglichst stabil sein.

Die Erkenntnisse der Studie 3 werden zum Abschluss der Hauptuntersuchungen ins Feld übertragen. Die vierte Studie prüft, ob Moduseffekte auftreten, die sich darin äußern, dass CR oder anderes weniger sorgfältiges Antwortverhalten bei der Teilnahme an Onlinebefragungen mit dem Smartphone tatsächlich häufiger sind. Dies würde die Antwortqualität der Smartphone-Teilnehmer*innen wesentlich verringern und insbesondere infolge größerer Messfehler durch Befragte den TSE systematisch erhöhen. Auch Item-Nonresponse könnte häufiger auftreten, wenn gar nicht geantwortet wird. Im Between-Subjects-Design werden gezielt CR-Tendenzen sowie die generelle Antwortqualität gerätweise verglichen. In den Analysen wird der Effekt des selbstgewählten Ausfüllgeräts vom Ausfüllkontext möglichst getrennt.

5.4.1 Methode

Die vorliegende Studie zählt nach Kenntnis des Autors zu den ersten Untersuchungen, die den Einfluss des Smartphones als Ausfüllgerät auf CR analysieren. Sie erweitert zudem den Forschungsstand zu Moduseffekten des Smartphones bezüglich der Antwortqualität, indem auch generelle Antwortqualitätseinschränkungen berücksichtigt werden, die aus dem Verhalten der Teilnehmer*innen (z. B. durch verschiedene Grade des Satisficings) resultieren könnten und sich in klassischen Indikatoren der Antwortqualität (siehe Kapitel 3.2.3) widerspiegeln würden. Durch den Einbezug von CR kann der Beitrag des Smartphones zum TSE über ein großes Spektrum möglicher Antwortqualitätseinschränkungen untersucht werden, wobei besonders starke Einschränkungen, die aus CR hervorgehen, vordergründig betrachtet werden. Unter Berücksichtigung wichtiger Kontrollvariablen und Kontexteinflüsse prüft die Studie 4 somit, ob

die potenzielle Anfälligkeit des Smartphones für Aufmerksamkeitsablenkung, die sich in Studie 3 zeigt, während der Teilnahme im Feld persistiert.

Entsprechend wird der Einfluss des Ausfüllgeräts auf unaufmerksames Antworten unter Einbezug relevanter Kontrollvariablen regressionsanalytisch geprüft. Um selbst marginale Unterschiede zwischen Ausfüllgeräten im Zweifel aufzudecken, und in Ermangelung möglicher Vorbildstudien des Smartphoneeinflusses auf CR, wird für das Regressionsgewicht der dichotomen Variable *Gerät* (Smartphone vs. PC/Laptop) eine niedrige kleine Effektstärke angelegt ($f^2 = 0,01$). Die erforderliche Stichprobengröße für Studie 4 wurde a-priori anhand des Programms G*Power (Faul et al. 2007; Faul et al. 2009) ermittelt und beträgt $N = 614$ Proband*innen. Zeigen sich für diese Stichprobengröße keine Unterschiede zwischen den Ausfüllgeräten, würde dies gegen Moduseffekte des Smartphones sprechen und dieses als Ausfüllgerät substantiell stützen. Treten hingegen systematische Unterschiede auf, müssten künftig die Rahmenbedingungen des Smartphones als Ausfüllgerät gewissenhaft geprüft und ggf. angepasst werden, um den Moduseffekten zu begegnen.

5.4.1.1 Versuchsablauf

Die Onlinebefragung wurde auf der Befragungsplattform SoSci Survey (Leiner 2023) angelegt. Die Feldphase der Befragung umfasste das letzte Quartal des Jahres 2020 (Oktober bis Dezember). Die Akquise wurde breit gestreut und kombinierte persönliche Ansprachen im Umfeld des Forschungsteams des Methodenlehrstuhls der wirtschaftswissenschaftlichen Fakultät an der BUW sowie im Netzwerk Studierender und wissenschaftlicher Hilfskräfte, die das Forschungsprojekt unterstützten. Zudem wurde die Rekrutierung in der späteren Feldphase auf die Online-Studienbörse SurveyCircle (SurveyCircle 2023) ausgeweitet. Bei SurveyCircle inserieren Forscher*innen eigene Befragungsprojekte und nehmen wiederum an anderen Befragungen teil, um für sich selbst Proband*innen zu gewinnen. Der utilitaristische Ansatz von Studienbörsen ist anfälliger für Antwortqualitätseinschränkungen (z. B. Cheung et al. 2017; Fleischer et al. 2015; Kennedy et al. 2020), sodass SurveyCircle bewusst integriert wurde, um diverse und praxisnahe Motivationsstrukturen der Teilnehmer*innen adäquat abzubilden. Durch den späteren Einbezug der Studienbörse in die Rekrutierung können Aufmerksamkeitseffekte des Ausfüllgeräts von etwaigen Besonderheiten der Studienbörse für die Antwortqualität zeitlich entkoppelt analysiert werden. Die Onlinebefragung beinhaltete insgesamt 175 Items und einen längeren Text. Für die Beantwortung benötigten die Teilnehmer*innen durchschnittlich 25,34 Minuten, womit die mindestens angestrebte Befragungslänge von 20 Minuten erreicht wurde. Unter allen Befragten wurden drei Entschädigungen (je 25 Euro) als Teilnahmedank verlost.

5.4.1.2 Stichprobe

Die Befragung wurde von insgesamt 1038 Personen aufgerufen, von denen 97 die Befragung nicht begannen oder unmittelbar abbrachen. Für die Analyse von Careless Responding war das Ausfüllen zumindest bis zu einer bestimmten Stelle nötig, ab der das CR-Ausmaß frühestens festgestellt werden konnte. Die erste CR-Abfrage erfolgte auf Seite 8 des Fragebogens und 709 Personen bearbeiteten die Befragung bis dorthin. 24 weitere Personen mussten ausgeschlossen werden, da sie die Befragung lediglich durchklickten, ohne zu antworten, sodass für erste Analysen ein *Gesamtdatensatz* mit 685 Versuchspersonen verbleibt.

Weitere 50 Personen brachen die Befragung, nachdem sie Seite 8 erreicht hatten, auf einer der folgenden Seiten vorzeitig ab oder verfielen in ein Durchklicken ohne Antwortgabe. Die verbleibenden 635 Versuchspersonen bilden den *Hauptdatensatz* für die meisten CR-Analysen der Hypothesenprüfung. Da die Analysen auf CR und somit ungewöhnliche Antwortmuster zielen, wird auf eine datengetriebene Ausreißeranalyse verzichtet. Für den *Hauptdatensatz* lag der Altersschnitt bei 32,95 Jahren ($SD = 13,43$, Altersspanne = 16 bis 78 Jahre). 64,25% ($n = 408$) gaben ihr Geschlecht als weiblich, 33,54% als männlich ($n = 213$) und 0,63% ($n = 4$) als divers an (10 fehlende Werte). Damit nahmen signifikant mehr weibliche als männliche Personen teil ($\chi^2(1) = 61,232, p < 0,001$). Im *Hauptdatensatz* nutzten 323 Proband*innen den PC/Laptop (50,87%), 271 Teilnehmer*innen das Smartphone (42,68%) und 41 das Tablet für die Teilnahme (6,46%). Tablet-Teilnehmer*innen werden für Zusatzauswertungen im Datensatz belassen, gehen aber in die Hypothesenprüfung des Geräteffekts für CR nicht ein. Die Geräthäufigkeit der Untersuchungsgruppen (Smartphone vs. PC/Laptop) unterscheidet sich signifikant ($\chi^2(1) = 4,552, p = 0,033$), wobei mehr Befragte den PC/Laptop nutzten. Das Geschlecht verteilt sich dabei auf beide Gerätgruppen im *Hauptdatensatz* gleich ($\chi^2(1) = 2,072, p = 0,150$).

Analog zur Beschreibung des *Hauptdatensatzes* wird die Beschreibung des *Gesamtdatensatzes* in **Anhang F** vorgenommen. In diesem findet sich kein Unterschied der Häufigkeit der Ausfüllgeräte ($\chi^2(1) = 2,129, p = 0,145$). Die Analysen werden größtenteils auf den aussagekräftigeren *Hauptdatensatz* angewendet, da in diesem für alle Befragten häufiger Werte für alle Kontrollvariablen der finalen Regressionsmodelle vorliegen. Nur situativ erfolgt die Hypothesenprüfung anhand des Gesamtdatensatzes, wenn nicht alle Kontrollvariablen benötigt werden.

5.4.1.3 Fragebogenbeschreibung und Operationalisierung der Hypothesen

Der Erhebung wurde ein umfangreicher Pretest mit Expert*innen des methodischen Fachbereichs der Fakultät der Wirtschaftswissenschaft an der BUW vorgeschaltet. Dabei wurden mög-

liche Verständnisschwierigkeiten ermittelt und insbesondere die Breite und Güte der CR-Aufdeckung geprüft. Die nachfolgenden Darstellungen beziehen sich auf die endgültige Fragebogenversion. Variablenbezeichnungen werden kursiv dargestellt. Zu Beginn der Befragung wurden die Teilnehmer*innen über das vermeintliche Untersuchungsziel informiert. Als Cover-Thema der Befragung fungierte Umweltbewusstsein und -verhalten im Zuge der Covid-19 Pandemie. Damit wurde ein starker Aktualitätsbezug (Herbst/Winter 2020) hergestellt, da die steigenden weltweiten Klimaproteste der Fridays-for-Future Bewegung (vgl. Reinhardt 2019; Sommer et al. 2019; Wallis und Loy 2021) sowie die gesellschaftliche Verankerung von Umweltthemen (z. B. Gatersleben, Murtagh und Abrahamse 2014; Lange und Dewitte 2019) in Verbindung mit den zu diesem Zeitpunkt omnipräsenten Herausforderungen der Covid-19 Pandemie (vgl. Ciotti et al. 2020; Velavan und Meyer 2020; World Health Organization 2020) einbezogen wurden. Das Ausfüllgerät als zentrales Element der Untersuchung wurde frei gewählt, durch SoSci Survey automatisch aufgezeichnet sowie durch die Selbstangabe der Teilnehmer*innen abgesichert. In der Einführung bestätigten die Versuchspersonen, dass ihre Daten gemäß der europäischen DSGVO zu rein wissenschaftlichen Zwecken verarbeitet und genutzt werden können. Die Befragten wurden gebeten (S. 2), falls sie am Smartphone teilnahmen, den Landscape-Modus zu nutzen, für den die gerätübergreifend gleiche Darstellung der Befragungsinhalte garantiert wurde. Im Anschluss wurden die soziodemografischen Variablen *Geschlecht*, *Alter*, *Bildungsabschluss* und *Erwerbsstatus* abgefragt (S. 3).

Im Sinne des Untersuchungsziels war es von vorrangigem Interesse, eine längere, aber kohärente Befragung zu gewährleisten. Die verwendeten Inventare wurden daher passend zur Grundthematik ausgewählt und sollten zugleich eine möglichst eindeutige CR-Aufdeckung erlauben. Auf den folgenden beiden Seiten (S. 4+5) diente die angepasste Skala „*Naturverträgliches Handeln*“ (Scherhorn et al. 1999) als Aufwärmphase. Die Proband*innen gaben auf einer teilverbalisierten, numerisch verankerten, fünfstufigen Skala (1 = „bisher nie“ bis 5 = „immer“) an, wie häufig sie bestimmte umweltrelevante Verhaltensweisen (z. B. „*Recyclingpapier nutzen*“) ausführen. Zur Ausweitung der Aufwärmphase wurden die zwölf Items der Originalskala um sieben weitere, selbst erstellte ergänzt (z. B. „*Bevorzugt regionale Produkte konsumieren*“). Die interne Konsistenz über alle 19 Items hinweg war angemessen ($\alpha = 0,775$).

Die folgende Seite (S. 6) präsentierte Textauszüge der Studie „Umweltbewusstsein in Deutschland 2018“ (Rubik et al. 2019) des Bundesministeriums für Umwelt, Naturschutz und nukleare Sicherheit. Die ausgewählten Textausschnitte definieren den Begriff des „Sozialen Milieus“ und grenzen fünf Milieus (Traditionelle Milieus, Etablierte Milieus, Bürgerlicher Mainstream, Prekäre Milieus, Kritisch-kreative Milieus) voneinander ab (Rubik et al. 2019, S.

73-82). Die Proband*innen wurden gebeten, den Text sorgfältig zu lesen, und im Anschluss wurden sie zu den Textinhalten befragt. Die inhaltlichen Rückfragen zum Text wurden auf der siebten Seite präsentiert ohne die Möglichkeit, zum Text zurückzukehren. Anhand der Anzahl korrekt gelöster *Rückfragen* wird die Hypothese **H4.5.2** (*Am Smartphone antworten Proband*innen auf inhaltliche Rückfragen seltener korrekt als am PC/Laptop*) untersucht. Die Fragen variierten dabei im Schwierigkeitsgrad und dem erforderlichen Textverständnis. Einerseits wurden rudimentäre und schnell ersichtliche Informationen (*Jahr der Studie*; Erkennen eines Milieus, das im Text nicht vorgestellt wird = *falsches Milieu*) und zudem komplexere Informationen abgefragt, die eine tiefere Auseinandersetzung mit dem Text erfordern (*inhaltliche Unterscheidung* der vorgestellten Milieus; *Anzahl untersuchter Milieus* in der Studie). Eine *offene Frage* vervollständigte die Seite, bei der der eingeführte Begriff des „Milieus“ als Freitext eigenständig definiert werden sollte, worüber die Hypothese **H4.5.1** (*Am Smartphone werden offene Fragen weniger ausführlich beantwortet als am PC/Laptop*) geprüft wird.

Auf Seite 8 schloss ein komplexer, auf die Thematik angepasster IMC (siehe *Abbildung 12*) den Befragungsabschnitt ab. Ein längerer Einleitungstext instruierte die Befragten, die folgende Frage (sich selbst einem der Milieus des Textes zuzuordnen) nicht zu beantworten. Um dabei die bewusste Nicht-Antwort von einer unaufmerksamen Nicht-Antwort trennen zu können, sollte zudem die zweite Frage der Seite („*Wie gut trifft die Beschreibung des Milieus, dem Sie sich gerade zugeordnet haben, tatsächlich auf Sie zu*“) mit „trifft voll und ganz zu“ beantwortet werden. Das Bestehen des IMC (*Check bestanden*) operationalisiert die Hypothese **H4.1.2** (*Am Smartphone bestehen Proband*innen einen Instruktions-Manipulationscheck (IMC) seltener als am PC/Laptop*).

Auf den nächsten Fragebogenseiten (S. 9-11) beantworteten die Befragten die weiterentwickelte „Skala zur Messung von zentralen Kenngrößen des Umweltbewusstseins“ (Geiger und Holzauer 2020). Diese unterscheidet die drei Messgrößen *Umweltaffekt*, *Umweltkognition* und *Umweltverhalten*. Zur reflektiven Erfassung der *Umweltkognition* (acht Items, $\alpha = 0,799$) und des *Umweltaffekts* (sieben Items $\alpha = 0,661$) gaben die Proband*innen auf einer vollverbalisierten, fünfstufigen Zustimmungsskala (1 = „stimme überhaupt nicht zu“ bis 5 = „stimme voll und ganz zu“) ihre Einstellung zu bestimmten Aussagen zur Umweltsituation an (Beispiel *Umweltkognition*: „*Wir sollten nicht mehr Ressourcen verbrauchen als nachwachsen können*“; Beispiel *Umweltaffekt*: „*Der Klimawandel bedroht auch die Lebensgrundlagen hier in Deutschland*“). Zusätzlich wurde auf beiden Seiten jeweils ein auf den Befragungsinhalt leicht angepasstes Bogus-Item integriert, das auf derselben vollverbalisierten, fünfstufigen Zustimmungsskala beantwortet wurde. Das erste Bogus-Item (*Bogus1*, S. 9) lautete „*Ich saß noch nie in einem*

Die Zuordnung zu einem bestimmten sozialen Milieu ist nicht immer ganz eindeutig. Eine sehr wichtige Rolle spielt aus Sicht der Forschung der erste Eindruck, der beim Lesen entsteht. Hier ordnet man sich häufig unmittelbar einer bestimmten Gruppe zu. Das führt schnell dazu, dass weniger differenziert über die tatsächliche Passung nachgedacht wird. Dabei überdecken insbesondere soziodemografische Elemente die Frage, ob auch Bedürfnisse und Einstellungen zur eigenen Person passen. Die Entscheidung wird schnell und impulsiv getroffen. Potentiell wichtige Informationen werden gar nicht mehr wirklich aufgenommen. Vor diesem Hintergrund möchten wir an dieser Stelle prüfen, ob Sie sich die Zeit nehmen, die Instruktionen zu lesen, oder möglicherweise Ihrem ersten Impuls folgen. Bitte ignorieren Sie daher die folgende Frage und nehmen Sie keine Zuordnung zu einem Milieu für sich vor, indem Sie keine der dargestellten Optionen auswählen. Bitte aktivieren Sie abschließend lediglich die Option "trifft voll und ganz zu" bei der letzten Frage auf der Seite ("Wie gut trifft die Beschreibung des Milieus, dem Sie sich zugeordnet haben, tatsächlich auf Sie zu?") und klicken Sie dann auf "Weiter".

Welchem der untenstehenden Milieus würden Sie sich am ehesten zuordnen, wenn Sie eine Entscheidung treffen müssten? Orientieren Sie sich dabei weniger an soziodemografischen Merkmalen wie Alter und Geschlecht. Antworten Sie so intuitiv und gleichzeitig gewissenhaft wie möglich.

Traditionelles Milieu

Etabliertes Milieu

Bürgerlicher Mainstream

Prekäres Milieu

Kritisch-kreatives Milieu

Wie gut trifft die Beschreibung des Milieus, dem Sie sich zugeordnet haben, tatsächlich auf Sie zu?

	trifft überhaupt nicht zu	trifft eher nicht zu	trifft eher zu	trifft voll und ganz zu
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 12. Instruktions-Manipulationscheck (IMC) der Studie 4

Rot umrandet ist die konkrete Ausfüllanweisung

Quelle: Eigener Fragebogen, SoSci Survey (Leiner 2023)

fahrenden Auto“ (Dunn et al. 2018, S. 120) und das zweite (*Bogus2*, S. 10) „Ich bin schon in jedes Land der Welt geflogen“ (Meade und Craig 2012, S. 441). Zudem wurde mit sieben formativen Items (S. 11) das *Umweltverhalten* in verschiedenen Bereichen auf einer fünfstufigen, vollverbalisierten Häufigkeitsskala (1 = „nie“ bis 5 = „immer“) abgefragt (z. B. „Ich kaufe Lebensmittel aus kontrolliert-biologischem Anbau“).

Die Seiten 12 und 13 präsentierten den „Environmental Portrait Value Questionnaire“ (Bouman, Steg und Kiers 2018), der eine umweltspezifische Adaptation des „Portrait Value Questionnaire“ (Schwartz et al. 2001) verkörpert. Die 17 Items der Skala beschreiben, angepasst an das eigene Geschlecht, Personen und deren Einstellungen (z. B. „Es ist ihm/ihr wichtig, dass jeder Mensch die gleichen Chancen haben sollte“). Die Teilnehmer*innen wurden angehalten, einen Ähnlichkeitsvergleich anhand der siebenstufigen, numerisch verankerten, endpunktverbalisierten Skala des Originalinventars von 1 = „sehr unähnlich“ bis 7 = „sehr ähnlich“ zwischen sich selbst und der jeweils beschriebenen Person vorzunehmen. Der Environmental Portrait Value Questionnaire gliedert sich in vier Dimensionen: *Biosphäre* (vier Items, $\alpha = 0,833$), *Altruismus* (fünf Items, $\alpha = 0,740$), *Hedonismus* (drei Items, $\alpha = 0,842$) und *Egoismus* (fünf Items, $\alpha = 0,669$). Um Konsistenzen auch skalenübergreifend beispielsweise im Sinne der

H4.4.1 (*Am Smartphone antworten die Teilnehmer*innen weniger konsistent auf psychometrische Synonyme als am PC/Laptop*) und **H4.4.2** (*... Antonyme...*) ableiten zu können, wurde eine weitere Werteskala (S. 14) angeschlossen. Das „Short Schwartz’s Value Survey“ (Short-SVS) (Boer 2014) listet insgesamt zehn verschiedene Werte inklusive Definitionen auf, wobei auf einer vollverbalisierten, sechsstufigen Wichtigkeitsskala (1 = „überhaupt nicht wichtig“ bis 6 = „sehr wichtig“) beurteilt wird, welche Relevanz die Werte im eigenen Leben haben.

Neben den umweltbezogenen Größen wurde als letztes etabliertes Inventar ohne unmittelbaren CR-Bezug – analog zu Studie 3 (siehe Kapitel 5.3.1.3) – das BFI-2 eingesetzt (Danner et al. 2016; Danner et al. 2019). Das BFI-2 erfasst mit jeweils zwölf Items (insgesamt 60 Items) die fünf Persönlichkeitsdimensionen *Verträglichkeit* ($\alpha = 0,742$), *Negative Emotionalität* ($\alpha = 0,834$), *Offenheit für Erfahrungen* ($\alpha = 0,787$), *Gewissenhaftigkeit* ($\alpha = 0,814$) und *Extraversion* ($\alpha = 0,810$). Alle Items werden auf einer vollverbalisierten, fünfstufigen Zustimmungsskala (1 = „stimme überhaupt nicht zu“ bis 5 = „stimme voll und ganz zu“) beantwortet. Auf den Fragebogenseiten 15 bis 20 wurden jeweils zehn BFI-Items dargeboten. Zudem wurden im Verlauf des BFI-2 die drei Bogus-Items „*Ich habe noch nie gelacht*“ (*Bogus3*, S. 16) und „*Ich lüge immer*“ (*Bogus4*, S. 19) (Dunn et al. 2018, S. 120) sowie „*Ich kann zwei Kilometer in zwei Minuten laufen*“ (*Bogus5*, S. 20) (Huang et al. 2015a, S. 303) integriert. Den Abschluss des Blocks bildete ein letztes Bogus-Item „*Ich nehme gerade an einer Befragung teil*“ (*Bogus6*, S. 20) nach Meade und Craig (2012, S. 441). Alle Bogus-Items (*Bogus1* - *Bogus6*) wurden möglichst unauffällig und wenig irritierend integriert, auf derselben Zustimmungsskala beantwortet und bilden die Grundlage der Hypothesenprüfung der **H4.1.1** (*Am Smartphone werden Bogus-Items seltener korrekt beantwortet als am PC/Laptop*). Zudem wurde als erweitertes Prüfkriterium der Antwortkonsistenz das Item „*Ich bleibe auch bei Rückschlägen zuversichtlich*“ zu Beginn des BFI (S. 15) abgefragt und gegen Ende der Skala (S. 19) wiederholt. Für die Untersuchung der Hypothese **H4.4.3** (*Am Smartphone erkennen die Proband*innen seltener als am PC/Laptop, dass Skalenumkehr vorliegt, und antworten dadurch inkonsistenter*) wurde auf der vierten Seite der BFI-Items (S. 18) die Skalierung einmalig unangekündigt gedreht und ab der Folgeseite (S. 19) wieder im Ursprungsformat verwendet. Die Zustimmungsskala verlief auf Seite 18 nicht von links nach rechts aufsteigend, sondern absteigend.

Den Auftakt des letzten Befragungsteils zum Befragungsverhalten und zur Befragungssituation bildete auf Seite 21 die Skala „Selbstberichtete Antwortstile“ (Maniaci und Rogge 2014, S. 64). Diese erfasst generelle Tendenzen von Befragten für unaufmerksames Antwortverhalten mit zehn Items (z. B. „*Wie häufig... überspringen Sie Teile der Instruktion?*“) auf einer siebenstufigen, teilverbalisierten Antwortskala von 1 = „nie“ über 4 = „etwa in der Hälfte der Fälle“

bis 7 = „immer“. Aus den zehn Items werden die vier Dimensionen *Selbstberichtetes CR* (vier Items, $\alpha = 0,833$), *Antworthast* (vier Items, $\alpha = 0,833$), *Antwortmuster* (zwei Items, $\rho = 0,607$) und *Instruktionen überspringen* (zwei Items, $\rho = 0,680$) gebildet. Für Zwei-Item-Skalen ist als Konsistenzmaß statt CA der Spearman-Brown Koeffizient (ρ) besser geeignet (Eisinga et al. 2013). Abschließend (S. 22) wurde der Befragungskontext (*Ausfüllort*, *Anwesenheit Dritter*, *Körperhaltung* beim Ausfüllen, *Ablenkung*, *Befragungsgerät*, *Screenshot-Strategien* während der Textaufgaben, *Multi-Tasking* während der Befragung) abgefragt. Die zugehörigen Items wurden als inhaltsvalide Einzelitems entsprechend des C-OAR-SE Ansatzes erstellt (vgl. Ros-siter 2002). Die *Anwesenheit Dritter* wurde dichotomisiert, wobei unterschieden wurde, ob die Proband*innen allein (= 1) oder nicht allein (= 0) waren. Das *Multi-Tasking* wurde auf einer fünfstufigen, vollverbalisierten Intensitätsskala erfasst (1 = „überhaupt nicht“, 5 = „(fast) die gesamte Zeit über“). Die selbst wahrgenommene *Ablenkung* während der Teilnahme wurde über das Item „Bitte geben Sie das Ausmaß an, in dem Sie durch externe Reize (akustisch, visuell etc.) während der Befragung abgelenkt wurden“ erhoben, das auf einer siebenstufigen, numerisch verankerten, endpunktverbalisierten Ablenkungsskala von 1 = „sehr geringe Ablenkung“ bis 7 = „sehr starke Ablenkung“ beantwortet wurde. Zudem beurteilten die Proband*innen, wie schon in den Studien 1 und 3 des Forschungsprogramms, die Befragung nach *Interesse* sowie wahrgenommener *Befragungsdauer* und vergaben eine *Gesamtbewertung* für die Befragung. Dies geschah jeweils auf fünfstufigen (*Interesse*, 1 = „gar nicht interessant“ bis 5 = „sehr interessant“; *Gesamtbewertung*, 1 = „sehr schlecht“ bis 5 = „sehr gut“) bzw. siebenstufigen (*Befragungsdauer*, 1 = „sehr kurz“ bis 7 = „sehr lang“), endpunktverbalisierten, numerisch verankerten Antwortskalen (de Bruijne und Wijnant 2013, S. 490-492).

Zuletzt wurde die *situative Aufmerksamkeit* als Grundlage der **H4.2.1** (*Bei Proband*innen, die mit dem Smartphone teilnehmen, ist die selbstberichtete Aufmerksamkeit geringer als bei Proband*innen, die mit dem PC/Laptop teilnehmen*) über Selbstangaben zur Gewissenhaftigkeit des Antwortens erhoben. Die Befragten beantworteten die Items „Ich habe nicht so sehr darauf geachtet, was die einzelnen Fragen eigentlich meinten“, „Ich habe den Fragebogen ausgefüllt, ohne mir wirklich Gedanken (über mich) zu machen“ „Ich habe die Fragen nicht gewissenhaft beantwortet“ (Huang et al. 2012, S. 107), ergänzt um das umgekehrte, selbst erstellte Item „Ich habe den Fragebogen nach bestem Wissen und Gewissen sorgfältig ausgefüllt“ auf einer fünfstufigen, vollverbalisierten Skala (1 = „trifft überhaupt nicht zu“ bis 5 = „trifft voll und ganz zu“). Die *situative Aufmerksamkeit* ergab sich als Skalenmittelwert der vier Items ($\alpha = 0,713$). Zuletzt wurde das *UseMe-Item* „Ihrer ehrlichen Meinung nach: Sollten wir Ihre Daten aus dieser Studie in unseren Analysen nutzen?“ (Brühlmann et al. 2020, S. 4; Meade und Craig

2012, S. 442) als letzter direkter CR-Indikator der **H4.2.2** (*Proband*innen, die mit dem Smartphone teilnehmen, geben häufiger an, dass ihre Daten für die Auswertung nicht genutzt werden sollten*) abgefragt. Zum Abschluss der Befragung konnten die Versuchspersonen ihre Vermutung über das Untersuchungsziel sowie Kritik oder Anmerkungen jeweils als Freitextantwort äußern. Die Verabschiedung beinhaltete ein Debriefing, das über den Hintergrund der Befragung aufklärte. Zusätzlich konnte an der Verlosung der drei Geldgewinne (je 25 Euro) teilgenommen werden. Hierfür wurden die Proband*innen auf eine Befragungsseite der Software LimeSurvey (Limesurvey GmbH 2023) umgeleitet, um die Kontaktdaten für die Gewinnbenachrichtigung von ihren anderen Angaben im Hauptfragebogen zu lösen. Der gesamte Fragebogen der Studie 4 ist in **Anhang G** einsehbar.

Die post-hoc Detektionsmethoden werden situationsadäquat auf die Befragungsinhalte angewendet. Für die Antwortkonsistenzprüfung der Hypothesen **H4.4.1** und **H4.4.2** werden psychometrische *Synonyme* und *Antonyme* für alle Items identifiziert, die auf mindestens fünfstufigen Skalen beantwortet werden und somit als quasimetrisch angenommen werden können (Harpe 2015; Urban und Mayerl 2018, S. 301-302). Zudem werden Long Strings skalenübergreifend ermittelt. Dabei dient die durchschnittliche *Stringlänge* der Prüfung der **H4.3.2** (*Am Smartphone sind Long Strings durchschnittlich länger als am PC/Laptop*) und die **H4.3.1** (*Am Smartphone treten in den Antwortmustern der Proband*innen häufiger sehr lange Long Strings auf als am PC/Laptop*) wird über die maximale Länge des Long Strings jeder Person (*LongString*) und die CR-Klassifikation anhand dieses Long Strings (*LongStringCR*) geprüft. Die Aufdeckung monotonen Antwortens wird um die Ermittlung der *IRV* für die **H4.3.3** (*Am Smartphone ist die individuelle Response Validität (IRV) geringer als am PC/Laptop*) erweitert, die auf das BFI-2 angewendet wird. Für die datengetriebene CR-Aufdeckung ist das BFI-2 bewährt (Brühlmann et al. 2020) und post-hoc Detektionsmethoden eignen sich besonders in späteren Befragungsteilen (z. B. Dunn et al. 2018; Herzog und Bachman 1981). In allen Skalen wurde die Reihenfolge der Items, wenn möglich, dem Originalinventar entnommen und nicht rotiert, um Long Strings und Antwortmuster akkurat identifizieren zu können.

Die Operationalisierung der Hypothesen wird ergänzt durch klassische Antwortqualitätsindikatoren. Die Berechnung der *Mahalanobis-Distanz* der **H4.6a** (*Proband*innen, die mit dem Smartphone teilnehmen, weisen höhere Mahalanobis-Distanzwerte auf als Proband*innen, die mit dem PC/Laptop teilnehmen*) bezieht sich ebenfalls auf alle mindestens quasimetrischen Befragungsisems. Die Antwortzeit (*RT*) wird herangezogen zur Untersuchung der Hypothesen **H4.5.3** (*Proband*innen, die mit dem Smartphone teilnehmen, werden häufiger aufgrund sehr schneller Antwortzeiten als Careless Responder klassifiziert als Proband*innen, die mit dem*

PC/Laptop teilnehmen) und **H4.5.4** (*Am Smartphone benötigen Teilnehmer*innen generell mehr Zeit als am PC/Laptop, um die Befragung durchzuführen*). Die Analysen werden abgeschlossen über die Betrachtung der Anzahl der *Missings* als Ausdruck der Item-Nonresponse (**H4.6b**: *Für Proband*innen, die mit dem Smartphone teilnehmen, treten mehr Missings in den Antworten auf als für Proband*innen, die mit dem PC/Laptop teilnehmen*) und der Abbruchraten (*Abbruch*) für die **H4.6c** (*Proband*innen, die mit dem Smartphone teilnehmen, brechen die Befragungsteilnahme häufiger ab als Proband*innen, die mit dem PC/Laptop teilnehmen*).

5.4.2 Ergebnisse

Die Datenauswertung erfolgte hauptsächlich mit dem Auswertungsprogramm R (R Core Team 2023) und wurde situativ durch die Statistikprogramme SPSS (IBM Corp. 2021) sowie JASP (JASP Team 2023) ergänzt. Für alle Hypothesen fungiert die dichotome Variable *Gerät* (0 = PC/Laptop, 1 = Smartphone) als UV und wird in Form allgemeiner linearer Modelle (ALM) entweder regulär als multiple lineare Regression für mindestens quasimetrische AVn oder als Logit-Modell der multiplen logistischen Regression für dichotome AVn untersucht (Backhaus et al. 2021, S. 289-382; Bortz und Schuster 2016, S. 311-348). Generell werden, aufgrund des dichotomen Charakters der UV *Gerät* und einiger Kontrollvariablen, immer die nicht-standardisierten Regressionsgewichte (*b*) mitsamt der zugehörigen *z*-/*t*-Werte und dem resultierenden *p*-Wert berichtet. Für AVn, die auf Zähldaten basieren, wird die Poisson-Regression verwendet (vgl. Beaujean und Grant 2016; Winkelmann 2008).

5.4.2.1 Voraussetzungsprüfung und Voranalysen

Die Analysen prüfen, ob Careless Responding und daraus resultierende Einschränkungen der Antwortqualität am Smartphone häufiger vorkommen. Grundlage der Hypothesenprüfung ist der *Hauptdatensatz* und die Befunde werden stets, wenn möglich, am *Gesamtdatensatz* abgesichert und Abweichungen berichtet. Nur vereinzelt erfolgt die Hypothesenprüfung direkt anhand des *Gesamtdatensatzes*. Für reguläre ALMs sollten die Voraussetzungen der Linearität, Normalverteilung der AVn und Homoskedastizität als Varianzhomogenität der Residuen erfüllt sein, wobei zudem geringe Multikollinearität wünschenswert ist (Bortz und Schuster 2016, S. 311-348). Die Normalverteilung kann aufgrund der Stichprobengröße über den zentralen Grenzwertsatz angenommen werden (Bortz und Schuster 2016, S. 586). Alle AVn, die auf mindestens fünfstufigen Ratingskalen basieren, können als quasimetrisch behandelt werden (Harpe 2015; Urban und Mayerl 2018, S. 301-302). Skalenmittelwerte und Zähldaten können ebenfalls im Rahmen der parametrischen Regressionsanalyse analysiert werden. Die Poisson-Regression

für Zähldaten erfordert zudem Äquidispersion, sodass bei Über- oder Unterdispersion auf die Quasipoisson-Regression ausgewichen wird (Winkelmann 2008, S. 45-50). Für das Logit-Modell sollten die Beobachtungen unabhängig sein, möglichst große Beobachtungsgruppen (größer 25) vorliegen und möglichst geringe Multikollinearität auftreten (Backhaus et al. 2021, S. 379). In den Analysen werden die Varianzhomogenität der Residuen (für reguläre und Poisson-ALMs) und die Multikollinearität (für alle Analysen) geprüft. Bei starken Voraussetzungsverletzungen wird, wenn möglich, als nicht-parametrische Alternative der Kruskal-Wallis-Test für die relevante Gruppenvariable *Gerät* herangezogen (Leonhart 2022, S. 274-276).

In Vorbereitung der Hypothesenprüfung werden zunächst die Teilnahmegruppen für den *Hauptdatensatz* verglichen. Dies geschieht über *t*-Tests für unabhängige Stichproben, für die die Voraussetzungen gelten, dass die Beobachtungen unabhängig sind, die AVn in den Gruppen normalverteilt sind, Varianzhomogenität vorliegt und die AVn mindestens intervallskaliert sind (Bühner und Ziegler 2017, S. 301). Des Weiteren kommen für Häufigkeiten χ^2 -Tests zur Anwendung, die unabhängige Beobachtungen benötigen, die eindeutig einer Bedingung oder Bedingungskombination zuzuordnen sind (Bühner und Ziegler 2017, S. 353). In allen Analysen werden die Voraussetzungen geprüft und sind erfüllt, wenn nicht anders angegeben, wobei erneut der zentrale Grenzwertsatz für die Normalverteilungsannahme greift. Liegt Varianzheterogenität vor, wird der Welch-*t*-Wert berichtet. Aufgrund des Feldcharakters der Studie wird für die UV *Gerät* eine Überprüfung der Variablen *Bildschirmbreite* ($F(1, 592) = 296,630, p < 0,001$), *Bildschirmhöhe* ($F(1, 592) = 26,053, p < 0,001$) und *Fragebogenbreite* ($F(1, 592) = 63,172, p < 0,001$) vorgeschaltet, für die jeweils Varianzheterogenität vorliegt. Die Ergebnisse des Welch-*t*-Tests unterstützen die Aufteilung in die beiden Gruppen der UV *Gerät*, da die Geräte der Smartphone-Teilnehmer*innen signifikant kleiner sind: geringere *Bildschirmbreite* ($t(342,79) = 70,361, p < 0,001$), geringere *Bildschirmhöhe* ($t(568,88) = 17,004, p < 0,001$) und geringere Breite der Fragebogendarstellung (*Fragebogenbreite*) ($t(304,90) = 110,430, p < 0,001$).

Als soziodemografische Kontrollvariablen gehen *Alter* und *Geschlecht* in die Hypothesenprüfung ein (siehe Kapitel 3.1.2 und 4). In der vorliegenden Stichprobe sind Proband*innen, die mit dem Smartphone an der Befragung teilnehmen, jünger (keine Varianzhomogenität: $F(1, 592) = 7,308, p = 0,007$; $t(585,80) = 1,984, p = 0,040$). Insgesamt nehmen auch mehr weibliche Personen an der Befragung teil, wobei sich das Geschlecht auf beide Gerätgruppen gleich verteilt (siehe Kapitel 5.4.1.2).

In der Hypothesenprüfung soll der Geräteffekt vom Teilnahmekontext entkoppelt analysiert werden. Smartphone-Teilnehmer*innen berichten tendenziell mehr *Ablenkung* während der

Teilnahme als Befragte, die den PC/Laptop nutzen ($t(588) = -1,875, p = 0,061$). *Ablenkung* ist auch theoretisch eine relevante Kontrollvariable und wird daher in das Regressionsmodell einbezogen. Da während der Teilnahme mit dem Smartphone häufiger Bystander (*Anwesenheit Dritter*) anwesend waren ($\chi^2(1) = 14,161, p < 0,001$), wird auch diese Variable ins Modell aufgenommen. Der Teilnahmeort wird hingegen nicht berücksichtigt, da lediglich 70 Personen nicht im eigenen Zuhause teilnahmen²⁰. Da für *Multi-Tasking* keine Gruppenunterschiede auftreten und *Multi-Tasking* über *Ablenkung* hinaus wenig eigene Erklärungskraft verspricht, wird *Multi-Tasking* ebenfalls nicht in die Analysen einbezogen²¹. Das regressionsanalytische Grundmodell prüft somit den Einfluss der UV *Gerät* unter Einbezug der Kontrollvariablen *Alter*, *Geschlecht*, *Ablenkung* und *Anwesenheit Dritter*. In der Hypothesenprüfung zeigt sich für alle letztlichen Modellvariablen²² keine Multikollinearität. Die VIF-Werte sind stets kleiner als 1,16, sodass auch konservative Grenzwerte (VIF = 5) (Backhaus et al. 2021, S. 100) deutlich nicht erreicht werden. Zudem liegt stets Homoskedastizität vor.

5.4.2.2 Hypothesenprüfung

Den Auftakt der Hypothesenprüfung bildet die Analyse der Abbruchraten (**H4.6c**). Für diese können die Kontrollvariablen nicht einbezogen werden, da sie erst im Verlauf der Befragung erhoben wurden. Befragungsabbrüche als klassisches Antwortqualitätsmaß indizieren als Aufmerksamkeitsindikator nicht ausschließlich CR, da häufig bis zum Abbruch aufmerksam und gewissenhaft teilgenommen wird (Peytchev 2009). Entsprechend wird die Prüfung der **H4.6c** der regressionsanalytischen Untersuchung vorgeschaltet. Proband*innen, die mit dem Smartphone (229 Abbrüche – 45,80%) teilnehmen, brechen die Befragung häufiger ab als Personen, die den PC/Laptop (155 Abbrüche – 32,02%) nutzen ($\chi^2(1) = 19,396, p < 0,001$). Der Befund bestätigt die **H4.6c**, da Befragte die Befragung (unabhängig von den initialen Aufrufzahlen) an beiden Geräten auch gleich häufig wirklich aufnehmen ($\chi^2(1) = 0,106, p = 0,745$) und demnach lediglich tatsächliche Abbrüche im Befragungsverlauf am Smartphone häufiger auftreten.

Die Analyse der **H4.1.2** zum IMC kann anhand des *Gesamtdatensatzes* erfolgen, da der IMC relativ früh in der Befragung präsentiert wurde. Für die Analyse wird für alle Proband*innen die dichotome Variable *Check bestanden* (0 = nicht bestanden, 1 = bestanden) gebildet. Der Check gilt als bestanden, wenn der Ausfüllanweisung gefolgt wird. Dafür dürfen sich die Befragten keinem Milieu zuordnen und müssen zusätzlich, wie instruiert, in der Folgefrage die

²⁰ Herbst und Winter des Jahres 2020 als Hochphase der Covid-19 Pandemie limitieren die Diversität der Kontextfaktoren aufgrund der flächendeckenden Kontaktminimierung.

²¹ In keiner der Analysen verändern sich die Befundmuster, wenn *Multi-Tasking* zusätzlich berücksichtigt wird.

²² Dies bezieht bereits die weiteren Modellvariablen *Check bestanden* und *Ziel erkannt* ein (siehe Kapitel 5.4.2.2).

Antwortoption „stimme voll und ganz zu“ auswählen. Die Analyse des multiplen logistischen Regressionsmodells zeigt, dass Proband*innen, die mit dem Smartphone teilnehmen, abweichend von der Hypothese **H4.1.2** den Manipulationsscheck signifikant häufiger bestehen als PC/Laptop-Teilnehmer*innen ($b = 0,379$, $z = 2,173$, $p = 0,030$). Zugleich führt ein hohes Maß an *Ablenkung* eher dazu, dass der Check signifikant weniger oft bestanden wird ($b = -0,144$, $z = -2,147$, $p = 0,032$) und auch Ältere (*Alter*) bestehen den Check seltener ($b = -0,025$, $z = -3,425$, $p < 0,001$). Instruierte Ausfüllanweisungen, wie der IMC, können Proband*innen grundsätzlich dafür sensibilisieren, dass Aufmerksamkeit erfasst wird, und somit CR im weiteren Befragungsverlauf reduzieren (z. B. Anduiza und Galais 2017; Clifford und Jerit 2015; Hauser und Schwarz 2015; Oppenheimer et al. 2009). Aufgrund der möglichen Sensibilisierung, die in der Folge gerätübergreifend zu aufmerksamerem Antworten führen könnte, und aufgrund des kontraintuitiven Gerätunterschieds wird das Regressionsmodell für alle CR-Analysen, auf die der IMC wirken könnte, um die Variable *Check bestanden* ergänzt, damit weiterhin gezielt die reinen Geräteeffekte betrachtet werden.

Sollten Befragte durch den IMC oder in der weiteren Folge etwa durch Bogus-Items das Untersuchungsziel erkennen, könnte dies gesondert das Antwortverhalten beeinflussen und CR reduzieren. Zum Abschluss der Befragung konnten die Versuchspersonen eine Vermutung über das Untersuchungsziel äußern. Die Variable *Ziel erkannt* (0 = nicht erkannt, 1 = Ziel erkannt) wird mit 1 kodiert, wenn die Befragten im weitesten Sinne angeben, dass Aufmerksamkeit im Forschungsfokus liegt. Generell wird das Untersuchungsziel von den Proband*innen deutlich öfter nicht erkannt ($n = 493$) als erkannt ($n = 138$) ($\chi^2(1) = 199,720$, $p < 0,001$) – am Smartphone identifizieren die Proband*innen das Forschungsziel aber signifikant häufiger korrekt als am PC/Laptop ($\chi^2(1) = 6,823$, $p = 0,009$). Daher wird die Variable *Ziel erkannt* ebenfalls in die Hypothesenprüfung integriert und erweitert gemeinsam mit der Variable *Check bestanden* das bisherige Grundmodell (*Gerät, Alter, Geschlecht, Ablenkung, Anwesenheit Dritte*). Im Folgenden werden, wann immer möglich, die Effekte des Modells mit allen Variablen berichtet.

Neben dem IMC wird die Hypothese **H4.1** zu den Infrequency-Indikatoren von CR über Bogus-Items (**H4.1.1**) geprüft. Alle Bogus-Items werden auf einer fünfstufigen Zustimmungsskala beantwortet und bei aufmerksamer Teilnahme ist jeweils nur die klare Ablehnung respektive Zustimmung plausibel. Da einige Bogus-Items erst im späteren Verlauf der Befragung auftreten und alle Kontrollvariablen möglichst vollständig einbezogen werden sollen, wird, wie hergeleitet, der *Hauptdatensatz* für die Analysen genutzt. Das Bestehen der sechs Bogus-Items (*Bogus1 - Bogus6*) wird dichotomisiert (0 = nicht bestanden, 1 = bestanden). Es wird eine strengere Bestehensbedingung angelegt, bei der nur die extreme, strikte Ablehnung (Antwortoption

1) oder Zustimmung (5) als Bestehen interpretiert wird, und zusätzlich eine mildere Variante des Bestehens bei grundsätzlicher Ablehnung (Antwortoptionen 1 oder 2) oder Zustimmung (4 oder 5) angeschlossen. Berichtet werden die Ergebnisse der strengen Versionen und etwaige Abweichungen bei der Untersuchung der milderen Bestehensbedingung. Final wird die Summe der richtig beantworteten Bogus-Items als Variable *Bogus_total* angelegt und ebenfalls geprüft. Das ALM wird mit allen Kontrollvariablen als Logit-Modell auf alle Bogus-Items einzeln und als Poisson-Regression auf die Summenvariablen angewendet. Dabei können Bodeneffekte für die Bogus-Items ausgeschlossen werden. Die geringste Häufigkeit des Nicht-Bestehens zeigt sich beim Item *Bogus1* ($n = 29$), während das Item *Bogus4* gar von $n = 159$ Versuchspersonen nicht bestanden wird. Bei der Einzelauswertung aller Bogus-Items zeigt sich für die Variable *Bogus4* (nur für die strenge Version) ein kontraintuitiver signifikanter Geräteeffekt ($b = 0,496$, $z = 2,420$, $p = 0,016$), wobei dieses Item am Smartphone eher bestanden wird. Für alle weiteren Items tritt kein Geräteinfluss auf. Die Variablen *Ablenkung* (signifikant in 66,67% der Analysen; $b_{max} = -0,351$, $z = -3,355$, $p < 0,001$) und *Check bestanden* (signifikant in 75% der Analysen; $b_{max} = 0,744$, $z = 3,082$, $p = 0,002$) zeigen jedoch konstant einen signifikanten Einfluss auf das Bestehen der Bogus-Items (dargestellt sind jeweils die größten Effekte über alle Analysen). Die Summenvariable *Bogus_total* erlaubt die robuste Hypothesenprüfung der **H4.1.1** und bestätigt die Einzelbefunde dahingehend, dass kein Geräteeffekt ($b = 0,020$, $t = 1,073$, $p = 0,284$) vorliegt, aber geringere *Ablenkung* ($b = -0,024$, $t = -3,443$, $p < 0,001$), das Bestehen des IMC (*Check bestanden*) ($b = 0,086$, $t = 4,527$, $p < 0,001$) und das Erkennen des Untersuchungsziels (*Ziel erkannt*) ($b = 0,059$, $t = 2,630$, $p = 0,009$) ein häufigeres Bestehen der Bogus-Items signifikant vorhersagen. Zudem findet sich für jüngere (*Alter*) Personen ($b = 0,002$, $t = 2,806$, $p = 0,005$) und Männer (*Geschlecht*) ($b = -0,042$, $t = -2,158$, $p = 0,031$) eine geringere Anzahl bestandener Bogus-Items. In dem Analysemodell fällt der Test auf Äquidispersion ($z = -22,299$, $p < 0,001$) signifikant aus, sodass aufgrund der vorliegenden Unterdispersion auf die Quasipoisson-Regression ausgewichen wird.

Für die inhaltlichen Fragen zum präsentierten Text (**H4.5.2**) wird zunächst für jedes Item separat evaluiert, ob dieses korrekt beantwortet wurde (0 = falsche Antwort, 1 = korrekte Antwort), und für die Hypothesenprüfung die Anzahl der korrekten Antworten als Variable *Rückfragen* erstellt. Da die inhaltlichen Fragen vor dem IMC in der Befragung angeordnet sind, kann das Untersuchungsziel noch nicht erkannt worden sein, sodass die Prädiktoren *Check bestanden* und *Ziel erkannt* nicht eingehen können. Dies ermöglicht letztmalig, die Analysen auf den *Gesamtdatensatz* zu beziehen. Für die Zählvariable *Rückfragen* liegt Unterdispersion vor ($z = -13,095$, $p < 0,001$), weshalb die Quasipoisson-Regression angewendet wird. Die Befunde

bestätigen die einseitig formulierte Hypothese **H4.5.2**: Smartphone-Teilnehmer*innen beantworten inhaltliche *Rückfragen* zum Text seltener korrekt ($b = -0,700$, $t = -1,780$, $p = 0,038$). Auch höhere *Ablenkung* führt zu weniger korrekten Antworten ($b = -0,052$, $t = -3,449$, $p < 0,001$), jüngere Menschen (*Alter*) antworten häufiger korrekt ($b = -0,004$, $t = -2,502$, $p = 0,013$) und bei *Anwesenheit Dritter* werden kontraintuitiv mehr korrekte Antworten gegeben ($b = -0,106$, $t = -2,033$, $p = 0,043$). Der Vorteil des PCs/Laptops manifestiert sich auch bei den einzelnen inhaltlichen Rückfragen: Für zwei der vier Fragen findet sich ein signifikanter Einfluss des *Geräts* im Logit-Modell der Regression. Dies betrifft die kognitiv herausfordernden Fragen, die ein tieferes Textverständnis und ausdauerndere Aufmerksamkeit beim Lesen erfordern. Informationen, die später im Text genannt werden (*Anzahl untersuchter Milieus*; $b = -0,521$, $z = -2,931$, $p = 0,004$) oder komplexer sind (*inhaltliche Unterscheidung* der Milieus; $b = -0,461$, $z = -2,607$, $p = 0,009$), werden am PC/Laptop besser erinnert, während für sehr einfache Textinformationen (*falsches Milieu* erkennen oder *Jahr der Studie*) keine Geräteeffekte auftreten. Die letzte inhaltliche Rückfrage bildet als *offene Frage* die Grundlage der Hypothese **H4.5.1**. Die Befragten werden gebeten, den Milieubegriff, der im Text vorgestellt wurde, mit eigenen Worten zu definieren, wobei die Buchstabenzahl der Freitextantwort für die Hypothesenprüfung herangezogen wird. Die Anzahl der Buchstaben wird, da es sich dabei um eine Zählvariable handelt, ebenfalls als Poisson-Regression ausgewertet. Aufgrund der verletzten Äquidispersion ($z = 11,033$, $p < 0,001$) – in diesem Fall liegt Überdispersion vor – werden die Kennwerte der Quasipoisson-Regression berichtet. Hypothesenkonform zeigt sich für die *offene Frage*, dass Proband*innen am PC/Laptop signifikant ausführlicher antworten ($b = -0,244$, $t = -3,503$, $p < 0,001$). Für stärker abgelenkte (*Ablenkung*) Personen ($b = -0,121$, $t = -4,357$, $p < 0,001$), Frauen (*Geschlecht*) ($b = 0,162$, $t = 2,307$, $p = 0,021$) und bei *Anwesenheit Dritter* ($b = -0,235$, $t = -2,584$, $p = 0,010$) fällt die Antwort signifikant kürzer aus.

Alle nachfolgenden Analysen beziehen sich auf den *Hauptdatensatz* und integrieren stets sämtliche Kontrollvariablen. Die **H4.5.4** prüft dabei zunächst unabhängig von CR als reguläres ALM, inwiefern die Bearbeitung bei vergleichbarem Ausfüllverhalten am Smartphone mehr Zeit erfordert als am PC/Laptop. Für diese Analyse müssen sehr lange und sehr kurze Antwortzeiten (mehr als zwei Standardabweichungen vom Mittelwert abweichend) zunächst ausgeschlossen werden, da diese die gruppenspezifischen RTs sehr stark verzerren. Für extrem lange Antwortzeiten ist außerdem noch unklar, inwiefern sie aus Sicht der Antwortqualität problembehaftet sind (Read et al. 2022). Sehr kurze RTs hingegen signalisieren eindeutig CR und verzerren die tatsächlich benötigte Bearbeitungsdauer nach unten. Die Analysen verzeichnen, dass das Ausfüllgerät die *RT* signifikant dergestalt beeinflusst, dass am Smartphone tatsächlich

länger gebraucht wird ($b = 76,644$, $t = 1,872$, $p = 0,031$). Die *RT* ist zudem länger für ältere (*Alter*) Befragte ($b = 16,280$, $t = 9,904$, $p < 0,001$), bei *Anwesenheit Dritter* ($b = -221,850$, $t = -4,020$, $p < 0,001$), für Männer (*Geschlecht*) ($b = 91,947$, $t = 2,149$, $p = 0,032$), für Personen, die den IMC bestehen (*Check bestanden*) ($b = 217,964$, $t = 5,216$, $p < 0,001$), und für Befragte, die das Untersuchungsziel erkennen (*Ziel erkannt*) ($b = 228,225$, $t = 4,549$, $p < 0,001$).

Anschließend wird über den „zwei Sekunden pro Item“-Ansatz (siehe Kapitel 3.3.3.3) auf sehr kurze *RTs* als CR-Maß gezielt (**H4.5.3**). Aufmerksames Antworten in weniger als zwei Sekunden pro Item ist nicht plausibel und geringere *RTs* signalisieren CR. Dabei verläuft die Klassifikation als CR in zwei Schritten: zum einen bezogen auf die Gesamtzeit (*RT_gesamt*) und andererseits bezogen auf die benötigte Zeit pro Fragebogenseite (*RT_Seite*). Einbezogen werden alle Fragebogenseiten, die ausschließlich Items beinhalten, was 162 Items auf 17 Fragebogenseiten einschließt. Bei der dichotome Variable *RT_gesamt* (0 = kein CR, 1 = CR) sprechen also Antwortzeiten unter 324 Sekunden für CR. Insgesamt werden 13 Personen als unaufmerksam klassifiziert (8 am PC/Laptop, 5 am Smartphone). Die kleinen Fallzahlen erschweren die regressionsanalytische Auswertung des Logit-Modells, das für *Gerät* keinen signifikanten Effekt findet ($b = 0,179$, $z = 0,273$, $p = 0,785$). Stärkere *Ablenkung* jedoch wirkt signifikant auf die globale Klassifikation als „unaufmerksam“ ($b = 0,541$, $z = 2,870$, $p = 0,004$). Der zweite Analyseschritt prüft seitenweise, ob CR vorliegt, indem für jede Seite ermittelt wird, ob die *RT* geringer ist als das doppelte der Anzahl der Items dieser Seite. Diese seitenweise Prüfung der *RT* wird auch in aktuellen Forschungsbeiträgen empfohlen (Baumgartner und Weijters 2022, S. 125; Bowling et al. 2021b). Die Variable *RT_Seite* (0 = nie CR, 1 = mindestens auf einer Seite CR) markiert $n = 60$ Proband*innen (43 am PC/Laptop, 17 am Smartphone) als unaufmerksam. Hier zeigt sich im Logit-Modell entgegen der Hypothesenformulierung, dass CR signifikant häufiger am PC/Laptop auftritt ($b = -0,723$, $z = -2,295$, $p = 0,022$). Erneut ist die zu schnelle Teilnahme bei hoher *Ablenkung* wahrscheinlicher ($b = 0,212$, $z = 2,105$, $p = 0,035$). Die Fallzahlen (< 25 pro Gruppe) sind weiterhin recht gering, weshalb zur finalen Absicherung und robusten Prüfung der **H4.5.3** der Anteil der Seiten (*RT%*), auf denen entsprechend des Cut-Off-Werts zu schnell geantwortet wird, für jede Person berechnet und als AV über ein reguläres ALM geprüft wird. Hier tritt kein Geräteffekt mehr auf ($b = -0,008$, $z = -0,771$, $p = 0,441$). Der Einfluss der *Ablenkung* auf zu schnelles Antworten persistiert jedoch ($b = 0,015$, $z = 3,709$, $p < 0,001$) und auch das Regressionsgewicht von *Alter* ($b = -0,001$, $z = -2,983$, $p = 0,003$), *Check bestanden* ($b = -0,035$, $z = -3,210$, $p = 0,001$) und *Ziel erkannt* ($b = -0,028$, $z = -2,177$, $p = 0,030$) ist jeweils signifikant.

Die Unterhypothesen **H4.2.1** und **H4.2.2** der **H4.2** adressieren CR über Selbstberichte (reguläre ALMs). Für den Skalenwert der *situativen Aufmerksamkeit* (**H4.2.1**) liegt kein Geräteeffekt vor ($b = 0,057, t = 1,144, p = 0,253$). Die selbstberichtete *situative Aufmerksamkeit* ist aber geringer bei hoher *Ablenkung* ($b = 0,120, t = 6,484, p < 0,001$), bei *Anwesenheit Dritter* ($b = 0,150, t = 2,256, p = 0,025$) und bei Nicht-Bestehen des IMC (*Check bestanden*) ($b = -0,229, t = -4,491, p < 0,001$). Auch für die vier Dimensionen *Antworthast* ($b = 0,075, t = 0,686, p = 0,493$), *Selbstberichtetes CR* ($b = 0,134, t = 1,763, p = 0,079$), *Antwortmuster* ($b = -0,019, t = -0,223, p = 0,824$) und *Instruktionen überspringen* ($b = 0,032, t = 0,270, p = 0,788$) der Skala Selbstberichtete Antwortstile (Maniaci und Rogge 2014, S. 64), die generelles Antwortverhalten bei der Befragungsteilnahme erhebt, treten keine Effekte der Variable *Gerät* auf.

Für das *UseMe-Item* (**H4.2.2**) geben nur 7,12% der Proband*innen an, dass ihre Daten nicht genutzt werden sollten. Aufgrund der geringen Häufigkeiten ($n = 21$, PC/Laptop = 13, Smartphone = 8) müssen die Ergebnisse des Logit-ALM mit Vorsicht behandelt werden. Analog zu Ergebnissen der **H4.2.1** zeigt sich ebenfalls kein Effekt des *Geräts* ($b = 0,188, z = 0,477, p = 0,694$), während die höhere *Ablenkung* signifikant ($b = -0,324, z = -2,216, p = 0,027$) und *Check bestanden* bei Nicht-Bestehen tendenziell ($b = 1,052, z = 1,788, p = 0,074$) eher mit Ablehnung des *UseMe-Items* einhergehen. Somit bewährt sich die **H4.2** einheitlich nicht.

Bei der Analyse der Anzahl der *Missings* (**H4.6b**) wird aufgrund des Zählcharakters die Poisson-Regression verwendet, wobei Äquidispersion gegeben ist ($z = 1,091, p = 0,275$). Für *Gerät* findet sich kein Einfluss auf die *Missings* ($b = 0,038, z = 0,559, p = 0,576$). Für größere *Ablenkung* ($b = 0,062, z = 2,551, p = 0,011$), das Nicht-Bestehen des IMCs (*Check bestanden*) ($b = -0,231, z = -3,198, p = 0,001$) und wenn das Untersuchungsziel nicht erkannt (*Ziel erkannt*) wird ($b = -0,404, z = -4,206, p < 0,001$), treten hingegen signifikant mehr *Missings* auf.

Die nachfolgenden Analysen beziehen sich auf die post-hoc Detektionsmethoden, für die die Prüfvariablen zuvor aus dem Datenmaterial abgeleitet werden. Für die datengeleitete Beurteilung von CR wird das R-Package „Careless“ (Yentes und Wilhelm 2021) verwendet, in dem die Ermittlung von Long Strings (*LongString* und *Stringlänge*), die Berechnung der IRV, die Identifikation psychometrischer *Synonyme* und *Antonyme* und die Berechnung der *Mahalanobis-Distanz* implementiert sind. Monotones Antwortverhalten (**H4.3**) kann über Long Strings und die IRV als Varianz individuellen Antwortens aufgedeckt werden. Zur Ermittlung der Long Strings (**H4.3.1** und **H4.3.2**) wird das Package Careless auf alle aufeinanderfolgenden Items der Befragung angewendet, die mindestens mit einer quasimetrischen, fünfstufigen Antwortskala hinterlegt sind. Dies geschieht, bevor umgekehrte oder verneinte Items umkodiert werden, sodass entsprechend der diversen Itemtypen und Skalenvielfalt der vorliegenden Studie große

Long Strings für aufmerksames Antwortverhalten nicht plausibel sind. Es resultieren für jede*n Teilnehmer*in zwei Variablen: die Variable *LongString* (H4.3.1), die die Länge des längsten Strings als Abfolge identischer aufeinanderfolgender Antworten ausgibt, und die Variable *Stringlänge* (H4.3.2) als durchschnittliche Länge aufeinanderfolgender gleicher Antworten. Für die Prüfung der H4.3.1 wird aufgrund des Zählcharakters das Regressionsmodell im Poisson-Format (Äquidispersion: $z = 1,050$, $p = 0,917$) zunächst auf die Variable *LongString* angewendet. Dabei zeigt sich, entgegen der Formulierung der Hypothese, nach Ausschluss zweier sehr großer verzerrender Extremwerte tendenziell ein Effekt des *Geräts* dahingehend, dass am PC/Laptop der maximale *LongString* größer ist ($b = -0,070$, $t = -1,768$, $p = 0,077$). Keine weitere Modellvariable birgt hier einen signifikanten Einfluss. Anhand der Variable *LongString* erfolgt zudem die CR-Klassifikation (siehe Kapitel 3.3.3.3) über den Scree-Ansatz (Hong et al. 2020; Johnson 2005; Meade und Craig 2012). *Abbildung 13* stellt die Klassifikation in der vorliegenden Studie 4 dar. Im Diagramm wird für jede Länge des maximalen Long Strings die Häufigkeit der Proband*innen, die diese maximale Stringlänge aufweisen, als Datenpunkt eingetragen. Alle Punkte werden anschließend verbunden. Der Knick in der entstehenden Verteilung liegt recht deutlich bei $x = 7$, sodass Strings, die sieben gleiche Antworten in Folge oder mehr darstellen, ungewöhnliches Antwortverhalten signalisieren und für CR sprechen (siehe *Abbildung 13*). Durch diese eindeutige CR-Klassifikation ist die präzise Prüfung der H4.3.1 möglich. Die

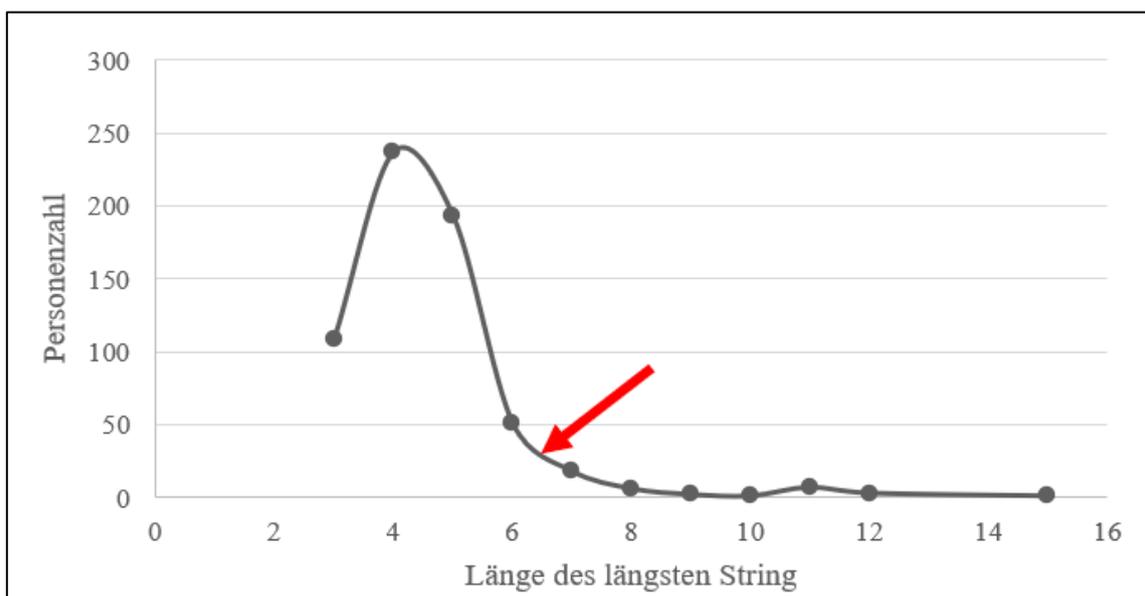


Abbildung 13. Klassifikation als CR über den Long String-Ansatz

Der Scree-Ansatz (vgl. Cattell 1966) der Faktorenanalyse wird auf die CR-Klassifikation über Long Strings übertragen. Die Länge des längsten Long Strings wird auf der x-Achse gegen die Anzahl an Personen (y-Achse) abgetragen, deren längster String diese maximale Länge aufweist. Der rote Pfeil markiert die Knickstelle in der Verteilung, die anzeigt, dass Personen, die häufiger als sechsmal aufeinanderfolgend gleich antworten, eher unaufmerksam sind.

neue dichotome Variable *LongStringCR* (0 = kein CR, 1 = CR) kennzeichnet Personen, deren maximale Anzahl gleicher aufeinanderfolgender Antworten für CR spricht und wird über das Logit-ALM geprüft. Analog zum maximalen Long String (*LongString*) werden PC/Laptop-Teilnehmer*innen beim Scree-Ansatz tendenziell eher als unaufmerksam klassifiziert ($b = -0,696, z = -1,874, p = 0,061$). Die Hypothese **H4.3.1** kann somit abgelehnt werden. Erneut tritt kein Einfluss der Kontrollvariablen auf. Ähnlich sind die Ergebnisse für die *Stringlänge* (**H4.3.2**). Auch hier zeigt das reguläre ALM, dass die durchschnittliche *Stringlänge* am Smartphone tendenziell, und anders als vermutet, kürzer ist ($b = -0,032, t = -1,750, p = 0,081$). Personen, die den IMC bestehen (*Check bestanden*), weisen im Schnitt eine geringere *Stringlänge* auf ($b = -0,044, t = -2,326, p = 0,020$).

Als letzte Unterhypothese monotonen Antwortens (**H4.3**) prüft die **H4.3.3** die *IRV* im regulären ALM. Die *IRV* wird auf die 60 Items des BFI angewendet, bevor umgekehrte Items umkodiert werden, sodass geringe *IRV*-Werte monotonen Antworten indizieren. Erneut ist der Befund nonkonform zur Hypothesenformulierung: Geringere *IRV*-Werte (eher CR) sind am PC/Laptop tendenziell wahrscheinlicher ($b = 0,029, t = 1,665, p = 0,096$). Dabei mündet stärkere *Ablenkung* in geringerer *IRV* ($b = -0,023, t = -3,551, p < 0,001$). Interessante Einblicke liefert die detailliertere Analyse hinsichtlich des Zeitverlaufs. Auf den ersten beiden Seiten des BFI zeigt sich deutlich kein Geräteffekt für die *IRV* ($b = 0,011, t = 0,613, p = 0,540$). Auf den letzten beiden Seiten antworten die Proband*innen am PC/Laptop hingegen tendenziell monotoner ($b = 0,036, t = 1,809, p = 0,071$). Dabei wird die Skalenumkehr auf Seite 18, die unabhängig vom Zeitverlauf Einfluss auf die Antworten nehmen könnte, bewusst von der Monotonieprüfung ausgenommen. In beiden Fällen liegt ein signifikanter Einfluss der *Ablenkung* vor ($b_{Seite1+2} = -0,023, t = -3,551, p < 0,001$; $b_{Seite5+6} = -0,032, t = -4,268, p < 0,001$).

Über *IRV*-Betrachtungen erfolgt auch die Prüfung der Hypothese **H4.4.3** als Auftakt der Konsistenzanalysen (**H4.4**). Auf der vierten Seite der BFI-Items (S. 18) wurde die Skala einmalig unangekündigt umgekehrt. Unaufmerksame Personen, denen diese Skalenumkehr nicht auffällt, sollten weniger konsistent antworten. Zur Hypothesenprüfung wird die Variable *Skalenumkehr* entwickelt. Alle umgekehrten und alle verneinten BFI-Items werden zunächst umkodiert und im Anschluss wird für alle fünf BFI-Dimensionen (*Verträglichkeit, Negative Emotionalität, Offenheit, Gewissenhaftigkeit* und *Extraversion*) separat die *IRV* dimensionsspezifisch pro Versuchsperson berechnet. Bei Antwortkonsistenz sollte die *IRV* jeder Dimension nach der Umkodierung eher gering sein. Die Variable *Skalenumkehr* ergibt sich als Summe der *IRV*-Werte der fünf Dimensionen – hohe Werte deuten an, dass die Skalenumkehr eher nicht erkannt wurde, während niedrige Werte für aufmerksame Teilnahme sprechen. Das reguläre

ALM findet keinen Effekt der Variable *Gerät* ($b = 0,064$, $t = 0,713$, $p = 0,476$). Ältere (*Alter*) Personen ($b = -0,013$, $t = -3,790$, $p < 0,001$) und Personen, die das Untersuchungsziel (*Ziel erkannt*) erkennen ($b = -0,231$, $t = -2,104$, $p = 0,036$), antworten aber konsistenter und nehmen damit die Skalenumkehr eher wahr.

Die Ermittlung psychometrischer *Synonyme* (H4.4.1) und *Antonyme* (H4.4.2) erfolgt ebenfalls über das R-Package „Careless“. Zunächst werden für alle Items der Befragung paarweise Korrelationen berechnet, um datengetrieben diejenigen Itempaare zu identifizieren, für die sehr hohe positive (Synonyme) und negative (Antonyme) Korrelationen zu erwarten sind. Der vorgeschlagene Cut-Off-Wert zur Identifikation der Synonyme bzw. Antonyme beträgt $r \geq |0,6|$ (siehe Kapitel 3.3.3.3). In der Studie 4 werden darüber 18 Itempaare als psychometrische Synonyme identifiziert. Derselbe Cut-Off-Wert liefert für die Antonyme hingegen nur ein einziges Itempaar. Für umgekehrte und verneinte Items sind weniger eindeutige Korrelationsstrukturen nicht ungewöhnlich (z. B. Barnette 2000; Kam et al. 2021), weshalb der Cut-Off-Wert der Antonyme in der vorliegenden Studie nach unten korrigiert wird. Zwölf Itempaare zeigen zumindest eine Korrelation von $r < -0,4$ und bilden daher die psychometrischen Antonyme. Für die Hypothesenprüfung werden zunächst zum einen die Synonyme und andererseits die Antonyme paarweise auf zwei Vektoren aufgeteilt. Für jede Versuchsperson ergeben sich aus der Korrelation der Synonym-Vektoren und der Antonym-Vektoren zwei Korrelationswerte, aus denen die AVn *Synonyme* und *Antonyme* gebildet werden. Bei aufmerksamem Antworten sollten die Korrelationswerte der Proband*innen im Betrag groß sein. Die Hypothesenprüfung über das reguläre ALM zeigt weder für psychometrische *Synonyme* ($b = -0,006$, $t = -0,564$, $p = 0,573$) noch für psychometrische *Antonyme* ($b = -0,004$, $t = -0,188$, $p = 0,851$) einen Effekt des *Geräts*. In beiden Fällen senkt *Ablenkung* ($b_{Synonyme} = -0,010$, $t = -2,511$, $p = 0,012$; $b_{Antonyme} = 0,043$, $t = 5,191$, $p < 0,001$) die Antwortkonsistenz signifikant und Proband*innen, die den IMC bestehen (*Check bestanden*), antworten konsistenter ($b_{Synonyme} = 0,029$, $t = 2,573$, $p = 0,010$; $b_{Antonyme} = -0,088$, $t = -3,844$, $p < 0,001$). Vertiefend für die Konsistenzprüfung werden die Antworten auf das Item „*Ich bleibe auch bei Rückschlägen zuversichtlich*“ analysiert, das zu Beginn des BFI-Blocks präsentiert und gegen Ende erneut identisch abgefragt wurde. Bei konsistentem Antworten sollten diese beiden Items sehr ähnlich beantwortet werden. Dafür wird die Differenz aus den Antworten gebildet und der Betrag als neue Variable *Rückschläge* angelegt, wobei höhere Differenzen eher CR signalisieren. Analog zu den sonstigen Konsistenzanalysen findet das reguläre ALM auch hier keinen Geräteffekt ($b = 0,023$, $t = 0,500$, $p = 0,617$).

Den Abschluss der Hypothesenprüfung bildet die Analyse der *Mahalanobis-Distanz*, die ebenfalls anhand des R-Packages „Careless“ für jede Person über alle Items berechnet wird, die

auf mindestens fünfstufigen Skalen beantwortet werden. Das reguläre ALM zeigt einen signifikanten Geräteeffekt, wobei Proband*innen am Smartphone höhere *Mahalanobis-Distanzen* ($b = 5,941, t = 1,809, p = 0,036$) aufweisen. Höhere *Ablenkung* geht mit höherer *Mahalanobis-Distanz* einher ($b = 4,335, t = 3,542, p < 0,001$) und für Personen, die den IMC bestehen (*Check bestanden*), sind die *Mahalanobis-Distanzen* geringer ($b = -15,104, t = -4,479, p < 0,001$). *Tabelle 11* stellt die Größe aller Geräteinflüsse aus den Hypothesenprüfungen und die jeweilige Bedeutung aller Modellvariablen gesammelt dar.

5.4.2.3 Zusatzbetrachtungen

Die ausführliche Hypothesenprüfung wird durch einige Zusatzanalysen erweitert. Zunächst werden die Befunde abgesichert, indem in allen Analysemodellen die Wechselwirkungen *Gerät*Ablenkung* und *Gerät*Anwesenheit Dritter* für den Interaktionseffekt des Kontexts mit dem Ausfüllgerät ergänzt werden. Der VIF dieser zusätzlichen Interaktionseffekte signalisiert mit Werten $VIF > 5$, dass die Interaktionseffekte Multikollinearität in das Regressionsmodell induzieren. Dies ist bei Linearkombinationen beteiligter Modellvariablen erwartbar und senkt die Aussagekraft. Für den Interaktionseffekt *Gerät*Anwesenheit Dritter* finden sich keine signifikanten Einflüsse. Der Interaktionseffekt *Gerät*Ablenkung* verändert die Befunde der Hypothesenprüfung nur in einem Fall. Konkret fällt durch den Einbezug *Gerät*Ablenkung* der Effekt der UV *Gerät* für die AV *Check bestanden* nicht mehr signifikant aus ($b = 0,015, z = 0,464, p = 0,642$). Der isolierte Einfluss der Variable *Ablenkung* bleibt hingegen nach Einbezug des Interaktionseffekts weiter signifikant ($b = -0,195, z = -2,097, p = 0,036$). In allen Analysen besitzen die Interaktionseffekte selbst keine signifikanten Einflüsse.

Ergänzt wird die Hypothesenprüfung auch mit Blick auf die Studienbörse SurveyCircle. Vor Öffnung der Befragung für die Teilnahme über SurveyCircle ist das Smartphone das etwas häufigere Ausfüllgerät ($n = 152$) im Vergleich zum PC/Laptop ($n = 118$). Dieser Effekt dreht sich für den Gesamtbefragungszeitraum um, wobei der PC/Laptop ($n = 205$) nach dem Datum der Inserierung bei SurveyCircle deutlich häufiger als Ausfüllgerät eingesetzt wird als das Smartphone ($n = 119$). Dabei handelt es sich vermutlich um ein Artefakt, da bei SurveyCircle darauf geachtet wurde, dass die genutzten Ausfüllgeräte der Befragten, die über SurveyCircle gewonnen werden konnten, sich in etwa ausgleichen²³. Dafür spricht ebenfalls, dass die Geräteverteilung im *Gesamtdatensatz* insgesamt gleich ist ($\chi^2(1) = 2,129, p = 0,145$). Am Smartphone

²³ Eine eindeutige Klassifikation als SurveyCircle-Teilnehmer*in war ohne die Erhebung personenbezogener Daten, von der aus Datenschutzgründen abgesehen wurde, nicht möglich. Entsprechend können Teilnehmer*innen nach dem Inserat bei SurveyCircle sowohl von der Studienbörse stammen als auch anders rekrutiert worden sein.

H4. ...																	
	1.1	1.2	2.1	2.2	3.1	3.2	3.3	4.1	4.2	4.3	5.1	5.2	5.3	5.4	6.a	6.b	6.c
	Bogus-Items	IMC	CR-Skala	Use-Me	Long String	Stringlänge	IRV	Synonyme	Antonyme	Skalenumkehr	offene Frage	Rückfragen	RT%	RT	Mahalanobis	Missings	Abbruch
<i>Gerät</i>	<i>p</i> = 0,284	<i>p</i> = 0,030	<i>p</i> = 0,253	<i>p</i> = 0,694	<i>p</i> = 0,077	<i>p</i> = 0,081	<i>p</i> = 0,096	<i>p</i> = 0,573	<i>p</i> = 0,851	<i>p</i> = 0,476	<i>p</i> < 0,001	<i>p</i> = 0,038	<i>p</i> = 0,441	<i>p</i> = 0,031	<i>p</i> = 0,036	<i>p</i> = 0,576	<i>p</i> < 0,001
<i>Alter</i>	<i>p</i> = 0,005	<i>p</i> < 0,001	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> < 0,001	<i>p</i> > 0,05	<i>p</i> = 0,013	<i>p</i> > 0,05	<i>p</i> < 0,001	<i>p</i> > 0,05	<i>p</i> > 0,05	
<i>Geschlecht</i>	<i>p</i> = 0,031	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> = 0,021	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> = 0,032	<i>p</i> > 0,05	<i>p</i> > 0,05	
<i>Ablenkung</i>	<i>p</i> < 0,001	<i>p</i> = 0,032	<i>p</i> < 0,001	<i>p</i> = 0,027	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> < 0,001	<i>p</i> = 0,012	<i>p</i> < 0,001	<i>p</i> > 0,05	<i>p</i> < 0,001	<i>p</i> < 0,001	<i>p</i> = 0,004	<i>p</i> > 0,05	<i>p</i> < 0,001	<i>p</i> = 0,011	
<i>Anwesenheit Dritter</i>	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> = 0,025	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> = 0,010	<i>p</i> = 0,043	<i>p</i> > 0,05	<i>p</i> < 0,001	<i>p</i> > 0,05	<i>p</i> > 0,05	
<i>Check bestanden</i>	<i>p</i> < 0,001		<i>p</i> < 0,001	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> = 0,020	<i>p</i> > 0,05	<i>p</i> = 0,010	<i>p</i> < 0,001	<i>p</i> > 0,05			<i>p</i> > 0,05	<i>p</i> < 0,001	<i>p</i> < 0,001	<i>p</i> = 0,001	
<i>Ziel erkannt</i>	<i>p</i> = 0,009		<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> > 0,05	<i>p</i> = 0,036			<i>p</i> > 0,05	<i>p</i> = 0,009	<i>p</i> > 0,05	<i>p</i> < 0,001	

Tabelle 11. Übersicht über die Ergebnisse der Hypothesenprüfung der Studie 4

Für alle Modellvariablen aller Hypothesenprüfungen sind die *p*-Werte der Regressionsgewichte aufgelistet. Für die UV Gerät werden die Rohwerte vollständig berichtet, in allen anderen Fällen nur signifikante *p*-Werte < 0,05. Alle signifikanten Effekte sind **fett** gedruckt. In der ersten Zeile sind hypothesenkonform gerichtete Effekte **grün** und nonkonforme Effekte **rot** gefärbt.

brechen Teilnehmer*innen die Befragung aber häufiger ab (**H4.6c**), was auch die Ungleichverteilung der Geräte im *Hauptdatensatz* ($\chi^2(1) = 4,552, p = 0,033$) relativiert. Um einen systematischen Effekt der Studienbörse auszuschließen oder, falls er vorliegt, aufzudecken, werden alle Analysen nur für die Datensätze wiederholt, die bereits vor der Inserierung der Befragung bei SurveyCircle erhoben wurden. Die uneinheitlichen Befunde zu exklusiven Geräteeffekten bleiben für diese reduzierte Stichprobe ($n = 270$) größtenteils gewahrt. Wie schon beim Interaktionseffekt *Gerät*Ablenkung* verändert sich das Befundmuster aber für die *AV Check bestanden*. Der Nachteil des PCs/Laptops löst sich auf und der Geräteinfluss ist deutlich nicht signifikant ($b = -0,039, z = -0,148, p = 0,882$). Auch die häufigere Klassifikation als CR über Long Strings (*LongStringCR*) ($b = -0,769, z = -1,167, p = 0,243$), der seitenweise Anteil (*RT%*) sehr schnellen Antwortens ($b = 0,024, z = 1,492, p = 0,137$), die *Stringlänge* ($b = -0,018, t = -0,800, p = 0,425$) und die *IRV* als Ausdruck der Antwortkonsistenz im Befragungsteil zum BFI ($b = 0,006, t = 0,279, p = 0,781$) werden vor dem Studienstart bei SurveyCircle nicht durch das Ausfüllgerät beeinflusst. Der vermeintliche Nachteil des PCs/Laptops, der sich genau für diese Analysen in der Hypothesenprüfung zeigt, tritt also erst nach der Einstellung der Studie auf der Plattform SurveyCircle auf. Ansonsten bleiben die Effekte überwiegend bestehen, wobei speziell die Befunde der inhaltlichen *Rückfragen* zum Text und der Antwort auf die *offene Frage* interessant sind. *Rückfragen* werden am Smartphone weiterhin signifikant seltener korrekt beantwortet und der Effekt ist sogar rein numerisch verstärkt ($b = -0,143, t = -2,431, p = 0,016$). Erneut kann dabei unterschieden werden zwischen simpleren *Rückfragen* (keine signifikanten Geräteeffekte) und komplexeren *Rückfragen* (signifikante Geräteeffekte), die eine gründlichere Auseinandersetzung mit dem Text verlangen. Auch der Einfluss des Ausfüllgeräts auf die Antwortlänge bei der *offenen Frage* bleibt signifikant und kennzeichnet diese am Smartphone als kürzer ($b = -0,345, t = -3,239, p = 0,001$).

Die Zusatzanalysen betrachten anknüpfend noch die Antwortgüte der Freitextantwort für alle Datensätze. Die Proband*innen waren aufgefordert, den Milieubegriff eigenständig zu definieren. Für die Auswertung der Güte wird eine dichotomisierte Variable (*Antwortgüte*, 0 = keine/falsche Antwort, 1 = grundsätzlich korrekte Antwort) unabhängig von der Antwortlänge erstellt und über das Logit-ALM ausgewertet. Für den *Gesamtdatensatz* zeigt sich eine marginalsignifikant höhere *Antwortgüte* am PC/Laptop ($b = -0,337, z = -1,896, p = 0,058$). Diese ist auch höher bei geringerer *Ablenkung* ($b = -0,227, z = -3,455, p < 0,001$). Ohne Einbezug der Proband*innen, die nach Einstellung der Studie bei SurveyCircle teilnahmen, verstärkt sich der Effekt noch. Hier tritt nun ein deutlich signifikanter Einfluss der Variable *Gerät* auf die *Antwortgüte* auf ($b = -0,805, z = -2,958, p = 0,003$), wobei PC/Laptop-Teilnehmer*innen

gehaltvoller antworten. Zur Absicherung eines tatsächlichen Effekts des Ausfüllgeräts wird abschließend verglichen, ob die Befragten auf der Textseite gerätweise unterschiedlich viel Zeit verbringen. Dafür werden alle Personen ($n = 3$) ausgeschlossen, die auf der Textseite länger als zehn Minuten verweilen, da dies für Abwesenheit oder längere Unterbrechungen spricht und solche extremen Ausreißer die Analysen verzerren. Der t -test für unabhängige Stichproben (alle Voraussetzungen erfüllt) findet keinen signifikanten Unterschied ($t(623) = 0,099$, $p = 0,921$) und zeigt, dass Befragte am PC/Laptop ($\bar{x} = 140,15$ Sekunden, $SD = 108,58$) und Smartphone ($\bar{x} = 139,29$ Sekunden, $SD = 105,70$) im Mittel ähnlich viel Zeit auf der Textseite verbringen.

An die Hypothesenprüfung wird final noch eine differenzierte Auseinandersetzung mit den Bogus-Items angeschlossen. Bei dieser Betrachtung werden zwei Typen inhaltlich unterschieden: Für einige der Bogus-Items ist der Interpretationsspielraum klar begrenzt, da sie sich auf Einzelereignisse beziehen und dementsprechend pauschal nicht oder nur extrem unwahrscheinlich abgelehnt werden können bzw. ihnen zugestimmt werden kann. In der vorliegenden Befragung betrifft dies die Items „*Ich saß noch nie in einem fahrenden Auto*“ (*Bogus1*), „*Ich habe noch nie gelacht*“ (*Bogus3*) und „*Ich nehme gerade an einer Befragung teil*“ (*Bogus6*), die als Itemtyp „*Einzelereignis*“ klassifiziert werden. Die Bogus-Items „*Ich bin schon in jedes Land der Welt geflogen*“ (*Bogus2*), „*Ich lüge immer*“ (*Bogus4*) und „*Ich kann zwei Kilometer in zwei Minuten laufen*“ (*Bogus5*) sind interpretationsoffener und bilden die Itemkategorie „*Interpretationsspielraum*“. Sehr reisefreudige sowie sportliche Personen könnten sich auch bei gewissenhafter Antwort schwertun, die Items komplett abzulehnen, da sie schon in mehr Länder als viele ihrer Mitmenschen geflogen sind, oder überaus sportlich sind und daher vergleichsweise schnell laufen. Auch Personen, die nicht stets die Wahrheit sagen, könnten *Bogus4* zumindest teilweise zustimmen. *Tabelle 12* listet die Durchfallquoten der sechs Bogus-Items getrennt nach Strengegrad und Itemtyp auf. Darin zeigen sich deskriptive Unterschiede zwischen den Durchfallquoten der beiden Itemkategorien, die besonders stark ausfallen, wenn das Bestehen des Items nur bei strikter Ablehnung/Zustimmung attestiert wird (strenge Bestehensprüfung). Bei strenger Prüfung bestehen 4,581 bis 8,202% der Versuchspersonen die Items des Typs *Einzelereignis* nicht, während die Items mit *Interpretationsspielraum* deutlich häufiger nicht bestanden werden (19,085 bis 25,198%). Bei milder Bestehensprüfung ist die Spannweite der Durchfallquoten über alle Itemtypen aber deutlich geringer (3,011 bis 7,886%). Zur Analyse dieser Anteilsunterschiede wird eine mixed-ANOVA durchgeführt mit dem *Itemtyp* als Zwischensubjektfaktor (0 = *Einzelereignis*, 1 = *Interpretationsspielraum*) und der *Strenge der Prüfung* (0 = mild, 1 = streng) als Innersubjektfaktor (Field 2017, S. 591-594). Die Ergebnisse zeigen, dass sowohl die Faktoren *Itemtyp* ($F(1, 4) = 78,990$, $p < 0,001$, $\eta^2_{\text{partial}} = 0,952$) und *Strenge der*

Prüfung ($F(1, 4) = 43,524, p = 0,003, \eta^2_{\text{partial}} = 0,916$) als auch die Interaktion der beiden Faktoren ($F(1, 4) = 20,994, p = 0,010, \eta^2_{\text{partial}} = 0,840$) signifikant sind. Dabei treten keine zusätzlichen Effekte des Ausfüllgeräts auf.

Item	Itemtyp	Durchfallquote streng	Durchfallquote nicht streng
<i>Bogus1</i> : „Ich saß noch nie in einem fahrenden Auto“	Einzelereignis	4,581	3,791
<i>Bogus3</i> : „Ich habe noch nie gelacht“	Einzelereignis	6,656	3,011
<i>Bogus6</i> : „Ich nehme gerade an einer Befragung teil“	Einzelereignis	8,202	5,047
<i>Bogus2</i> : „Ich bin schon in jedes Land der Welt geflogen“	Interpretations-spielraum	19,968	7,765
<i>Bogus4</i> : „Ich lüge immer“	Interpretations-spielraum	25,198	6,498
<i>Bogus5</i> : „Ich kann zwei Kilometer in zwei Minuten laufen“	Interpretations-spielraum	19,085	7,886

Tabelle 12. Durchfallquoten der Bogus-Items in Studie 4 aufgeteilt nach Itemtyp

5.4.3 Diskussion

Die Aufmerksamkeit als Ausdruck der Teilnahmemotivation ist die wesentliche Grundlage der Antwortqualität in (Online-)Befragungen (siehe Kapitel 3.2.1.2). Gerätspezifische Aufmerksamkeitsunterschiede würden den TSE erhöhen und womöglich zu Moduseffekten durch das Smartphone führen. Die Ergebnisse der dritten Studie der vorliegenden Forschungsreihe attestieren dem Smartphone im Laborversuch eine potenziell größere Anfälligkeit für Unaufmerksamkeit in Onlinebefragungen, da das Blickverhalten am Smartphone unsteter und sprunghafter ist. Die vierte Studie knüpft an diesen Befund an und prüft, ob Befragte am Smartphone im Feld tatsächlich vermehrt unaufmerksam antworten und entsprechend des Survey-Response Modells (Tourangeau et al. 2000, S. 7-17) anfälliger für vom Iteminhalt losgelöstes Antworten sind. Am Smartphone könnte dann häufiger CR auftreten, sodass der Messfehler durch Befragte steigt. Dabei ist die Prüfung des Antwortverhaltens und speziell von möglichem CR sehr breit angelegt und orientiert sich an aktuellen Literaturempfehlungen (z. B. vgl. Baumgartner und Weijters 2022; Curran 2016; Hong et al. 2020; Meade und Craig 2012). Zudem wird Item-Nonresponse als weiterer Teilbereich der Antwortqualität betrachtet, um etwaige Moduseffekte hinsichtlich der Antwortqualität so weit wie möglich erschöpfend untersuchen zu können.

In den Ergebnissen der vierten Studie finden sich nur vereinzelt signifikante Unterschiede im Antwortverhalten und der Antwortqualität zwischen Smartphone- und PC/Laptop-Teilnehmer*innen. Generell zeichnen die Befunde das Bild, dass das Smartphone nicht anfälliger für

CR ist und sprechen somit gegen Moduseffekte. Die CR-Raten in Studie 4 sind indes substantiell, ohne dass ein pauschaler Geräteffekt für die Antwortqualität auftritt. Dadurch können Bodeneffekte als Erklärung für Nulleffekte des Geräteinflusses ausgeschlossen werden.

Für klassische Infrequency-Indikatoren können die Hypothesen (**H4.1.1** und **H4.1.2**) klar abgelehnt werden und die Befundrichtung ist teilweise sogar umgekehrt, da der IMC (**H4.1.2**) am PC/Laptop entgegen der Hypothesenformulierung seltener bestanden wird. Sehr hohe Raten des Nicht-Bestehens des IMC (in der vorliegenden Studie > 50%) treten gerade bei komplexeren IMCs häufig auf (Burns et al. 2014; Clifford und Jerit 2015; Leiner 2019; Morren und Paas 2020) und sind nicht ausschließlich ein Indikator für klassisches CR, sondern auch für andere, weniger starke Aufmerksamkeitseinschränkungen und zudem abhängig von kognitiven Fähigkeiten (siehe Kapitel 3.3.3.2). Die vorliegende Studie bestätigt zusätzlich einen eigenen Einfluss des IMC auf das Antwortverhalten dahingehend, dass IMCs die Aufmerksamkeit erhöhen können (z. B. Anduiza und Galais 2017; Clifford und Jerit 2015; Oppenheimer et al. 2009). Bei Personen, die den IMC bestehen, tritt in der Folge deutlich weniger CR auf und das IMC-Bestehen als Kontrollvariable im Regressionsmodell der Hypothesenprüfung übt fast durchgängig einen signifikanten Einfluss auf die AVn aus. Ähnlich wirkt Einsicht in das Untersuchungsziel. Erkennen Versuchspersonen, dass die Aufmerksamkeit während der Teilnahme überprüft wird (etwa als Folge des IMC), zeigen sie weniger CR. Im Rahmen der Hypothesenprüfung ist auch der Einfluss des Ziel-Erkennens auf CR-Maße wiederholt signifikant.

Für die **H4.1.1** der Bogus-Items treten keine Unterschiede zwischen den Ausfüllgeräten auf. Am Smartphone wird auf Bogus-Items genauso häufig richtig geantwortet wie am PC/Laptop. Die übergeordnete Hypothese **H4.1** kann aufgrund der klaren Ablehnung der **H4.1.1** und **H4.1.2** nicht angenommen werden. Infrequency-Ansätze liefern keine Nachweise für häufigeres CR am Smartphone. Die Zusatzauswertungen demonstrieren aber unabhängig vom Geräteffekt, dass die Formulierung und Integration von Bogus-Items überlegt geschehen sollte. Diese Einordnung ist relevant, da die Anteile des Nicht-Bestehens der Bogus-Items stark variieren. Gerade wenn mehrstufige Antwortskalen hinterlegt sind (Berry et al. 2019; Dunn et al. 2018; Huang et al. 2015a; Olatunji et al. 2007), sollten Interpretationsspielräume bei der Beantwortung der Bogus-Items möglichst reduziert und stattdessen vorrangig Einzelereignisse abgefragt werden. Andernfalls sind die Raten des Nicht-Bestehens deutlich höher und es kann nicht sorgfältig getrennt werden, ob diese aus unaufmerksamem Antworten oder dem Bedürfnis nach gewissenhaftem Antworten resultieren. Aufmerksame Personen, die insgesamt sehr wenig schlafen, könnten beispielsweise das Item „Ich schlafe weniger als eine Stunde pro Nacht“

(Meade und Craig 2012, S. 441) nicht konsequent ablehnen, da sie im Vergleich mit anderen Personen deutlich weniger schlafen und diesen Unterschied gerne ausdrücken möchten.

Auch im selbstberichteten CR (**H4.2**) zeigen sich keine Unterschiede zwischen den Ausfüllgeräten. Die berichtete situative Aufmerksamkeit (**H4.2.1**) und die Häufigkeit der Antwort auf das UseMe-Item, dass die Daten eher nicht für die Auswertung genutzt werden sollten (**H4.2.2**), fallen am Smartphone und am PC/Laptop gleich aus. Für sich selbst berichteten Befragte demnach am Smartphone gleich häufig Unaufmerksamkeit wie am PC/Laptop, sodass auch dieser Befund eher gegen Moduseffekte des Smartphones als Ausfüllgerät spricht.

Zur Aufdeckung weiterer CR-Formen, die sich in den Datenstrukturen zeigen, werden die Tendenzen zu monotonem Antworten (**H4.3**) und die Antwortkonsistenz (**H4.4**) gerätweise verglichen. Die Ergebnisse der Studie 4 widersprechen beiden Hypothesen und finden keine Nachteile des Smartphones. Monotones Antworten (**H4.3**) tritt, entgegen der Hypothesenformulierung, in Form von Long Strings, die Straightlining signalisieren, am PC/Laptop sogar tendenziell häufiger auf (**H4.3.1**). Auch die durchschnittliche Stringlänge (**H4.3.2**) und die maximale Stringlänge fallen am PC/Laptop tendenziell höher aus. Ergänzend ist auch die individuelle Antwortvarianz (IRV) am PC/Laptop tendenziell geringer als am Smartphone (**H4.3.3**). Der erwartete Nachteil des Smartphones kann demzufolge für monotonen Antworten nicht bestätigt werden. Entsprechend der vorgestellten Ergebnisse der Studie 4 scheint monotonen Antworten stattdessen insgesamt sogar am PC/Laptop eher wahrscheinlich.

Für den IMC und monotonen Antworten dreht sich der Geräteeffekt also scheinbar um und kennzeichnet eher den PC/Laptop als anfällig für unaufmerksames Antworten. Die Zusatzanalysen ordnen diesen vermeintlichen Geräteeffekt jedoch ein und konstatieren, dass das Ausfüllgerät auch in diesen Bereichen nicht pauschal auf CR und allgemein auf die Antwortqualität Einfluss nimmt. Die Effekte entstehen erst nach Öffnung der Studie für Befragte, die über die Studienbörse SurveyCircle rekrutiert werden. In der gesonderten Überprüfung nur für Personen, die zuvor teilnehmen, verschwinden die Effekte: Der IMC wird hier gerätübergreifend gleich häufig bestanden und monotonen Antwortverhalten tritt gerätübergreifend gleich oft auf. In Studienbörsen zeigt sich häufiger geringe Antwortqualität (z. B. Cheung et al. 2017; Fleischer et al. 2015; Kennedy et al. 2020). Da die Gerätverteilung bei der Rekrutierung über SurveyCircle konstant gehalten wurde, ist die teilweise geringere Antwortqualität am PC/Laptop aber nicht auf die häufigere Teilnahme mit dem PC/Laptop über SurveyCircle zurückzuführen. Die vorliegenden Befunde lassen jedoch vermuten, dass Befragte, die über Studienbörsen mit dem PC/Laptop teilnehmen, häufiger strikt darauf bedacht sein könnten, den persönlichen Aufwand bei der Teilnahme so gering wie möglich zu halten. Versuchspersonen, die die

Befragung möglichst schnell erledigen möchten, greifen dafür häufiger auf Speeding zurück (Zhang und Conrad 2014). Am PC/Laptop kann Straightlining durch monotone Klickbewegungen (z. B. mit Computermäusen, wobei nur minimale Cursorbewegungen für Straightlining nötig sind) besonders zeitsparend eingesetzt werden, um den Antwortaufwand zu minimieren, sodass die Zeitersparnis entsprechend der vorliegenden Studienbefunde dort ggf. eher über Long Strings oder monotones Antworten erreicht wird (Keusch und Yan 2017). Dies signalisiert aber keine allgemeinen Moduseffekte, da sich abseits solcher Studienbörsen keine klaren Straightlining-Unterschiede für verschiedene Ausfüllgeräte zeigen (Erens et al. 2019; Mavletova et al. 2018; Tourangeau et al. 2018). Straightlining nimmt gerade bei langen Befragungen in späten Befragungsteilen zu (Herzog und Bachman 1981; Schonlau und Toepoel 2015; Zhang und Conrad 2014); es scheint somit eher ein Ausdruck der Teilnahmemotivation zu sein, der nicht direkt mit dem Ausfüllgerät verknüpft ist. Die Ergebnisse der Studie 4 kennzeichnen entsprechend Smartphones und PCs/Laptops als ähnlich anfällig für monotones Antworten und Speeding als CR-Formen. Der PC/Laptop ist nur dann als etwas anfälliger, wenn Befragte ganz bewusst Aufwand vermeiden und die Antwortgabe verkürzen wollen, was in Studienbörsen häufiger ein vorrangiges Teilnahmeziel darstellt.

Die Konsistenz der Antworten (**H4.4**) ist am Smartphone ebenfalls vergleichbar mit der Antwortkonsistenz am PC/Laptop. Gerätübergreifend fällt die Antwortkonsistenz für psychometrische Synonyme (**H4.4.1**) und psychometrische Antonyme (**H4.4.2**) ähnlich aus. Sowohl am Smartphone als auch am PC/Laptop treten häufiger inkonsistente Antwortmuster auf, ohne dass die Konsistenz gerätspezifisch stärker gehemmt ist. Analog zur Literatur offenbart aber auch die vorliegende Studie, dass der Einsatz psychometrischer Antonyme zur CR-Klassifikation weniger performant ist, da die Korrelationsmuster für umgekehrte oder verneinte Items stärker variieren und mit relevanten Hintergrundvariablen wie Bildung oder Erfahrung verknüpft sind (Barnette 2000; Kam et al. 2021; Yan et al. 2018). Des Weiteren erkennen die Teilnehmer*innen in der Studie 4 die unangekündigte Skalenumkehr (**H4.4.3**) während der Abfrage der BFI-Items am Smartphone gleich häufig wie am PC/Laptop. Somit tritt RR als Form von CR, das sich insbesondere in der Antwortkonsistenz offenbaren kann (siehe Kapitel 3.3.1.2 und 3.3.3.3), eher gerätunabhängig auf. Die Konsistenzanalysen der Hypothesenprüfung und auch die punktuellen Zusatzanalysen, die ebenfalls keine Konsistenzunterschiede aufzeigen, konstatieren damit gerätübergreifend vergleichbare Antwortqualität, wobei der Messfehler durch Befragte an beiden betrachteten Ausfüllgeräten (PC/Laptop und Smartphone) ähnlich ausfällt.

Die Hypothesen **H4.1-H4.4** sprechen somit grundsätzlich dafür, dass durch Smartphones als Ausfüllgeräte in Onlinebefragungen keine nachteiligen Moduseffekte auftreten. Mit Blick auf

die Ergebnisse der dritten Studie des vorliegenden Forschungsprogramms stellt sich die Frage, warum das unstete Blickverhalten am Smartphone eher nicht zu Aufmerksamkeitseinschränkungen und damit zu Einbußen in der Antwortqualität führt. Am Smartphone kann im Alltag ein Verengungseffekt („Narrowing“) der Aufmerksamkeit auf das Gerät beobachtet werden, wobei die Nutzer*innen selbst extrem saliente externe Reize nicht wahrnehmen und den Fokus permanent auf das Smartphone richten (z. B. Chen und Pai 2018; Hyman et al. 2010; Melumad und Meyer 2020). Möglicherweise überträgt sich dies auch auf die Befragungsteilnahme. Unter häufigem Abschweifen des Blickes, das die Ergebnisse der dritten Studie für das Smartphone feststellen, muss die kognitive Leistung und die Informationsverarbeitung nicht zwingend leiden. Durch Rehearsal in Form von Refixationen können die Gedächtnisinhalte aktualisiert werden (z. B. Gwizdka und Zhang 2015, S. 814; Nikolaev et al. 2023; Wedel und Pieters 2000; Zelinsky et al. 2011), sodass bei Blicksprüngen (Sakkaden) ohne Informationsaufnahme (Rayner 1998) die Leistung durch Refixationen aufrechterhalten werden kann (Meghanathan, Nikolaev und van Leeuwen 2019). Grundlegend dafür ist, dass Distraktoren, die durch den sprunghaften Blick und das Abweichen von der Befragung ins Sichtfeld geraten, die Aufmerksamkeit nicht binden. Aktuelle Forschungsbefunde zeigen klar und passend zu den vorliegenden Ergebnissen, dass es grundsätzlich gut möglich ist, potenzielle Ablenkungsreize zu inhibieren, die dann die Leistungsfähigkeit und die kognitive Verarbeitung nicht hemmen (Meghanathan et al. 2019; Oberauer 2019). Werden Ablenkungsreize im Sichtfeld bei der kurzzeitigen Blickverlagerung als irrelevant eingestuft, werden sie nicht aktiv verarbeitet (Meghanathan et al. 2019; Meghanathan et al. 2020). Bei Blickrückkehr zu den Befragungsinhalten ist die vorherige Gedächtnisspur dann noch intakt und die weitere Verarbeitung startet ohne Qualitätsverlust (Houtkamp und Roelfsema 2006; Meghanathan et al. 2019). Resultat ist dann lediglich ein leicht erhöhter Zeitaufwand für die Refixationen, aber keine Verringerung der Antwortqualität.

Zur Einordnung dieser Befundmuster wird die Diskussion zur **H4.5.4** vorgezogen, da diese genau auf längere RTs bei der Teilnahme mit dem Smartphone ausgerichtet ist. Die Ergebnisse der Studie 4 bestätigen gemäß der **H4.5.4**, dass Befragte am Smartphone signifikant länger für die Beantwortung derselben Befragungsinhalte benötigen. Dies deckt sich mit der Literatur, in der RTs am Smartphone auch stets größer ausfallen (z. B. Andreadis 2015a, S. 74; Couper und Peterson 2017; Drewes 2014, S. 378; Lambert und Miller 2015; Mavletova 2013; Revilla et al. 2016a; Struminskaya et al. 2015; Tourangeau et al. 2018). Die Synthese der Studienergebnisse der Studien 3 und 4 der vorliegenden Arbeit eröffnet dafür einen neuen Erklärungsansatz: Aufgrund des unsteteren Blickverhaltens am Smartphone könnten häufiger Refixationen benötigt

werden, um Rehearsal anzuregen und Leistungsabfälle in der kognitiven Verarbeitung zu prävenieren. Daher könnte am Smartphone bei gleicher Leistung für die Bearbeitung länger gebraucht werden, wodurch ggf. längere RTs am Smartphone in Onlinebefragungen entstehen.

Potenzielle Defizite des unsteten Blickverhaltens können demnach bei ausreichender Motivation, konzentriert teilzunehmen, aufgefangen werden (Herten, Otto und Wolf 2017; Meghanathan et al. 2019). Sind Ablenkungsreize jedoch sehr stark bzw. dominant oder lang anhaltend, binden sie eher die Aufmerksamkeit (Meghanathan et al. 2019; Meghanathan et al. 2020; Pashler 1988), die sich dann auch nachhaltig von der Befragung entfernen kann. Die Studienergebnisse untermauern diesen Wirkmechanismus: In mehr als 80% der geprüften Regressionsmodelle zeigt sich ein signifikanter Einfluss der Variable Ablenkung. Bei stärkerer Ablenkung tritt gerätübergreifend häufiger CR auf, das heißt, die Aufmerksamkeit sinkt, wodurch sich die Antwortqualität teilweise drastisch reduziert. Die Befunde unterstreichen also, dass der Teilnahmekontext systematisch auf das Antwortverhalten wirkt (z. B. Höhne et al. 2020a; Pan et al. 2014). Ablenkungsreize (z. B. Lärm, Bystander, Multi-Tasking) reduzieren ganz allgemein die Antwortqualität in Befragungen (z. B. Aquilino 1993; Sendelbah et al. 2016; Wenz 2021b; Zwarun und Hall 2014) und sind bereits in Onlinebefragungen häufiger und breiter gefächert (van Selm und Jankowski 2006; Ward et al. 2014; Wharton et al. 2003). Die Diversität des Ausfüllkontexts ist am Smartphone nochmals potenziert (z. B. Antoun et al. 2017; Höhne und Schlosser 2019; Lütters 2017, S. 68; Mavletova und Couper 2013) und Proband*innen berichten am Smartphone auch tatsächlich häufiger Ablenkungen während der Teilnahme (z. B. Ansolabehere und Schaffner 2015; de Bruijne und Oudejans 2015, S. 141; Revilla et al. 2016a; Toninelli und Revilla 2016a, S. 4). Der bedeutende Mehrwert der vorliegenden Forschungsreihe liegt darin, dass der Geräteffekt des Smartphones isoliert und dadurch entkoppelt von etwaigen Kontexteinflüssen untersucht wird. Die vorliegenden Befunde und die vorherigen Ausführungen ordnen diese Kontextvielfalt ein und legen nahe, dass Unterschiede in der Antwortqualität zwischen verschiedenen Ausfüllgeräten (z. B. Höhne und Schlosser 2019; Kim und Sundar 2014; Revilla und Couper 2018; Schlosser und Silber 2020, S. 278; Struminskaya et al. 2015) eher nicht auf das Gerät selbst, sondern eher auf dominante Ablenkungsreize zurückzuführen sind, die am Smartphone häufiger auftreten. Werden Hintergrundvariablen und Kontextfaktoren kleinschrittig kontrolliert, verschwinden Antwortqualitätsunterschiede zu meist (Andreadis 2015b; Clement, Severin und Shamshiri-Petersen 2020; Cook 2014; Tourangeau et al. 2018). Auch in Studie 4 weisen alle einbezogenen Kontrollvariablen vor allem kontextueller Art (Ablenkung, Anwesenheit Dritter) aber auch soziodemografischer (Alter, Geschlecht) wiederholt signifikante Einflüsse auf das Antwortverhalten auf. Primär schwächen

dabei Ablenkungsreize und dominante Kontextfaktoren die Antwortqualität. Diese wirken an den verschiedenen Ausfüllgeräten ähnlich stark und es treten daher keine systematischen Interaktionen der Kontextfaktoren mit dem Ausfüllgerät auf.

Die letzte explizite CR-Hypothese (**H4.5**) adressiert die Verarbeitungstiefe am Smartphone und offenbart zum Abschluss der Hypothesenprüfung einen möglichen kritischen Bereich, in dem ein Moduseffekt des Smartphones als Ausfüllgerät vorliegen könnte. Analog zum Forschungsstand (Erens et al. 2019; Mavletova 2013; Revilla et al. 2016a; Wells et al. 2014) antworten die Befragten am Smartphone signifikant weniger ausführlich (**H4.5.1**) auf offene Fragen. Auch die Güte der kürzeren Antworten ist dabei geringer. Gleichzeitig werden in Studie 4 inhaltliche Rückfragen (**H4.5.2**) zum dargestellten Text am Smartphone ebenfalls seltener korrekt beantwortet. Gerade bei solchen Aufgaben mit größerer kognitiver Herausforderung ist die Leistung am Smartphone mitunter schwächer (Antoun et al. 2018; Barr et al. 2015). Eine mögliche Erklärung könnte sein, dass Inhalte am Smartphone als herausfordernder erlebt werden können (z. B. Raptis et al. 2013, S. 133; Schlosser und Mays 2018; Wenz 2021b). Insbesondere offene Fragen und inhaltliche Rückfragen erfordern als eher komplexe Befragungsinhalte ein höheres Maß an stabiler kognitiver Verarbeitung. Die Kapazität des visuellen Arbeitsgedächtnisses ist jedoch gerade beim sinnentnehmenden Lesen stark limitiert (Baddeley 2003; Daneman und Carpenter 1980) und wird durch Aufmerksamkeitsverlagerungen bedeutend gehemmt (Steindorf und Rummel 2020; Zhang et al. 2021).

Das unstete Blickverhalten, das in Studie 3 am Smartphone bei der Befragungsteilnahme beobachtet wird, könnte genau für diese Befunde ein möglicher Erklärungsansatz sein. Unstetes Blickverhalten schwächt die Leseleistung (Radach, Inhoff und Heller 2002, S. 141; Radach und Kennedy 2004; Rayner 2009) und die Kapazitätsgrenzen sind abhängig von der Komplexität der Informationen (Alvarez und Cavanagh 2004). Bei hoher kognitiver Herausforderung kann die Unstetigkeit des Blicks nicht mehr ausgeglichen werden, während dieser Ausgleich für einfache Aufgaben noch möglich ist (Meghanathan et al. 2019). Die Ergebnisse der Studie 4 bekräftigen diesen Erklärungsansatz: Einfachere Fragen werden gerätübergreifend gleich häufig korrekt beantwortet, während der Nachteil des Smartphones sich erst für Fragen offenbart, die ein tieferes Textverständnis erfordern. Für die **H4.5.1** und **H4.5.2** ändern sich die Befundmuster auch nicht durch den Ausschluss der SurveyCircle-Teilnehmer*innen, was die Anfälligkeit für geringere Antwortqualität am Smartphone bei komplexen Befragungsinhalten untermauert. Denn obwohl die Versuchspersonen die Textseite am Smartphone genauso lange betrachten wie am PC/Laptop, antworten sie trotzdem weniger gehaltvoll und ausführlich. Unstetes Blick-

verhalten am Smartphone kann demnach möglicherweise bei komplexen Inhalten und zusammenhängenden Texten nicht immer vollständig ausgeglichen werden und letztlich zu Antwortqualitätseinschränkungen in Onlinebefragungen führen. Dadurch entsteht vermutlich ein Moduseffekt, der den Messfehler durch die Befragten für die beschriebenen Anwendungsfälle erhöht. Für die **H4.5.3** zeigt sich daran anknüpfend, dass Speeding am Smartphone hingegen nicht häufiger auftritt, das heißt, nicht schneller als 2 Sekunden pro Item (seitenweise und insgesamt) geantwortet wird. Dieser Befund überrascht mit Blick auf die bisherigen Ausführungen nicht, da Straightlining eher als Speeding-Strategie am PC/Laptop genutzt wird, während diese Antwortform in den Befunden am Smartphone sehr selten ist. Die geringere Verarbeitungstiefe am Smartphone, die die Befunde der Hypothesenprüfung der **H4.5.1** und **H4.5.2** zeigen, kann demnach nicht auf Speeding als CR-Form zurückgeführt werden und die Antwortqualitätsreduktionen bei eher komplexen Inhalten treten auf, obwohl die Befragten insgesamt am Smartphone länger für die Teilnahme benötigen (**H4.5.4**). Zusammengefasst kann für die **H4.5** daher konstatiert werden, dass komplexe Befragungsinhalte und tiefgreifendes Textverständnis sowie offene Fragen, die ausführliche und gehaltvolle Antworten erfordern, einen möglichen Problembereich für das Smartphone als Ausfüllgerät darstellen, für den sich womöglich ein Moduseffekt zeigt und in dem der Messfehler durch Befragte steigt.

Für klassische Antwortqualitätsindikatoren über reines CR hinaus sind die Befunde der Hypothesenprüfung gemischt. Zum einen wird die Vermutung, dass Befragungen am Smartphone häufiger abgebrochen werden, eindeutig bestätigt (**H4.6c**). Dies deckt sich mit der Literatur, in der sich ebenfalls höhere Abbruchraten am Smartphone finden lassen (z. B. Buskirk und Andrus 2014; Cook 2014; Keusch und Yan 2017; Lambert und Miller 2015; Poggio et al. 2015). Befragungsabbrüche treten dabei schon unabhängig vom Gerät häufiger bei komplexen oder herausfordernden Befragungsinhalten sowie offenen Fragen auf (z. B. Erens et al. 2019; Peytchev 2009; Tuschl, Meister und Laube 2016, S. 213). In Studie 4 zeigt sich zusätzlich, dass die Abbruchraten am Smartphone erst im Verlauf der Befragung höher sind als am PC/Laptop. Die Abbruchraten am Smartphone steigen in der Befragung deutlich stärker als am PC/Laptop nach den inhaltlichen Rückfragen zum langen Text, der offenen Frage und dem herausfordernden IMC. Hier könnte ein ähnlicher Erklärungsansatz wie bei der Hypothese **H4.5** wirken: Das unstete Blickverhalten (Studie 3) und der zeitlich begrenzte sowie variable Nutzungskontext des Smartphones bei der sonstigen Verwendung im Alltag (z. B. Deng et al. 2019; Kim et al. 2005; Kim et al. 2017) könnten identische Inhalte in Befragungen am Smartphone herausfordernder wirken lassen als am PC/Laptop und den Abbruch fördern. Da die Qualität der Antworten bis zum Abbruch allerdings nicht zwingend schlechter ist (Peytchev 2009) und technisch bedingte

Befragungsabbrüche im Zuge der technologischen Entwicklung immer seltener werden (Drewes 2014, S. 373; Wenz 2021a, S. 111), offenbaren diese Befunde allerdings eher keinen negativen Geräteinfluss des Smartphones auf die Qualität der abgegebenen Antworten vor der Abbruchsentscheidung.

Missings (**H4.6b**) treten in der Studie 4 wiederum am Smartphone nicht häufiger auf, wobei die Befundlage hierfür diverser ist: Einige Quellen berichten, dass am Smartphone häufiger klassische Item-Nonresponse in Form fehlender Werte auftritt als am PC/Laptop (Couper et al. 2017, S. 148; Lugtig und Toepoel 2016), während viele andere Studien keine Unterschiede finden (z. B. Andreadis 2015b; Revilla und Couper 2018; Toepoel und Lugtig 2014; Tourangeau et al. 2018). Missings treten am Smartphone dabei besonders häufig bei offenen Fragen und komplexen Inhalten auf (Lambert und Miller 2015; Struminskaya et al. 2015), was im Einklang mit den Befunden der Studie 4 zur Verarbeitungstiefe (**H4.5**) die Anfälligkeit des Smartphones für Antwortqualitätsreduktionen bei herausfordernden Befragungselementen erneut unterstreicht. Die Ergebnisse der Studie 4 stützen aber konkret für Missings – zumindest bei regulären, weniger komplexen Inhalten – eher die Sichtweise, dass diese am Smartphone nicht pauschal häufiger auftreten und für die Item-Nonresponse eher keine klaren Moduseffekte entstehen. Die Befragung wird zwar häufiger komplett abgebrochen, aber nicht häufiger ohne Antwortgabe weitergeführt. Dies ist wichtig für die korrekte Klassifikation von Item-Nonresponse. Werden auch fehlende Werte nach Abbruch der Befragung als Missings behandelt, ist dies nicht korrekt, da sie kein Ausdruck der Item-Nonresponse sind, sondern systematisch gar nicht mehr geantwortet wird.

Eine letzte Gerätesonderheit des Smartphones zeigt sich hypothesenkonform in höheren Mahalanobis-Distanzen (**H4.6a**). Die Datenstrukturen zwischen Smartphone- und PC/Laptop-Nutzer*innen sind demnach nicht komplett uniform, auch wenn sich hinsichtlich der Aufmerksamkeitsleistung überwiegend keine Unterschiede in Studie 4 zeigen. Die Relevanz dieses Befundes ist aber unklar, da die theoretische Fundierung der Mahalanobis-Distanz als CR-Maß noch aussteht (Arias et al. 2020; Baumgartner und Weijters 2022, S. 136-139). Hohe Mahalanobis-Distanzen signalisieren lediglich, dass bestimmte Antwortreihen im Vergleich zum gesamten Datensatz ungewöhnlich sind. In der vorliegenden Studie sind die Raten von CR aber mitunter recht hoch, sodass die Interpretation ungewöhnlichen Verhaltens erschwert wird. Bei sehr hohen CR-Raten könnte sich die Signalwirkung der Mahalanobis-Distanz umkehren, sodass eher aufmerksame Befragte auffällig und damit ungewöhnlich sind. Dagegen spricht, dass in der vorliegenden Arbeit auch selbstberichtete hohe Ablenkung mit hohen Mahalanobis-Distanzen einhergeht, während Personen, die den IMC bestehen, niedrigere Distanzen aufweisen.

Gleichwohl erschwert die teilweise hohe CR-Prävalenz, wie ausgeführt, die Einordnung. Die gefundenen Unterschiede könnten außerdem auch auf nicht kontrollierte Hintergrundvariablen zurückzuführen sein und somit nicht exklusiv CR signalisieren. Künftige Forschungsprojekte könnten davon profitieren, die Bedeutung der Mahalanobis-Distanz als CR-Maß zu sondieren und die Einsatzmöglichkeiten für die Beurteilung der generellen Antwortqualität zu validieren.

Die Befunde für die Mahalanobis-Distanzen demonstrieren, dass Studie 4 auch mit einigen Limitationen behaftet ist. Trotz des breiten Ansatzes der CR-Aufdeckung könnte CR vereinzelt übersehen werden. Hier könnten fortgeschrittene statistische Verfahren (z. B. Person-Fit-Maße, Machine Learning) die Aufdeckung noch weiter vertiefen (vgl. Arthur et al. 2021; Conijn et al. 2019; Goldammer et al. 2020; Schroeders et al. 2022). Auch der Einbezug von Paradata, gerade am Smartphone, könnte weitere Einblicke ermöglichen (Gogami et al. 2021; Kern et al. 2021). In der Literatur zeigen sich zudem Zusammenhänge von CR und spezifischen Personeneigenschaften (z. B. implizite Aggression und Benevolenz) (DeSimone et al. 2020; McKay et al. 2018), die bei der Analyse in Studie 4 nicht über die Persönlichkeitseigenschaften der Big Five hinaus einbezogen werden. Zudem wird das Kontinuum der Antwortqualität nicht in seiner gesamten Breite berücksichtigt und der Fokus vorrangig auf geringe Antwortqualität gelegt. Womöglich könnte hier eine differenziertere Analyse verschiedener Grade der Antwortqualität die Prüfung des Einsatzes von Smartphones entweder weiter absichern oder etwaige weniger kraftvolle Qualitätseinschränkungen aufdecken.

Abschließend hätte die eindeutige Identifikation der Teilnehmer*innen, die über die Studienbörse SurveyCircle rekrutiert wurden, möglicherweise tiefere Dateneinsicht ermöglicht. Künftig könnten separate aber inhaltsgleiche Befragungen für Studienbörsen und andere Rekrutierungswege angelegt und erst im Nachgang ineinander überführt werden. Ein weiterer wesentlicher Kritikpunkt liegt bei Studie 4 im Befragungszeitpunkt. Im Herbst/Winter des Jahres 2020 erlaubten die Herausforderungen der Covid-19 Pandemie, sogar gesetzlich reguliert, nur wenig Kontakt mit Mitmenschen und der Bewegungsspielraum der Bevölkerung war erheblich eingeschränkt, sodass sowohl die sozialen Einflüsse als auch die örtliche Flexibilität des Ausfüllkontexts wesentlich begrenzt waren. In zukünftigen Studien könnten die Untersuchungen möglicher Interaktionseffekte von Kontext und Gerät, die in der vorliegenden Studie nicht gefunden werden konnten, ausgeweitet werden. Diese gezielte Betrachtung der gerätweise variierenden Kontextvielfalt könnte die Erkenntnislage weiter stärken.

6 Abschließende Diskussion

Das vorliegende Dissertationsprojekt untersucht anhand eines methodisch breit angelegten Forschungsprogramms die Forschungsfrage:

„Entstehen durch das Smartphone als Ausfüllgerät Moduseffekte in Onlinebefragungen und welche Folgen hat dies ggf. für den Einsatz des Smartphones als Ausfüllgerät?“

Die Studienreihe strebt die möglichst umfassende Prüfung etwaiger Moduseffekte im TSE-Framework an, die aus dem reinen Geräteinfluss des Smartphones als Ausfüllgerät resultieren. Die Annäherung an die Antwort auf die Forschungsfrage erfolgt schrittweise, indem zunächst zwei Voruntersuchungen die Prüfung möglicher Moduseffekte des Smartphones als Ausfüllgerät vorbereiten. Im Anschluss erweitern die Hauptuntersuchungen die Prüfung möglicher Moduseffekte des Smartphones, in denen zwei Studien aus verschiedenen Perspektiven Einflüsse des Smartphones auf die Antwortqualität beleuchten. Die folgenden Unterkapitel geben anhand der empirischen Ergebnisse eine versuchsweise Antwort auf die Fragestellung (Kapitel 6.1), ordnen die Relevanz der Befunde der Studienreihe für die empirische Wirtschafts- und Sozialforschung ein (Kapitel 6.2) und zeigen die Limitationen sowie künftigen Forschungsbedarf auf (Kapitel 6.3), bevor ein Fazit die Arbeit abschließt (Kapitel 6.4).

6.1 Zusammenfassung der Haupterkenntnisse

In modernen, wissenschaftlichen Onlinebefragungen ist das Smartphone ein reguläres und beliebtes Ausfüllgerät (z. B. Bosnjak et al. 2017, S. 55; Callegaro 2013; Gummer et al. 2019; Revilla 2017; Toepoel und Lugtig 2014). Onlinebefragungen wurden zwar nicht systematisch für die Teilnahme mit dem Smartphone geöffnet, aber die Diversität der Ausfüllgeräte wächst parallel zur technologischen Entwicklung aufgrund der universellen Einsetzbarkeit moderner internetfähiger Geräte. Moduseffekte des Smartphones könnten die Validität der erhobenen Daten hemmen, wenn der Einsatz des Smartphones das Antwortverhalten verändert und dadurch den Total Survey Error erhöht (vgl. Bautista 2012; Biemer 2010; Faulbaum 2022; Groves und Lyberg 2010; Johnson 2005; Weisberg 2009).

Die ersten beiden Studien des Forschungsprogramms trennen entsprechend der Empfehlungen von Tourangeau (2017, S. 117) für die Analyse von Moduseffekten etwaige Einflüsse des Ausfüllgeräts auf die Antwortqualität von etwaigen Einflüssen auf die Unit-Nonresponse. Die erste Studie der vorliegenden Arbeit überprüft hierzu Moduseffekte des Smartphones für sensible Themenbereiche, da sensible Themen für Moduseffekte (z. B. unaufrichtiges Antworten und bewusste Antwortverweigerung) besonders empfänglich sind (z. B. Barnett 1998; Jäckle et

al. 2010; Tourangeau et al. 1997a; Tourangeau und Yan 2007; Weisberg 2009, S. 278). Aufgrund der Individualität und persönlichen Bedeutung (siehe auch Kapitel 2.3, insbesondere 2.3.2) des Smartphones (z. B. Carter et al. 2016, S. 1440; Jung 2014; Melumad und Pham 2020) könnten Personen bei sensiblen Themen an diesem ggf. sogar ehrlicher und offener antworten (siehe Kapitel 3.1). Moduseffekte würden dann bei sensiblen Themen eher einen Vorteil des Smartphones als Ausfüllgerät anzeigen. Die erste Studie schließt als Laborstudie Kontexteffekte der Teilnahme, die den Selbstbericht hemmen könnten, bei der Untersuchung explizit aus.

Studie 1: Die Studie zeigt, dass das Smartphone bei sensiblen Fragen die Antwortqualität nicht erhöht. Unabhängig vom Ausfüllgerät beantworten Versuchspersonen sensible Fragen gleich aufrichtig (Messfehler durch Befragte) und auch eher gleich häufig (Item-Nonresponse). Auch die Tendenz, sozial erwünscht zu antworten, wird nicht vom Ausfüllgerät beeinflusst. Es treten bei sensiblen Fragen also keine Moduseffekte am Smartphone auf, die den TSE beeinflussen.

Die zweite Voruntersuchung (Studie 2) löst sich von sensiblen Fragen und nähert sich möglichen Moduseffekten für die Unit-Nonresponse, indem die Hintergründe der Teilnahmeentscheidung an Onlinebefragungen mit verschiedenen Ausfüllgeräten betrachtet werden. Dafür wird geprüft, ob diese Geräte unterschiedlich gut geeignet sind, befragungsrelevante Bedürfnisse für die Teilnahme an Onlinebefragungen zu erfüllen. Die Geräteigenschaften und die Flexibilität der Befragungsteilnahme könnten das Smartphone als Ausfüllgerät für verschiedene Bedürfnisdimensionen entweder eher disqualifizieren oder eher prädestinieren. Hierdurch womöglich variierende gerätspezifische Unit-Nonresponseraten könnten dann den TSE erhöhen, wenn Befragte, denen bestimmte Dimensionen wichtig sind, bevorzugt mit dem Gerät teilnehmen, das für diese Dimensionen besser geeignet ist.

Studie 2: Die Ergebnisse unterstreichen, dass Smartphones und PCs/Laptops nicht als identische Ausfüllgeräte wahrgenommen werden. Die wahrgenommene Eignung des Smartphones für Onlinebefragungen hinsichtlich der Befriedigung nutzungsrelevanter Bedürfnisse unterscheidet sich von der Einschätzung anderer Ausfüllgeräte. Allerdings variiert die Eignungseinschätzung verschiedener Ausfüllgeräte interindividuell stark. Die Studienergebnisse demonstrieren, dass das individuell tatsächlich gewählte Ausfüllgerät in der eigenen Wahrnehmung für bedürfnisrelevante Teilnahmebedingungen besser geeignet erscheint als in der Wahrnehmung der Nutzer*innen anderer Geräte und dass kein stabiler Prognosewert für die Gerätwahl besteht.

Wenn das Ausfüllgerät selbst gewählt wird, findet dementsprechend eine natürliche Kontrolle der Teilnahmebedürfnisse statt. Unit-Nonresponse, die daraus entstehen könnte, dass befragungsrelevante Bedürfnisse von einem bestimmten Ausfüllgerät nicht saturiert werden, ist

demnach nur zu erwarten, wenn die Teilnahme mit dem Smartphone gar nicht möglich ist. Da die Präferenz für das Smartphone analog zum Bedürfnis nach selbstbestimmter Gerätwahl steigt (z. B. de Bruijne und Wijnant 2013; Krämer 2017, S. 100; Poulková et al. 2019, S. 305; Revilla et al. 2016b), könnte also eher nicht der Einsatz, sondern stattdessen der Ausschluss des Smartphones als Ausfüllgerät ggf. zu Unit-Nonresponse führen. Die Befunde legen nahe, dass auch die persönliche Bedeutung des Smartphones nicht gegen den Einsatz als Ausfüllgerät in Onlinebefragungen spricht und hierdurch eher keine Moduseffekte induziert werden. Möglicherweise ist die Bindung an das eigene Smartphone nicht im jeweils von der Person verwendeten Gerät selbst begründet, sondern entsteht eher aus den genutzten Funktionen, die das Gerät bereitstellt (z. B. Melumad und Pham 2020; Salehan und Negahban 2013; Thorsteinsson und Page 2014). Die Voruntersuchungen zeigen zusammenfassend, dass zum einen eher keine positiven Moduseffekte des Smartphones für die Antwortqualität bei sensiblen Fragen auftreten und andererseits keine gerätspezifisch variierenden Unit-Nonresponseraten aufgrund von Moduseffekten des Smartphones zu erwarten sind.

Für die breite Prüfung im TSE-Framework wird in den Hauptuntersuchungen die Antwortqualität am Smartphone stärker in den Blick genommen und die Betrachtung über sensible Themen hinaus erweitert. Die Voruntersuchungen betrachten lediglich mögliche positive Moduseffekte des Smartphones. Für die Unterbereiche der Antwortqualität (Messfehler durch Befragte und Item-Nonresponse) zeigen sich in der Forschungsliteratur jedoch wiederholt Unterschiede, wenn Befragte mit dem Smartphone teilnehmen: Diese sprechen eher für Moduseffekte, die den TSE erhöhen würden (siehe Kapitel 3.2.3). Die Grundlage der Antwortqualität in Onlinebefragungen bildet die Aufmerksamkeit als Folge der Motivation während der Teilnahme (siehe Kapitel 3.2.1 und 3.3). Eine Laborstudie, die das Blickverhalten am Smartphone mit dem Blickverhalten am PC/Laptop während der Befragungsteilnahme vergleicht, eröffnet die Hauptuntersuchungen und betreibt Grundlagenforschung dazu, ob möglicherweise Aufmerksamkeitsdefizite, die die Antwortqualität hemmen könnten, am Smartphone wahrscheinlicher sind.

Studie 3: Die Studienbefunde zeichnen ein klares Bild, das das Smartphone als potenziell anfälliger für unaufmerksames Antworten charakterisiert. Das Blickverhalten ist am Smartphone unsteter: Die Proband*innen blicken am Smartphone häufiger von der Befragung weg, die Fixationszeiten sind kürzer, es werden mehr Refixationen in die Befragung hinein ausgeführt und der Blick auf die Befragungsinhalte ist deutlich weniger stabil.

Solches unstetes Blickverhalten ist mit geringerer Aufmerksamkeit assoziiert (z. B. Just und Carpenter 1976; Meißner und Oll 2019; Rayner 2009; Tatler et al. 2005). Die potenzielle Anfälligkeit des Smartphones für Unaufmerksamkeit in Onlinebefragungen wird in der vorliegenden Arbeit final mit einer breit angelegten Feldstudie zu Careless Responding (CR) geprüft. Diese vierte Studie der Forschungsreihe untersucht CR sowie weiteres für die Antwortqualität kritisches Antwortverhalten und erfasst darüber indirekt die Aufmerksamkeit der Befragten. Dadurch können etwaige Moduseffekte des Smartphones hinsichtlich der Antwortqualität geprüft werden, die aus dem unsteten Blickverhalten folgen könnten. Die Studie 4 stellt die Kernstudie des vorgestellten Forschungsprogramms dar und liefert auch die umfangreichsten Erkenntnisse.

Studie 4: Entgegen der Vermutungen und der Befunde der dritten Studie tritt CR am Smartphone nicht häufiger auf als am PC/Laptop. Proband*innen nehmen an beiden Ausfüllgeräten gleichermaßen gewissenhaft sowie aufmerksam an Onlinebefragungen teil und die jeweils erhobenen Daten zeugen gerätübergreifend von ähnlicher Antwortqualität. Trotz des unsteteren Blickverhaltens gelingt es Proband*innen am Smartphone, während der Befragungsteilnahme überwiegend aufmerksam zu bleiben. Die potenziellen Defizite können ausgeglichen werden, solange mögliche Distraktoren als irrelevant eingestuft werden (Meghanathan et al. 2019; Meghanathan et al. 2020). Nach der Blickabkehr stoßen Refixationen dann Rehearsal als Update-Funktion der Gedächtnisinhalte an, ohne dass die Aufmerksamkeitsleistung leidet (Baddeley 2003; Meghanathan et al. 2019; Tatler et al. 2005). Dieser Ausgleich ist aber nicht uneingeschränkt möglich, sondern die Unstetigkeit des Blicks am Smartphone führt bei eher komplexen und herausfordernden Befragungsinhalten zur Verringerung der Antwortqualität. Unstetes Blickverhalten erschwert gerade aufmerksames Lesen (Radach et al. 2002, S. 141; Radach und Kennedy 2004; Rayner 2009), sodass das Smartphone hier anfällig für Verzerrungen ist. Bei hoher kognitiver Beanspruchung, wie etwa bei der Verarbeitung längerer, zusammenhängender Texte und bei komplexen Informationen, kann unstetes Blickverhalten kaum adäquat ausgeglichen werden (Alvarez und Cavanagh 2004; Meghanathan et al. 2019). Die Studienergebnisse bestätigen dies und demonstrieren, dass die Leseleistung und der damit verbundene Abruf von Gedächtnisinhalten am Smartphone bei komplexen Inhalten erschwert sein kann. Analog zur Literatur werden zudem auch offene Fragen am Smartphone weniger ausführlich beantwortet bei zugleich geringerer Antwortgüte (Lambert und Miller 2015; Revilla et al. 2016a; Wells et al. 2014).

Das breite Forschungsprogramm der vorliegenden Arbeit kennzeichnet zusammenfassend das Smartphone als adäquates Ausfüllgerät in Onlinebefragungen. Dabei wird der Forschungsstand wesentlich erweitert, indem Moduseffekte zuerst im Allgemeinen und vertieft spezifisch abgestimmt auf Konsequenzen für die Antwortqualität in den Blick genommen werden. Dadurch wird eine aussagekräftige Betrachtung des Einflusses auf den TSE möglich. Überwiegend zeigen sich keine kritischen Moduseffekte, sodass sich das Smartphone als Ausfüllgerät in Onlinebefragungen weitestgehend bewährt. Bei längeren Texten und der Verarbeitung komplexer Informationen sowie bei differenzierten Antworten auf offene Fragen kann die Antwortqualität am Smartphone aber mitunter leiden, da der Messfehler durch Befragte und somit der TSE ansteigen können. Dies deutet auf einen bereichsspezifischen Moduseffekt des Smartphones für komplexe Inhalte und lange Texte hin. Das Wissen um das Ausfüllgerät kann in diesem Bereich statistische Analyse stärken, um auch gerätübergreifend robuste Schlussfolgerungen aus den erhobenen Daten ableiten zu können. Aus der Synthese der Ergebnisse der Studien 2 und 4 ergibt sich regulierend, dass Teilnehmer*innen das Smartphone zwar eher als Komfortquelle während der Teilnahme einschätzen können (Studie 2), aber dadurch nicht weniger motiviert bei der Beantwortung sind (Studie 4). Für Item-Nonresponse finden sich sogar überhaupt keine stabilen Moduseffekte des Smartphones.

Aus den Analysen der vorliegenden Arbeit kann vielmehr abgeleitet werden, dass Ablenkung während der Teilnahme eine wesentliche Gefahr für CR als Ursache reduzierter Antwortqualität darstellt. Dominante und anhaltende Ablenkungsreize während der Befragungsteilnahme reduzieren die Antwortqualität deutlich (z. B. Meghanathan et al. 2020; Oberauer 2019; Sendelbah et al. 2016; Wenz 2021b) und erhöhen somit den TSE wesentlich, da der Messfehler durch Befragte steigt. Dies tritt in den vorgestellten Befunden des Forschungsprogramms sehr deutlich zu Tage. Es zeigt sich allerdings kein Interaktionseffekt zwischen dem Ausfüllgerät und der Ablenkung. Ablenkungen sind aufgrund der größeren Kontextdiversität bei der Teilnahme mit dem Smartphone zwar häufiger (z. B. de Bruijne und Oudejans 2015, S. 136; Höhne et al. 2020a; Kim et al. 2005; Kim et al. 2017), wirken aber nicht stärker auf die Antwortqualität und Aufmerksamkeit als am PC/Laptop. Dieser Befund unterstreicht nachdrücklich das Vorgehen der vorliegenden Arbeit: Im Forschungsprogramm wurde stets großer Aufwand betrieben, um den Geräteeffekte zu isolieren und entkoppelt von Kontexteinflüssen zu untersuchen.

6.2 Implikationen für die empirische Wirtschafts- und Sozialforschung

Onlinebefragungen stellen in der modernen Forschung und insbesondere in der empirischen Wirtschafts- und Sozialforschung eine der beliebtesten und verbreitetsten Methoden der Datenerhebung dar (z. B. Couper 2005, 2008, S. 62; Hulland et al. 2018). Als wichtigste Grunderkenntnis der vorliegenden Arbeit können Smartphones als Ausfüllgeräte in Onlinebefragungen größtenteils bedenkenlos eingesetzt werden, ohne selbst Moduseffekte zu erzeugen, die den TSE erhöhen. Die freie Gerätwahl bei der Teilnahme sollte also in Onlinebefragungen stets ermöglicht werden, um den Bedürfnissen der Proband*innen gerecht zu werden (Axinn et al. 2015; Krämer 2017, S. 100; Revilla et al. 2016b) und etwaigen Problemen der Coverage entgegen zu wirken (Couper et al. 2018; Wenz 2021a, S. 101). Es empfiehlt sich jedoch, das Ausfüllgerät stets mit zu erheben, um es ggf. bei vertiefenden Analysen als Kontrollvariable in die Datenauswertung einbeziehen zu können (z. B. Zhang et al. 2017). Insbesondere wenn längere Textpassagen Teil von Fragebogen sind oder Antworten auf offene Fragen ausgewertet werden, ist das Wissen um das Ausfüllgerät relevant, da sich hier im Forschungsprogramm der einzige Bereich zeigt, in dem Moduseffekte auftreten könnten, die die Antwortqualität am Smartphone punktuell hemmen.

Für die Gestaltung von Onlinebefragungen ergeben sich aus den Befunden der vorliegenden Arbeit einige konkrete Konsequenzen. Die Matrixdarstellung von Items ist am Smartphone in Onlinebefragungen grundsätzlich möglich: Trotz der geringeren Gerätgröße und des vermehrten Scrollings zeigen die Befunde, dass der Einsatz von Matrizen am Smartphone eher unproblematisch ist, solange die adäquate Darstellung (z. B. im Landscape-Modus) gewährleistet wird (de Bruijne 2015, S. 135; Tourangeau et al. 2017). Da dieselben Inhalte am Smartphone aber herausfordernder wirken können als am PC/Laptop, forciert dies das Postulat, Befragungen möglichst niedrighschwellig, schlank und einfach zu konzipieren (z. B. Ansolabehere und Schaffner 2014; Eberl 2016, S. 220; Krosnick 2017, S. 445-451; van Selm und Jankowski 2006). Diese Forderung wird durch den diversen Nutzungskontext und dadurch, dass konzentrierte Nutzungsphasen am Smartphone in der Regel kürzer sind, verstärkt (Deng et al. 2019; Kim et al. 2017; Yang et al. 2012).

Eine der Hauptgefahren für die Antwortqualität in Onlinebefragungen resultiert hingegen aus dem unkontrollierten Befragungskontext. Ablenkung während der Teilnahme reduziert die Antwortqualität deutlich und schmälert die Aussagekraft der Befunde (z. B. Oberauer 2019; Sendelbah et al. 2016; Wenz 2021b; Zwarun und Hall 2014). Die höhere Frequenz von Ablenkungsreizen am Smartphone konfrontiert die empirische Wirtschafts- und Sozialforschung

durch die Möglichkeit der Smartphoneteilnahme mit dem Gebot, Kontextbedingungen der Teilnahme stets zu erfassen und, wenn möglich, Ablenkungsreize zu reduzieren. Onlinebefragungen sollten nach reiz- und störungsarmen Ausfüllsituationen streben, um suboptimales Antwortverhalten zu minimieren.

Abseits der Rolle des Smartphones als Ausfüllgerät ergeben sich aus den vorgestellten Befunden auch Empfehlungen für die Stichprobenakquise in der fragebogenbasierten, empirischen Wirtschafts- und Sozialforschung. Die Resultate der vierten Studie indizieren, dass Verzerrungen durch den Einbezug von Studienbörsen entstehen können. Proband*innen, die über Studienbörsen teilnehmen, neigen eher zu weniger gewissenhaftem Ausfüllverhalten (Cheung et al. 2017; Fleischer et al. 2015; Kennedy et al. 2020). Bei der Nutzung von Studienbörsen sollten daher möglichst akkurat die zugrundeliegenden Motivationsstrukturen betrachtet und hoher Aufwand betrieben werden, um geringe Antwortqualität aufzudecken (z. B. über die Prüfung auf CR, Antwortstile oder Antwortheuristiken).

Der zweite, größere Forschungsbereich, aus dem wichtige Implikationen folgen, ist Careless Responding (CR) selbst. CR ist in allen Befragungen und somit auch in der empirischen Wirtschafts- und Sozialforschung eine wesentliche Grundgefahr für valide Ergebnisse (z. B. Abbey und Meloy 2017; Arias et al. 2020; DeSimone et al. 2020). Im besten Fall kann CR weitestgehend präveniert, mindestens sollte es aber aufgedeckt werden (Arthur et al. 2021; Goldammer et al. 2020; Ward und Meade 2023). Die Aufdeckungsmöglichkeiten für CR wurden inzwischen schon mehrfach ausführlich zusammengefasst (z. B. Baumgartner und Weijters 2022; Curran 2016; Hong et al. 2020; Meade und Craig 2012) und werden stetig weiterentwickelt (z. B. Niesen et al. 2016; Schroeders et al. 2022). Aus den Erkenntnissen der vorliegenden Arbeit kann vorsichtig abgeleitet werden, dass IRIs eine praktikablere Alternative für die Aufdeckung von CR darstellen als IMCs. IRIs kombinieren den Vorteil, als kognitive Bodenschwelle zu fungieren und die Aufmerksamkeit im Zuge der CR-Aufdeckung gleichzeitig zu fördern, mit vergleichsweise geringerer kognitiver Herausforderung: Sie sind somit in der Lage, CR eindeutiger zu erkennen (Anduiza und Galais 2017; Curran und Hauser 2019; Gummer et al. 2021; Kam und Chan 2018). Die sehr hohen Raten des Nicht-Bestehens des IMC (> 50%) in der vorliegenden Arbeit können höchstwahrscheinlich nicht ausschließlich als Ausdruck von CR bewertet werden.

Als weitere Erkenntnis sollte die Auswahl und Konzeption von Bogus-Items stets reflektiert erfolgen (Curran und Hauser 2019). Der Charakter der Items sollte möglichst eindeutig sein. Für Bogus-Items, die Interpretationsspielraum offenlassen, ist es nicht möglich, klare Grenzen aufmerksamen Antwortens festzulegen, da auch aufmerksame Befragte scheinbar unplausibel

antworten können. Daher sollten Bogus-Items bevorzugt werden, die sich auf Einzelereignisse beziehen und für die nur eine Antwort tatsächlich korrekt ist (siehe Kapitel 5.4.2.3).

6.3 Limitationen und Ausblick

Der Forschungsbeitrag der vorliegenden Arbeit ist auch mit einigen Limitationen verbunden, die bei der Einordnung der Befunde berücksichtigt werden müssen und künftige Forschungsbedarfe eröffnen. Um das Forschungsziel zu erreichen, wurden Kontexteffekte teilweise ausgeschlossen oder zumindest explizit kontrolliert. Aufgrund der schier unbegrenzten Kontextvielfalt der Smartphonennutzung könnten im Forschungsprogramm womöglich relevante Komponenten des Teilnehmerrahmens übersehen worden sein, die für die Antwortqualität aber relevant sein könnten. In Zukunft sollte daher auch das Nutzungserleben während der Befragungsteilnahme noch gezielter adressiert werden, um weitere womöglich bedeutsame Drittvariablen aufzudecken. Mit diesen könnten die Analysen weiter vertieft werden.

Künftige Studien könnten auch die Analyse des Antwortverhaltens und die Aufmerksamkeitsmessung methodisch weiterentwickeln. Statt das Blickverhalten lediglich bezogen auf generelle Befragungsgrenzen zu analysieren, könnten künftige Eye-Tracking Untersuchungen die Blickparameter kleinschrittiger unterteilen (z. B. Rayner 2009) oder auch den Vergleich des Blickverhaltens für einzelne Fragebogenelemente anstreben. Dabei könnten die Untersuchungen zusätzlich von der Aufzeichnung der Hirnaktivität während der Teilnahme profitieren, da dies tiefere Einblicke in Aufmerksamkeitsstrukturen verspricht (Duchowski 2017, S. 251) und so die Analyse der Hintergründe der Antwortqualität weiter schärfen kann.

Da der Fokus in der vorliegenden Arbeit auf Moduseffekten des Smartphones liegt, ist die Prüfung auf den Vergleich zum PC/Laptop begrenzt und legt die möglichst gleiche Bildschirmdarstellung als Vergleichsgrundlage an. Darstellungsunterschiede wirken in Befragungen unmittelbar auf das Antwortverhalten (siehe Kapitel 2.4.1), sodass zusätzlich künftig noch starker Forschungsbedarf für gerätübergreifende und gerätspezifische Darstellungsoptimierung besteht. Weiterführende Erkenntnisse könnten es Forscher*innen ermöglichen, die Befragung dynamisch an das Ausfüllgerät anzupassen, wenn sich je nach Gerät unterschiedliche Darstellungsmuster besser eignen (Antoun et al. 2018). Für Schieberegler z. B. wurde dabei schon gezeigt, dass sie als Antwortskalenformat speziell am Smartphone eher nicht funktional sind (Antoun et al. 2018; Funke 2016; Toepoel 2016, S. 126). Es wäre lohnend, die Erkenntnislage auszuweiten, indem die Gelingensbedingungen des Smartphoneeinsatzes in Onlinebefragungen ebenso wie mögliche Grenzbereiche aufgedeckt und dann austariert werden.

Die vorliegende Arbeit schließt Tablets als Ausfüllgeräte von der Betrachtung aus. Auch wenn Tablets als Ausfüllgeräte in der wissenschaftlichen Befragungspraxis eine untergeordnete Rolle spielen (Mavletova und Couper 2014; Poulová et al. 2019, S. 305; Schlosser und Silber 2020, S. 266), könnte der Einfluss als mobiles Ausfüllgerät eigene Moduseffekte erzeugen. Dies könnte separat vor allem bezüglich der Besonderheiten der Antwortqualität bei langen Texten sowie der Antwort auf offene Fragen untersucht werden, für die sich Moduseffekte des Smartphones vorliegend andeuten. Abseits davon könnte es ergiebig sein, bei der Stichprobenakquise auf reine Studierendensamples zu verzichten (vgl. Henrich et al. 2010; Peterson 2001). Für alle Studien des Forschungsprogramms könnte darüber hinaus eine weite Öffnung der Stichproben dabei helfen, die Befunde der vorliegenden Arbeit einerseits zu replizieren und sie andererseits verlässlich generalisieren zu können.

Das Potenzial des Smartphones als Erhebungsinstrument ist auch abseits klassischer Onlinebefragung beträchtlich (Miller 2012). Smartphones als alltägliche Begleiter ermöglichen die unmittelbare, kontextübergreifende Datenübermittlung, deren Datenreichtum durch die Sensorenvielfalt des Smartphones noch erweitert wird (z. B. Höhne et al. 2020a; Höhne und Schlosser 2019; Keusch et al. 2022; Mimura et al. 2015). Die Möglichkeiten reichen von Bewegungsdaten (z. B. GPS-Signale) über Paradata bis hin zu ganzheitlichen Tracking-Daten des menschlichen Organismus. Hier besteht noch weiterer Forschungsbedarf, um die Rahmenbedingungen des sinnvollen Smartphoneeinsatzes als wissenschaftliches Erhebungsinstrument zu eruieren. Aufgrund des persönlichen Charakters des Smartphones könnte es beispielsweise künftig auch für qualitative Textinterviews als Antwortgerät fungieren, unter Ausschluss des artifiziellen Charakters persönlicher Interviewsituationen (Conrad et al. 2017b, S. 314; Wakenhut und Wieseke 2017, S. 108-112).

Eine weitere, final denkbare Möglichkeit, wie Smartphones die Befragungsforschung enorm bereichern könnten, eröffnen Forschungs-Apps (Keusch und Conrad 2022; Roberts et al. 2022; Struminskaya et al. 2020a), die in der vorliegenden Arbeit ausgeklammert wurden. Grundsätzlich sollten Forschungs-Apps einen vertrauenswürdigen Charakter aufweisen (Martin und Shilton 2016; Revilla et al. 2016b; Sutanto et al. 2013) und niedrighschwellig genutzt werden können (Robinson et al. 2013; Wenz, Jäckle und Couper 2019). Hier könnten künftige Forschungsvorhaben die Gelingensbedingungen von Forschungs-Apps sowie die Nutzungsbereitschaft potenzieller Teilnehmer*innen adressieren (Struminskaya et al. 2020b) und diese jeweils möglichst auf die optimale Darstellung am Smartphone abstellen. Forschungs-Apps bieten entsprechend

der Erkenntnisse der vorliegenden Arbeit für zukünftige Forschungsvorhaben eine sehr interessante Möglichkeit, da die vorliegende Arbeit Moduseffekte des Smartphones, die dem Gerätkarakter entstammen, in Onlinebefragungen weitestgehend eher ausschließt.

6.4 Fazit

Die moderne wissenschaftliche Befragungsforschung ist aufgrund der dynamischen technologischen Entwicklung, die sowohl die Gestaltungsmöglichkeiten als auch die Diversität der Teilnahmeformate erhöht, mit stetig wachsenden Herausforderungen konfrontiert. Eine perfekte Befragung zu konzipieren, stellt sich als praktisch unmöglich dar, weshalb umfassende Einblicke in Gelingensbedingungen und potenzielle Fehlerquellen bei der Erstellung der Befragung benötigt werden (Bradburn und Sudman 2004, S. 40; MacKenzie und Podsakoff 2012). Das vorliegende Dissertationsprojekt zieht für die aktuelle Befragungspraxis das Fazit, dass Smartphones in Onlinebefragungen eher keine absoluten Moduseffekte erzeugen, die permanent den Total Survey Error erhöhen. Es kann konkludiert werden, dass das Smartphone ein valides Ausfüllgerät darstellt und die Aufmerksamkeitsleistung während der Teilnahme als Grundlage der Antwortqualität am Smartphone überwiegend adäquat ist. Die Befunde negieren nicht gänzlich die Besonderheiten des Smartphones als Gerät, sondern demonstrieren vielmehr, dass das Wissen um das Ausfüllgerät und dessen Eigenschaften gerade bei längeren Befragungen, die einen höheren kognitiven Aufwand verlangen, angezeigt ist und für die Datenauswertung sehr hilfreich sein kann. Für längere Textpassagen, ausführliche offene Fragen oder herausfordernde Gedächtnisaufgaben in Onlinebefragungen kann bei der Teilnahme am Smartphone ein Risiko für Reduktionen der Antwortqualität entstehen, sodass in diesen Fällen die statistische Kontrolle des Ausfüllgeräts stets empfehlenswert ist.

7 Literaturverzeichnis

- Abbey, James D. und Margaret G. Meloy (2017), „Attention by design: Using attention checks to detect inattentive respondents and improve data quality,” *Journal of Operations Management*, 53-56 (1), 63-70.
- Adler, Rachel F. und Raquel Benbunan-Fich (2012), „Juggling on a high wire: Multitasking effects on performance,” *International Journal of Human-Computer Studies*, 70 (2), 156-168.
- Ajzen, Icek (1985), „From intentions to actions: A theory of planned behavior,” in *Action control. From cognition to behavior. SSSP Springer Series in Social Psychology*, Julius Kuhl und Jürgen Beckmann, Hrsg. Berlin, Heidelberg: Springer, 11-39.
- Ajzen, Icek (1991), „The theory of planned behavior,” *Organizational Behavior and Human Decision Processes*, 50 (2), 179-211.
- Al Baghal, Tarek (2019), „The effect of online and mixed-mode measurement of cognitive ability,” *Social Science Computer Review*, 37 (1), 89-103.
- Alarcon, Gene M. und Michael A. Lee (2022), „The relationship of insufficient effort responding and response styles: An online experiment,” *Frontiers in Psychology*, 12, 784375.
- Alvarez, George A. und Patrick Cavanagh (2004), „The capacity of visual short-term memory is set both by visual information load and by number of objects,” *Psychological Science*, 15 (2), 106-111.
- Andreadis, Ioannis (2015a), „Comparison of response times between desktop and smartphone users,” in *Mobile research methods. Opportunities and challenges of mobile research methodologies*, Daniele Toninelli, Robert Pinter und Pablo de Pedraza, Hrsg. London: Ubiquity Press, 63-79.
- Andreadis, Ioannis (2015b), „Web surveys optimized for smartphones: Are there differences between computer and smartphone users?,” *methods, data, analyses*, 9 (2), 213-238.
- Anduiza, Eva und Carol Galais (2017), „Answering without reading: IMCs and strong satisficing in online surveys,” *International Journal of Public Opinion Research*, 29 (3), 497-519.
- Ansolabehere, Stephen und Brian F. Schaffner (2014), „Does survey mode still matter? Findings from a 2010 multi-mode comparison,” *Political Analysis*, 22 (3), 285-303.
- Ansolabehere, Stephen und Brian F. Schaffner (2015), „Distractions: The incidence and consequences of interruptions for survey respondents,” *Journal of Survey Statistics and Methodology*, 3 (2), 216-239.
- Antoun, Christopher (2015), „Who are the internet users, mobile Internet users, and mobile-mostly internet users?: Demographic differences across internet-use subgroups in the U.S.,” in *Mobile research methods. Opportunities and challenges of mobile research methodologies*, Daniele Toninelli, Robert Pinter und Pablo de Pedraza, Hrsg. London: Ubiquity Press, 99 - 117.
- Antoun, Christopher und Alexandru Cernat (2020), „Factors affecting completion times: A comparative analysis of smartphone and PC web surveys,” *Social Science Computer Review*, 38 (4), 477-489.
- Antoun, Christopher, Mick P. Couper und Frederick G. Conrad (2017), „Effects of mobile versus PC web on survey response quality,” *Public Opinion Quarterly*, 81 (1), 280-306.
- Antoun, Christopher, Jonathan Katz, Josef Argueta und Lin Wang (2018), „Design heuristics for effective smartphone questionnaires,” *Social Science Computer Review*, 36 (5), 557-574.
- Aquilino, William S. (1993), „Effects of spouse presence during the interview on survey responses concerning marriage,” *Public Opinion Quarterly*, 57 (3), 358.
- Aquilino, William S., Debra L. Wright und Andrew J. Supple (2000), „Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use,” *Substance Use & Misuse*, 35 (6-8), 845-867.
- Arias, Víctor B., Luis E. Garrido, Cristina Jenaro, Agustín Martínez-Molina und Benito Arias (2020), „A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data,” *Behavior Research Methods*, 52 (6), 2489-2505.

- Arn, Birgit, Stefan Klug und Janusz Kolodziejcki (2015), „Evaluation of an adapted design in a multi-device online panel: A DemoSCOPE case study,” *methods, data, analyses*, 9 (2), 185-212.
- Arndt, Aaron D., John B. Ford, Barry J. Babin und Vinh Luong (2022), „Collecting samples from online services: How to use screeners to improve data quality,” *International Journal of Research in Marketing*, 39 (1), 117-133.
- Arthur, Winfred, Ellen Hagen und Felix George (2021), „The lazy or dishonest respondent: Detection and prevention,” *Annual Review of Organizational Psychology and Organizational Behavior*, 8 (1), 105-137.
- Augner, Christoph und Gerhard W. Hacker (2012), „Associations between problematic mobile phone use and psychological parameters in young adults,” *International Journal of Public Health*, 57 (2), 437-441.
- Aust, Frederik, Birk Diedenhofen, Sebastian Ullrich und Jochen Musch (2013), „Seriousness checks are useful to improve data validity in online research,” *Behavior Research Methods*, 45 (2), 527-535.
- Axinn, William G., Heather H. Gatny und James Wagner (2015), „Maximizing data quality using mode switching in mixed-device survey design: Nonresponse bias and models of demographic behavior,” *methods, data, analyses*, 9 (2), 163-184.
- Babakhani, Nazila, Leonard J. Paas und Sara Dolnicar (2022), „Do instructional manipulation checks measure inattention or miscomprehension?,” *International Journal of Social Research Methodology*, 1-12.
- Backhaus, Klaus, Bernd Erichson, Sonja Gensler, Rolf Weiber und Thomas Weiber (2021), *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Springer eBook Collection*. Wiesbaden: Springer Gabler.
- Bäckström, Martin und Fredrik Björklund (2019), „Is reliability compromised towards the end of long personality inventories?,” *European Journal of Psychological Assessment*, 35 (1), 14-21.
- Baddeley, Alan D. (1992), „Working memory: The interface between memory and cognition,” *Journal of Cognitive Neuroscience*, 4 (3), 281-288.
- Baddeley, Alan D. (2000), „The episodic buffer: A new component of working memory?,” *Trends in Cognitive Sciences*, 4 (11), 417-423.
- Baddeley, Alan D. (2003), „Working memory: Looking back and looking forward,” *Nature reviews. Neuroscience*, 4 (10), 829-839.
- Baddeley, Alan D. und Graham Hitch (1974), „Working memory,” in *Psychology of learning and motivation. Advances in research and theory. Psychology of Learning and Motivation*, Nr. 8, Gordon H. Bower, Hrsg. Burlington: Elsevier, 47-89.
- Bandura, Albert (2012), *Self-efficacy. The exercise of control*. New York City: Freeman.
- Barnett, Julie (1998), „Sensitive questions and response effects: An evaluation,” *Journal of Managerial Psychology*, 13 (1/2), 63-76.
- Barnette, J. Jackson (2000), „Effects of stem and likert Response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems,” *Educational and Psychological Measurement*, 60 (3), 361-370.
- Barr, Nathaniel, Gordon Pennycook, Jennifer A. Stolz und Jonathan A. Fugelsang (2015), „The brain in your pocket: Evidence that smartphones are used to supplant thinking,” *Computers in Human Behavior*, 48, 473-480.
- Baumgartner, Hans und Jan-Benedict E. Steenkamp (2001), „Response styles in marketing research: A cross-national investigation,” *Journal of Marketing Research*, 38 (2), 143-156.
- Baumgartner, Hans und Bert Weijters (2012), „Commentary on “Common Method Bias in marketing: Causes, mechanisms, and procedural remedies”” *Journal of Retailing*, 88 (4), 563-566.
- Baumgartner, Hans und Bert Weijters (2019), „Measurement in marketing,” *Foundations and Trends in Marketing*, 12 (4), 278-400.

- Baumgartner, Hans und Bert Weijters (2022), „How to identify careless responders in surveys,” in *Measurement in marketing. Review of Marketing Research*, Nr. 19, Hans Baumgartner und Bert Weijters, Hrsg. Bingley: Emerald Publishing Limited, 121-141.
- Baumgartner, Hans, Bert Weijters und Rik Pieters (2018), „Misresponse to survey questions: A conceptual framework and empirical test of the effects of reversals, negations, and polar opposite core concepts,” *Journal of Marketing Research*, 55 (6), 869-883.
- Bautista, René (2012), „An overlooked approach in survey research: Total Survey Error,” in *Handbook of survey methodology for the social sciences*, Lior Gideon, Hrsg. New York City: Springer New York, 37-49.
- Beach, Daniel A. (1989), „Identifying the random responder,” *The Journal of Psychology*, 123 (1), 101-103.
- Beaujean, A. Alexander und Morgan B. Grant (2016), „Tutorial on using regression models with count outcomes using R,” *Practical Assessment, Research, and Evaluation*, 21 (1-4), 1-18.
- Becker, Rolf und Ralph Günther (2004), „Selektives Antwortverhalten bei Fragen zum delinquenten Handeln. Eine empirische Studie über die Wirksamkeit der ‚sealed envelope technique‘ bei selbst berichteter Delinquenz mit Daten des ALLBUS 2000,” *ZUMA-Nachrichten*, 28 (54), 39-59.
- Beggs, James K. (1992), „On the social nature of nonsocial perception: The mere ownership effect,” *Journal of Personality and Social Psychology*, 62 (2), 229-237.
- Belson, William A. (1966), „The effects of reversing the presentation order of verbal rating scales,” *Journal of Advertising Research*, 6, 30-37.
- Bentler, Peter M., Douglas N. Jackson und Samuel Messick (1971), „Identification of content and style: A two-dimensional interpretation of acquiescence,” *Psychological Bulletin*, 76 (3), 186-204.
- Berinsky, Adam J., Michele F. Margolis und Michael W. Sances (2014), „Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys,” *American Journal of Political Science*, 58 (3), 739-753.
- Berry, David T. R., Martha W. Wetter, Ruth A. Baer, Lene Larsen, Cynthia Clark und Keith Monroe (1992), „MMPI-2 random responding indices: Validation using a self-report methodology,” *Psychological Assessment*, 4 (3), 340-345.
- Berry, Katherine, Rusham Rana, Austin Lockwood, Lee Fletcher und Daniel Pratt (2019), „Factors associated with inattentive responding in online survey research,” *Personality and Individual Differences*, 149, 157-159.
- Bethlehem, Jelke (2015), „On the quality of web panels,” in *Survey measurements. Techniques, data quality and sources of error*, Uwe Engel, Hrsg. Frankfurt am Main: Campus-Verlag, 112-129.
- Bhattacharya, Sudip, Md Abu Bashar, Abhay Srivastava und Amarjeet Singh (2019), „NOMOPHOBIA: NO MOBILE PHONE PHOBIA,” *Journal of Family Medicine and Primary Care*, 8 (4), 1297-1300.
- Biemer, Paul P. (2010), „Total Survey Error: Design, implementation, and evaluation,” *Public Opinion Quarterly*, 74 (5), 817-848.
- Billieux, Joel (2012), „Problematic use of the mobile phone: A literature review and a pathways model,” *Current Psychiatry Reviews*, 8 (4), 299-307.
- Blondel, Vincent D., Adeline Decuyper und Gautier Krings (2015), „A survey of results on mobile phone datasets analysis,” *EPJ Data Science*, 4 (1), 1-55.
- Bock, Beth C., Ryan Lantini, Herpreet Thind, Kristen Walaska, Rochelle K. Rosen, Joseph L. Fava, Nancy P. Barnett und Lori A. Scott-Sheldon (2016a), „The mobile phone affinity scale: Enhancement and refinement,” *JMIR mHealth and uHealth*, 4 (4), e134.
- Bock, Beth C., Herpreet Thind, Joseph L. Fava, Kristen Walaska, Nancy P. Barnett, Rochelle K. Rosen, Regina Traficante und Ryan Lantini (2016b), „Development of the mobile phone attachment scale,” in *Proceedings of the 49th annual Hawaii international conference on system sciences. 5-8 January 2016, Kauai, Hawaii*, Tung X. Bui und Ralph H. Sprague, Hrsg. Piscataway, NJ: IEEE, 3401-3407.

- Boer, Diana (2014), „SSVS-G. Short Schwartz's Value Survey - German,” in *Psychologische und sozialwissenschaftliche Kurzskalen. Standardisierte Erhebungsinstrumente für Wissenschaft und Praxis*, Elmar Brähler, Christoph J. Kemper und Markus Zenger, Hrsg. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 299-302.
- Booth-Kewley, Stephanie, Jack E. Edwards und Paul Rosenfeld (1992), „Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference?,” *Journal of Applied Psychology*, 77 (4), 562-566.
- Bortz, Jürgen und Christof Schuster (2016), *Statistik für Human- und Sozialwissenschaftler. Extras online. Springer-Lehrbuch*. Berlin, Heidelberg: Springer.
- Bosnjak, Michael, Robert Bauer und Kai W. Weyandt (2017), „Mixed devices in online surveys: Prevalence, determinants, and consequences,” in *Mobile Research. Grundlagen und Zukunftsaussichten für die Mobile Marktforschung*, Axel Theobald, Hrsg. Wiesbaden: Springer Fachmedien, 53-65.
- Bosnjak, Michael und Tracy L. Tuten (2001), „Classifying response behaviors in web-based surveys,” *Journal of Computer-Mediated Communication*, 6 (3), JMC636.
- Bouman, Thijs, Linda Steg und Henk A. L. Kiers (2018), „Measuring values in environmental research: A test of an environmental portrait value questionnaire,” *Frontiers in Psychology*, 9, 564.
- Bowling, Ann (2005), „Mode of questionnaire administration can have serious effects on data quality,” *Journal of Public Health*, 27 (3), 281-291.
- Bowling, Nathan A., Anthony M. Gibson, Joseph W. Houpt und Cheyna K. Brower (2021a), „Will the questions ever end? Person-level increases in careless responding during questionnaire completion,” *Organizational Research Methods*, 24 (4), 718-738.
- Bowling, Nathan A. und Jason L. Huang (2018), „Your attention please! Toward a better understanding of research participant carelessness,” *Applied Psychology*, 67 (2), 227-230.
- Bowling, Nathan A., Jason L. Huang, Caleb B. Bragg, Steve Khazon, Mengqiao Liu und Caitlin E. Blackmore (2016), „Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality,” *Journal of Personality and Social Psychology*, 111 (2), 218-229.
- Bowling, Nathan A., Jason L. Huang, Cheyna K. Brower und Caleb B. Bragg (2021b), „The quick and the careless: The construct validity of page time as a measure of insufficient effort responding to surveys,” *Organizational Research Methods*, 26 (2), 323-352.
- Bradburn, Norman M. und Seymour Sudman (2004), „The current status of questionnaire research,” in *Measurement errors in surveys. Wiley Series in Probability and Statistics*, Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz und Seymour Sudman, Hrsg. Hoboken: Wiley-Interscience, 27-40.
- Brandimarte, Laura, Alessandro Acquisti und George Loewenstein (2013), „Misplaced confidences: Privacy and the control paradox,” *Social Psychological and Personality Science*, 4 (3), 340-347.
- Brasel, S. Adam und James Gips (2014), „Tablets, touchscreens, and touchpads: How varying touch interfaces trigger psychological ownership and endowment,” *Journal of Consumer Psychology*, 24 (2), 226-233.
- Breitsohl, Heiko und Corinna Steidelmüller (2018), „The impact of insufficient effort responding detection methods on substantive responses: Results from an experiment testing parameter invariance,” *Applied Psychology*, 67 (2), 284-308.
- Brick, J. Michael und Douglas Williams (2013), „Explaining rising nonresponse rates in cross-sectional surveys,” *The ANNALS of the American Academy of Political and Social Science*, 645 (1), 36-59.
- Britton, Bruce K. (1980), „Use of cognitive capacity in reading: Effects of processing information from text for immediate recall and retention,” *Journal of Reading Behavior*, 12 (2), 129-137.
- Brooke, John (1996), „SUS: A 'quick and dirty' usability scale,” in *Usability evaluation in industry*, Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester und Ian L. McClelland, Hrsg. Boca Raton: CRC Press, an imprint of Taylor and Francis, 189-194.

- Brower, Cheyna K. (2020), „What are you looking at? Using eye-tracking to provide insight into careless responding,” Dissertation, Wright State University, Dayton.
- Brühlmann, Florian, Serge Petralito, Lena F. Aeschbach und Klaus Opwis (2020), „The quality of data collected online: An investigation of careless responding in a crowdsourced sample,” *Methods in Psychology*, 2, 100022.
- Buchanan, Elizabeth A. und Erin E. Hvizdak (2009), „Online survey tools: Ethical and methodological concerns of human research ethics committees,” *Journal of Empirical Research on Human Research Ethics*, 4 (2), 37-48.
- Buchanan, Tom (2000), „Potential of the internet for personality research,” in *Psychological experiments on the internet*, Michael H. Birnbaum, Hrsg. San Diego: Academic Press, 121-140.
- Bühner, Markus und Matthias Ziegler (2017), *Statistik für Psychologen und Sozialwissenschaftler. Studium Psychologie*. Hallbergmoos: Pearson Deutschland GmbH.
- Burns, Gary N., Neil D. Christiansen, Megan B. Morris, David A. Periard und John A. Coaster (2014), „Effects of applicant personality on resume evaluations,” *Journal of Business and Psychology*, 29 (4), 573-591.
- Buskirk, Trent D. und Charles H. Andrus (2012), „Smart surveys for smart phones: Exploring various approaches for conducting online mobile surveys via smartphones,” *Survey Practice*, 5 (1), 1-11.
- Buskirk, Trent D. und Charles H. Andrus (2014), „Making mobile browser surveys smarter: Results from a randomized experiment comparing online surveys completed via computer or smartphone,” *Field Methods*, 26 (4), 322-342.
- Buskirk, Trent D., Ted Saunders und Joey Michaud (2015), „Are sliders too slick for surveys? An experiment comparing slider and radio button scales for smartphone, tablet and computer based surveys,” *methods, data, analyses*, 9 (2), 229-260.
- Busse, Britta und Marek Fuchs (2012), „The components of landline telephone survey coverage bias. The relative importance of no-phone and mobile-only populations,” *Quality and Quantity*, 46 (4), 1209-1225.
- Callegaro, Mario (2013), „From mixed-mode to multiple devices: Web surveys, smartphone surveys and apps: Has the respondent gone ahead of us in answering surveys?,” *International Journal of Market Research*, 55 (2), 317-320.
- Campbell, Donald T. und Julian C. Stanley (1963), *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin
- Campbell, Scott W. (2007), „Perceptions of mobile phone use in public settings: A cross-cultural comparison,” *International Journal of Communication*, 1, 738-757.
- Campbell, Scott W. und Yong J. Park (2008), „Social implications of mobile telephony: The rise of personal communication society,” *Sociology Compass*, 2 (2), 371-387.
- Caplan, David und Gloria S. Waters (1999), „Verbal working memory and sentence comprehension,” *Behavioral and Brain Sciences*, 22 (1), 77-94.
- Carrier, L. Mark, Nancy A. Cheever, Larry D. Rosen, Sandra Benitez und Jennifer Chang (2009), „Multitasking across generations: Multitasking choices and difficulty ratings in three generations of Americans,” *Computers in Human Behavior*, 25 (2), 483-489.
- Carter, Michelle und Varun Grover (2015), „Me, my self, and I (T),” *MIS Quarterly*, 39 (4), 931-958.
- Carter, Michelle, Varun Grover und Jason B. Thatcher (2016), „Mobile phone identity,” in *Encyclopedia of e-commerce development, implementation, and management*, In Lee, Hrsg. Hershey: Business Science Reference, 1435-1447.
- Cattell, Raymond B. (1966), „The scree test for the number of factors,” *Multivariate Behavioral Research*, 1 (2), 245-276.

- Chae, Minhee und Jinwoo Kim (2004), „Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones,” *Behaviour & Information Technology*, 23 (3), 165-181.
- Chen, Henian, Patricia Cohen und Sophie Chen (2010), „How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological Studies,” *Communications in Statistics - Simulation and Computation*, 39 (4), 860-864.
- Chen, Li-Qun, Xing Xie, Xin Fan, Wei-Ying Ma, Hong-Jiang Zhang und He-Qin Zhou (2003), „A visual attention model for adapting images on small displays,” *Multimedia Systems*, 9 (4), 353-364.
- Chen, Ping-Ling und Chih-Wei Pai (2018), „Pedestrian smartphone overuse and inattentive blindness: An observational study in Taipei, Taiwan,” *BMC Public Health*, 18 (1), 1342.
- Cheung, Janelle H., Deanna K. Burns, Robert R. Sinclair und Michael Sliter (2017), „Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations,” *Journal of Business and Psychology*, 32 (4), 347-361.
- Ciotti, Marco, Massimo Ciccozzi, Alessandro Terrinoni, Wen-Can Jiang, Cheng-Bin Wang und Sergio Bernardini (2020), „The COVID-19 pandemic,” *Critical Reviews in Clinical Laboratory Sciences*, 57 (6), 365-388.
- Clark, Michael E., Ronald J. Girona und Robert W. Young (2003), „Detection of back random responding: Effectiveness of MMPI-2 and Personality Assessment Inventory validity indices,” *Psychological Assessment*, 15 (2), 223-234.
- Clement, Sanne L., Majbritt C. Severin und Ditte Shamshiri-Petersen (2020), „Device effects on survey response quality. A comparison of smartphone, tablet and PC responses on a cross sectional probability sample,” *Survey Methods: Insights from the field*, Special Issue: ‘Advancements in Online and Mobile Survey Methods’, 1-17.
- Clifford, Scott und Jennifer Jerit (2015), „Do attempts to improve respondent attention increase social desirability bias?,” *Public Opinion Quarterly*, 79 (3), 790-802.
- Conijn, Judith M., Gunhild Franz, Wilco H. M. Emons, Edwin de Beurs und Ingrid V. E. Carlier (2019), „The assessment and impact of careless responding in routine outcome monitoring within mental health care,” *Multivariate Behavioral Research*, 54 (4), 593-611.
- Conrad, Frederick G., Mick P. Couper, Roger Tourangeau und Chan Zhang (2017a), „Reducing speeding in web surveys by providing immediate feedback,” *Survey Research Methods*, 11 (1), 45-61.
- Conrad, Frederick G., Michael F. Schober, Christopher Antoun, Andrew L. Hupp und H. Yanna Yan (2017b), „Text Interviews on Mobile Devices,” in *Total Survey Error in practice*, Paul P. Biemer, Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker und Brady T. West, Hrsg. Hoboken: John Wiley & Sons, 299-318.
- Cook, William A. (2014), „Is mobile a reliable platform for survey taking?,” *Journal of Advertising Research*, 54 (2), 141-148.
- Costa, Paul T. und Robert R. McCrae (2008), „The Revised NEO Personality Inventory (NEO-PI-R),” in *Personality measurement and testing. The SAGE handbook of personality theory and assessment*, Nr. 2, Gregory J. Boyle, Gerald Matthews und Donald Saklofske, Hrsg. Los Angeles: Sage, 179-198.
- Cottrill, Caitlin D., Francisco C. Pereira, Fang Zhao, Inês F. Dias, Hock B. Lim, Moshe E. Ben-Akiva und P. C. Zegras (2013), „Future mobility survey: Experience in developing a smartphone-based travel survey in Singapore,” *Transportation Research Record: Journal of the Transportation Research Board*, 2354 (1), 59-67.
- Couper, Mick P. (2000), „Review: Web surveys: A review of issues and approaches,” *Public Opinion Quarterly*, 64 (4), 464-494.
- Couper, Mick P. (2005), „Technology trends in survey data collection,” *Social Science Computer Review*, 23 (4), 486-501.

- Couper, Mick P. (2008), „Technology and the survey interview/questionnaire,” in *Envisioning the survey interview of the future. Wiley series in survey methodology*, Frederick G. Conrad und Michael F. Schober, Hrsg. Hoboken: Wiley-Interscience, 58-76.
- Couper, Mick P. (2013), „Is the sky falling? New technology, changing media, and the future of surveys,” *Survey Research Methods*, 7 (3), 145-156.
- Couper, Mick P., Christopher Antoun und Aigul Mavletova (2017), „Mobile web surveys,” in *Total Survey Error in practice*, Paul P. Biemer, Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker und Brady T. West, Hrsg. Hoboken: John Wiley & Sons, 133-154.
- Couper, Mick P., Garret Gremel, William G. Axinn, Heidi Guyer, James Wagner und Brady T. West (2018), „New options for national population surveys: The implications of internet and smartphone coverage,” *Social Science Research*, 73, 221-235.
- Couper, Mick P. und Gregg J. Peterson (2017), „Why do web surveys take longer on smartphones?,” *Social Science Computer Review*, 35 (3), 357-377.
- Couper, Mick P., Roger Tourangeau, Frederick G. Conrad und Scott D. Crawford (2004), „What they see is what we get: Response options for web surveys,” *Social Science Computer Review*, 22 (1), 111-127.
- Couper, Mick P., Michael W. Traugott und Mark J. Lamias (2001), „Web survey design and administration,” *Public Opinion Quarterly*, 65 (2), 230-253.
- Credé, Marcus (2010), „Random responding as a threat to the validity of effect size estimates in correlational research,” *Educational and Psychological Measurement*, 70 (4), 596-612.
- Cunningham, Sheila J., David J. Turk, Lynda M. Macdonald und C. Neil Macrae (2008), „Yours or mine? Ownership and memory,” *Consciousness and Cognition*, 17 (1), 312-318.
- Curran, Paul G. (2016), „Methods for the detection of carelessly invalid responses in survey data,” *Journal of Experimental Social Psychology*, 66, 4-19.
- Curran, Paul G. und Kelsey A. Hauser (2019), „I’m paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items,” *Journal of Research in Personality*, 82, 103849.
- Daneman, Meredyth und Patricia A. Carpenter (1980), „Individual differences in working memory and reading,” *Journal of Verbal Learning and Verbal Behavior*, 19 (4), 450-466.
- Danner, Daniel, Beatrice Rammstedt, Matthias Bluemke, Clemens Lechner, Sabrina Berres, Thomas Knopf, Christopher J. Soto und Oliver P. John (2019), „Das Big Five Inventar 2,” *Diagnostica*, 65 (3), 121-132.
- Danner, Daniel, Beatrice Rammstedt, Matthias Bluemke, Lisa Treiber, Sabrina Berres, Christopher J. Soto und Oliver P. John (2016), „Die deutsche Version des Big Five Inventory 2 (BFI-2),” *Zusammenstellung sozialwissenschaftlicher Items und Skalen*, ZIS - GESIS Leibniz-Institut für Sozialwissenschaften, Mannheim.
- Davel, Coriena (2017), „The mobile phone as an extension of the self: A study among adolescents in a secondary school,” Dissertation, University of South Africa, Pretoria.
- Davis, Fred D. (1989), „Perceived usefulness, perceived ease of use, and user acceptance of information technology,” *MIS Quarterly*, 13 (3), 319-340.
- Davis, Fred D. (1993), „User acceptance of information technology: System characteristics, user perceptions and behavioral impacts,” *International Journal of Man-Machine Studies*, 38 (3), 475-487.
- Davis, Fred D., Richard P. Bagozzi und Paul R. Warshaw (1989), „User acceptance of computer technology: A comparison of two theoretical models,” *Management Science*, 35 (8), 982-1003.
- de Bruijne, Marika (2015), *Designing web surveys for the multi-device internet. PhD thesis series / Center for Economic Research, Tilburg University*, Nr. 456. Tilburg: Tilburg University.

- de Bruijne, Marika und Marije Oudejans (2015), „Online surveys and the burden of mobile responding,” in *Survey measurements. Techniques, data quality and sources of error*, Uwe Engel, Hrsg. Frankfurt am Main: Campus-Verlag, 130-145.
- de Bruijne, Marika und Arnaud Wijnant (2013), „Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computer-assisted web survey,” *Social Science Computer Review*, 31 (4), 482-504.
- de Bruijne, Marika und Arnaud Wijnant (2014), „Mobile response in web panels,” *Social Science Computer Review*, 32 (6), 728-742.
- de Leeuw, Edith D. (2005), „To mix or not to mix data collection modes in surveys,” *Journal of Official Statistics*, 21 (5), 233-255.
- de Leeuw, Edith D. und Joop J. Hox (2015), „Survey mode and mode effects,” in *Survey measurements. Techniques, data quality and sources of error*, Uwe Engel, Hrsg. Frankfurt am Main: Campus-Verlag, 22-34.
- de Maesschalck, Roy, Delphine Jouan-Rimbaud und Désiré L. Massart (2000), „The Mahalanobis distance,” *Chemometrics and Intelligent Laboratory Systems*, 50 (1), 1-18.
- de Reuver, Mark, Shahrokh Nikou und Harry Bouwman (2016), „Domestication of smartphones and mobile applications: A quantitative mixed-method study,” *Mobile Media & Communication*, 4 (3), 347-370.
- Deng, Tao, Shaheen Kanthawala, Jingbo Meng, Wei Peng, Anastasia Kononova, Qi Hao, Qin hao Zhang und Prabu David (2019), „Measuring smartphone usage and task switching with log tracking and self-reports,” *Mobile Media & Communication*, 7 (1), 3-23.
- Denison, Alexander J. und Brenton M. Wiernik (2022), „Careless response processes are heterogeneous: Comment on Goldammer et al. (2020),” *Meta-Psychology*, 6, 2637.
- DeSimone, Justin A., H. Kristl Davison, Jeremy L. Schoen und Mark N. Bing (2020), „Insufficient effort responding as a partial function of implicit aggression,” *Organizational Research Methods*, 23 (1), 154-180.
- DeSimone, Justin A., Alice J. DeSimone, Peter D. Harms und Dustin Wood (2018), „The differential impacts of two forms of insufficient effort responding,” *Applied Psychology*, 67 (2), 309-338.
- DeSimone, Justin A. und Peter D. Harms (2018), „Dirty data: The effects of screening respondents who provide low-quality data in survey research,” *Journal of Business and Psychology*, 33 (5), 559-577.
- DeSimone, Justin A., Peter D. Harms und Alice J. DeSimone (2015), „Best practice recommendations for data screening,” *Journal of Organizational Behavior*, 36 (2), 171-181.
- DeStatis (2021), „Private Haushalte in der Informationsgesellschaft – Nutzung von Informations- und Kommunikationstechnologien 2020,” (Zugriff am September 28, 2023), [verfügbar unter https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/IT-Nutzung/Publikationen/Downloads-IT-Nutzung/private-haushalte-ikt-2150400207004.pdf?__blob=publicationFile].
- Deutskens, Elisabeth, Ko de Ruyter, Martin Wetzels und Paul Oosterveld (2004), „Response rate and response quality of internet-based surveys: An experimental study,” *Marketing Letters*, 15 (1), 21-36.
- DeviceAtlas (2019), „Viewport, resolution, diagonal screen size and DPI for the most popular smartphones,” (Zugriff am September 28, 2023), [verfügbar unter <https://deviceatlas.com/blog/viewport-resolution-diagonal-screen-size-and-dpi-most-popular-smartphones>].
- Díaz de Rada, Vidal und Juan A. Domínguez-Álvarez (2014), „Response quality of self-administered questionnaires: A comparison between paper and web questionnaires,” *Social Science Computer Review*, 32 (2), 256-269.
- Dillon, Andrew, John Richardson und Cliff McKnight (1990), „The effects of display size and text splitting on reading lengthy text from screen,” *Behaviour & Information Technology*, 9 (3), 215-227.

- Dolnicar, Sara, Christian Laesser und Katrina Matus (2009), „Online versus paper: Format effects in tourism surveys,” *Journal of Travel Research*, 47 (3), 295-316.
- Doty, D. Harold und William H. Glick (1998), „Common Methods Bias: Does common methods variance really bias results?,” *Organizational Research Methods*, 1 (4), 374-406.
- Douglas, Susan P. und C. Samuel Craig (2007), „Collaborative and iterative translation: An alternative approach to back translation,” *Journal of International Marketing*, 15 (1), 30-43.
- Drew, Trafton, Sage E. P. Boettcher und Jeremy M. Wolfe (2017), „One visual search, many memory searches: An eye-tracking investigation of hybrid search,” *Journal of Vision*, 17 (11), 5.
- Drewes, Frank (2014), „An empirical test of the impact of smartphones on panel-based online data collection,” in *Online panel research. A data quality perspective*. Wiley Online Library, Mario Callegaro, Hrsg. Chichester: Wiley, 367-386.
- Drouin, Michelle, Daren H. Kaiser und Daniel A. Miller (2012), „Phantom vibrations among undergraduates: Prevalence and associated psychological characteristics,” *Computers in Human Behavior*, 28 (4), 1490-1496.
- Duchnicky, Robert L. und Paul A. Kolers (1983), „Readability of text scrolled on visual display terminals as a function of window size,” *Human Factors*, 25 (6), 683-692.
- Duchowski, Andrew T. (2017), *Eye tracking methodology. Theory and practice*. Cham: Springer.
- Dufau, Stephane, Jon A. Duñabeitia, Carmen Moret-Tatay, Aileen McGonigal, David Peeters, F-Xavier Alario, David A. Balota, Marc Brysbaert, Manuel Carreiras, Ludovic Ferrand, Maria Ktori, Manuel Perea, Kathy Rastle, Olivier Sasburg, Melvin J. Yap, Johannes C. Ziegler und Jonathan Grainger (2011), „Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science,” *PLOS ONE*, 6 (9), e24974.
- Duffy, Bobby, Kate Smith, George Terhanian und John Bremer (2005), „Comparing data from online and face-to-face surveys,” *International Journal of Market Research*, 47 (6), 615-639.
- Dunaway, Johanna, Kathleen Searles, Mingxiao Sui und Newly Paul (2018), „News attention in a mobile era,” *Journal of Computer-Mediated Communication*, 23 (2), 107-124.
- Dunn, Alexandra M., Eric D. Heggestad, Linda R. Shanock und Nels Theilgard (2018), „Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences,” *Journal of Business and Psychology*, 33 (1), 105-121.
- Eastman, Jacqueline K., Rajesh Iyer, Sara Liao-Troth, Donnie F. Williams und Mitch Griffin (2014), „The role of involvement on millennials’ mobile technology behaviors: The moderating impact of status consumption, innovation, and opinion leadership,” *Journal of Marketing Theory and Practice*, 22 (4), 455-470.
- Eberl, Markus (2016), „Shorter Smarter Surveys,” in *Marktforschung der Zukunft - Mensch oder Maschine? Bewährte Kompetenzen in neuem Kontext*, Bernhard Keller, Hans-Werner Klein und Stefan Tuschl, Hrsg. Wiesbaden: Springer Gabler, 217-230.
- Edwards, Jeffrey R. (2019), „Response invalidity in empirical research: Causes, detection, and remedies,” *Journal of Operations Management*, 65 (1), 62-76.
- Eisinga, Rob, Manfred t. Grotenhuis und Ben Pelzer (2013), „The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?,” *International Journal of Public Health*, 58 (4), 637-642.
- Embretson, Susan E. und Steven P. Reise (2013), *Item response theory for psychologists. Multivariate applications book series*. New York: Psychology Press.
- Engel, Uwe (2015), „Introduction,” in *Survey measurements. Techniques, data quality and sources of error*, Uwe Engel, Hrsg. Frankfurt am Main: Campus-Verlag, 9-23.
- Erens, Bob, Debbie Collins, Tommaso Manacorda, Jennifer Gosling, Nicholas Mays, David Reid und William Taylor (2019), „Comparing data quality from personal computers and mobile devices in an online survey among professionals,” *Social Research Practice* (7), 15-26.

- Evans, Joel R. und Anil Mathur (2018), „The value of online surveys: A look back and a look ahead,” *Internet Research*, 28 (4), 854-887.
- Faul, Franz, Edgar Erdfelder, Axel Buchner und Albert-Georg Lang (2009), „Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses,” *Behavior Research Methods*, 41 (4), 1149-1160.
- Faul, Franz, Edgar Erdfelder, Albert-Georg Lang und Axel Buchner (2007), „G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences,” *Behavior Research Methods*, 39 (2), 175-191.
- Faulbaum, Frank (2022), „Total Survey Error,” in *Handbuch Methoden der empirischen Sozialforschung. Handbuch*, Nina Baur und Jörg Blasius, Hrsg. Wiesbaden: Springer VS, 567-584.
- Featherman, Mauricio S. und Paul A. Pavlou (2003), „Predicting e-services adoption: A perceived risk facets perspective,” *International Journal of Human-Computer Studies*, 59 (4), 451-474.
- Field, Andy (2017), *Discovering statistics using IBM SPSS statistics*. London: SAGE Publications Ltd.
- Fischer, Josef A., Hendrik Hüttermann und Simon Werther (2021), „Employee Experience Questionnaire (EXQ): Fragebogen zur Messung von Zufriedenheit, Commitment und Engagement,” *Zusammenstellung sozialwissenschaftlicher Items und Skalen*, ZIS - GESIS Leibniz-Institut für Sozialwissenschaften, Mannheim.
- Fleischer, Avi, Alan D. Mead und Jialin Huang (2015), „Inattentive responding in MTurk and other online samples,” *Industrial and Organizational Psychology*, 8 (2), 196-202.
- Fredrickson, Barbara L. (2001), „The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions,” *American Psychologist*, 56 (3), 218-226.
- Fricker, Scott, Mirta Galesic, Roger Tourangeau und Ting Yan (2005), „An experimental comparison of web and telephone surveys,” *Public Opinion Quarterly*, 69 (3), 370-392.
- Friedman, Hershey H. und Taiwo Amoo (1999), „Rating the rating scales,” *The Journal of Marketing Management*, 9 (3), 114-123.
- Fuller, Wayne A. (2004), „Regression estimation in the presence of measurement error,” in *Measurement errors in surveys. Wiley Series in Probability and Statistics*, Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz und Seymour Sudman, Hrsg. Hoboken: Wiley-Interscience, 617-635.
- Fullwood, Chris, Sally Quinn, Linda K. Kaye und Charlotte Redding (2017), „My virtual friend: A qualitative analysis of the attitudes and experiences of smartphone users: Implications for smartphone attachment,” *Computers in Human Behavior*, 75, 347-355.
- Funke, Frederik (2016), „A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales,” *Social Science Computer Review*, 34 (2), 244-254.
- Gabaix, Xavier (2019), „Behavioral inattention,” in *Handbook of behavioral economics. Foundations and applications. Handbooks in economics*, B. Douglas Bernheim, Stefano Della Vigna und David I. Laibson, Hrsg. Amsterdam, Oxford: North-Holland an imprint of Elsevier, 261-343.
- Galesic, Mirta (2006), „Dropouts on the web: Effects of interest and burden experienced during an online survey,” *Journal of Official Statistics*, 22 (2), 313-328.
- Galesic, Mirta und Michael Bosnjak (2009), „Effects of questionnaire length on participation and indicators of response quality in a web survey,” *Public Opinion Quarterly*, 73 (2), 349-360.
- Galesic, Mirta, Roger Tourangeau, Mick P. Couper und Frederick G. Conrad (2008), „Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding,” *Public Opinion Quarterly*, 72 (5), 892-913.
- Gatersleben, Birgitta, Niamh Murtagh und Wokje Abrahamse (2014), „Values, identity and pro-environmental behaviour,” *Contemporary Social Science*, 9 (4), 374-392.
- Geiger, Sonja und Brigitte Holzhauer (2020), *Weiterentwicklung einer Skala zur Messung von zentralen Kenngrößen des Umweltbewusstseins*. Dessau-Roßlau: Umweltbundesamt.

- Gerich, Joachim (2009), „Multimediale Elemente in der Computerbasierten Datenerhebung,“ in *Umfrageforschung. Herausforderungen und Grenzen. Österreichische Zeitschrift für Soziologie Sonderheft*, Nr. 9, Martin Weichbold, Hrsg. Wiesbaden: VS Verlag für Sozialwissenschaften, 107-129.
- Ghose, Anindya, Avi Goldfarb und Sang P. Han (2013), „How is the mobile internet different? Search costs and local activities,“ *Information Systems Research*, 24 (3), 613-631.
- Gibson, Anthony M. und Nathan A. Bowling (2020), „The effects of questionnaire length and behavioral consequences on careless responding,“ *European Journal of Psychological Assessment*, 36 (2), 410-420.
- Gnambs, Timo und Kai Kaspar (2015), „Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis,“ *Behavior Research Methods*, 47 (4), 1237-1259.
- Gogami, Masaki, Yuki Matsuda, Yutaka Arakawa und Keiichi Yasumoto (2021), „Detection of careless responses in online surveys using answering behavior on smartphone,“ *IEEE Access*, 9, 53205-53218.
- Goldammer, Philippe, Hubert Annen, Peter L. Stöckli und Klaus Jonas (2020), „Careless responding in questionnaire measures: Detection, impact, and remedies,“ *The Leadership Quarterly*, 31 (4), 101384.
- Goodman, Joseph K., Cynthia E. Cryder und Amar Cheema (2013), „Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples,“ *Journal of Behavioral Decision Making*, 26 (3), 213-224.
- Göritz, Anja S. (2004), „The impact of material incentives on response quantity, response quality, sample composition, survey outcome and cost in online access panels,“ *International Journal of Market Research*, 46 (3), 327-345.
- Gosling, Samuel D. und Winter Mason (2015), „Internet research in psychology,“ *Annual Review of Psychology*, 66 (1), 877-902.
- Grable, John E., Wookjae Heo und Abed Rabbani (2021), „Characteristics of random responders in a financial risk-tolerance questionnaire,“ *Journal of Financial Services Marketing*, 26 (1), 1-9.
- Greenleaf, Eric A. (1992a), „Improving rating scale measures by detecting and correcting bias components in some response styles,“ *Journal of Marketing Research*, 29 (2), 176-188.
- Greenleaf, Eric A. (1992b), „Measuring extreme response style,“ *Public Opinion Quarterly*, 56 (3), 328-351.
- Groves, Robert M. (2006), „Nonresponse rates and nonresponse bias in household surveys,“ *Public Opinion Quarterly*, 70 (5), 646-675.
- Groves, Robert M., Robert B. Cialdini und Mick P. Couper (1992), „Understanding the decision to participate in a survey,“ *Public Opinion Quarterly*, 56 (4), 475-495.
- Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer und Roger Tourangeau (2009), *Survey Methodology*. Hoboken: John Wiley & Sons, Inc.
- Groves, Robert M. und Lars E. Lyberg (2010), „Total Survey Error: Past, present, and future,“ *Public Opinion Quarterly*, 74 (5), 849-879.
- Guerreiro, João, Paulo Rita und Duarte Trigueiros (2015), „Attention, emotions and cause-related marketing effectiveness,“ *European Journal of Marketing*, 49 (11/12), 1728-1750.
- Gummer, Tobias, Jan K. Höhne, Tobias Rettig, Joss Roßmann und Mirjan Kummerow (2023), „Is there a growing use of mobile devices in web surveys? Evidence from 128 web surveys in Germany,“ *Quality and Quantity* (Online-Vorabveröffentlichung), 1-21.
- Gummer, Tobias und Tanja Kunz (2022), „Relying on external information sources when answering knowledge questions in web surveys,“ *Sociological Methods & Research*, 51 (2), 816-836.
- Gummer, Tobias, Franziska Quöß und Joss Roßmann (2019), „Does increasing mobile device coverage reduce heterogeneity in completing web surveys on smartphones?,“ *Social Science Computer Review*, 37 (3), 371-384.
- Gummer, Tobias und Joss Roßmann (2015), „Explaining interview duration in web surveys: A multi-level approach,“ *Social Science Computer Review*, 33 (2), 217-234.

- Gummer, Tobias, Joss Roßmann und Henning Silber (2021), „Using instructed response items as attention checks in web surveys: Properties and implementation,” *Sociological Methods & Research*, 50 (1), 238-264.
- Gunter, Barrie, David Nicholas, Paul Huntington und Peter Williams (2002), „Online versus offline research: Implications for evaluating digital media,” *Aslib Proceedings*, 54 (4), 229-239.
- Gwizdka, Jacek und Yinglong Zhang (2015), „Differences in eye-tracking measures between visits and revisits to relevant and irrelevant web pages,” in *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. ACM Digital Library*, Ricardo Baeza-Yates, Hrsg. New York City: ACM, 811-814.
- Hair Jr., Joseph F., Dana E. Harrison und Jeffrey J. Risher (2018), „Marketing research in the 21st century: Opportunities and challenges,” *Brazilian Journal of Marketing - Revista Brasileira de Marketing*, Special Issue 17 (5), 666-699.
- Hargittai, Eszter (2009), „An update on survey measures of web-oriented digital literacy,” *Social Science Computer Review*, 27 (1), 130-137.
- Harpe, Spencer E. (2015), „How to analyze Likert and other rating scale data,” *Currents in Pharmacy Teaching and Learning*, 7 (6), 836-850.
- Hartmann, Petra (1995), „Response behavior in interview settings of limited privacy,” *International Journal of Public Opinion Research*, 7 (4), 383-390.
- Hauser, David J. und Norbert Schwarz (2015), „It’s a trap! Instructional manipulation checks prompt systematic thinking on “tricky” tasks,” *SAGE Open*, 5 (2), 1-6.
- Hauser, David J., Aashna Sunderrajan, Madhuri Natarajan und Norbert Schwarz (2017), „Prior exposure to instructional manipulation checks does not attenuate survey context effects driven by satisficing or gricean norms,” *methods, data, analyses*, 10 (2), 195-220.
- Heberlein, Thomas A. und Robert Baumgartner (1978), „Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature,” *American Sociological Review*, 43 (4), 447.
- Heerwegh, Dirk (2006), „An investigation of the effect of lotteries on web survey response rates,” *Field Methods*, 18 (2), 205-220.
- Heerwegh, Dirk und Geert Loosveldt (2008), „Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality,” *Public Opinion Quarterly*, 72 (5), 836-846.
- Henrich, Joseph, Steven J. Heine und Ara Norenzayan (2010), „The weirdest people in the world?,” *The Behavioral and Brain Sciences*, 33 (2-3), 61-83.
- Herten, Nadja, Tobias Otto und Oliver T. Wolf (2017), „The role of eye fixation in memory enhancement under stress - An eye tracking study,” *Neurobiology of Learning and Memory*, 140, 134-144.
- Herzing, Jessica M. E. (2019), „Mobile web surveys,” *FORS Guide No. 01, Version 1.0*, Swiss Centre of Expertise in the Social Sciences FORS, Lausanne.
- Herzog, A. Regula und Jerald G. Bachman (1981), „Effects of questionnaire length on response quality,” *Public Opinion Quarterly*, 45 (4), 549-559.
- Höhne, Jan K. und Timo Lenzner (2015), „Investigating response order effects in web surveys using eye tracking,” *Psihologija*, 48 (4), 361-377.
- Höhne, Jan K., Melanie A. Revilla und Stephan Schlosser (2020a), „Motion instructions in surveys: Compliance, acceleration, and response quality,” *International Journal of Market Research*, 62 (1), 43-57.
- Höhne, Jan K. und Stephan Schlosser (2018), „Investigating the adequacy of response time outlier definitions in computer-based web surveys using Paradata SurveyFocus,” *Social Science Computer Review*, 36 (3), 369-378.

- Höhne, Jan K. und Stephan Schlosser (2019), „SurveyMotion: What can we learn from sensor data about respondents' completion and response behavior in mobile web surveys?,” *International Journal of Social Research Methodology*, 22 (4), 379-391.
- Höhne, Jan K., Stephan Schlosser, Mick P. Couper und Annelies G. Blom (2020b), „Switching away: Exploring on-device media multitasking in web surveys,” *Computers in Human Behavior*, 111, 106417.
- Holland, Carol A. und Patrick M. A. Rabbitt (1990), „Autobiographical and text recall in the elderly: An investigation of a processing resource deficit,” *The Quarterly Journal of Experimental Psychology Section A*, 42 (3), 441-470.
- Holm, Sture (1979), „A simple sequentially rejective multiple test procedure.,” *Scandinavian Journal of Statistics*, 6 (2), 65-70.
- Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka und Joost van de Weijer (2011), *Eye tracking. A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Hong, Maxwell R. und Ying Cheng (2019), „Clarifying the effect of test speededness,” *Applied Psychological Measurement*, 43 (8), 611-623.
- Hong, Maxwell R., Jeffrey T. Steedle und Ying Cheng (2020), „Methods of detecting insufficient effort responding: Comparisons and practical recommendations,” *Educational and Psychological Measurement*, 80 (2), 312-345.
- Honsel, Kai (2012), *Integrated Usage Mining. Eine Methode zur Analyse des Benutzerverhaltens im Web*. Gabler research. Wiesbaden: Gabler.
- Hough, Leaetta M., Newell K. Eaton, Marvin D. Dunnette, John D. Kamp und Rodney A. McCloy (1990), „Criterion-related validities of personality constructs and the effect of response distortion on those validities,” *Journal of Applied Psychology*, 75 (5), 581-595.
- Houtkamp, R. und P. R. Roelfsema (2006), „The effect of items in working memory on the deployment of attention and the eyes during visual search,” *Journal of Experimental Psychology: Human Perception and Performance*, 32 (2), 423-442.
- Huang, Jason L., Nathan A. Bowling, Mengqiao Liu und Yuhui Li (2015a), „Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions,” *Journal of Business and Psychology*, 30 (2), 299-311.
- Huang, Jason L., Paul G. Curran, Jessica Keeney, Elizabeth M. Poposki und Richard P. DeShon (2012), „Detecting and deterring insufficient effort responding to surveys,” *Journal of Business and Psychology*, 27 (1), 99-114.
- Huang, Jason L. und Justin A. DeSimone (2021), „Insufficient effort responding as a potential confound between survey measures and objective tests,” *Journal of Business and Psychology*, 36 (5), 807-828.
- Huang, Jason L., Mengqiao Liu und Nathan A. Bowling (2015b), „Insufficient effort responding: Examining an insidious confound in survey data,” *Journal of Applied Psychology*, 100 (3), 828-845.
- Huang, Yunhui, Lei Wang und Junqi Shi (2009), „When do objects become more attractive? The individual and interactive effects of choice and ownership on object evaluation,” *Personality & Social Psychology Bulletin*, 35 (6), 713-722.
- Huddleston, Patricia, Bridget K. Behe, Stella Minahan und R. Thomas Fernandez (2015), „Seeking attention: An eye tracking study of in-store merchandise displays,” *International Journal of Retail & Distribution Management*, 43 (6), 561-574.
- Hulland, John, Hans Baumgartner und Keith M. Smith (2018), „Marketing survey research best practices: Evidence and recommendations from a review of JAMS articles,” *Journal of the Academy of Marketing Science*, 46 (1), 92-108.

- Hyman, Ira E., S. Matthew Boss, Breanne M. Wise, Kira E. McKenzie und Jenna M. Caggiano (2010), „Did you see the unicycling clown? Inattention blindness while walking and talking on a cell phone,” *Applied Cognitive Psychology*, 24 (5), 597-607.
- IBM Corp. (2021), *IBM SPSS Statistics for Windows, Version 28.0*. Armonk: IBM Corp.
- iMotions (2021), *iMotions 8.1*. Kopenhagen: iMotions A/S.
- Islam, Nayeem und Roy Want (2014), „Smartphones: Past, present, and future,” *IEEE Pervasive Computing*, 13 (4), 89-92.
- Jäckle, Annette, Caroline Roberts und Peter Lynn (2010), „Assessing the effect of data collection mode on measurement,” *International Statistical Review*, 78 (1), 3-20.
- Jackson, Douglas N. (1977), *Jackson vocational interest survey: Manual*. London (Ontario): Research Psychologists Press.
- Jacob, Robert J. K. und Keith S. Karn (2003), „Eye tracking in human-computer interaction and usability research,” in *The mind's eye. Cognitive and applied aspects of eye movement research; [... from papers presented at the 11th European Conference on Eye Movements (August 22 - 25 2001, Turku Finland)*, Jukka Hyönä, Hrsg. Amsterdam: Elsevier, 573-605.
- Janke, Stefan und Angelika Glöckner-Rist (2012), „Deutsche Version der Satisfaction with Life Scale (SWLS),” *Zusammenstellung sozialwissenschaftlicher Items und Skalen*, ZIS - GESIS Leibniz-Institut für Sozialwissenschaften, Mannheim.
- Jann, Ben (2015), „Asking sensitive questions: Overview and introduction,” in *Improving survey methods. Lessons from recent research. European Association of Methodology series*, Uwe Engel, Ben Jann, Peter Lynn, Annette Scherpenzeel und Patrick J. Sturgis, Hrsg. New York City: Routledge Taylor & Francis Group, 101-105.
- Jarodzka, Halszka und Saskia Brand-Gruwel (2017), „Tracking the reading eye: Towards a model of real-world reading,” *Journal of Computer Assisted Learning*, 33 (3), 193-201.
- JASP Team (2023), *JASP (Version 0.16.4)*. Amsterdam.
- Jiménez, Fernando R. und Kevin E. Voss (2014), „An alternative approach to the measurement of emotional attachment,” *Psychology & Marketing*, 31 (5), 360-370.
- Jockisch, Maike (2009), „Das Technologieakzeptanzmodell,” in *„Das ist gar kein Modell!“. Unterschiedliche Modelle und Modellierungen in Betriebswirtschaftslehre und Ingenieurwissenschaften. Springer eBook Collection Business and Economics*, Gerhard Bandow und Hartmut H. Holzmüller, Hrsg. Wiesbaden: Gabler, 233-254.
- Johnson, John A. (2005), „Ascertaining the validity of individual protocols from web-based personality inventories,” *Journal of Research in Personality*, 39 (1), 103-129.
- Johnston, Lloyd D. und Patrick M. O'Malley (1997), „The recanting of earlier reported drug use by young adults,” *NIDA Research Monograph*, 167, 59-80.
- Joinson, Adam (1999), „Social desirability, anonymity, and internet-based questionnaires,” *Behavior Research Methods, Instruments, & Computers*, 31 (3), 433-438.
- Jones, Matt, George Buchanan und Harold Thimbleby (2003), „Improving web search on small screen devices,” *Interacting with Computers*, 15 (4), 479-495.
- Jones, Matt, Gary Marsden, Norliza Mohd-Nasir, Kevin Boone und George Buchanan (1999), „Improving web interaction on small displays,” *Computer Networks*, 31 (11-16), 1129-1137.
- Jones, Matthew R. und Helena Karsten (2008), „Giddens's structuration theory and information systems research,” *MIS Quarterly*, 32 (1), 127-157.
- Joo, Jihyuk und Yoonmo Sang (2013), „Exploring Koreans' smartphone usage: An integrated model of the technology acceptance model and uses and gratifications theory,” *Computers in Human Behavior*, 29 (6), 2512-2518.
- Jung, Yoonhyuk (2014), „What a smartphone is to me: Understanding user values in using smartphones,” *Information Systems Journal*, 24 (4), 299-321.

- Just, Marcel A. und Patricia A. Carpenter (1976), „Eye fixations and cognitive processes,” *Cognitive Psychology*, 8 (4), 441-480.
- Just, Marcel A. und Patricia A. Carpenter (1980), „A theory of reading: From eye fixations to comprehension,” *Psychological Review*, 87 (4), 329-354.
- Kam, Chester C. S. (2019), „Careless responding threatens factorial analytic results and construct validity of personality measure,” *Frontiers in Psychology*, 10, 1258.
- Kam, Chester C. S. und Gabriel H. Chan (2018), „Examination of the validity of instructed response items in identifying careless respondents,” *Personality and Individual Differences*, 129, 83-87.
- Kam, Chester C. S. und John P. Meyer (2015), „How careless responding and acquiescence response bias can influence construct dimensionality,” *Organizational Research Methods*, 18 (3), 512-541.
- Kam, Chester C. S., John P. Meyer und Shaojing Sun (2021), „Why do people agree with both regular and reversed items? A logical response perspective,” *Assessment*, 28 (4), 1110-1124.
- Katz, James E. und Satomi Sugiyama (2005), „Mobile phones as fashion statements: The co-creation of mobile communication’s public meaning,” in *Mobile communications. Re-negotiation of the social sphere. Computer supported cooperative work*, Nr. 31, Rich Ling und Per E. Pedersen, Hrsg. London: Springer, 63-81.
- Kemper, Christoph J., Constanze Beierlein, Doreen Bensch, Anastassiya Kovaleva und Beatrice Rammstedt (2012), „Eine Kurzsкала zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens: Die Kurzsкала Soziale Erwünschtheit-Gamma (KSE-G),” *GESIS Working Papers* Nr. 25, GESIS - Leibnitz-Institut für Sozialwissenschaften, Mannheim.
- Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Philip D. Waggoner, Ryan Jewell und Nicholas J. G. Winter (2020), „The shape of and solutions to the MTurk quality crisis,” *Political Science Research and Methods*, 8 (4), 614-629.
- Kern, Christoph, Jan K. Höhne, Stephan Schlosser und Melanie A. Revilla (2021), „Completion conditions and response behavior in smartphone surveys: A prediction approach using acceleration data,” *Social Science Computer Review*, 39 (6), 1253-1271.
- Keusch, Florian und Frederick G. Conrad (2022), „Using smartphones to capture and combine self-reports and passively measured behavior in social research,” *Journal of Survey Statistics and Methodology*, 10 (4), 863-885.
- Keusch, Florian, Alexander Wenz und Frederick G. Conrad (2022), „Do you have your smartphone with you? Behavioral barriers for measuring everyday activities with smartphone sensors,” *Computers in Human Behavior*, 127, 107054.
- Keusch, Florian und Ting Yan (2017), „Web versus mobile web: An experimental study of device effects and self-selection effects,” *Social Science Computer Review*, 35 (6), 751-769.
- Kim, Hoyoung, Jinwoo Kim und Yeonsoo Lee (2005), „An empirical study of use contexts in the mobile internet, focusing on the usability of information architecture,” *Information Systems Frontiers*, 7 (2), 175-186.
- Kim, Jaewon, Paul Thomas, Ramesh Sankaranarayana, Tom Gedeon und Hwan-Jin Yoon (2015), „Eye-tracking analysis of user behavior and performance in web search on large and small screens,” *Journal of the Association for Information Science and Technology*, 66 (3), 526-544.
- Kim, Jaewon, Paul Thomas, Ramesh Sankaranarayana, Tom Gedeon und Hwan-Jin Yoon (2016), „Understanding eye movements on mobile devices for better presentation of search results,” *Journal of the Association for Information Science and Technology*, 67 (11), 2607-2619.
- Kim, Jimin, Goya Choi, Younghoon Chang und Myeong-Cheol Park (2017), „Use contexts of personal computing devices: Determinants of use context changes,” *Industrial Management & Data Systems*, 117 (10), 2431-2451.

- Kim, Ki J. und S. Shyam Sundar (2014), „Does screen size matter for smartphones? Utilitarian and hedonic effects of screen size on smartphone adoption,” *Cyberpsychology, Behavior and Social Networking*, 17 (7), 466-473.
- Kim, Yujin, Jennifer Dykema, John Stevenson, Penny Black und D. Paul Moberg (2019), „Straightlining: Overview of measurement, comparison of indicators, and effects in mail-web mixed-mode surveys,” *Social Science Computer Review*, 37 (2), 214-233.
- King, Anna L. S., Alexandre M. Valença, Adriana C. O. Silva, Tathiana Baczynski, Marcele R. Carvalho und Antonio. E. Nardi (2013), „Nomophobia: Dependency on virtual environments or social phobia?,” *Computers in Human Behavior*, 29 (1), 140-144.
- Kleck, Gary und Kelly Roberts (2012), „What survey modes are most effective in eliciting self-reports of criminal or delinquent behavior?,” in *Handbook of survey methodology for the social sciences*, Lior Gideon, Hrsg. New York City: Springer New York, 417-439.
- Konok, Veronika, Ákos Pogány und Ádám Miklósi (2017), „Mobile attachment: Separation from the mobile phone induces physiological and behavioural stress and attentional bias to separation-related stimuli,” *Computers in Human Behavior*, 71, 228-239.
- Kowler, Eileen, Eric Anderson, Barbara Doshier und Erik Blaser (1995), „The role of attention in the programming of saccades,” *Vision Research*, 35 (13), 1897-1916.
- Krajbich, Ian und Antonio Rangel (2011), „Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions,” *Proceedings of the National Academy of Sciences of the United States of America*, 108 (33), 13852-13857.
- Krämer, Andreas (2017), „Das Design von Rating-Skalen in Zeiten von Mobile Research,” in *Mobile Research. Grundlagen und Zukunftsaussichten für die Mobile Marktforschung*, Axel Theobald, Hrsg. Wiesbaden: Springer Fachmedien, 87-104.
- Krebs, Dagmar und Jan K. Höhne (2021), „Exploring scale direction effects and response behavior across PC and smartphone surveys,” *Journal of Survey Statistics and Methodology*, 9 (3), 477-495.
- Kreuter, Frauke, Stanley Presser und Roger Tourangeau (2008), „Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity,” *Public Opinion Quarterly*, 72 (5), 847-865.
- Krosnick, Jon A. (1991), „Response strategies for coping with the cognitive demands of attitude measures in surveys,” *Applied Cognitive Psychology*, 5 (3), 213-236.
- Krosnick, Jon A. (1999), „Survey research,” *Annual Review of Psychology*, 50 (1), 537-567.
- Krosnick, Jon A. (2017), „Questionnaire design,” in *The Palgrave handbook of survey research*, David L. Vannette und Jon A. Krosnick, Hrsg. Cham: Palgrave Macmillan US (Springer), 439-455.
- Krugman, Dean M., Richard J. Fox, James E. Fletcher und Tina H. Rojas (1994), „Do adolescents attend to warnings in cigarette advertising? An eye-tracking approach,” *Journal of Advertising Research*, 34 (6), 39-53.
- Krumpal, Ivar (2013), „Determinants of social desirability bias in sensitive surveys: A literature review,” *Quality and Quantity*, 47 (4), 2025-2047.
- Krumpal, Ivar und Anatol-Fiete Näher (2012), „Entstehungsbedingungen sozial erwünschten Antwortverhaltens: Eine experimentelle Onlinestudie zum Einfluss des Wordings und des Kontexts bei unangenehmen Fragen,” *Soziale Welt*, 63 (1), 65-89.
- Kulas, John T., Rachael Klahr und Lindsey Knights (2019), „Confound it! Social desirability and the “reverse-scoring” method effect,” *European Journal of Psychological Assessment*, 35 (6), 855-867.
- Kurtz, John E. und Catherine L. Parrish (2001), „Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R,” *Journal of Personality Assessment*, 76 (2), 315-332.
- Laeng, Bruno, Takashi Suegami und Samira Aminihajibashi (2016), „Wine labels: An eye-tracking and pupillometry study,” *International Journal of Wine Business Research*, 28 (4), 327-348.

- Lagun, Dmitry, Chih-Hung Hsieh, Dale Webster und Vidhya Navalpakkam (2014), „Towards better measurement of attention and satisfaction in mobile search,” in *Proceedings of the 37th international ACM SIGIR conference on research & development in information retrieval*. ACM Digital Library, Shlomo Geva, Hrsg. New York City: ACM, 113-122.
- Lai, Meng-Lung, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia W.-Y. Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang und Chin-Chung Tsai (2013), „A review of using eye-tracking technology in exploring learning from 2000 to 2012,” *Educational Research Review*, 10, 90-115.
- Lamb, Roberta und Rob Kling (2003), „Reconceptualizing users as social actors in information systems research,” *MIS Quarterly*, 27 (2), 197.
- Lambert, Amber D. und Angie L. Miller (2015), „Living with smartphones: Does completion device affect survey responses?,” *Research in Higher Education*, 56 (2), 166-177.
- Lane, Wilburn und Chris Manner (2011), „The impact of personality traits on smartphone ownership and use,” *International Journal of Business and Social Science*, 2 (17), 22-28.
- Lange, Florian und Siegfried Dewitte (2019), „Measuring pro-environmental behavior: Review and recommendations,” *Journal of Environmental Psychology*, 63, 92-100.
- Langhaug, Lisa F., Lorraine Sherr und Frances M. Cowan (2010), „How to improve the validity of sexual behaviour reporting: Systematic review of questionnaire delivery modes in developing countries,” *Tropical Medicine & International Health*, 15 (3), 362-381.
- Lee, Hana, Sunwoong Kim, Mick P. Couper und Youngje Woo (2019), „Experimental comparison of PC web, smartphone web, and telephone surveys in the new technology era,” *Social Science Computer Review*, 37 (2), 234-247.
- Lee, Inseong, Jaesoo Kim und Jinwoo Kim (2005), „Use contexts for the mobile internet: A longitudinal study monitoring actual use of mobile internet services,” *International Journal of Human-Computer Interaction*, 18 (3), 269-292.
- Lee, JooWon und Jae-Hyeon Ahn (2012), „Attention to banner ads and their effectiveness: An eye-tracking approach,” *International Journal of Electronic Commerce*, 17 (1), 119-137.
- Lee, Raymond M. (1993), *Doing research on sensitive topics*. London: Sage.
- Lehmann, Erich L. (1993), „The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two?,” *Journal of the American Statistical Association*, 88 (424), 1242-1249.
- Leiner, Dominik J. (2019), „Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys,” *Survey Research Methods*, 13 (3), 229-248.
- Leiner, Dominik J. (2023), *SoSci Survey (Version 3.4.20)*. München.
- Lenzner, Timo, Lars Kaczmirek und Mirta Galesic (2011), „Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension,” *International Journal of Public Opinion Research*, 23 (3), 361-373.
- Leonhart, Rainer (2022), *Lehrbuch Statistik. Einstieg und Vertiefung*. Bern: Hogrefe AG.
- Lepp, Andrew, Jacob E. Barkley und Aryn C. Karpinski (2014), „The relationship between cell phone use, academic performance, anxiety, and satisfaction with life in college students,” *Computers in Human Behavior*, 31, 343-350.
- Liebe, Ulf, Klaus Glenk, Malte Oehlmann und Jürgen Meyerhoff (2015), „Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behaviour in web surveys?,” *Journal of Choice Modelling*, 14, 17-31.
- Limesurvey GmbH (2023), *LimeSurvey: An open source survey tool*. Hamburg: LimeSurvey GmbH.
- Liu, Mengqiao, Nathan A. Bowling, Jason L. Huang und Tiana Kent (2013), „Insufficient effort responding to surveys as a threat to calidity. The perceptions and practices of SIOP members,” *The Industrial-Organizational Psychologist*, 51 (1), 32-39.

- Liu, Mingnan, Sunghye Lee und Frederick G. Conrad (2015), „Comparing extreme response styles between agree-disagree and item-specific scales,” *Public Opinion Quarterly*, 79 (4), 952-975.
- Liu, Yong und Hongxiu Li (2011), „Exploring the impact of use context on mobile hedonic services adoption: An empirical study on mobile gaming in China,” *Computers in Human Behavior*, 27 (2), 890-898.
- Loftus, Geoffrey R. (1972), „Eye fixations and recognition memory for pictures,” *Cognitive Psychology*, 3 (4), 525-551.
- Logie, Robert H., Gesualdo M. Zucco und Alan D. Baddeley (1990), „Interference with visual short-term memory,” *Acta Psychologica*, 75 (1), 55-74.
- Luck, Steven J. und Edward K. Vogel (2013), „Visual working memory capacity: From psychophysics and neurobiology to individual differences,” *Trends in Cognitive Sciences*, 17 (8), 391-400.
- Lugtig, Peter und Vera Toepoel (2016), „The use of PCs, smartphones, and tablets in a probability-based panel survey: Effects on survey measurement error,” *Social Science Computer Review*, 34 (1), 78-94.
- Lütters, Holger (2017), „Auf der Suche nach neuen Fragetypen für die mobile Marktforschung,” in *Mobile Research. Grundlagen und Zukunftsaussichten für die Mobile Marktforschung*, Axel Theobald, Hrsg. Wiesbaden: Springer Fachmedien, 67-85.
- Lyberg, Lars E. und Diana M. Stukel (2017), „The roots and evolution of the Total Survey Error concept,” in *Total Survey Error in practice*, Paul P. Biemer, Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker und Brady T. West, Hrsg. Hoboken: John Wiley & Sons, 1-22.
- Lynn, Peter (1998), „Data collection mode effects on responses to attitudinal questions,” *Journal of Official Statistics*, 14 (1), 1-14.
- Ma, Wei J., Masud Husain und Paul M. Bays (2014), „Changing concepts of working memory,” *Nature Neuroscience*, 17 (3), 347-356.
- MacKenzie, Scott B. und Philip M. Podsakoff (2012), „Common Method Bias in marketing: Causes, mechanisms, and procedural remedies,” *Journal of Retailing*, 88 (4), 542-555.
- Magraw-Mickelson, Zoe, Harry H. Wang und Mario Gollwitzer (2022), „Survey mode and data quality: Careless responding across three modes in cross-cultural contexts,” *International Journal of Testing*, 22 (2), 121-153.
- Mahalanobis, Prasanta C. (1936), „On the generalized distance in statistics,” *Proceedings of the National Institute of Science in India*, 12, 49-55.
- Manfreda, Katja L., Michael Bosnjak, Jernej Berzelak, Iris Haas und Vasja Vehovar (2008), „Web surveys versus other survey modes: A meta-analysis comparing response rates,” *International Journal of Market Research*, 50 (1), 79-104.
- Maniaci, Michael R. und Ronald D. Rogge (2014), „Caring about carelessness: Participant inattention and its effects on research,” *Journal of Research in Personality*, 48, 61-83.
- Mannetti, Lucia, Antonio Pierro und Stefano Livi (2002), „Explaining consumer conduct: From planned to self-expressive behavior,” *Journal of Applied Social Psychology*, 32 (7), 1431-1451.
- Manzo, Amber N. und Jennifer M. Burke (2012), „Increasing response rate in web-based/internet surveys,” in *Handbook of survey methodology for the social sciences*, Lior Gideon, Hrsg. New York City: Springer New York, 327-343.
- Marjanovic, Zdravko, Ronald Holden, Ward Struthers, Robert Cribbie und Esther Greenglass (2015), „The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders,” *Personality and Individual Differences*, 84, 79-83.
- Martin, Kirsten und Katie Shilton (2016), „Why experience matters to privacy: How context-based experience moderates consumer privacy expectations for mobile applications,” *Journal of the Association for Information Science and Technology*, 67 (8), 1871-1882.

- Mavletova, Aigul (2013), „Data quality in PC and mobile web surveys,” *Social Science Computer Review*, 31 (6), 725-743.
- Mavletova, Aigul und Mick P. Couper (2013), „Sensitive topics in PC web and mobile web surveys: Is there a difference?,” *Survey Research Methods*, 7 (3), 191-205.
- Mavletova, Aigul und Mick P. Couper (2014), „Mobile web survey design: Scrolling versus paging, SMS versus e-mail invitations,” *Journal of Survey Statistics and Methodology*, 2 (4), 498-518.
- Mavletova, Aigul und Mick P. Couper (2015), „A meta-analysis of breakoff rates in mobile web surveys,” in *Mobile research methods. Opportunities and challenges of mobile research methodologies*, Daniele Toninelli, Robert Pinter und Pablo de Pedraza, Hrsg. London: Ubiquity Press, 81-98.
- Mavletova, Aigul und Mick P. Couper (2016), „Grouping of items in mobile web questionnaires,” *Field Methods*, 28 (2), 170-193.
- Mavletova, Aigul, Mick P. Couper und Daniil Lebedev (2018), „Grid and item-by-item formats in PC and mobile web surveys,” *Social Science Computer Review*, 36 (6), 647-668.
- Mayfrank, Lothar, M. Mobashery, Hubert Kimmig und Burkhart Fischer (1986), „The role of fixation and visual attention in the occurrence of express saccades in man,” *European Archives of Psychiatry and Neurological Sciences*, 235 (5), 269-275.
- McAlister, Leigh und Edgar Pessemier (1982), „Variety seeking behavior: An interdisciplinary review,” *Journal of Consumer Research*, 9 (3), 311-322.
- McCallum, Ethan B. und Zoë D. Peterson (2012), „Investigating the impact of inquiry mode on self-reported sexual behavior: Theoretical considerations and review of the literature,” *Journal of Sex Research*, 49 (2-3), 212-226.
- McCrae, Robert R. und Paul T. Costa (1987), „Validation of the five-factor model of personality across instruments and observers,” *Journal of Personality and Social Psychology*, 52 (1), 81-90.
- McGonagle, Alyssa K., Jason L. Huang und Benjamin M. Walsh (2016), „Insufficient effort survey responding: An under-appreciated problem in work and organisational health psychology research,” *Applied Psychology*, 65 (2), 287-321.
- McGrath, Robert E., Matthew Mitchell, Brian H. Kim und Leaetta M. Hough (2010), „Evidence for response bias as a source of error variance in applied assessment,” *Psychological Bulletin*, 136 (3), 450-470.
- McKay, Alexander S., Donna M. Garcia, John P. Clapper und Kenneth S. Shultz (2018), „The attentive and the careless: Examining the relationship between benevolent and malevolent personality traits with careless responding in online surveys,” *Computers in Human Behavior*, 84, 295-303.
- Meade, Adam W. und S. Bartholomew Craig (2012), „Identifying careless responses in survey data,” *Psychological Methods*, 17 (3), 437-455.
- Meghanathan, Radha N., Andrey R. Nikolaev und Cees van Leeuwen (2019), „Refixation patterns reveal memory-encoding strategies in free viewing,” *Attention, Perception, & Psychophysics*, 81 (7), 2499-2516.
- Meghanathan, Radha N., Cees van Leeuwen, Marcello Giannini und Andrey R. Nikolaev (2020), „Neural correlates of task-related refixation behavior,” *Vision Research*, 175, 90-101.
- Meijer, Rob R., A. Susan M. Niessen und Jorge N. Tendeiro (2016), „A practical guide to check the consistency of item response patterns in clinical research through person-fit statistics: Examples and a computer program,” *Assessment*, 23 (1), 52-62.
- Meißner, Martin und Josua Oll (2019), „The promise of eye-tracking methodology in organizational research: A taxonomy, review, and future avenues,” *Organizational Research Methods*, 22 (2), 590-617.
- Melumad, Shiri (2017), „The distinct psychology of smartphone usage,” Dissertation, Columbia University, New York City.

- Melumad, Shiri, J. Jeffrey Inman und Michel T. Pham (2019), „Selectively emotional: How smartphone use changes user-generated content,” *Journal of Marketing Research*, 56 (2), 259-275.
- Melumad, Shiri und Robert Meyer (2020), „Full disclosure: How smartphones enhance consumer self-disclosure,” *Journal of Marketing*, 84 (3), 28-45.
- Melumad, Shiri und Michel T. Pham (2020), „The smartphone as a pacifying technology,” *Journal of Consumer Research*, 47 (2), 237-255.
- Meyvis, Tom und Stijn M. J. van Osselaer (2018), „Increasing the power of your study by increasing the effect size,” *Journal of Consumer Research*, 44 (5), 1157-1173.
- Miller, Geoffrey (2012), „The smartphone psychology manifesto,” *Perspectives on Psychological Science*, 7 (3), 221-237.
- Miller, Joel (2008), „Impact of situational factors on survey measured fear of crime,” *International Journal of Social Research Methodology*, 11 (4), 307-325.
- Miller, Norman und Donald T. Campbell (1959), „Recency and primacy in persuasion as a function of the timing of speeches and measurements,” *Journal of Abnormal Psychology*, 59 (1), 1-9.
- Mimura, Koki, Hirohisa Kishino, Genta Karino, Etsuko Nitta, Aya Senoo, Kentaro Ikegami, Tetsuya Kunikata, Hideo Yamanouchi, Shun Nakamura, Kan Sato und Mamiko Koshiba (2015), „Potential of a smartphone as a stress-free sensor of daily human behaviour,” *Behavioural Brain Research*, 276, 181-189.
- Mohr, Gisela und Andreas Müller (2004), „Depressivität im nichtklinischen Kontext,” *Zusammenstellung sozialwissenschaftlicher Items und Skalen*, ZIS - GESIS Leibniz-Institut für Sozialwissenschaften, Mannheim.
- Molyneux, Logan (2018), „Mobile news consumption,” *Digital Journalism*, 6 (5), 634-650.
- Morren, Meike und Leonard J. Paas (2020), „Short and long instructional manipulation checks: What do they measure?,” *International Journal of Public Opinion Research*, 32 (4), 790-800.
- Mueller-Hanson, Rose A., Eric D. Heggstad und George C. Thornton III (2006), „Individual differences in impression management: An exploration of the psychological processes underlying faking,” *Psychology Science*, 48 (3), 288-312.
- Mutz, Michael und Sylvia Kämpfer (2011), „...und nun zum Wetter: Beeinflusst die Wetterlage die Einschätzung von politischen und wirtschaftlichen Sachverhalten? / ...and Now the Weather: Does Weather Influence the Assessment of Political and Economic Issues?,” *Zeitschrift für Soziologie*, 40 (4), 208-226.
- Neuert, Cornelia E. und Timo Lenzner (2016), „Incorporating eye tracking into cognitive interviewing to pretest survey questions,” *International Journal of Social Research Methodology*, 19 (5), 501-519.
- Newman, Jessica C., Don C. Des Jarlais, Charles F. Turner, Jay Gribble, Phillip Cooley und Denise Paone (2002), „The differential effects of face-to-face and computer interview modes,” *American Journal of Public Health*, 92 (2), 294-297.
- Neyman, Jerzy und Egon S. Pearson (1933), „IX. On the problem of the most efficient tests of statistical hypotheses,” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231 (694-706), 289-337.
- Nichols, David S., Roger L. Greene und Peter Schmolck (1989), „Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials,” *Journal of Clinical Psychology*, 45 (2), 239-250.
- Niessen, A. Susan M., Rob R. Meijer und Jorge N. Tendeiro (2016), „Detecting careless respondents in web-based questionnaires: Which method to use?,” *Journal of Research in Personality*, 63, 1-11.
- Nikolaev, Andrey R., Benedikt V. Ehinger, Radha N. Meghanathan und Cees van Leeuwen (2023), „Planning to revisit: Neural activity in refixation precursors,” *Journal of Vision*, 23 (7), 2.
- North, David, Kevin A. Johnston und Jacques Ophoff (2014), „The use of mobile phones by South African university students,” *Issues in Informing Science and Information Technology*, 11, 115-138.

- Oberauer, Klaus (2019), „Working memory and attention - A conceptual analysis and review,” *Journal of Cognition*, 2 (1), 36.
- Oberski, Daniel L. (2012), „Comparability of survey measurements,” in *Handbook of survey methodology for the social sciences*, Lior Gideon, Hrsg. New York City: Springer New York, 477-498.
- Olatunji, Bunmi O., Nathan L. Williams, David F. Tolin, Jonathan S. Abramowitz, Craig N. Sawchuk, Jeffrey M. Lohr und Lisa S. Elwood (2007), „The disgust scale: Item analysis, factor structure, and suggestions for refinement,” *Psychological Assessment*, 19 (3), 281-297.
- Ong, Anthony D. und David J. Weiss (2000), „The impact of anonymity on responses to sensitive questions,” *Journal of Applied Social Psychology*, 30 (8), 1691-1708.
- Ophir, Yaakov, Itay Sisso, Christa S. C. Asterhan, Refael Tikochinski und Roi Reichart (2020), „The Turker blues: Hidden factors behind increased depression rates among Amazon’s Mechanical Turkers,” *Clinical Psychological Science*, 8 (1), 65-83.
- Oppenheimer, Daniel M., Tom Meyvis und Nicolas Davidenko (2009), „Instructional manipulation checks: Detecting satisficing to increase statistical power,” *Journal of Experimental Social Psychology*, 45 (4), 867-872.
- Orquin, Jacob L. und Simone Mueller Loose (2013), „Attention and choice: A review on eye movements in decision making,” *Acta Psychologica*, 144 (1), 190-206.
- Osborne, Jason W. und Elaine Waters (2002), „Four assumptions of multiple regression that researchers should always test,” *Practical Assessment, Research & Evaluation*, 8 (1), 2.
- Paas, Leonard J. und Meike Morren (2018), „Please do not answer if you are reading this: Respondent attention in online panels,” *Marketing Letters*, 29 (1), 13-21.
- Pan, Bing, Arch G. Woodside und Fang Meng (2014), „How contextual cues impact response and conversion rates of online surveys,” *Journal of Travel Research*, 53 (1), 58-68.
- Panova, Tayana und Xavier Carbonell (2018), „Is smartphone addiction really an addiction?,” *Journal of Behavioral Addictions*, 7 (2), 252-259.
- Park, C. Whan, Deborah J. Macinnis, Joseph Priester, Andreas B. Eisingerich und Dawn Iacobucci (2010), „Brand attachment and brand attitude strength: Conceptual and empirical differentiation of two critical brand equity drivers,” *Journal of Marketing*, 74 (6), 1-17.
- Park, Chang S. und Barbara K. Kaye (2019), „Smartphone and self-extension: Functionally, anthropomorphically, and ontologically extending self via the smartphone,” *Mobile Media & Communication*, 7 (2), 215-231.
- Park, Jaehyun und Sung H. Han (2013), „Defining user value: A case study of a smartphone,” *International Journal of Industrial Ergonomics*, 43 (4), 274-282.
- Park, Yangil und Jengchung V. Chen (2007), „Acceptance and adoption of the innovative use of smartphone,” *Industrial Management & Data Systems*, 107 (9), 1349-1365.
- Park, Yong J. (2015), „My whole world’s in my palm! The second-level divide of teenagers’ mobile use and skill,” *New Media & Society*, 17 (6), 977-995.
- Pashler, Harold (1988), „Familiarity and visual change detection,” *Perception & Psychophysics*, 44 (4), 369-378.
- Paulhus, Delroy L. (1991), „Measurement and control of response bias,” in *Measures of personality and social psychological attitudes. Measures of social psychological attitudes*, Nr. 1, John P. Robinson, Hrsg. San Diego: Academic Press, 17-59.
- Paulhus, Delroy L. (2002), „Socially desirable responding: The evolution of a construct,” in *The role of constructs in psychological and educational measurement*, Henry I. Braun, Douglas N. Jackson und David E. Wiley, Hrsg. Mahwah: Lawrence Erlbaum Associates, 61-84.
- Paulhus, Delroy L. und Simine Vazire (2007), „The self-report method,” in *Handbook of research methods in personality psychology*, Richard W. Robins, R. Chris Fraley und Robert F. Krueger, Hrsg. New York City: The Guilford Press, 224-239.

- Peterson, Gregg J. (2012), „Unintended mobile respondents,” Paper presented at the annual Council of American Survey Research Organizations (CASRO) technology conference, New York City.
- Peterson, Gregg J., Jamie Griffin, John LaFrance und JiaoJiao Li (2017), „Smartphone participation in web surveys,” in *Total Survey Error in practice*, Paul P. Biemer, Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker und Brady T. West, Hrsg. Hoboken: John Wiley & Sons, 203-233.
- Peterson, Robert A. (2001), „On the use of college students in social science research: Insights from a second-order meta-analysis,” *Journal of Consumer Research*, 28 (3), 450-461.
- Peytchev, Andy (2009), „Survey breakoff,” *Public Opinion Quarterly*, 73 (1), 74-97.
- Peytchev, Andy (2011), „Breakoff and unit nonresponse across web surveys,” *Journal of Official Statistics*, 27 (1), 33-47.
- Peytchev, Andy (2013), „Consequences of survey nonresponse,” *The ANNALS of the American Academy of Political and Social Science*, 645 (1), 88-111.
- Peytchev, Andy und Craig A. Hill (2010), „Experiments in mobile web survey design: Similarities to other modes and unique considerations,” *Social Science Computer Review*, 28 (3), 319-335.
- Phillips, William A. (1974), „On the distinction between sensory storage and short-term visual memory,” *Perception & Psychophysics*, 16 (2), 283-290.
- Piccoli, Gabriele und Myle Ott (2014), „Impact of mobility and timing on user-generated content,” *MIS Quarterly Executive*, 13 (3), 147-157.
- Pieters, Rik und Luk Warlop (1999), „Visual attention during brand choice: The impact of time pressure and task motivation,” *International Journal of Research in Marketing*, 16 (1), 1-16.
- Pinsoeneault, Terry B. (2007), „Detecting random, partially random, and nonrandom Minnesota Multiphasic Personality Inventory-2 protocols,” *Psychological Assessment*, 19 (1), 159-164.
- Pinter, Robert, Daniele Toninelli und Pablo de Pedraza (2015), „Mobile research methods: Possibilities and issues of a new promising way of conducting research,” in *Mobile research methods. Opportunities and challenges of mobile research methodologies*, Daniele Toninelli, Robert Pinter und Pablo de Pedraza, Hrsg. London: Ubiquity Press, 1-10.
- Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee und Nathan P. Podsakoff (2003), „Common Method Biases in behavioral research: A critical review of the literature and recommended remedies,” *Journal of Applied Psychology*, 88 (5), 879-903.
- Podsakoff, Philip M., Scott B. MacKenzie und Nathan P. Podsakoff (2012), „Sources of method bias in social science research and recommendations on how to control it,” *Annual Review of Psychology*, 63 (1), 539-569.
- Poggio, Teresio, Michael Bosnjak und Kai W. Weyandt (2015), „Survey participation via mobile devices in a probability-based online-panel: Prevalence, determinants, and implications for nonresponse,” *Survey Practice*, 8 (1), 1-7.
- Porter, Stephen R. und Michael E. Whitcomb (2003), „The impact of contact type on web survey response rates,” *Public Opinion Quarterly*, 67 (4), 579-588.
- Posner, Michael I. (1980), „Orienting of attention,” *Quarterly Journal of Experimental Psychology*, 32 (1), 3-25.
- Poulová, Petra, Blanka Klímová und Daniela Pulkrábková (2019), „Use of mobile devices - A survey study,” in *Smart education and e-Learning 2019. Springer eBooks Intelligent Technologies and Robotics*, Nr. 144, Vladimir L. Uskov, Robert J. Howlett und Lakhmi C. Jain, Hrsg. Singapur: Springer, 303-312.
- Poynter, Ray (2015), „The utilization of mobile technology and approaches in commercial market research,” in *Mobile research methods. Opportunities and challenges of mobile research methodologies*, Daniele Toninelli, Robert Pinter und Pablo de Pedraza, Hrsg. London: Ubiquity Press, 11-20.

- Prüfer, Peter, Lisa Vazansky und Darius Wystup (2003), *Antwortskalen im ALLBUS und ISSP: eine Sammlung. ZUMA-Methodenbericht*, Nr. 11. Mannheim: ZUMA.
- R Core Team (2023), *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.
- Rad, Mostafa S., Alison J. Martingano und Jeremy Ginges (2018), „Toward a psychology of Homo sapiens: Making psychological science more representative of the human population,” *Proceedings of the National Academy of Sciences of the United States of America*, 115 (45), 11401-11405.
- Radach, Ralph, Albrecht Inhoff und Dieter Heller (2002), „The role of attention and spatial selection in fluent reading,” in *Basic functions of language, reading and reading disability. Neuropsychology and Cognition*, Nr. 20, Evelin Witruk, Angela D. Friederici und Thomas Lachmann, Hrsg. Boston: Springer, 137-153.
- Radach, Ralph und Alan Kennedy (2004), „Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research,” *European Journal of Cognitive Psychology*, 16 (1-2), 3-26.
- Raento, Mika, Antti Oulasvirta und Nathan Eagle (2009), „Smartphones: An emerging tool for social scientists,” *Sociological Methods & Research*, 37 (3), 426-454.
- Ran, Shan, Mengqiao Liu, Lisa A. Marchiondo und Jason L. Huang (2015), „Difference in response effort across sample types: Perception or reality?,” *Industrial and Organizational Psychology*, 8 (2), 202-208.
- Raptis, Dimitrios, Nikolaos Tselios, Jesper Kjeldskov und Mikael B. Skov (2013), „Does size matter? Investigating the impact of mobile phone screen size on users’ perceived usability, effectiveness and efficiency,” in *Proceedings of the 15th international conference on human-computer interaction with mobile devices and services. ACM Conferences*, Michael Rohs, Hrsg. New York City: ACM, 127-136.
- Ravaja, Niklas, Kari Kallinen, Timo Saari und Liisa Keltikangas-Jarvinen (2004), „Suboptimal exposure to facial expressions when viewing video messages from a small screen: Effects on emotion, attention, and memory,” *Journal of Experimental Psychology: Applied*, 10 (2), 120-13.
- Rayner, Keith (1998), „Eye movements in reading and information processing: 20 years of research,” *Psychological Bulletin*, 124 (3), 372-422.
- Rayner, Keith (2009), „Eye movements and attention in reading, scene perception, and visual search,” *Quarterly Journal of Experimental Psychology*, 62 (8), 1457-1506.
- Read, Blair, Lukas Wolters und Adam J. Berinsky (2022), „Racing the clock: Using response time as a proxy for attentiveness on self-administered surveys,” *Political Analysis*, 30 (4), 550-569.
- Reeves, Byron, Annie Lang, Eun Y. Kim und Deborah Tatar (1999), „The effects of screen size and message content on attention and arousal,” *Media Psychology*, 1 (1), 49-67.
- Reichle, Erik D., Alexander Pollatsek, Donald L. Fisher und Keith Rayner (1998), „Toward a model of eye movement control in reading,” *Psychological Review*, 105 (1), 125-157.
- Reingold, Eyal M. (2014), „Eye tracking research and technology: Towards objective measurement of data quality,” *Visual Cognition*, 22 (3), 635-652.
- Reinhardt, Sibylle (2019), „Fridays For Future – Moral und Politik gehören zusammen,” *GWP – Gesellschaft, Wirtschaft, Politik*, 68 (2), 159-162.
- Reips, Ulf-Dietrich (2000), „The web experiment method,” in *Psychological experiments on the internet*, Michael H. Birnbaum, Hrsg. San Diego: Academic Press, 89-117.
- Reise, Steven P., Dale S. Kim, Maxwell Mansolf und Keith F. Widaman (2016), „Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively re-weighted least squares to the Rosenberg self-esteem scale,” *Multivariate Behavioral Research*, 51 (6), 818-838.

- Reuband, Karl-Heinz (1987), „Unerwünschte Dritte beim Interview,“ *Zeitschrift für Soziologie*, 16 (4), 303-308.
- Reutskaja, Elena, Rosemarie Nagel, Colin F. Camerer und Antonio Rangel (2011), „Search dynamics in consumer choice under time pressure: An eye-tracking study,“ *American Economic Review*, 101 (2), 900-926.
- Revilla, Melanie A. (2017), „Analyzing survey characteristics, participation, and evaluation across 186 surveys in an online opt-in panel in Spain,“ *methods, data, analyses*, 11 (2), 135-162.
- Revilla, Melanie A. und Mick P. Couper (2018), „Comparing grids with vertical and horizontal item-by-item formats for PCs and smartphones,“ *Social Science Computer Review*, 36 (3), 349-368.
- Revilla, Melanie A. und Carlos Ochoa (2016), „Open narrative questions in PC and smartphones: Is the device playing a role?,“ *Quality and Quantity*, 50 (6), 2495-2513.
- Revilla, Melanie A. und Carlos Ochoa (2017), „Ideal and maximum length for a web survey,“ *International Journal of Market Research*, 59 (5), 557-565.
- Revilla, Melanie A., Daniele Toninelli und Carlos Ochoa (2016a), „PCs versus smartphones in answering web surveys: Does the device make a difference?,“ *Survey Practice*, 9 (4), 1-6.
- Revilla, Melanie A., Daniele Toninelli und Carlos Ochoa (2017), „An experiment comparing grids and item-by-item formats in web surveys completed through PCs and smartphones,“ *Telematics and Informatics*, 34 (1), 30-42.
- Revilla, Melanie A., Daniele Toninelli, Carlos Ochoa und Germán Loewe (2015), „Who has access to mobile devices in an online opt-in panel? An analysis of potential respondents for mobile surveys,“ in *Mobile research methods. Opportunities and challenges of mobile research methodologies*, Daniele Toninelli, Robert Pinter und Pablo de Pedraza, Hrsg. London: Ubiquity Press, 119-139.
- Revilla, Melanie A., Daniele Toninelli, Carlos Ochoa und Germán Loewe (2016b), „Do online access panels need to adapt surveys for mobile devices?,“ *Internet Research*, 26 (5), 1209-1227.
- Rieger, Diana, Dorothee Hefner und Peter Vorderer (2017), „Mobile recovery? The impact of smartphone use on recovery experiences in waiting situations,“ *Mobile Media & Communication*, 5 (2), 161-177.
- Ritter, Lois A. und Valerie M. Sue (2007), „Introduction to using online surveys,“ *New Directions for Evaluation*, 115, 5-14.
- Roberts, Caroline, Emily Gilbert, Nick Allum und Léila Eisner (2019), „Research synthesis: Satisficing in surveys: A systematic review of literature,“ *Public Opinion Quarterly*, 83 (3), 598-626.
- Roberts, Caroline, Jessica M. E. Herzing, Marc A. Manjon, Philip Abbet und Daniel Gatica-Perez (2022), „Response burden and dropout in a probability-based online panel study – A comparison between an app and browser-based design,“ *Journal of Official Statistics*, 38 (4), 987-1017.
- Robinson, Tim, Thomas Cronin, Haider Ibrahim, Mark Jinks, Timothy Molitor, Joshua Newman und Jonathan Shapiro (2013), „Smartphone use and acceptability among clinical medical students: A questionnaire-based study,“ *Journal of Medical Systems*, 37 (3), 9936.
- Rolstad, Sindre, John Adler und Anna Rydén (2011), „Response burden and questionnaire length: Is shorter better? A review and meta-analysis,“ *Value in Health*, 14 (8), 1101-1108.
- Rosbergen, Edward, Rik Pieters und Michel Wedel (1997), „Visual attention to advertising: A segment-level analysis,“ *Journal of Consumer Research*, 24 (3), 305-314.
- Rosenberger, Robert (2015), „An experiential account of phantom vibration syndrome,“ *Computers in Human Behavior*, 52, 124-131.
- Rossiter, John R. (2002), „The C-OAR-SE procedure for scale development in marketing,“ *International Journal of Research in Marketing*, 19 (4), 305-335.
- Roßmann, Joss (2017), *Satisficing in Befragungen. Theorie, Messung und Erklärung*. Wiesbaden: Springer VS.

- Rubik, Frieder, Ria Müller, Richard Harnisch, Brigitte Holzhauser, Michael Schipperges und Sonja Geiger (2019), „Umweltbewusstsein in Deutschland 2018: Ergebnisse einer repräsentativen Bevölkerungsumfrage,“ Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit (BMU); Umweltbundesamt, Berlin, Dessau-Roßlau.
- Salehan, Mohammad und Arash Negahban (2013), „Social networking on smartphones: When mobile phones become addictive,“ *Computers in Human Behavior*, 29 (6), 2632-2639.
- Sánchez-Fernández, Juan, Francisco Muñoz-Leiva und Francisco J. Montoro-Ríos (2012), „Improving retention rate and response quality in web-based surveys,“ *Computers in Human Behavior*, 28 (2), 507-514.
- Sauer, Vera J., Sabrina C. Eimler, Sanaz Maafi, Michael Pietrek und Nicole C. Krämer (2015), „The phantom in my pocket: Determinants of phantom phone sensations,“ *Mobile Media & Communication*, 3 (3), 293-316.
- Scherhorn, Gerhard, Hendrik Haas, Frank Hellenthal und Sabine Seibold (1999), „Naturverträgliches Handeln,“ *Zusammenstellung sozialwissenschaftlicher Items und Skalen*, ZIS - GESIS Leibniz-Institut für Sozialwissenschaften, Mannheim.
- Schlosser, Stephan und Anja Mays (2018), „Mobile and dirty: Does using mobile devices affect the data quality and the response process of online surveys?,“ *Social Science Computer Review*, 36 (2), 212-230.
- Schlosser, Stephan und Henning Silber (2020), „Endgerätespezifische und darstellungsabhängige Bearbeitungszeit- und Antwortverhaltensunterschiede in Webbefragungen,“ in *Grundlagen - Methoden - Anwendungen in den Sozialwissenschaften. Festschrift für Steffen-M. Kühnel*, Anja Mays, André Dingelstedt, Verena Hambauer, Stephan Schlosser, Florian Berens, Jürgen Leibold und Jan K. Höhne, Hrsg. Wiesbaden, Heidelberg: Springer VS, 263-282.
- Schmidt, Katharina, Tobias Gummer und Joss Roßmann (2020), „Effects of respondent and survey characteristics on the response quality of an open-ended attitude question in web surveys,“ *methods, data, analyses*, 14 (1), 3-34.
- Schmitt, Neal und Daniel M. Stuits (1985), „Factors defined by negatively keyed items: The result of careless respondents?,“ *Applied Psychological Measurement*, 9 (4), 367-373.
- Schober, Michael F., Frederick G. Conrad, Christopher Antoun, Patrick Ehlen, Stefanie Fail, Andrew L. Hupp, Michael Johnston, Lucas Vickers, H. Yanna Yan und Chan Zhang (2015), „Precision and disclosure in text and voice interviews on smartphones,“ *PLOS ONE*, 10 (6), e0128337.
- Schonlau, Matthias und Vera Toepoel (2015), „Straightlining in web survey panels over time,“ *Survey Research Methods*, 9 (2), 125-137.
- Schouten, Barry, Jan van den Brakel, Bart Buelens, Jan van der Laan und Thomas Klausch (2013), „Disentangling mode-specific selection and measurement bias in social surveys,“ *Social Science Research*, 42 (6), 1555-1570.
- Schroeders, Ulrich, Christoph Schmidt und Timo Gnams (2022), „Detecting careless responding in survey data using stochastic gradient boosting,“ *Educational and Psychological Measurement*, 82 (1), 29-56.
- Schultz, P. Wesley und Florian G. Kaiser (2012), „Promoting pro-environmental behavior,“ in *The Oxford handbook of environmental and conservation psychology. Oxford library of psychology*, Susan D. Clayton, Hrsg. Oxford, New York City: Oxford University Press, 556-580.
- Schultz, P. Wesley und Lynette Zelezny (1999), „Values as predictors of environmental attitudes: Evidence for consistency across 14 countries,“ *Journal of Environmental Psychology*, 19 (3), 255-265.
- Schwartz, Shalom H., Gila Melech, Arielle Lehmann, Steven Burgess, Mari Harris und Vicki Owens (2001), „Extending the cross-cultural validity of the theory of basic human values with a different method of measurement,“ *Journal of Cross-Cultural Psychology*, 32 (5), 519-542.

- Schwarz, Andrew und Wynn Chin (2007), „Looking forward: Toward an understanding of the nature and definition of IT acceptance,” *Journal of the Association for Information Systems*, 8 (4), 230-243.
- Sendelbah, Anže, Vasja Vehovar, Ana Slavec und Andraž Petrovčič (2016), „Investigating respondent multitasking in web surveys using paradata,” *Computers in Human Behavior*, 55, 777-787.
- Shadish, William R., Thomas D. Cook und Donald T. Campbell (2002), *Experimental and quasi-experimental designs for generalized causal inference*. Boston, Mass.: Houghton Mifflin
- Shen, Kelly, Anthony R. McIntosh und Jennifer D. Ryan (2014), „A working memory account of refixations in visual search,” *Journal of Vision*, 14 (14), 11.
- Shi, Savannah W., Michel Wedel und Rik Pieters (2013), „Information acquisition during online decision making: A model-based exploration using eye-tracking data,” *Management Science*, 59 (5), 1009-1026.
- Singh, Ranjit K., Anja S. Göritz und Klaus Moser (2015), „Methoden der psychologischen Marktforschung,” in *Wirtschaftspsychologie. Springer-Lehrbuch*, Klaus Moser, Hrsg. Berlin, Heidelberg: Springer, 161-178.
- Skierkowski, Dorothy und Rebecca M. Wood (2012), „To text or not to text? The importance of text messaging among college-aged youth,” *Computers in Human Behavior*, 28 (2), 744-756.
- Smith, Tom W. und Jibum Kim (2015), „A review of survey data-collection modes: With a focus on computerizations,” *Sociological Theory and Methods*, 30 (2), 185-200.
- Sommer, Jana, Birk Diedenhofen und Jochen Musch (2017), „Not to be considered harmful: Mobile-device users do not spoil data quality in web surveys,” *Social Science Computer Review*, 35 (3), 378-387.
- Sommer, Moritz, Dieter Rucht, Sebastian Haunss und Sabrina Zajak (2019), „Fridays for Future. Profil, Entstehung und Perspektiven der Protestbewegung in Deutschland,” *ipb working paper series*, Institut für Protest- und Bewegungsforschung, Berlin
- Spink, Amanda, Charles Cole und Mary Waller (2008), „Multitasking behavior,” *Annual Review of Information Science and Technology*, 42 (1), 93-118.
- Srivastava, Lara (2005), „Mobile phones and the evolution of social behaviour,” *Behaviour & Information Technology*, 24 (2), 111-129.
- Stald, Gitte (2008), „Mobile identity: Youth, identity, and mobile communication media,” in *Youth, identity, and digital media. The John D. and Catherine T. Macarthur Foundation series on digital media and learning*, David Buckingham, Hrsg. Cambridge: MIT Press, 143-164.
- Stapleton, Carey E. (2013), „The smart(phone) way to collect survey data,” *Survey Practice*, 6 (2), 1-7.
- Statcounter (2023), „Desktop vs mobile vs tablet market share worldwide | Statcounter global stats,” (Zugriff am September 28, 2023), [verfügbar unter <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet>].
- Stedman, Richard C., Nancy A. Connelly, Thomas A. Heberlein, Daniel J. Decker und Shorna B. Allred (2019), „The end of the (research) world as we know it? Understanding and coping with declining response rates to mail surveys,” *Society & Natural Resources*, 32 (10), 1139-1154.
- Steenkamp, Jan-Benedict E., Martijn G. de Jong und Hans Baumgartner (2010), „Socially desirable response tendencies in survey research,” *Journal of Marketing Research*, 47 (2), 199-214.
- Steg, Linda, Goda Perlaviciute, Ellen van der Werff und Judith Lurvink (2014), „The significance of hedonic values for environmentally relevant attitudes, preferences, and actions,” *Environment and Behavior*, 46 (2), 163-192.
- Steger, Diana, Ulrich Schroeders und Timo Gnamb (2020), „A meta-analysis of test scores in proctored and unproctored ability assessments,” *European Journal of Psychological Assessment*, 36 (1), 174-184.
- Steger, Diana, Ulrich Schroeders und Oliver Wilhelm (2021), „Caught in the act: Predicting cheating in unproctored knowledge assessment,” *Assessment*, 28 (3), 1004-1017.

- Steinbrecher, Markus, Joss Roßmann und Jan E. Blumenstiel (2015), „Why do respondents break off web surveys and does it matter? Results from four follow-up surveys,” *International Journal of Public Opinion Research*, 27 (2), 289-302.
- Steindorf, Lena und Jan Rummel (2020), „Do your eyes give you away? A validation study of eye-movement measures used as indicators for mindless reading,” *Behavior Research Methods*, 52 (1), 162-176.
- Stets, Jan E. und Chris F. Biga (2003), „Bringing identity theory into environmental sociology,” *Sociological Theory*, 21 (4), 398-423.
- Stöber, Joachim (1999), „Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität,” *Diagnostica*, 45 (4), 173-177.
- Stocké, Volker (2004), „Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit: Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion,” *Zeitschrift für Soziologie*, 33 (4), 303-320.
- Strahilevitz, Michal A. und George Loewenstein (1998), „The effect of ownership history on the valuation of objects,” *Journal of Consumer Research*, 25 (3), 276-289.
- Strick, Madelijn, Rob W. Holland, Rick van Baaren und Ad van Knippenberg (2010), „Humor in the eye tracker: Attention capture and distraction from context cues,” *The Journal of General Psychology*, 137 (1), 37-48.
- Struminskaya, Bella, Peter Lugtig, Florian Keusch und Jan K. Höhne (2020a), „Augmenting surveys with data from sensors and apps: Opportunities and challenges,” *Social Science Computer Review*, Online-Vorabveröffentlichung.
- Struminskaya, Bella, Vera Toepoel, Peter Lugtig, Marieke Haan, Annemieke Luiten und Barry Schouten (2020b), „Understanding willingness to share smartphone-sensor data,” *Public Opinion Quarterly*, 84 (3), 725-759.
- Struminskaya, Bella, Kai W. Weyandt und Michael Bosnjak (2015), „The effects of questionnaire completion using mobile devices on data quality. Evidence from a probability-based general population panel,” *methods, data, analyses*, 9 (2), 261-292.
- Stüber, Eva (2013), *Personalisierung im Internethandel. Die Akzeptanz von Kaufempfehlungen in der Bekleidungsbranche. Springer Gabler Research*. Wiesbaden: Springer Gabler.
- SurveyCircle (2023), *Forschungswebseite SurveyCircle*. Veröffentlicht 2016. Mannheim.
- Sutanto, Juliana, Elia Palme, Chuan-Hoo Tan und Chee W. Phang (2013), „Addressing the personalization-privacy paradox: An empirical assessment from a field experiment on smartphone users,” *MIS Quarterly*, 37 (4), 1141-1164.
- Taber, Keith S. (2018), „The use of Cronbach’s Alpha when developing and reporting research instruments in science education,” *Research in Science Education*, 48 (6), 1273-1296.
- Taddicken, Monika (2009), „Methodeneffekte von Web-Befragungen: Soziale Erwünschtheit vs. Soziale Entkontextualisierung,” in *Umfrageforschung. Herausforderungen und Grenzen. Österreichische Zeitschrift für Soziologie Sonderheft*, Nr. 9, Martin Weichbold, Hrsg. Wiesbaden: VS Verlag für Sozialwissenschaften, 85-104.
- Tatler, Benjamin W., Iain D. Gilchrist und Michael F. Land (2005), „Visual memory for objects in natural scenes: From fixations to object files,” *The Quarterly Journal of Experimental Psychology Section A*, 58 (5), 931-960.
- Terzi, Tayfun (2017), „Detecting semi-plausible response patterns,” Dissertation, London School of Economics and Political Science (LSE), London.
- Theobald, Axel (2017), „Rekrutierung, Motivation und Verhalten von Befragungsteilnehmern,” in *Praxis Online-Marktforschung*, Axel Theobald, Hrsg. Wiesbaden: Springer Fachmedien, 291-357.

- Theobald, Axel, João F. Baigger und Jochen Knöller (2017), „Mobile first – Gestalterische Umsetzung mobil optimierter Online-Befragungen,” in *Mobile Research. Grundlagen und Zukunftsaussichten für die Mobile Marktforschung*, Axel Theobald, Hrsg. Wiesbaden: Springer Fachmedien, 15-29.
- Thornton, Bill, Alyson Faires, Maija Robbins und Eric Rollins (2014), „The mere presence of a cell phone may be distracting,” *Social Psychology*, 45 (6), 479-488.
- Thorsteinsson, Gisli und Tom Page (2014), „User attachment to smartphones and design guidelines,” *International Journal of Mobile Learning and Organisation*, 8 (3-4), 201-215.
- Toepoel, Vera (2016), „Buttons of balken, klikken of slepen : wat werkt er nu het beste op mobiele telefoons, tablets of PCs?,” in *Ontwikkelingen in het Marktonderzoek : Jaarboek 2016 Markt Onderzoek Associatie*, Nr. 41, Fred. Bronner, Elisabeth Brügggen, Paul Dekker, Edith D. de Leeuw, Leonard J. Paas, Ale Smidts und Jaap E. Wieringa, Hrsg. Haarlem: Spaar en Hout, 119-136.
- Toepoel, Vera und Peter Lugtig (2014), „What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of internet users,” *Social Science Computer Review*, 32 (4), 544-560.
- Toepoel, Vera und Peter Lugtig (2015), „Online surveys are mixed-device surveys. Issues associated with the use of different (mobile) devices in web surveys,” *methods, data, analyses*, 9 (2), 155-162.
- Toepoel, Vera, Corrie Vis, Marcel Das und Arthur van Soest (2009), „Design of web questionnaires: The effects of the number of items per screen,” *Sociological Methods & Research*, 37 (3), 371-392.
- Toninelli, Daniele und Melanie A. Revilla (2016a), „Is the smartphone participation affecting the web survey experience?,” in *Proceedings of the 48th scientific meeting of the Italian Statistical Society*, Società Italiana di Statistica, Hrsg. Salerno: SIS (Società Italiana di Statistica), 1-6.
- Toninelli, Daniele und Melanie A. Revilla (2016b), „Smartphones vs PCs: Does the device affect the web survey experience and the measurement error for sensitive topics? - A replication of the Mavletova & Couper’s 2013 experiment,” *Survey Research Methods*, 10 (2), 153-169.
- Toninelli, Daniele und Melanie A. Revilla (2020), „How mobile device screen size affects data collected in web surveys,” in *Advances in questionnaire design, development, evaluation, and testing*, Paul C. Beatty, Debbie Collins, Lyn Kaye, Jose-Luis Padilla, Gordon B. Willis und Amanda Wilmot, Hrsg. Hoboken: Wiley, 349-373.
- Tourangeau, Roger (1987), „Attitude measurement: A cognitive perspective,” in *Social information processing and survey methodology. Recent Research in Psychology*, Hans- J. Hippler, Norbert Schwarz und Seymour Sudman, Hrsg. New York City: Springer, 149-162.
- Tourangeau, Roger (2017), „Mixing modes: Tradeoffs among coverage, nonresponse, and coverage error,” in *Total Survey Error in practice*, Paul P. Biemer, Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker und Brady T. West, Hrsg. Hoboken: John Wiley & Sons, 115-132.
- Tourangeau, Roger, Aaron Maitland, Gonzalo Rivero, Hanyu Sun, Douglas Williams und Ting Yan (2017), „Web surveys by smartphone and tablets: Effects on survey responses,” *Public Opinion Quarterly*, 81 (4), 896-929.
- Tourangeau, Roger, Kenneth A. Rasinski, Jared B. Jobe, Tom W. Smith und William F. Pratt (1997a), „Sources of error in a survey on sexual behavior,” *Journal of Official Statistics*, 13 (4), 341-365.
- Tourangeau, Roger, Lance J. Rips und Kenneth A. Rasinski (2000), *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tourangeau, Roger und Tom W. Smith (1996), „Asking sensitive questions: The impact of data collection mode, question format, and question context,” *Public Opinion Quarterly*, 60 (2), 275-304.
- Tourangeau, Roger, Tom W. Smith und Kenneth A. Rasinski (1997b), „Motivation to report sensitive behaviors on surveys: Evidence from a bogus pipeline experiment,” *Journal of Applied Social Psychology*, 27 (3), 209-222.

- Tourangeau, Roger, Hanyu Sun, Ting Yan, Aaron Maitland, Gonzalo Rivero und Douglas Williams (2018), „Web surveys by smartphones and tablets: Effects on data quality,” *Social Science Computer Review*, 36 (5), 542-556.
- Tourangeau, Roger und Ting Yan (2007), „Sensitive questions in surveys,” *Psychological Bulletin*, 133 (5), 859-883.
- Trau, Raymond N. C., Charmine E. J. Härtel und Günter F. Härtel (2013), „Reaching and hearing the invisible: Organizational research on invisible stigmatized groups via web surveys,” *British Journal of Management*, 24 (4), 532-541.
- Tuschl, Stefan, Sandra Meister und Sarah Laube (2016), „Alles nur ein Spiel?,” in *Marktforschung der Zukunft - Mensch oder Maschine? Bewährte Kompetenzen in neuem Kontext*, Bernhard Keller, Hans-Werner Klein und Stefan Tuschl, Hrsg. Wiesbaden: Springer Gabler, 189-215.
- Type/Code (2023), „Screen sizes | Viewport sizes and pixel densities for popular devices,” (Zugriff am September 28, 2023), [verfügbar unter <https://screensiz.es/phone>].
- Ulitzsch, Esther, Seyma N. Yildirim-Erbasli, Guher Gorgun und Okan Bulut (2022), „An explanatory mixture IRT model for careless and insufficient effort responding in self-report measures,” *The British Journal of Mathematical and Statistical Psychology*, 75 (3), 668-698.
- Urban, Dieter und Jochen Mayerl (2018), *Angewandte Regressionsanalyse: Theorie, Technik und Praxis*. Wiesbaden: Springer VS.
- van Selm, Martine und Nicholas W. Jankowski (2006), „Conducting online surveys,” *Quality and Quantity*, 40 (3), 435-456.
- van Vaerenbergh, Yves und Troy D. Thomas (2013), „Response styles in survey research: A literature review of antecedents, consequences, and remedies,” *International Journal of Public Opinion Research*, 25 (2), 195-217.
- Vannette, David L. und Jon A. Krosnick (2014), „Answering questions: A comparison of survey satisficing and mindlessness,” in *The Wiley Blackwell handbook of mindfulness*, Amanda Ie, Christelle T. Ngnoumen und Ellen J. Langer, Hrsg. Malden, Mass.: Wiley Blackwell, 312-327.
- Velavan, Thirumalaisamy P. und Christian G. Meyer (2020), „The COVID-19 epidemic,” *Tropical Medicine & International Health*, 25 (3), 278-280.
- Venkatesh, Viswanath und Hillol Bala (2008), „Technology Acceptance Model 3 and a research agenda on interventions,” *Decision Sciences*, 39 (2), 273-315.
- Venkatesh, Viswanath, Michael G. Morris, Gordon B. Davis und Fred D. Davis (2003), „User acceptance of information technology: Toward a unified view,” *MIS Quarterly*, 27 (3), 425-478.
- Venkatesh, Viswanath, James Y. L. Thong und Xin Xu (2012), „Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology,” *MIS Quarterly*, 36 (1), 157-178.
- Verkasalo, Hannu (2009), „Contextual patterns in mobile service usage,” *Personal and Ubiquitous Computing*, 13 (5), 331-342.
- Vicente, Paula und Inês Lopes (2013), „Mobile phone surveys to measure consumer behavior: Effects on data quality,” in *Back to basics: Consumer-centric marketing or target-centric marketing*, Pedro Ferreira und André Vieira, Hrsg. Porto: IPAM - Marketing School for Business, 140-146.
- Vincent, Jane (2006), „Emotionale Bindungen im Zeichen des Mobiltelefons,” in *Daumenkultur. Das Mobiltelefon in der Gesellschaft. Kultur- und Medientheorie*, Peter Glotz, Stefan Bertschi, Chris Locke und Henning Thies, Hrsg. Bielefeld: Transcript-Verlag, 135-142.
- Vorderer, Peter, Nicola Krömer und Frank M. Schneider (2016), „Permanently online – Permanently connected: Explorations into university students’ use of social media and mobile smart devices,” *Computers in Human Behavior*, 63, 694-703.
- Voss, Nathaniel M. (2021), „The effects of careless survey responding on the fit of latent variable models: A simulation study,” Dissertation, Kansas State University, Manhattan.

- Voss, Nathaniel M. (2023), „The effects of careless responding on the fit of confirmatory factor analysis and item response theory models,” *Behavior Research Methods*, Online-Vorabveröffentlichung, 1-23.
- Wakenhut, Ruth A. und Dirk Wieseke (2017), „Wie sich qualitative Forschung durch den Einsatz mobiler Geräte verändert,” in *Mobile Research. Grundlagen und Zukunftsaussichten für die Mobile Marktforschung*, Axel Theobald, Hrsg. Wiesbaden: Springer Fachmedien, 105-118.
- Wallis, Hannah und Laura S. Loy (2021), „What drives pro-environmental activism of young people? A survey study on the Fridays For Future movement,” *Journal of Environmental Psychology*, 74, 101581.
- Walsh, Shari P. und Katherine M. White (2007), „Me, my mobile, and I: The role of self- and prototypical identity influences in the prediction of mobile phone behavior,” *Journal of Applied Social Psychology*, 37 (10), 2405-2434.
- Walsh, Shari P., Katherine M. White, Stephen Cox und Ross M. Young (2011), „Keeping in constant touch: The predictors of young Australians’ mobile phone involvement,” *Computers in Human Behavior*, 27 (1), 333-342.
- Walsh, Shari P., Katherine M. White und Ross M. Young (2008), „Over-connected? A qualitative exploration of the relationship between Australian youth and their mobile phones,” *Journal of Adolescence*, 31 (1), 77-92.
- Walsh, Shari P., Katherine M. White und Ross M. Young (2010), „Needing to connect: The effect of self and others on young people’s involvement with their mobile phones,” *Australian Journal of Psychology*, 62 (4), 194-203.
- Wang, Biao (2016), „A study of factors influencing mobile internet user behavior based on the business scenario of mobile entertainment applications,” *International Journal of Simulation: Systems, Science & Technology*, 17 (14), 19.
- Wang, Chun und Gongjun Xu (2015), „A mixture hierarchical model for response times and response accuracy,” *The British Journal of Mathematical and Statistical Psychology*, 68 (3), 456-477.
- Wang, Chun, Gongjun Xu, Zhuoran Shang und Nathan Kuncel (2018), „Detecting aberrant behavior and item preknowledge: A comparison of mixture modeling method and residual method,” *Journal of Educational and Behavioral Statistics*, 43 (4), 469-501.
- Wang, Dan, Zheng Xiang und Daniel R. Fesenmaier (2014), „Adapting to the mobile world: A model of smartphone use,” *Annals of Tourism Research*, 48, 11-26.
- Ward, Mary K. und Adam W. Meade (2018), „Applying social psychology to prevent careless responding during online surveys,” *Applied Psychology*, 67 (2), 231-263.
- Ward, Mary K. und Adam W. Meade (2023), „Dealing with careless responding in survey data: Prevention, identification, and recommended best practices,” *Annual Review of Psychology*, 74, 577-596.
- Ward, Mary K., Adam W. Meade, Christopher M. Allred, Gabriel Pappalardo und J. William Stoughton (2017), „Careless response and attrition as sources of bias in online survey assessments of personality traits and performance,” *Computers in Human Behavior*, 76, 417-430.
- Ward, Mary K. und Samuel B. Pond (2015), „Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys,” *Computers in Human Behavior*, 48, 554-568.
- Ward, Peter, Taralyn Clark, Ramon Zabriskie und Trevor Morris (2014), „Paper/pencil versus online data collection: An exploratory study,” *Journal of Leisure Research*, 46 (1), 84-105.
- Wedel, Michel und Rik Pieters (2000), „Eye fixations on advertisements and memory for brands: A model and findings,” *Marketing Science*, 19 (4), 297-312.
- Wedel, Michel und Rik Pieters (2008), „A review of eye-tracking research in marketing,” in *Review of marketing research. Review of Marketing Research*, Nr. 4, Naresh K. Malhotra, Hrsg. Armonk, London: M.E. Sharpe, 123-147.

- Weijters, Bert, Hans Baumgartner und Niels Schillewaert (2013), „Reversed item bias: An integrative model,” *Psychological Methods*, 18 (3), 320-334.
- Weijters, Bert, Maggie Geuens und Niels Schillewaert (2009), „The proximity effect: The role of inter-item distance on reverse-item bias,” *International Journal of Research in Marketing*, 26 (1), 2-12.
- Weijters, Bert, Niels Schillewaert und Maggie Geuens (2008), „Assessing response styles across modes of data collection,” *Journal of the Academy of Marketing Science*, 36 (3), 409-422.
- Weisberg, Herbert F. (2009), *The Total Survey Error approach. A guide to the new science of survey research*. Chicago: University of Chicago Press.
- Wells, Tom, Justin T. Bailey und Michael W. Link (2013), „Filling the void: Gaining a better understanding of tablet-based surveys,” *Survey Practice*, 6 (1), 1-9.
- Wells, Tom, Justin T. Bailey und Michael W. Link (2014), „Comparison of smartphone and online computer survey administration,” *Social Science Computer Review*, 32 (2), 238-255.
- Wenz, Alexander (2021a), „Completing web surveys on mobile devices: Does screen size affect data quality?,” in *Sozialwissenschaftliche Datenerhebung im digitalen Zeitalter. Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute*, Tobias Wolbring, Heinz Leitgöb und Frank Faulbaum, Hrsg. Wiesbaden, Heidelberg: Springer VS, 101-121.
- Wenz, Alexander (2021b), „Do distractions during web survey completion affect data quality? Findings from a laboratory experiment,” *Social Science Computer Review*, 39 (1), 148-161.
- Wenz, Alexander, Annette Jäckle und Mick P. Couper (2019), „Willingness to use mobile technologies for data collection in a probability household panel,” *Survey Research Methods*, 13 (1), 1-22.
- Wharton, Christopher M., Jeffrey S. Hampl, Rick Hall und Donna M. Winham (2003), „PCs or paper-and pencil: Online surveys for data collection,” *Journal of the American Dietetic Association*, 103 (11), 1458-1460.
- Whelan, Thomas J. (2007), „Anonymity and confidentiality: Do survey respondents know the difference?,” Poster presented at the 30th annual meeting of the Society of Southeastern Social Psychologists, Durham.
- Wiersma, Wybo (2013), „The validity of surveys: Online and offline,” *Oxford Internet Institute*, 18 (3), 321-340.
- Wilcox, Rand R. und Harvey J. Keselman (2003), „Modern robust data analysis methods: Measures of central tendency,” *Psychological Methods*, 8 (3), 254-274.
- Wilkinson, Leland (1999), „Statistical methods in psychology journals: Guidelines and explanations,” *American Psychologist*, 54 (8), 594-604.
- Winkelmann, Rainer (2008), *Econometric analysis of count data*. Berlin, Heidelberg: Springer.
- Wise, Steven L. und Xiaojing Kong (2005), „Response time effort: A new measure of examinee motivation in computer-based tests,” *Applied Measurement in Education*, 18 (2), 163-183.
- Wolter, Felix, Hrsg. (2012), *Heikle Fragen in Interviews. Eine Validierung der Randomized Response-Technik*. Wiesbaden: Springer VS.
- Wood, Dustin, Peter D. Harms, Graham H. Lowman und Justin A. DeSimone (2017), „Response speed and response consistency as mutually validating indicators of data quality in online samples,” *Social Psychological and Personality Science*, 8 (4), 454-464.
- Woods, Carol M. (2006), „Careless responding to reverse-worded items: Implications for confirmatory factor analysis,” *Journal of Psychopathology and Behavioral Assessment*, 28 (3), 189-193.
- World Health Organization (2020), „Coronavirus disease (COVID-19),” Technical documents, World Health Organization, Genf.
- Wu, Meng-Jia, Kelly Zhao und Francisca Fils-Aime (2022), „Response rates of online surveys in published research: A meta-analysis,” *Computers in Human Behavior Reports*, 7, 100206.

- Yammarino, Francis J., Steven J. Skinner und Terry L. Childers (1991), „Understanding mail survey response behavior: A meta-analysis,” *Public Opinion Quarterly*, 55 (4), 613-639.
- Yan, Ting, Florian Keusch und Lirui He (2018), „The impact of question and scale characteristics on scale direction effects,” *Survey Practice*, 11 (2), 1-10.
- Yang, Shuiqing, Yaobin Lu, Sumeet Gupta und Yuzhi Cao (2012), „Does context matter? The impact of use context on mobile internet adoption,” *International Journal of Human-Computer Interaction*, 28 (8), 530-541.
- Yentes, Richard D. und Francisco Wilhelm (2021), *careless. Procedures for computing indices of careless responding. R package version 1.2.0*.
- York, Sue und Ray Poynter (2017), „Global mobile market research in 2017,” in *Mobile Research. Grundlagen und Zukunftsaussichten für die Mobile Marktforschung*, Axel Theobald, Hrsg. Wiesbaden: Springer Fachmedien, 1-14.
- Zelinsky, Gregory J., Lester C. Loschky und Christopher A. Dickinson (2011), „Do object refixations during scene viewing indicate rehearsal in visual working memory?,” *Memory & Cognition*, 39 (4), 600-613.
- Zhang, Chan und Frederick G. Conrad (2014), „Speeding in web surveys: The tendency to answer very fast and its association with straightlining,” *Survey Research Methods*, 8 (2), 127-135.
- Zhang, Han (2020), „Mind-wandering: What can we learn from eye movements?,” Dissertation, University of Michigan, Ann Arbor.
- Zhang, Han, Nicola C. Anderson und Kevin F. Miller (2021), „Refixation patterns of mind-wandering during real-world scene perception,” *Journal of Experimental Psychology: Human Perception and Performance*, 47 (1), 36-52.
- Zhang, XiaoChi, Lars Kuchinke, Marcella L. Woud, Julia Velten und Jürgen Margraf (2017), „Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ,” *Computers in Human Behavior*, 71, 172-180.
- Zwarun, Lara und Alice Hall (2014), „What’s going on? Age, distraction, and multitasking during online survey taking,” *Computers in Human Behavior*, 41, 236-244.

Anhang

Anhang A. Fragebogen der ersten Studie

Teil 1 – Online-Fragebogen

Seite 1



0% ausgefüllt

Work-Life-Balance von Studierenden der Universität Wuppertal

Hierlich willkommen bei unserer Studie zum Thema: „Work-Life-Balance von Studierenden der Universität Wuppertal“. Wir möchten Ihnen danken, dass Sie sich die Zeit genommen haben, an dieser Untersuchung teilzunehmen. Ziel der Studie ist es, mehr über die Bedürfnisse, Einstellungen und Verhaltensweisen der Studierendenschaft zu erfahren. Es hat sich gezeigt, dass auch Studierende immer mehr Wert auf die Work-Life-Balance legen und Universität und Studium daher versuchen sollten, dem daraus entstehenden Anspruch gerecht zu werden. Aus diesem Grund bezieht sich die Befragung sowohl auf Ihr studentisches als auch alltägliches oder privates Leben. Um die Bedürfnisse der Studierenden besser abschätzen zu können, wollen wir uns ein möglichst umfassendes Bild der Lebensrealitäten und der persönlichen Hintergründe der Studentinnen und Studenten verschaffen. Zu diesem Zweck spricht die vorliegende Befragung teilweise auch sehr persönliche Themenbereiche an.

Die Untersuchung besteht aus zwei kurzen Befragungen. Zunächst werden Sie einen Online-Fragebogen ausfüllen (Befragung Teil 1) – im Anschluss wird Ihnen dann noch eine kurze schriftliche Befragung (Befragung Teil 2) vorgelegt. Insgesamt wird die Teilnahme etwa 10-15 Minuten in Anspruch nehmen. Bitte lassen Sie sich in dieser Zeit nicht von äußeren Einflüssen ablenken.

Beachten Sie, dass die Teilnahme an der Untersuchung freiwillig ist und Sie die Befragung jederzeit ohne Angabe von Gründen abbrechen können, ohne dass Ihnen Nachteile entstehen. Die Untersuchung ist vollkommen anonym. Ihre Angaben können von niemandem und zu keiner Zeit mit Ihrer Person in Verbindung gebracht werden. Entsprechend der neuen Datenschutzgrundverordnung werden Ihre Angaben und Daten gewissenhaft geschützt.

Nach Beendigung der Studie haben Sie die Möglichkeit, an einem Gewinnspiel teilzunehmen, bei dem drei Amazon-Gutscheine im Wert von jeweils 20 Euro unter allen Teilnehmerinnen und Teilnehmern verlost werden. Die in diesem Kontext erhobenen personenbezogenen Daten werden separat erfragt und von den Projektmitarbeitenden vertraulich behandelt. Bei Fragen zur Untersuchung können Sie sich gerne an die Versuchsleiter*innen wenden.

Hiermit stimme ich zu, dass meine Angaben im Rahmen der neuen Datenschutzgrundverordnung (DSGVO) zu rein wissenschaftlichen Zwecken verwendet werden dürfen.

Ich stimme zu.

Weiter

Mats Schumacher, M.Sc.
Lehrstuhl für Methoden der empirischen Wirtschafts- und Sozialforschung
Bergische Universität Wuppertal

Seite 2



10% ausgefüllt

Damit der zweite Fragebogenteil den Angaben aus der Online-Befragung zugeordnet werden kann, ist es wichtig, dass Sie einen individuellen Code generieren, der von den Forschern nicht auf Ihre Person zurückgeführt werden kann.

Bitte generieren Sie Ihren Code, indem Sie die folgenden Angaben vervollständigen.

Die ersten beiden Buchstaben setzen sich aus dem **ersten** und **letzten** Buchstaben des Vornamens Ihrer Mutter zusammen. (z.B. AE bei Anne)

Der **Monat** Ihres Geburtsdatums als zweistellige Ziffer. (z.B. 02 bei 26.02.1994)

Die **letzten beiden** Buchstaben Ihres Geburtsorts. (z.B. AL bei Wuppertal)

Zurück

Weiter



20% ausgefüllt

Im Folgenden möchten wir Ihnen gerne ein paar Fragen zu Ihrem studentischen und alltäglichen Leben stellen. Die Fragen decken sowohl allgemeine als auch private Themenbereiche ab. Einige der Fragen könnten Sie als sehr persönlich wahrnehmen. **Da die Befragung vollkommen anonym und vertraulich ist und Ihre Angaben nicht Ihrer Person zugeordnet werden können**, möchten wir sie auch bei diesen Fragen um eine möglichst ehrliche Antwort bitten. Dabei gibt es keine richtigen oder falschen Antworten. Zunächst wird es um Ihr **studentisches Leben** gehen. Bitte beantworten Sie die Fragen ehrlich und vollständig. Sollten Sie sich bei einer Antwort nicht sicher sein, wählen Sie bitte die Antwortkategorie aus, die Ihnen am ehesten korrekt erscheint.

Bitte wählen Sie alle (Bildungs)Abschlüsse aus, über die Sie verfügen.

- Mittlere Reife
- Fachabitur
- Abitur
- abgeschlossene Berufsausbildung
- Bachelor
- Master
- Promotion
- Habilitation
- Sonstiges:

Bitte geben Sie an, in welchem Hochschulsesemester Sie sich insgesamt befinden.

Gemeint ist die Gesamtzahl aller Semester, die Sie an einer Hochschule eingeschrieben waren und studiert haben (z.B. 8 Semester bei 6 Semestern Bachelorstudium und 2 Semestern Masterstudium).

Semester

Zurück

Weiter



**BERGISCHE
UNIVERSITÄT
WUPPERTAL**

30% ausgefüllt

Im Folgenden möchten wir anhand einiger Kriterien von Ihnen wissen, wie Sie die **Bergische Universität Wuppertal** beurteilen. Sollten Sie sich bei einer Antwort nicht sicher sein, wählen Sie bitte die Antwortkategorie aus, die Ihnen am ehesten korrekt erscheint. Bitte beantworten Sie die Fragen ehrlich und vollständig.

Wie zufrieden sind Sie insgesamt mit dem Angebot der Mensen und Cafeterien der Bergischen Universität Wuppertal?

sehr unzufrieden	unzufrieden	eher unzufrieden	unentschieden	eher zufrieden	zufrieden	sehr zufrieden	kann ich nicht beurteilen	keine Angabe
<input type="radio"/>	<input type="radio"/>							

Wie zufrieden sind Sie mit der Betreuung durch das wissenschaftliche Fachpersonal in Ihrer Fakultät?

sehr unzufrieden	unzufrieden	eher unzufrieden	unentschieden	eher zufrieden	zufrieden	sehr zufrieden	kann ich nicht beurteilen	keine Angabe
<input type="radio"/>	<input type="radio"/>							

Wie zufrieden sind Sie mit dem Lehrangebot in Ihrem Studiengang?

sehr unzufrieden	unzufrieden	eher unzufrieden	unentschieden	eher zufrieden	zufrieden	sehr zufrieden	kann ich nicht beurteilen	keine Angabe
<input type="radio"/>	<input type="radio"/>							

Wie zufrieden sind Sie mit dem Hochschulsportangebot der Bergischen Universität Wuppertal?

sehr unzufrieden	unzufrieden	eher unzufrieden	unentschieden	eher zufrieden	zufrieden	sehr zufrieden	kann ich nicht beurteilen	keine Angabe
<input type="radio"/>	<input type="radio"/>							

Zurück

Weiter



40% ausgefüllt

Seit einigen Jahren berichten Studierende von steigendem Leistungsdruck im Studium. Insbesondere die nachfolgenden **Verhaltensweisen** werden im Zuge dieser Entwicklung immer häufiger beobachtet. Auch wenn Ihnen die Beantwortung einiger der folgenden Fragen möglicherweise unangenehm ist, würden wir uns über Ihre Antwort freuen. Sollten Sie sich bei einer Antwort nicht sicher sein, wählen Sie bitte die Antwortkategorie aus, die Ihnen am ehesten korrekt erscheint. Bitte beantworten Sie die Fragen ehrlich und vollständig.

Ich schiebe wichtige Aufgaben in meinem Studium (z.B. Lernen, Ausarbeitungen) vor mir her, statt sie sofort zu erledigen.

nie	sehr selten	selten	gelegentlich	oft	sehr oft	(fast) immer	keine Angabe
<input type="radio"/>							

Ich plagiiere (z.B. Nutzung von Texten anderer Autoren, ohne dies kenntlich zu machen) bewusst bei der Ausarbeitung von Texten, die ich während meines Studiums anfertigen muss (z.B. Hausarbeiten, Seminararbeiten und Abschlussarbeiten).

nie	sehr selten	selten	gelegentlich	oft	sehr oft	(fast) immer	keine Angabe
<input type="radio"/>							

Ich verschaffe mir bei Klausuren unerlaubte Vorteile (z. B. Verwendung von Spickzetteln, Abschauen bei Sitznachbarn).

nie	sehr selten	selten	gelegentlich	oft	sehr oft	(fast) immer	keine Angabe
<input type="radio"/>							

[Zurück](#)[Weiter](#)



50% ausgefüllt

Abseits des Studiums legen Studierende häufig großen Wert auf Ihr Privatleben und die Pflege sozialer Kontakte. Aus diesem Grund würden wir gerne etwas über Ihre private Situation erfahren. Die folgenden Fragen beziehen sich auf Ihr **Freizeitverhalten**. Auch wenn Ihnen die Beantwortung einiger der folgenden Fragen möglicherweise unangenehm ist, würden wir uns über Ihre Antwort freuen. Sollten Sie sich bei einer Antwort nicht sicher sein, wählen Sie bitte die Antwortkategorie aus, die Ihnen am ehesten korrekt erscheint. Bitte beantworten Sie die Fragen ehrlich und vollständig.

Ich treffe mich in einer durchschnittlichen Woche im Semester in meiner Freizeit mit Kommilitonen.

nie	sehr selten	selten	gelegentlich	oft	sehr oft	(fast) immer	keine Angabe
<input type="radio"/>							

Bitte geben Sie an, an wie vielen Tagen Sie in einer durchschnittlichen Woche in etwa Alkohol trinken.

nie	nicht wöchentlich	an 1-2 Tagen	an 3-4 Tagen	an 5-6 Tagen	jeden Tag	keine Angabe
<input type="radio"/>						

Wenn ich Alkohol trinke, dann behalte ich die Kontrolle über mein Verhalten.

nie	selten	oft	(fast) immer	keine Angabe
<input type="radio"/>				

[Zurück](#)[Weiter](#)



BERGISCHE
UNIVERSITÄT
WUPPERTAL

Seite 7

60% ausgefüllt

Im Rahmen dieser Befragung möchten wir auch den persönlichen Hintergrund der Studentinnen und Studenten kennenlernen. In diesem Kontext geht es darum, Bedürfnisstrukturen zu betrachten. Die folgenden Fragen beziehen sich daher auf Ihr **bisheriges Leben** und Ihre **Einstellungen**. Auch wenn Ihnen die Beantwortung einiger der folgenden Fragen möglicherweise unangenehm ist, würden wir uns über Ihre Antwort freuen. Sollten Sie sich bei einer Antwort nicht sicher sein, wählen Sie bitte die Antwortkategorie aus, die Ihnen am ehesten korrekt erscheint. Bitte beantworten Sie die Fragen ehrlich und vollständig.

Bitte beantworten Sie die folgenden Fragen in Bezug auf Ihr Handeln in der Vergangenheit.

	nein	ja	keine Angabe
Haben Sie jemals harte Drogen (z.B. Amphetamine, Kokain, Heroin) konsumiert (<i>hiermit sind nicht z.B. Alkohol/Marihuana gemeint</i>)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Haben Sie jemals alkoholisiert ein Fahrzeug geführt, wenn Sie eigentlich nicht mehr hätten fahren dürfen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Haben Sie sich jemals krankschreiben lassen, ohne wirklich krank zu sein?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Haben Sie schon einmal für eine sexuelle Handlung bezahlt oder Geld angenommen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Waren Sie Ihrer/m Partner*in oder einer/m früheren Partner*in stets treu?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Haben Sie sich jemals am nächsten Tag für das, was Sie unter Alkoholeinfluss getan haben, geschämt?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Haben Sie jemals Waren/Gegenstände aus einem Geschäft mitgenommen, ohne dafür zu bezahlen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter



70% ausgefüllt

Menschliche Einstellungen und Persönlichkeitsstrukturen spiegeln sich in bestimmten Verhaltensweisen wider. Die folgenden Aussagen zu Ihrem **Sozialverhalten** können auf Sie selbst mehr oder weniger zutreffen. Bitte geben Sie bei jeder Aussage an, wie sehr die Aussage auf Sie zutrifft. Auch wenn Ihnen die Beantwortung einiger der folgenden Fragen möglicherweise unangenehm ist, würden wir uns über Ihre Antwort freuen. Sollten Sie sich bei einer Antwort nicht sicher sein, wählen Sie bitte die Antwortkategorie aus, die Ihnen am ehesten korrekt erscheint. Bitte beantworten Sie die Fragen ehrlich und vollständig.

Die folgenden Aussagen können auf Sie selbst mehr oder weniger zutreffen. Bitte geben Sie bei jeder Aussage an, wie sehr die Aussage auf Sie zutrifft.

	trifft überhaupt nicht zu	trifft nicht zu	trifft eher nicht zu	unentschieden	trifft eher zu	trifft zu	trifft vollständig zu	keine Angabe
Wenn ich etwas versprochen habe, dann halte ich es auch.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn ich mich mit jemandem unterhalte, höre ich ihm immer aufmerksam zu.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Im Streit bleibe ich stets sachlich und objektiv.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich akzeptiere alle anderen Meinungen, auch wenn sie mit meiner eigenen nicht übereinstimmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn ich unterwegs bin, werfe ich Müll einfach in die Landschaft oder auf die Straße.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich zögere niemals, jemandem in einer Notlage beizustehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich helfe jemandem nur, wenn ich eine Gegenleistung erwarten kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich ernähre mich stets gesund.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Es kommt vor, dass ich andere bewusst ausnutze.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Auch wenn ich selbst gestresst bin, behandle ich andere immer freundlich und zuvorkommend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter



80% ausgefüllt

Ein letzter wichtiger Bereich der Work-Life-Balance betrifft neben dem Freizeitverhalten auch das **familiäre** und **persönliche Wohlbefinden**. Steigende universitäre Belastung und sonstige Verpflichtungen sorgen dafür, dass hier häufig wenig Zeit für bestimmte Aktivitäten bleibt. Auch wenn Ihnen die Beantwortung einiger der folgenden Fragen möglicherweise unangenehm ist, würden wir uns über Ihre Antwort freuen. Sollten Sie sich bei einer Antwort nicht sicher sein, wählen Sie bitte die Antwortkategorie aus, die Ihnen am ehesten korrekt erscheint. Bitte beantworten Sie die Fragen ehrlich und vollständig.

Bitte wählen Sie Ihren Familienstand aus.

[Bitte auswählen] ▾

Wie zufrieden sind Sie insgesamt mit Ihrem Privatleben?

sehr unzufrieden unzufrieden eher unzufrieden unentschieden eher zufrieden zufrieden sehr zufrieden keine Angabe

Wie häufig treiben Sie in einer durchschnittlichen Woche Sport?

Falls Sie keinen Sport treiben, geben Sie bitte „0“ an.

mal keine Angabe

Zum Abschluss der Fragen zu Ihrem Privatleben möchten wir Ihnen noch eine Frage zu Ihrem **Sexualleben** stellen. Uns ist bewusst, dass dies ein sehr privates und persönliches Thema ist. Gerade daher würden wir uns über Ihre ehrliche Antwort freuen.

Wie häufig hatten Sie im *letzten Monat* Geschlechtsverkehr?

mal keine Angabe

Zurück

Weiter



90% ausgefüllt

Zum Abschluss der Online-Befragung (Befragung Teil 1) möchten wir Ihnen gerne noch zwei Fragen zu Ihrer **Wahrnehmung der Befragung** stellen. Lesen Sie bitte die beiden folgenden Sätze und bestimmen Sie bei jeder Aussage, wie sehr die Aussage auf Sie zutrifft. Sollten Sie sich bei einer Antwort nicht sicher sein, wählen Sie bitte die Antwortkategorie aus, die Ihnen am ehesten korrekt erscheint. Bitte beantworten Sie die Fragen ehrlich und vollständig.

Ich bin sicher, dass diese Umfrage vertraulich ist.

trifft überhaupt nicht zu	trifft nicht zu	trifft eher nicht zu	unentschieden	trifft eher zu	trifft zu	trifft vollständig zu
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ich bin sicher, dass diese Umfrage anonym ist.

trifft überhaupt nicht zu	trifft nicht zu	trifft eher nicht zu	unentschieden	trifft eher zu	trifft zu	trifft vollständig zu
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[Zurück](#)[Weiter](#)

Vielen Dank für Ihre bisherige Teilnahme!

Der erste Teil der Befragung ist damit beendet. Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

Bitte informieren Sie den/die Versuchsleiter*in, dass Sie den Online-Teil der Befragung abgeschlossen haben.

Mats Schumacher, M.Sc.
Lehrstuhl für Methoden der empirischen Wirtschafts- und Sozialforschung
Bergische Universität Wuppertal

Wie wahrscheinlich ist es, dass Sie in Zukunft ein persönliches Smartphone nutzen werden?

Gar nicht wahr- scheinlich 1	2	3	4	5	6	Sehr wahr- scheinlich 7
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wenn Sie die freie Wahl hätten, wie hätten Sie den ersten Befragungsteil am liebsten ausgefüllt?

Bitte kreuzen Sie den entsprechenden Modus an.

<input type="checkbox"/> mit Laptop/PC	<input type="checkbox"/> als Papier- Fragebogen	<input type="checkbox"/> mit dem Smartphone
<input type="checkbox"/> mit einem Tablet	<input type="checkbox"/> Sonstiges: _____	

Wie würden Sie die Dauer der Befragung einschätzen?

Sehr kurz 1	2	3	4	Sehr lang 5
<input type="checkbox"/>				

Wie interessant fanden Sie die Befragung?

Gar nicht inte- ressant 1	2	3	4	Sehr interes- sant 5
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie würden Sie die Befragung insgesamt bewerten?

Sehr schlecht 1	2	3	4	Sehr gut 5
<input type="checkbox"/>				

Bitte wählen Sie Ihr Geschlecht aus.

männlich

weiblich

sonstiges

Bitte geben Sie Ihr Alter an.

Jahre

Wenn Sie ansonsten noch Anmerkungen (Anregungen, Kritik etc.) haben, können Sie diese im folgenden Feld eintragen.

Vielen Dank für Ihre Teilnahme. Wenn Sie an der Verlosung der drei Amazon-Gutscheine teilnehmen möchten, tragen Sie bitte Ihre E-Mail Adresse auf dem vorgefertigten Bogen, der separat abgeheftet und aufbewahrt wird, ein. Sollten Sie noch Fragen haben, wenden Sie sich gerne an die Versuchsleitung.

Anhang B. Fragebogen der zweiten Studie

Seite 1



0% ausgefüllt

Herzlich willkommen zur Studie zum Thema **Fragebogen-Design**. Vielen Dank, dass Sie sich die Zeit genommen haben, an dieser Untersuchung teilzunehmen.

Die Wissenschaft ist der „*Motor für Innovationen und damit Triebfeder für die wirtschaftliche Entwicklung*“¹. Wissenschaftliche Forschung ist also wichtig, um in unserer Gesellschaft auch in Zukunft die Lebensbedingungen optimieren zu können. Wesentliche Grundlage hierfür ist die Erhebung von Daten mit Hilfe von Umfragen. Bedauerlicherweise kommt es bei Umfragen aber häufig zu Problemen. Geringes Interesse und geringe Motivation führen dazu, dass Probanden beispielsweise die erstbeste Antwort wählen, komplett willkürlich antworten oder erst gar nicht an Befragungen teilnehmen. Ziel dieser Studie ist es, Bedürfnisse und Wünsche der Teilnehmer an Umfragen aufzudecken, um die Fragebogenkonstruktion künftig an diesen ausrichten zu können. Außerdem bin ich an Ihrer Wahrnehmung und Ihren Präferenzen hinsichtlich verschiedener Ausfüllgeräte interessiert.

Die Beantwortung der Studie wird etwa 8-10 Minuten in Anspruch nehmen. Bitte beantworten Sie die Fragen so aufmerksam und gewissenhaft wie möglich.

Beachten Sie, dass die Teilnahme an der Untersuchung freiwillig ist und Sie die Befragung jederzeit ohne Angabe von Gründen abbrechen können, ohne dass Ihnen Nachteile entstehen. Die Untersuchung ist vollkommen anonym. Ihre Angaben können von niemandem und zu keiner Zeit mit Ihrer Person in Verbindung gebracht werden. Entsprechend der neuen Datenschutzgrundverordnung werden Ihre Angaben und Daten gewissenhaft geschützt.

¹Konrad-Adenauer-Stiftung

Hiermit stimme ich zu, dass meine Angaben im Rahmen der aktuellen Datenschutzgrundverordnung (DSGVO) zu rein wissenschaftlichen Zwecken verwendet werden dürfen.

Ich stimme zu.

[Weiter](#)

Mats Schumacher, M. Sc., Bergische Universität Wuppertal – 2019

Seite 2



7% ausgefüllt

Sollten Sie mit dem Smartphone an dieser Umfrage teilnehmen, bitte ich Sie an dieser Stelle, das Smartphone zu drehen und in den Landscape-Modus zu wechseln. Dadurch wird eine übersichtliche Darstellungsform sichergestellt.

[Zurück](#)[Weiter](#)



13% ausgefüllt

Zunächst möchte ich gerne überprüfen, ob Sie die Voraussetzungen für die Teilnahme an dieser Studie erfüllen. Bitte beantworten Sie dazu diese Fragen.

Ich besitze mindestens ein Smartphone.

- Ja
 Nein

Ich besitze mindestens einen PC/Laptop.

- Ja
 Nein

Ich besitze mindestens einen Smart-TV.

- Ja
 Nein

Ich besitze mindestens ein Tablet.

- Ja
 Nein

Zurück

Weiter



20% ausgefüllt

Die folgenden Fragen beziehen sich auf Ihre Wahrnehmung bei der Teilnahme an wissenschaftlichen Studien. Bitte geben Sie an, wie sich die Ausfüllmedien PC/Laptop vs. Smartphone bezogen auf die folgenden Kriterien Ihrer Meinung nach unterscheiden. Sie müssen sich nicht für ein Gerät entscheiden, sondern können den Schieberegler entsprechend Ihrer Wahrnehmung positionieren, wobei jede Reglerposition eine gültige Antwort darstellt. Trifft ein Kriterium beispielsweise stärker, aber nicht ausschließlich auf das Smartphone zu, wählen Sie eine Position, die entsprechend Ihrer Wahrnehmung näher Richtung Smartphone liegt.

Auf welchem Gerät ist die Beantwortung eines Fragebogens Ihrer Ansicht nach zeitaufwendiger?

PC/Laptop Smartphone

Auf welchem Gerät empfinden Sie bei der Beantwortung eines Fragebogens eine höhere Sicherheit Ihrer Daten?

PC/Laptop Smartphone

Auf welchem Gerät empfinden Sie bei der Beantwortung eines Fragebogens eine höhere Vertraulichkeit Ihrer Angaben im Fragebogen?

PC/Laptop Smartphone

Auf welchem Gerät empfinden Sie bei der Beantwortung eines Fragebogens eine höhere Anonymität Ihrer Daten?

PC/Laptop Smartphone

[Zurück](#)[Weiter](#)

Anmerkung: Die Abfolge der Fragebogenseiten 4 und 5 wurde randomisiert.



27% ausgefüllt

Jetzt interessieren mich Ihre Bedürfnisse bei der Beantwortung von Fragebögen. Bitte ordnen Sie dazu die folgenden Aussagen Ihren Bedürfnissen entsprechend ein.

Mir ist wichtig, dass	überhaupt nicht wichtig	nicht wichtig	weder noch	wichtig	sehr wichtig
	1	2	3	4	5
die Beantwortung eines Fragebogens schnell geht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ich den Fragebogen jederzeit ausfüllen kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
eine Umfrage anonym ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
eine Umfrage vertraulich ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ich mich bei der Hinterlegung privater Daten in einer Umfrage sicher fühle.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ich den Fragebogen ortsungebunden ausfüllen kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter



33% ausgefüllt

Als Nächstes interessiert mich Ihre Einstellung zu (neuen) Technologien im Allgemeinen. Bitte geben Sie hier an, inwieweit Sie den folgenden Aussagen zustimmen.

	stimme überhaupt nicht zu	stimme nicht zu	stimme eher nicht zu	unent- schieden	stimme eher zu	stimme zu	stimme vollständig zu
	1	2	3	4	5	6	7
Ich fühle mich in der Lage, über neue Technologien auf dem Laufenden zu bleiben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin mir sicher, dass ich den Umgang mit neuen Technologien lernen kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter



40% ausgefüllt

Bei den untenstehenden Fragen geht es nun um eine bestimmte Technologie. Bitte beantworten Sie einige Fragen zu Ihrem Nutzungsverhalten bezogen auf Ihr persönliches **Smartphone**. Geben Sie dafür Ihre Zustimmung zu den folgenden Aussagen an.

	stimme überhaupt nicht zu 1	stimme nicht zu 2	stimme eher nicht zu 3	unent- schieden 4	stimme eher zu 5	stimme zu 6	stimme vollständig zu 7
Ich fühle mich wohl, wenn ich mein Smartphone in meiner Hand halte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn ich angespannt oder aufgebracht bin, hilft mir die Benutzung meines Smartphones dabei, mich zu entspannen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn ich sehe, dass andere Menschen ihr Smartphone benutzen, möchte ich mein Smartphone auch benutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Benutzung meines Smartphones hilft mir dabei, mit einer reizüberfluteten Umgebung umzugehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Benutzung meines Smartphones hilft mir dabei, mich in sozialen Situationen wohl zu fühlen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn mein Akku leer ist, ist es beinahe unerträglich, bis ich mein Smartphone aufladen kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter



47% ausgefüllt

Bei den untenstehenden Fragen geht es nun um eine bestimmte Technologie. Bitte beantworten Sie einige Fragen zu Ihrem Nutzungsverhalten bezogen auf Ihren persönlichen **PC/Laptop**. Geben Sie dafür Ihre Zustimmung zu den folgenden Aussagen an.

	stimme überhaupt nicht zu 1	stimme nicht zu 2	stimme eher nicht zu 3	unent- schieden 4	stimme eher zu 5	stimme zu 6	stimme vollständig zu 7
Wenn ich meinen PC/Laptop nicht zur Verfügung habe, ist es beinahe unerträglich, bis ich ihn wieder nutzen kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Benutzung meines PCs/Laptops hilft mir dabei, mit einer reizüberfluteten Umgebung umzugehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Benutzung meines PCs/Laptops hilft mir dabei, mich in sozialen Situationen wohl zu fühlen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, wenn ich am PC/Laptop sitze.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn ich sehe, dass andere Menschen ihren PC/Laptop benutzen, möchte ich meinen PC/Laptop auch benutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn ich angespannt oder aufgebracht bin, hilft mir die Benutzung meines PCs/Laptops dabei, mich zu entspannen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter

Anmerkung: Die Abfolge der Fragebogenseiten 7 und 8 wurde randomisiert.



53% ausgefüllt

Die folgenden Fragen beziehen sich auf das **Gerät, mit dem Sie diesen Online-Fragebogen beantworten**. Bitte geben Sie erneut Auskunft über Ihre Zustimmung zu den Aussagen.

	stimme überhaupt nicht zu 1	stimme nicht zu 2	stimme eher nicht zu 3	unent- schieden 4	stimme eher zu 5	stimme zu 6	stimme vollständig zu 7
Ich finde, dass mein Ausfüllgerät einfach zu benutzen ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich sehr selbstsicher, wenn ich mein Ausfüllgerät benutze.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich finde, dass die Nutzung meines Ausfüllgerätes umständlich ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe sehr schnell gelernt, wie man das Ausfüllgerät benutzt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich finde, dass mein Ausfüllgerät unnötig komplex ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter



60% ausgefüllt

Die folgenden Fragen beziehen sich auf Ihre Person.

Bitte geben Sie Ihr Geschlecht an.

- männlich
 weiblich

Bitte geben Sie Ihr Alter an.

Jahre

Zurück

Weiter



67% ausgefüllt

Welchen höchsten allgemeinen Schulabschluss haben Sie?

- Hauptschulabschluss
- Realschulabschluss
- Fachhochschulreife
- Allgemeine Hochschulreife (Abitur)
- Ich habe keinen Schulabschluss.
- Sonstiges

Sind Sie derzeit erwerbstätig? (Mehrfachauswahl möglich)**Ich bin**

- erwerbstätig in Vollzeit.
- erwerbstätig in Teilzeit.
- geringfügig beschäftigt (450€-Basis).
- Auszubildende/r.
- Schüler/in oder Student/in.
- zur Zeit nicht erwerbstätig.
- Sonstiges

Zurück

Weiter



73% ausgefüllt

Wie viele Personen leben in Ihrem Haushalt, Sie selbst eingeschlossen? Personen**Wie hoch ist Ihr persönliches monatliches Nettoeinkommen?** €

Zurück

Weiter



80% ausgefüllt

Auf welchem Gerät haben Sie diese Umfrage erstmalig geöffnet?

PC/Laptop

Smartphone

Tablet

Sonstiges, und zwar

Auf welchem Gerät haben Sie die Umfrage beantwortet?

PC/Laptop

Smartphone

Tablet

Sonstiges, und zwar

Zurück

Weiter



93% ausgefüllt

Bitte geben Sie abschließend kurz an, wie Sie die folgenden Aspekte dieser Umfrage wahrgenommen haben, indem Sie Ihre Zustimmung zu den Aussagen bewerten.

	stimme überhaupt nicht zu 1	stimme zu 2	nichtstimme nicht zu 3	eher nicht zu 4	unent- schieden 5	stimme eher zu 6	stimme vollständig zu 7
Diese Befragung war zeitaufwendig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe mich sicher gefühlt, private Daten bei dieser Umfrage zu hinterlegen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin mir sicher, dass diese Umfrage anonym ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin mir sicher, dass diese Umfrage vertraulich ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter



Vielen Dank für Ihre Teilnahme!

Sie können den Fragebogen jetzt schließen. Untenstehend finden Sie noch eine kurze Information über den Befragungshintergrund.

Tatsächlich stehen nicht nur die Bedürfnisse und Erwartungen für die Teilnahme an Befragungen im Forschungsfokus. Ein weiteres wesentliches Ziel der Studie ist es, ausgewählte und möglicherweise unbewusste, verhaltensbezogene Determinanten bezüglich Ihrer Bedeutung für die Wahl des Ausfüllgerätes dieser Befragung zu untersuchen. Konkret soll geprüft werden, welche Bedürfnisse und Einstellungen der Teilnehmer dazu beitragen, dass die Umfrage beispielsweise mit einem Smartphone bzw. einem PC/Laptop ausgefüllt wird.

Sollten Sie weiterführende Fragen haben, können Sie mich gerne dazu kontaktieren (xxx@uni-wuppertal.de).

Anhang C. Einwilligungserklärung Studie 3

Einwilligungserklärung:

Hiermit bestätige ich, dass ich freiwillig an der Eye-Tracking Studie des Lehrstuhls für Methoden der empirischen Wirtschafts- und Sozialforschung zum Thema „Relevanz verschiedener Persönlichkeitseigenschaften in der Befragungsforschung“ teilnehme. Verantwortlicher Ansprechpartner seitens des Lehrstuhls ist Mats Schumacher (Schumacher@wiwi.uni-wuppertal.de). Mir ist bewusst, dass ich die Befragung jederzeit ohne Angabe von Gründen abbrechen kann, ohne dass mir Nachteile entstehen.

Ich stimme zu, dass im Rahmen der Studie Daten über mich gesammelt und aufgezeichnet werden dürfen. Diese werden zu rein wissenschaftlichen Zwecken verwendet. Bei der Veröffentlichung der Studie in wissenschaftlichen Medien wird aus den Daten nicht hervorgehen, wer an dieser Untersuchung teilgenommen hat. Entsprechend der europäischen Datenschutzgrundverordnung werden meine Angaben und Daten gewissenhaft geschützt.

Mit meiner Teilnahme an der Studie erkläre ich mich einverstanden, dass Videoaufnahmen von mir angefertigt werden. Genau wie meine weiteren Angaben werden diese streng vertraulich behandelt und sind nur den Projektmitarbeitern zugänglich. Die Videoaufnahmen werden lediglich lokal auf dem Untersuchungsrechner gespeichert und unmittelbar nach der Verarbeitung gelöscht. Meine Angaben werden in anonymisierter Form archiviert und können so von niemandem und zu keiner Zeit mit meiner Person in Verbindung gebracht werden.

Ich weiß, dass ich jederzeit meine Einwilligungserklärung, ohne Angabe von Gründen, widerrufen kann, ohne dass dies für mich nachteilige Folgen hat. Dies kann ich mit einer E-Mail mit dem Betreff "Widerruf" an wiso@wiwi.uni-wuppertal.de tun, die Rechtmäßigkeit der Verarbeitung wird durch den Widerruf nicht rückwirkend beseitigt.

Nach Abschluss des Forschungsprojektes werden alle meine im Rahmen des Projektes erhobenen personenbezogenen Daten gelöscht werden.

Meine Zustimmung zur Studienteilnahme und den obenstehenden Ausführungen bestätige ich

(Name, Vorname)

mit meiner Unterschrift.

Anhang D. Fragebogen der dritten Studie

Seite 1



0% ausgefüllt

Herzlich willkommen zu unserer Eye-Tracking-Studie des Lehrstuhls für empirische Wirtschafts- und Sozialforschung der Bergischen Universität Wuppertal. Wir möchten Ihnen danken, dass Sie sich die Zeit nehmen, an dieser Studie teilzunehmen. In der wissenschaftlichen Forschung kommt Online-Befragungen eine besondere Bedeutung zu, da sie eine einmalige Möglichkeit bieten, menschliche Gedanken und Einstellungen direkt abzufragen und orts- und zeitungebunden zu erfassen. Um eine möglichst hohe Teilnahmemotivation und valide Ergebnisse erreichen zu können, ist es dabei wichtig, die Umfragen/Studien an den Präferenzen und Bedürfnissen der potentiellen Teilnehmer auszurichten. Unsere Studie hat das Ziel, den Zusammenhang zwischen diesen Präferenzen und Bedürfnissen und der Persönlichkeit zu untersuchen, um in Zukunft eine größtmögliche Passung anstreben zu können. Durch das Eye-Tracking ist es möglich, die Erkenntnisse zudem mit Ihrem Blickverhalten als Indikator für Wahrnehmungs- und Verarbeitungsmechanismen zu koppeln. So könnten Studien in Zukunft individuelle, persönlichkeitsgetreue Schwerpunkte setzen und entsprechend die Teilnehmer passgenauer ansprechen und abholen. Als modernes Erhebungsinstrument nutzen wir in dieser Studie das Smartphone.

Im Verlauf der Befragung werden wir Sie daher unter anderem bitten, einige Fragen zu Ihrer eigenen Person zu beantworten. Anschließend bieten wir Ihnen an, dass Sie sich bei Bedarf Ihr aus Ihren Angaben erstelltes komprimiertes Persönlichkeitsprofil ausgeben lassen können. Dabei werden Ihre Angaben mit den Normwerten einer großen repräsentativen Eichstichprobe verglichen. Die Beantwortung der Fragen wird etwa **10-15 Minuten** in Anspruch nehmen. Bitte beantworten Sie die Fragen so aufmerksam und gewissenhaft wie möglich.

Hiermit stimme ich zu, dass meine Angaben im Rahmen der neuen Datenschutzgrundverordnung (DSGVO) zu rein wissenschaftlichen Zwecken verwendet werden dürfen.

Ich stimme zu.

[Weiter](#)

M.Sc. Mats Schumacher, Bergische Universität Wuppertal – 2019

Anmerkung: Der Einleitungstext war in der PC-Bedingung ebenfalls auf das Ausfüllgerät angepasst.



Seite 2

9% ausgefüllt

Zunächst würden wir Sie bitten, einige Fragen zu Ihrer eigenen Person zu beantworten.

Welches Geschlecht haben Sie?

- Männlich
- Weiblich
- Divers

Wie alt sind Sie?

Ich bin Jahre alt.

Bitte wählen Sie den höchsten Bildungsabschluss aus, über den Sie verfügen.

- Ich habe keinen Schulabschluss.
- Hauptschulabschluss
- Mittlere Reife
- Fachabitur
- Abitur
- Bachelor
- Master
- Promotion
- Sonstiges:

Wie hoch ist ungefähr Ihr monatliches Nettoeinkommen?

Gemeint ist der Betrag, der sich aus allen Einkünften zusammensetzt und nach Abzug der Steuern und Sozialversicherungen übrig bleibt.

[Bitte auswählen] ▼

Weiter



BERGISCHE
UNIVERSITÄT
WUPPERTAL

18% ausgefüllt

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Würden Sie über sich z. B. sagen, dass Sie gerne Zeit mit anderen Menschen verbringen? Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils, teils	Stimme eher zu	Stimme voll und ganz zu
Ich gehe aus mir heraus, bin gesellig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin einfühlsam, warmherzig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin eher unordentlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bleibe auch in stressigen Situationen gelassen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin nicht sonderlich kunstinteressiert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin durchsetzungsfähig, energisch.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich begegne anderen mit Respekt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin bequem, neige zu Faulheit.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bleibe auch bei Rückschlägen zuversichtlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin vielseitig interessiert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitte überprüfen Sie es sorgfältig: Haben Sie jede Aussage beantwortet?

Weiter



BERGISCHE
UNIVERSITÄT
WUPPERTAL

27% ausgefüllt

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Würden Sie über sich z. B. sagen, dass Sie gerne Zeit mit anderen Menschen verbringen? Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils, teils	Stimme eher zu	Stimme voll und ganz zu
Ich schäume selten vor Begeisterung über.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich neige dazu, andere zu kritisieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin stetig, beständig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann launisch sein, habe schwankende Stimmungen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin erfinderisch, mir fallen raffinierte Lösungen ein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin eher ruhig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe mit anderen wenig Mitgefühl.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin systematisch, halte meine Sachen in Ordnung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich reagiere leicht angespannt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann mich für Kunst, Musik und Literatur begeistern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitte überprüfen Sie es sorgfältig: Haben Sie jede Aussage beantwortet?

Weiter



36% ausgefüllt

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Würden Sie über sich z. B. sagen, dass Sie gerne Zeit mit anderen Menschen verbringen? Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils, teils	Stimme eher zu	Stimme voll und ganz zu
Ich neige dazu, die Führung zu übernehmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe oft Streit mit anderen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich neige dazu, Aufgaben vor mir herzuschieben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin selbstsicher, mit mir zufrieden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich meide philosophische Diskussionen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin weniger aktiv und unternehmungslustig als andere.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin nachsichtig, verzeihe anderen leicht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin manchmal ziemlich nachlässig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin ausgeglichen, nicht leicht aus der Ruhe zu bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin nicht besonders einfallsreich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitte überprüfen Sie es sorgfältig: Haben Sie jede Aussage beantwortet?

Weiter



45% ausgefüllt

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Würden Sie über sich z. B. sagen, dass Sie gerne Zeit mit anderen Menschen verbringen? Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils, teils	Stimme eher zu	Stimme voll und ganz zu
Ich bin eher schüchtern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin hilfsbereit und selbstlos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich mag es sauber und aufgeräumt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich mache mir oft Sorgen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich weiß Kunst und Schönheit zu schätzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mir fällt es schwer, andere zu beeinflussen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin manchmal unhöflich und schroff.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin effizient, erledige Dinge schnell.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich oft bedrückt, freudlos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Es macht mir Spaß, gründlich über komplexe Dinge nachzudenken und sie zu verstehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitte überprüfen Sie es sorgfältig: Haben Sie jede Aussage beantwortet?

Weiter



BERGISCHE
UNIVERSITÄT
WUPPERTAL

55% ausgefüllt

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Würden Sie über sich z. B. sagen, dass Sie gerne Zeit mit anderen Menschen verbringen? Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils, teils	Stimme eher zu	Stimme voll und ganz zu
Ich bin voller Energie und Tatendrang.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin anderen gegenüber misstrauisch.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin verlässlich, auf mich kann man zählen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe meine Gefühle unter Kontrolle, werde selten wütend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin nicht sonderlich fantasievoll.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin gesprächig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Andere sind mir eher gleichgültig, egal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin eher der chaotische Typ, mache selten sauber.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich werde selten nervös und unsicher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich finde Gedichte und Theaterstücke langweilig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitte überprüfen Sie es sorgfältig: Haben Sie jede Aussage beantwortet?

Weiter



BERGISCHE
UNIVERSITÄT
WUPPERTAL

64% ausgefüllt

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Würden Sie über sich z. B. sagen, dass Sie gerne Zeit mit anderen Menschen verbringen? Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils, teils	Stimme eher zu	Stimme voll und ganz zu
In einer Gruppe überlasse ich lieber anderen die Entscheidung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin höflich und zuvorkommend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bleibe an einer Aufgabe dran, bis sie erledigt ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin oft deprimiert, niedergeschlagen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mich interessieren abstrakte Überlegungen wenig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin begeisterungsfähig und kann andere leicht mitreißen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Manchmal verhalte ich mich verantwortungslos, leichtsinnig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich reagiere schnell gereizt oder genervt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin originell, entwickle neue Ideen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitte überprüfen Sie es sorgfältig: Haben Sie jede Aussage beantwortet?

Weiter



73% ausgefüllt

Bevor wir zur Auswertung Ihres Persönlichkeitsprofils kommen, würden wir Sie bitten noch einige Fragen zu Ihren Erfahrungen und Bedürfnissen bezogen auf die Teilnahme an wissenschaftlichen Erhebungen zu beantworten.

Haben Sie zuvor schon an wissenschaftlichen Untersuchungen (z.B. Befragungen, Experimente) teilgenommen?

- Ja
 Nein

Falls ja, an wie vielen wissenschaftlichen Befragungen/Experimenten haben Sie in etwa in den letzten zwölf Monaten teilgenommen?

[Bitte auswählen]

Haben Sie zuvor schon an einer Eye-Tracking-Studie teilgenommen?

- Ja
 Nein

Wenn Sie die freie Wahl hätten, mit welchem Ausfüllgerät würden Sie am liebsten an einer Online-Befragung teilnehmen?

- Mit einem Smartphone
 Mit einem Tablet
 Mit einem Laptop
 Mit einem PC
 Mit einem anderen Gerät und zwar:

Wie sollte eine Online-Befragung aufgebaut bzw. gestaltet sein, damit Sie motiviert wären, teilzunehmen?

Hier können Sie auch in Stichworten antworten.

Möchten Sie sich ihr komprimiertes Persönlichkeitsprofil anzeigen lassen?

Diese Profil wird nicht gespeichert, sondern nur temporär für Sie erzeugt.

- Ja
 Nein

Weiter



Die Auswertung Ihres Persönlichkeitsprofils basiert auf der aktuellen Version des etablierten „Big Five Inventars“ (BF1 2) von Danner und Kollegen (2019). Dabei werden Ihre Angaben bezogen auf die fünf Persönlichkeitsdimensionen **Negative Emotionalität** (früher Neurotizismus), **Extraversion**, **Offenheit**, **Verträglichkeit** und **Gewissenhaftigkeit** zunächst aggregiert und anschließend zu den Werten einer repräsentativen Bevölkerungsstichprobe in Relation gesetzt. Dadurch kann jeweils eine Aussage über die Ausprägung der jeweiligen Dimension bezogen auf Ihre Person getroffen werden. Sollten Sie Ihr Geschlecht als männlich oder weiblich angegeben haben, bezieht sich die Auswertung auf eine Vergleichsgruppe bestehend aus Personen mit Ihrem angegebenen Geschlecht. Falls Sie Ihr Geschlecht als divers angegeben haben oder keine Angabe gemacht haben, dient die Gesamtbevölkerung als Referenzgruppe.

Beachten Sie, dass es sich bei dem untenstehenden Profil um ein komprimiertes Profil handelt, das weder einen Vollständigkeits- noch einen Wahrheitsanspruch erhebt. Es soll lediglich eine grobe Orientierung hinsichtlich der angesprochenen fünf Dimensionen basierend auf Ihren Angaben bieten. Alle untenstehenden Aussagen wurden dabei automatisiert Ihrer Selbstbeschreibung entsprechend generiert – darüber hinaus wurde keine Bewertung vorgenommen und auch keine weiteren potentiellen Bewertungskriterien einbezogen.

Extraversion:

Unter Extraversion wird in der Persönlichkeitsforschung verstanden, inwiefern das eigene Handeln und Denken nach außen gerichtet ist. Dies bezieht sich insbesondere auf den Austausch und das Handeln in sozialen Gruppen. Ist Extraversion gering ausgeprägt, wird von Introversion gesprochen.

Ihr Profil: Sie sind verglichen mit anderen Personen eher zurückhaltend im Umgang mit Menschen. Sie schätzen Ihre Unabhängigkeit und genießen es, auch mal allein sein zu können. In der Interaktion mit Gruppen nehmen Sie sich eher als andere Personen zurück.

Verträglichkeit:

Mit Verträglichkeit wird das interpersonelle Verhalten im Allgemeinen beschrieben. Somit bezieht sich Verträglichkeit unter anderem auf das Konfliktverhalten und Konformitätsstreben.

Ihr Profil: In Relation zu anderen Personen agieren Sie verhältnismäßig eher wettbewerbsorientiert. Es ist Ihnen tendenziell wichtiger, Ihre eigenen Interessen zu schützen, als um jeden Preis nach Harmonie zu streben. Sie versuchen tendenziell stärker als andere Personen, die Absicht im Handeln Ihres Gegenübers zu erkennen.

Negative Emotionalität:

Negative Emotionalität beschreibt wie Emotionen, insbesondere negative Emotionen, erlebt und verarbeitet werden. Dies meint, wie stark und häufig emotionale Reize sein müssen, damit sie sich auf den Gemütszustand auswirken.

Ihr Profil: Verglichen mit anderen Personen sind Sie eher ausgeglichen und erleben negative Emotionen weniger intensiv. Sie sind eher ruhig und kontrolliert, auch in Stresssituationen bleiben Sie meist eher gelassen.

Offenheit:

Die Dimension Offenheit beschreibt Interesse an neuen Erfahrungen, Erlebnissen und Eindrücken und in welchem Ausmaß diese gesucht werden. Sie beinhaltet eine Auseinandersetzung mit der Struktur bevorzugter Denk- und Verhaltensmuster.

Ihr Profil: Im Vergleich zu anderen Personen ziehen Sie Bekanntes und Bewährtes dem Neuen und Unbekannten eher vor. Konventionelle und etablierte Denk- und Verhaltensmuster sprechen Sie eher an als sprunghafte und ausgefallene. Sie sind verglichen mit anderen Personen tendenziell weniger künstlerisch interessiert und schätzen Routine im Alltag.

Gewissenhaftigkeit:

In der Dimension Gewissenhaftigkeit werden Persönlichkeitsaspekte wie Selbstkontrolle, Zielstrebigkeit und Organisationsfähigkeit zusammengefasst. Sie beschreibt Fokussierungs- und Konzentrationsfähigkeiten.

Ihr Profil: Im Vergleich zu anderen Personen verfolgen Sie Ihre Ziele eher gelassen. Ordentlichkeit und Organisation sind Ihnen nicht übermäßig wichtig. Phasen hoher Konzentration und Anstrengung halten Sie lieber kurz. Für Sie ist Penibilität bei der Lösung von Aufgaben vergleichsweise weniger wichtig als für andere Personen.

Weiter

Anmerkung: Darstellung eines beispielhaften Persönlichkeitsprofils



91% ausgefüllt

Abschließend würden wir Sie bitten, die Studie zu bewerten.

Gesamtbewertung

sehr schlecht 1 2 3 4 5 sehr gut

Wie würden Sie die Befragung insgesamt bewerten?

Interesse

gar nicht
interessant 1 2 3 4 5 sehr
interessant

Wie interessant fanden Sie die Befragung?

Dauer

sehr kurz 1 2 3 4 5 6 7 sehr lang

Wie würden Sie die Dauer der Befragung einschätzen?

Weiter



Vielen Dank für Ihre Teilnahme!

Die Befragung ist hiermit beendet. Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

Bitte informieren Sie den/die Versuchsleiter*in, dass Sie die Befragung abgeschlossen haben.

Anhang E. Weitere Analyseergebnisse der Studie 3

Ergebnisse der abhängigen *t*-Tests für die Beurteilung der Unterschiedlichkeit der Areas of Interest in Studie 3:

Variable	Smartphone SP_weit vs SP_eng			PC PC_weit vs. PC_eng		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>
BlickzeitAOI	6,453	53	* < 0,001	3,317	54	* 0,002
FixationenAOI	4,049	53	* < 0,001	1,659	54	0,103
Rückkehr	-7,511	53	* < 0,001	-4,136	54	* < 0,001
Refixationen	-7,790	53	* < 0,001	-2,512	54	* 0,015
Fixationszahl	6,363	53	* < 0,001	4,509	54	* < 0,001
Fixationsdauer	-5,685	53	* < 0,001	-4,141	54	* < 0,001

* = signifikant ($p < 0,05$)

Vollständige Ergebnistabelle der Zusatzanalysen zur Freitextantwort in Studie 3:

	Schritt 1: PC_weit vs. SP_eng		Schritt 2: PC_eng vs. SP_eng		Schritt 3: PC_weit vs. SP_weit	
	<i>t</i> (<i>df</i>)	<i>p</i>	<i>t</i> (<i>df</i>)	<i>p</i>	<i>t</i> (<i>df</i>)	<i>p</i>
BlickzeitAOI	-15,136 _(50,49)	* < 0,001	-15,290 _(50,55)	* < 0,001	-15,970 _(48,14)	* < 0,001
FixationenAOI	-10,035 _(80,50)	* < 0,001	-10,104 _(80,61)	* < 0,001	-10,621 _(78,79)	* < 0,001
Rückkehr	<u>4,375</u> ₍₈₄₎	<u>* < 0,001</u>	4,471 ₍₈₄₎	* < 0,001	5,798 ₍₈₄₎	* < 0,001
Refixationen	-0,504 ₍₈₄₎	0,616	-0,497 ₍₈₄₎	0,620	-0,091 ₍₈₄₎	0,928
Fixationszahl	<u>-7,558</u> _(62,62)	<u>* < 0,001</u>	-7,627 _(62,10)	* < 0,001	-9,047 _(66,21)	* < 0,001
Fixationsdauer	<u>4,032</u> ₍₈₄₎	<u>* < 0,001</u>	4,100 ₍₈₄₎	* < 0,001	4,036 ₍₈₄₎	* < 0,001

Für *t*-Werte, deren Freiheitsgradzahl mindestens eine Dezimalstelle besitzt, liegt keine Varianzhomogenität für die AV vor, sodass hier der *t*-Wert des Welch-Tests berichtet wird.

* = signifikant ($p < 0,05$); unterstrichen = im Haupttext berichtet

Anhang F. Stichprobenbeschreibung des Gesamtdatensatzes der Studie 4

Für den *Gesamtdatensatz* lag der Altersschnitt bei 32,92 Jahren ($SD = 13,41$; Altersspanne = 16 bis 78 Jahre). 64,67% ($n = 443$) gaben ihr Geschlecht als weiblich, 33,14% als männlich ($n = 227$) und 0,58% ($n = 4$) als divers an (11 fehlende Werte). Damit nahmen signifikant mehr weibliche als männliche Personen teil ($\chi^2(1) = 69,636, p < 0,001$). Im *Gesamtdatensatz* nutzten 340 Proband*innen den PC/Laptop (49,64%), 303 Teilnehmer*innen das Smartphone (44,23%) und 42 das Tablet für die Teilnahme (6,13%). Tablet-Teilnehmer*innen werden für Zusatzauswertungen im Datensatz belassen, gehen aber in die Hypothesenprüfung des Geräteffekts für CR nicht ein. Die Geräthäufigkeit der Untersuchungsgruppen (Smartphone vs. PC/Laptop) unterscheidet sich nicht ($\chi^2(1) = 2,129, p = 0,145$) und auch das Geschlecht verteilt sich auf beide Gerätgruppen gleich ($\chi^2(1) = 2,962, p = 0,085$).

Anhang G. Fragebogen der vierten Studie

Seite 1



0% ausgefüllt

Herzlich willkommen zur Studie mit dem Thema „Umweltbewusstsein, -einstellungen und persönliche Werte“ des Lehrstuhls für Methoden der empirischen Wirtschafts- und Sozialforschung der Bergischen Universität Wuppertal. Wir danken Ihnen, dass Sie sich die Zeit nehmen, an dieser Befragung teilzunehmen.

Umwelt- und Klimaschutz gehören sowohl auf institutioneller als auch auf persönlicher Ebene zu den größten Herausforderungen der heutigen Zeit. Diesem Thema wurde in den vergangenen Monaten und Jahren – auch durch die Fridays for Future Bewegung – viel Aufmerksamkeit geschenkt. Aktuell ist noch unklar, ob die Relevanz und Wahrnehmung diese Themas durch die Corona-Pandemie beeinflusst wird oder nicht.

Ziel dieser Untersuchung ist es daher, einen Einblick in die Zusammenhänge zwischen Umweltverhalten und -bewusstsein auf der einen und individuellen Bedürfnissen, Werten und persönlichen Charakteristiken auf der anderen Seite zu erhalten. Dies könnte die Vorhersage zukünftiger Entwicklungen (z.B. beim Verbraucher*innenverhalten) ermöglichen und ggf. dazu beitragen, Handlungsempfehlungen und Maßnahmen für Nachhaltigkeit und Klimaschutz abzuleiten.

Im Verlauf der Befragung werden wir Sie darum bitten, einige Fragen zu Ihrer eigenen Person, Ihren Werten und Einstellungen zu beantworten. Bitte beantworten Sie die Fragen so aufmerksam und gewissenhaft wie möglich. Die Teilnahme an der Befragung dauert in etwa **25-35 Minuten**.

Beachten Sie, dass die Teilnahme an der Untersuchung freiwillig ist und Sie die Befragung jederzeit ohne Angabe von Gründen verweigern oder abbrechen können, ohne dass Ihnen Nachteile entstehen. Die Untersuchung ist vollkommen anonym. Ihre Angaben können von niemandem und zu keiner Zeit mit Ihrer Person in Verbindung gebracht werden. Mit einer E-Mail mit dem Betreff "Widerruf" an wiso@wiwi.uni-wuppertal.de kann die Einwilligung in die Teilnahme jederzeit widerrufen werden, die Rechtmäßigkeit der Verarbeitung wird durch den Widerruf nicht rückwirkend beseitigt. Entsprechend der europäischen Datenschutzgrundverordnung werden Ihre Angaben und Daten gewissenhaft geschützt. Bei eventuellen Rückfragen können Sie sich an Herrn Mats Schumacher, M.Sc. (Schumacher@wiwi.uni-wuppertal.de) wenden.

Im Anschluss an die Befragung können Sie an einer Verlosung teilnehmen, bei der **drei Best-Choice/WUNSCH-Gutscheine** im Wert von jeweils **25 Euro** unter allen Teilnehmer*innen verlost werden.

Hiermit stimme ich zu, dass meine Angaben im Rahmen der europäischen Datenschutzgrundverordnung (DSGVO) zu rein wissenschaftlichen Zwecken verwendet werden dürfen.

Ich stimme zu.

Weiter

M.Sc. Mats Schumacher
Bergische Universität Wuppertal – 2020

Seite 2



4% ausgefüllt

Sollten Sie mit dem Smartphone an dieser Umfrage teilnehmen, bitten wir Sie, das Smartphone zu drehen und ins Querformat (Landscape-Modus) zu wechseln, falls nicht schon geschehen. Dadurch wird eine übersichtliche Darstellung des Fragebogens gewährleistet.

Weiter



7% ausgefüllt

Zunächst würden wir Sie bitten, einige Fragen zu Ihrer eigenen Person zu beantworten.

Wie alt sind Sie?

Ich bin Jahre alt.

Welches Geschlecht haben Sie?

- weiblich
- männlich
- divers

Welchen höchsten Schulabschluss oder Hochschulabschluss haben Sie?

Wenn Sie noch Schüler*in sind: Welchen Abschluss streben Sie an?

- Schule beendet ohne Abschluss
- Hauptschulabschluss/Volksschulabschluss oder Polytechnische Oberschule mit Abschluss 8./9. Klasse
- Mittlere Reife/Realschulabschluss oder Polytechnische Oberschule mit Abschluss 10. Klasse
- Abitur oder Fachabitur, Abschluss einer Fachoberschule (Hochschul- bzw. Fachhochschulreife)
- Hochschulabschluss (Universität, Hochschule, Fachhochschule)
- Anderer Schulabschluss:

Sind Sie derzeit erwerbstätig? (Mehrfachauswahl möglich)**Ich bin**

- erwerbstätig in Vollzeit.
- erwerbstätig in Teilzeit.
- geringfügig beschäftigt (450€-Basis).
- Auszubildende*r.
- Schüler*in oder Student*in.
- Rentner*in
- zur Zeit nicht erwerbstätig.
-
- Sonstiges

[Weiter](#)



11% ausgefüllt

An dieser Stelle möchten wir gerne etwas über Ihren Alltag und Ihr Verhalten erfahren.

Nachfolgend werden einige Beispiele für alltägliche Verhaltensweisen mit Umweltbezug aufgelistet. Bitte markieren Sie, wie häufig Sie diese Verhaltensweisen ausführen.

	bisher nie 1	2	manch- mal 3	4	immer 5
Recyclingpapier benutzen.	<input type="radio"/>				
Biologisch abbaubare Reinigungsmittel verwenden.	<input type="radio"/>				
Öffentliche Verkehrsmittel benutzen.	<input type="radio"/>				
Mit Wasser sparsam umgehen.	<input type="radio"/>				
Getränke nicht in der Dose kaufen.	<input type="radio"/>				
Zum Einkaufen eine Tasche mitnehmen, anstatt sich im Laden eine Plastiktüte geben zu lassen.	<input type="radio"/>				
Biologisch angebaute Lebensmittel kaufen.	<input type="radio"/>				
Für Kurzstrecken auf das Auto verzichten.	<input type="radio"/>				
Beim Kleiderkauf auf umweltverträgliche Herstellung achten.	<input type="radio"/>				

Weiter



15% ausgefüllt

Weiterhin geht es um Ihren Alltag und Ihr Verhalten.

Nachfolgend werden einige Beispiele für alltägliche Verhaltensweisen mit Umweltbezug aufgelistet. Bitte markieren Sie, wie häufig Sie diese Verhaltensweisen ausführen.

	bisher nie 1	2	manch- mal 3	4	immer 5
Verpackungen im Laden zurücklassen.	<input type="radio"/>				
Das Angebot von Nachfüllpackungen nutzen.	<input type="radio"/>				
Wenn möglich auf Flugreisen verzichten.	<input type="radio"/>				
Bei der Ernährung auf Fleisch verzichten.	<input type="radio"/>				
Die häusliche Beheizung reduzieren.	<input type="radio"/>				
Wäsche an der Luft trocknen.	<input type="radio"/>				
Elektrogeräte komplett abschalten (Standby vermeiden).	<input type="radio"/>				
Produkte, wenn möglich, gebraucht kaufen.	<input type="radio"/>				
Bevorzugt regionale Produkte konsumieren.	<input type="radio"/>				
Kreuzfahrten vermeiden.	<input type="radio"/>				

Weiter



19% ausgefüllt

Das Umweltbundesamt ist seit seiner Gründung im Jahr 1974 die zentrale Umweltbehörde Deutschlands. Seit 1996 befragt es alle zwei Jahre Bürgerinnen und Bürger in einer repräsentativen Umfrage nach ihrer Einschätzung zum Zustand der Umwelt, ihrem eigenen umweltrelevanten Verhalten sowie zu aktuellen Themen der Umweltpolitik. Dargestellt sind Abschnitte aus der öffentlich zugänglichen Studie „Umweltbewusstsein in Deutschland 2018“ (Rubik et al. 2019). **Bitte lesen Sie die folgenden Abschnitte zu "Sozialen Milieus" aufmerksam. Im Nachgang werden Ihnen dazu einige Fragen gestellt.**

Auszüge aus der Studie *"Umweltbewusstsein in Deutschland 2018"*:

Für die Umweltpolitik und Umweltkommunikation ist es wichtig zu wissen, welche Einstellungen und Erwartungen unterschiedliche gesellschaftliche Gruppen haben. Daher wurde in die Umweltbewusstseinsstudie 2018 – wie schon in die Vorgängerstudien seit 2008 – ein Modell sozialer Milieus integriert. (...) Soziale Milieus fassen Gruppen von Menschen zusammen, die ähnliche Werthaltungen, Mentalitäten und Prinzipien der Lebensführung haben. Durch die Betrachtung der sozialen Milieus wird somit deutlich, welche unterschiedlichen Alltagskulturen und Lebenswelten in der Gesellschaft vorhanden sind. Neben der Grundorientierung im Alltag werden auch die soziale Lage und die generationale Prägung unterschiedlicher Altersgruppen bei der Identifikation und Beschreibung sozialer Milieus berücksichtigt.

Milieu und Lebensmotto	Beschreibung
Traditionelle Milieus <i>„Hoffentlich bleibt alles so, wie es ist“</i>	Bei den Angehörigen der traditionellen Milieus handelt es sich meist um ältere Personen über 70 Jahre. (...) Es finden sich unterschiedliche soziale Lagen. (...) Sicherheit und Stabilität ist den Traditionellen wichtig. Sparsamkeit und die Bereitschaft zu verzichten entsprechen ihrer Grundorientierung. Sie haben grundsätzlich den Wunsch, Gewohntes und Bewährtes zu bewahren (...). Viele der aktuellen Entwicklungen sind für sie Anlass zu großer Sorge.
Etablierte Milieus <i>„Auf das Erreichte stolz sein und es genießen“</i>	Mittlere und höhere Altersgruppen im Alter von 40 bis 70 Jahren sind in den etablierten Milieus am häufigsten anzutreffen. Männer sind etwas stärker vertreten als Frauen. Sie haben ein mittleres bis hohes Bildungsniveau – mittlere Reife, Abitur oder Hochschulabschluss – und verfügen meist über gehobene bis sehr hohe Einkommen. Die Angehörigen dieser Milieus sind ausgesprochen leistungs- und erfolgsorientiert. Sie legen Wert auf einen hohen Lebensstandard, den sie sich ihrer Einschätzung nach aufgrund des eigenen beruflichen Engagements auch verdient haben. Die persönliche und gesellschaftliche Zukunft sehen sie optimistisch.
Bürgerlicher Mainstream <i>„Dazugehören und integriert sein“</i>	Die Milieus des bürgerlichen Mainstreams befinden sich meist in mittleren sozialen Lagen, in mittleren Bildungs- und in mittleren Einkommensgruppen. Die Altersgruppen der 40- bis 70-Jährigen sind am stärksten vertreten. Bei den alltäglichen Prioritäten steht das Privat- und Familienleben ganz oben. (...) Dabei ist es ihnen wichtig, ihren gewohnten Lebensstandard und den sozialen Status in der Mitte der Gesellschaft zu erhalten. Dafür sind sie bereit, Leistung zu erbringen und sich anzustrengen. Jedoch nehmen Ängste vor einem möglichen sozialen Abstieg zu.
Prekäre Milieus <i>„Über die Runden kommen und nicht negativ auffallen“</i>	In den prekären Milieus sind vor allem Altersgruppen ab 40 Jahren vertreten. Sie haben überwiegend niedrige formale Bildungsabschlüsse und sehr geringe bis geringe Einkommen. (...) Auch alleinerziehende Frauen sind überrepräsentiert. (...) Die Menschen möchten (und müssen) die Dinge geregelt bekommen, den Job (soweit vorhanden) durch- und behalten, sich und die Familie versorgen und die alltäglichen Routinen bewältigen. Insgesamt sehen sie sich als Verliererinnen und Verlierer der aktuellen gesellschaftlichen Entwicklungen und blicken eher pessimistisch in die Zukunft.
Kritisch-kreative Milieus <i>„Die Dinge kritisch hinterfragen; verantwortlich und sinnvoll leben“</i>	Die kritisch-kreativen Milieus umfassen ein breites Altersspektrum von 30 bis 70 Jahren. Typisch sind eine mittlere oder höhere Formalbildung und mittlere wie gehobene Einkommensgruppen. Frauen sind in diesen Milieus mit einem Anteil von 60 Prozent deutlich stärker vertreten als Männer. Die Milieugehörigen zeigen sich aufgeklärt, weltoffen und tolerant. Sie streben nach Selbstverwirklichung sowie nach Unabhängigkeit von Normen und Konventionen. An gesellschaftlichen und kulturellen Themen zeigen sie sich sehr interessiert.

Neben den vorgestellten Milieus wurde die jüngere Generation der bis 30-Jährigen in drei weitere Milieus unterteilt:

Junge Idealistische, Junge Pragmatische und Junge Distanzierte.

Anmerkung: Die Angaben dieses Textes basieren auf der Studie „Umweltbewusstsein in Deutschland 2018“ des Bundesministeriums für Umwelt, Naturschutz und nukleare Sicherheit“. Die gesamte Studie kann unter [folgendem Link](#) eingesehen werden.

Weiter



22% ausgefüllt

Nachfolgend finden Sie einige inhaltliche Fragen zu den Textauszügen auf der vorherigen Seite. Bitte benutzen Sie NICHT den Zurück-Button des Browsers, da die Befragung dadurch abgebrochen wird. Wir möchten überprüfen, wie einprägsam die Informationen des Textes vermittelt wurden. Bitte beantworten Sie die untenstehenden Fragen dafür so gewissenhaft wie möglich. Die Schwierigkeit der Fragen variiert, sodass Sie möglicherweise nicht alle beantworten können. Sollten Sie sich an die betreffenden Inhalte nicht mehr erinnern können, können Sie die entsprechende(n) Frage(n) einfach überspringen. Bei den offenen Fragen genügt es, wenn Sie in prägnanten Stichpunkten antworten.

Wie viele Milieus werden in der Studie insgesamt unterschieden?

Gemeint ist die Gesamtanzahl der Milieus in der Studie, nicht die Anzahl der auf der vorherigen Seite in der Tabelle vorgestellten Milieus.

Auf welches Jahr bezieht sich die Studie, aus der die Textauszüge stammen?**Welchem Milieu wird das Motto „Über die Runden kommen und nicht negativ auffallen“ zugeordnet?****Welches der folgenden Milieus wurde auf der vorherigen Seite nicht in der Tabelle vorgestellt?**

- Etabliertes Milieu
- Skeptisch-innovatives Milieu
- Prekäres Milieu
- Traditionelles Milieu

Was sind soziale Milieus?

Bitte beschreiben Sie den Milieubegriff in eigenen Worten.

Weiter



26% ausgefüllt

Die Zuordnung zu einem bestimmten sozialen Milieu ist nicht immer ganz eindeutig. Eine sehr wichtige Rolle spielt aus Sicht der Forschung der erste Eindruck, der beim Lesen entsteht. Hier ordnet man sich häufig unmittelbar einer bestimmten Gruppe zu. Das führt schnell dazu, dass weniger differenziert über die tatsächliche Passung nachgedacht wird. Dabei überdecken insbesondere soziodemografische Elemente die Frage, ob auch Bedürfnisse und Einstellungen zur eigenen Person passen. Die Entscheidung wird schnell und impulsiv getroffen. Potentiell wichtige Informationen werden gar nicht mehr wirklich aufgenommen. Vor diesem Hintergrund möchten wir an dieser Stelle prüfen, ob Sie sich die Zeit nehmen, die Instruktionen zu lesen, oder möglicherweise Ihrem ersten Impuls folgen. Bitte ignorieren Sie daher die folgende Frage und nehmen Sie keine Zuordnung zu einem Milieu für sich vor, indem Sie keine der dargestellten Optionen auswählen. Bitte aktivieren Sie abschließend lediglich die Option "trifft voll und ganz zu" bei der letzten Frage auf der Seite ("Wie gut trifft die Beschreibung des Milieus, dem Sie sich zugeordnet haben, tatsächlich auf Sie zu?") und klicken Sie dann auf "Weiter".

Welchem der untenstehenden Milieus würden Sie sich am ehesten zuordnen, wenn Sie eine Entscheidung treffen müssten? Orientieren Sie sich dabei weniger an soziodemografischen Merkmalen wie Alter und Geschlecht. Antworten Sie so intuitiv und gleichzeitig gewissenhaft wie möglich.

Traditionelles Milieu

Etabliertes Milieu

Bürgerlicher Mainstream

Prekäres Milieu

Kritisch-kreatives Milieu

	trifft überhaupt nicht zu	trifft eher nicht zu	trifft eher zu	trifft voll und ganz zu
Wie gut trifft die Beschreibung des Milieus, dem Sie sich zugeordnet haben, tatsächlich auf Sie zu?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



30% ausgefüllt

Die Fragen auf dieser Seite beziehen sich auf Ihre generellen Einstellungen und Gedanken zur Umwelt und zu Umweltthemen.

Im Folgenden sehen Sie eine Reihe von Aussagen. Bitte geben Sie für jede Aussage an, in welchem Maße Sie ihr zustimmen.

Kreuzen Sie „kann ich nicht beantworten“ nur dann an, wenn Sie eine Aussage nicht beurteilen können.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils / teils	Stimme eher zu	Stimme voll und ganz zu	kann ich nicht beant- worten
Es gibt natürliche Grenzen des Wachstums, die unsere industrialisierte Welt längst erreicht hat.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Zugunsten der Umwelt sollten wir alle bereit sein, unseren derzeitigen Lebensstandard einzuschränken.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Es beunruhigt mich, wenn ich daran denke, in welchen Umweltverhältnissen zukünftige Generationen wahrscheinlich leben müssen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Der Klimawandel bedroht auch die Lebensgrundlagen hier in Deutschland.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Menschengemachte Umweltprobleme wie die Abholzung der Wälder oder das Plastik in den Weltmeeren empören mich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich saß noch nie in einem fahrenden Auto.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Für ein gutes Leben sind andere Dinge wichtig als Umwelt und Natur.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jede*r einzelne trägt Verantwortung dafür, dass wir nachfolgenden Generationen eine lebenswerte Umwelt hinterlassen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Umweltproblematik wird von vielen Umweltschützer*innen stark übertrieben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



33% ausgefüllt

Die Fragen auf dieser Seite beziehen sich auf Ihre generellen Einstellungen und Gedanken zur Umwelt und zu Umweltthemen.

Im Folgenden sehen Sie eine Reihe von Aussagen. Bitte geben Sie für jede Aussage an, in welchem Maße Sie ihr zustimmen.

Kreuzen Sie „kann ich nicht beantworten“ nur dann an, wenn Sie eine Aussage nicht beurteilen können.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils / teils	Stimme eher zu	Stimme voll und ganz zu	kann ich nicht beantworten
Ich freue mich über Initiativen, die nachhaltige Lebensweisen einfach ausprobieren (z.B. Ökodörfer, SlowFood-Bewegung).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wir sollten nicht mehr Ressourcen verbrauchen als nachwachsen können.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Es macht mich wütend, wenn ich sehe, wie Deutschland seine Klimaschutzziele verfehlt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin schon in jedes Land der Welt geflogen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mehr Umweltschutz bedeutet auch mehr Lebensqualität und Gesundheit für alle.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wir brauchen in Zukunft mehr Wirtschaftswachstum, auch wenn das die Umwelt belastet.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich ärgere mich, wenn mir Umweltschützer*innen vorschreiben wollen, wie ich leben soll.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wir müssen Wege finden, wie wir unabhängig vom Wirtschaftswachstum gut leben können.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



37% ausgefüllt

Nun geht es noch einmal um Ihr persönliches Verhalten bei Kauf und Nutzung von Produkten und Dienstleistungen. Denken Sie an Ihren Alltag und geben Sie an, wie oft Sie die folgenden Handlungen durchführen. Bitte geben Sie „nie“ an, wenn Sie das Genannte nie tun bzw. „immer“ wenn Sie es immer tun – mit den Ausprägungen dazwischen können Sie Ihre Antwort entsprechend Ihrem persönlichen Verhalten genauer abstimmen.

Kreuzen Sie „kann ich nicht beantworten“ dann an, wenn eine Frage auf Ihre momentane Lebenssituation nicht zutrifft (beispielsweise können Sie keine Angaben über Ihr Fahrverhalten machen, wenn Sie keinen Führerschein besitzen).

	nie	selten	gelegentlich	oft	immer	kann ich nicht beantworten
Ich kaufe Lebensmittel aus kontrolliert-biologischem Anbau.	<input type="radio"/>					
Beim Einkaufen wähle ich Produkte mit Umweltsiegel (z.B. Bioland, blauer Engel, Demeter, EU Biosiegel oder EU Ecolabel).	<input type="radio"/>					
Zu den Hauptmahlzeiten esse ich Fleisch.	<input type="radio"/>					
Für meine alltäglichen Wege benutze ich das Fahrrad, öffentliche Verkehrsmittel oder gehe zu Fuß.	<input type="radio"/>					
Beim Kauf von Haushaltsgeräten wähle ich besonders energieeffiziente Geräte (A+++ oder A++ Energieeffizienzsiegel).	<input type="radio"/>					
Beim Spenden gebe ich mein Geld an Umwelt- und Naturschutzgruppen.	<input type="radio"/>					
Ich engagiere mich aktiv für den Umwelt- und Naturschutz.	<input type="radio"/>					

Weiter



41% ausgefüllt

Auf dieser und der folgenden Seite möchten wir gerne etwas über Ihre persönlichen Einstellungen und Bedürfnisse erfahren. Dafür bitten wir Sie, sich selbst entsprechend der unten dargestellten Profile zu charakterisieren, indem Sie sich jeweils mit der dargestellten Person vergleichen.

Die folgenden Statements beschreiben die Ansichten einer fiktiven Person. Bitte benutzen Sie die dargestellten Antwortmöglichkeiten und geben Sie für jedes Statement an, wie ähnlich die beschriebene Person Ihnen ist.

Die Person ist mir...

sehr unähnlich 1 2 3 4 5 6 7 sehr ähnlich

Es ist ihm/ihr wichtig, Umweltverschmutzung zu vermeiden.	<input type="radio"/>						
Es ist ihm/ihr wichtig, dass jeder Mensch die gleichen Chancen haben sollte.	<input type="radio"/>						
Es ist ihm/ihr wichtig, die Handlungen anderer Menschen zu kontrollieren.	<input type="radio"/>						
Es ist ihm/ihr wichtig, sich um die Schwächeren zu kümmern.	<input type="radio"/>						
Es ist ihm/ihr wichtig, Spaß zu haben.	<input type="radio"/>						
Es ist ihm/ihr wichtig, die Umwelt zu schützen.	<input type="radio"/>						
Es ist ihm/ihr wichtig, Macht über andere zu haben.	<input type="radio"/>						
Es ist ihm/ihr wichtig, dass alle Personen gerecht behandelt werden.	<input type="radio"/>						

Weiter



44% ausgefüllt

Die folgenden Statements beschreiben die Ansichten einer fiktiven Person. Bitte benutzen Sie die dargestellten Antwortmöglichkeiten und geben Sie für jedes Statement an, wie ähnlich die beschriebene Person Ihnen ist.

Die Person ist mir...

sehr unähnlich 1 2 3 4 5 6 7 sehr ähnlich

Es ist ihm/ihr wichtig, die Natur zu achten.	<input type="radio"/>						
Es ist ihm/ihr wichtig, die Freuden des Lebens zu genießen.	<input type="radio"/>						
Es ist ihm/ihr wichtig, andere beeinflussen zu können.	<input type="radio"/>						
Es ist ihm/ihr wichtig, dass es keine Kriege und Konflikte gibt.	<input type="radio"/>						
Es ist ihm/ihr wichtig, Geld und Eigentum zu besitzen.	<input type="radio"/>						
Es ist ihm/ihr wichtig, Dinge zu tun, die ihm/ihr Freude bereiten.	<input type="radio"/>						
Es ist ihm/ihr wichtig, sich der Natur anzupassen und zu ihr zu passen.	<input type="radio"/>						
Es ist ihm/ihr wichtig, hart und ambitioniert zu arbeiten.	<input type="radio"/>						
Es ist ihm/ihr wichtig, anderen gegenüber hilfsbereit zu sein.	<input type="radio"/>						

Weiter

Anmerkung: Auf den Seiten 12 und 13 wurden die Statements auf das Geschlecht der Proband*innen angepasst. Dargestellt ist die neutrale Formulierung (divers oder fehlender Wert beim Geschlecht)



49% ausgefüllt

Verschiedene Menschen richten ihr Leben nach verschiedenen Werten aus. An dieser Stelle sind wir daran interessiert, welche Werte in Ihrem Leben eine Rolle spielen.

Bitte lesen Sie sich die folgenden Werte und ihre Beschreibungen durch. Kreuzen Sie dann jeweils an, wie wichtig diese Werte für Sie persönlich sind.

	überhaupt nicht wichtig	nicht wichtig	eher nicht wichtig	eher wichtig	wichtig	sehr wichtig
MACHT: Sozialer Status und Prestige, Kontrolle oder Dominanz über Leute und Ressourcen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
LEISTUNG: Persönlicher Erfolg durch die Demonstration von Kompetenz gemäß sozialer Maßstäbe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
HEDONISMUS: Vergnügen und sinnliche Belohnung des Selbst.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ANREGUNG: Aufregendes Leben, Reiz des Neuen und Herausforderungen im Leben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SELBSTBESTIMMUNG: Eigenständiges Denken und Verhalten, Kreieren und Erkunden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
UNIVERSALISMUS: Verständnis, Wertschätzung, Toleranz und Schutz des Wohles aller Menschen und der Natur.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SOZIALITÄT: Erhaltung und Verbesserung des Wohlergehens der Menschen, mit denen man regelmäßigen Kontakt hat.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TRADITION: Respekt, Verpflichtung und Akzeptanz von Bräuchen und Meinungen, die die Tradition oder Religion vorschreibt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
KONFORMITÄT: Zügelung von Verhalten oder Neigungen, die Andere verärgern oder schaden könnten und die soziale Erwartungen und Normen verletzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SICHERHEIT: Schutz, Harmonie und Stabilität der Gesellschaft, von Beziehungen und des Selbst.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



55% ausgefüllt

Bitte beantworten Sie die folgende(n) Frage(n) zu Ihrer eigenen Person. Wir möchten gerne von Ihnen wissen, wie Sie sich selbst einschätzen.

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils / teils	Stimme eher zu	Stimme voll und ganz zu
Ich gehe aus mir heraus, bin gesellig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin einfühlsam, warmherzig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin eher unordentlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bleibe auch in stressigen Situationen gelassen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin nicht sonderlich kunstinteressiert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin durchsetzungsfähig, energisch.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich begegne anderen mit Respekt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin bequem, neige zu Faulheit.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bleibe auch bei Rückschlägen zuversichtlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin vielseitig interessiert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



60% ausgefüllt

Bitte beantworten Sie die folgende(n) Frage(n) zu Ihrer eigenen Person. Wir möchten gerne von Ihnen wissen, wie Sie sich selbst einschätzen.

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils / teils	Stimme eher zu	Stimme voll und ganz zu
Ich schäume selten vor Begeisterung über.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich neige dazu, andere zu kritisieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin stetig, beständig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann launisch sein, habe schwankende Stimmungen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin erfinderisch, mir fallen raffinierte Lösungen ein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin eher ruhig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe noch nie gelacht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe mit anderen wenig Mitgefühl.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin systematisch, halte meine Sachen in Ordnung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich reagiere leicht angespannt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann mich für Kunst, Musik und Literatur begeistern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



65% ausgefüllt

Bitte beantworten Sie die folgende(n) Frage(n) zu Ihrer eigenen Person. Wir möchten gerne von Ihnen wissen, wie Sie sich selbst einschätzen.

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils / teils	Stimme eher zu	Stimme voll und ganz zu
Ich neige dazu, die Führung zu übernehmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe oft Streit mit anderen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich neige dazu, Aufgaben vor mir herzuschieben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin selbstsicher, mit mir zufrieden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich meide philosophische Diskussionen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin weniger aktiv und unternehmungslustig als andere.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin nachsichtig, vergebe anderen leicht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin manchmal ziemlich nachlässig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin ausgeglichen, nicht leicht aus der Ruhe zu bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin nicht besonders einfallsreich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



70% ausgefüllt

Bitte beantworten Sie die folgende(n) Frage(n) zu Ihrer eigenen Person. Wir möchten gerne von Ihnen wissen, wie Sie sich selbst einschätzen.

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme voll und ganz zu	Stimme eher zu	Teils / teils	Stimme eher nicht zu	Stimme überhaupt nicht zu
Ich bin eher schüchtern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin hilfsbereit und selbstlos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich mag es sauber und aufgeräumt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich mache mir oft Sorgen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich weiß Kunst und Schönheit zu schätzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mir fällt es schwer, andere zu beeinflussen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin manchmal unhöflich und schroff.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin effizient, erledige Dinge schnell.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich oft bedrückt, freudlos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Es macht mir Spaß, gründlich über komplexe Dinge nachzudenken und sie zu verstehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



75% ausgefüllt

Bitte beantworten Sie die folgende(n) Frage(n) zu Ihrer eigenen Person. Wir möchten gerne von Ihnen wissen, wie Sie sich selbst einschätzen.

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils / teils	Stimme eher zu	Stimme voll und ganz zu
Ich bin voller Energie und Tatendrang.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin anderen gegenüber misstrauisch.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin verlässlich, auf mich kann man zählen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich lüge immer.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe meine Gefühle unter Kontrolle, werde selten wütend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin nicht sonderlich fantasievoll.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin gesprächig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Andere sind mir eher gleichgültig, egal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin eher der chaotische Typ, mache selten sauber.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich werde selten nervös und unsicher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich finde Gedichte und Theaterstücke langweilig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



80% ausgefüllt

Bitte beantworten Sie die folgende(n) Frage(n) zu Ihrer eigenen Person. Wir möchten gerne von Ihnen wissen, wie Sie sich selbst einschätzen.

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

	Stimme überhaupt nicht zu	Stimme eher nicht zu	Teils / teils	Stimme eher zu	Stimme voll und ganz zu
In einer Gruppe überlasse ich lieber anderen die Entscheidung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin höflich und zuvorkommend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bleibe an einer Aufgabe dran, bis sie erledigt ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin oft deprimiert, niedergeschlagen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann zwei Kilometer in zwei Minuten laufen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mich interessieren abstrakte Überlegungen wenig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin begeisterungsfähig und kann andere leicht mitreißen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Manchmal verhalte ich mich verantwortungslos, leichtsinnig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich reagiere schnell gereizt oder genervt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bleibe auch bei Rückschlägen zuversichtlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin originell, entwickle neue Ideen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich nehme gerade an einer Befragung teil.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



85% ausgefüllt

Im Folgenden geht es um Ihr Verhalten in Online-Befragungen. Bei solchen Befragungen ist es durchaus normal, dass nicht alle Befragten den Fragebogen so gründlich ausfüllen, wie sich dies die Studienleitung wünscht. Bitte beantworten Sie die folgenden Fragen so ehrlich und gewissenhaft wie möglich.

Wie häufig...	nie			etwa in der Hälfte der Fälle			immer
lesen Sie jede Frage?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
schenken Sie jeder einzelnen Frage Beachtung?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
nehmen Sie sich so viel Zeit wie nötig, um eine Frage ehrlich zu beantworten?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
folgen Ihre Antworten in Frageblöcken bestimmten Mustern?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
nutzen Sie dieselben Antwortoptionen für einen Block von Fragen zu einem bestimmten Thema (statt jede Frage tatsächlich zu lesen)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
antworten Sie schnell, ohne nachzudenken?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
antworten Sie impulsiv, ohne nachzudenken?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
erledigen Sie Befragungen hastig?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
überfliegen Sie Instruktionen schnell?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
überspringen Sie Teile der Instruktion?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



90% ausgefüllt

An dieser Stelle möchten wir gerne noch etwas über die Situation erfahren, in der Sie an dieser Befragung teilgenommen haben. Bitte beantworten Sie dafür die folgenden Fragen.

An welchem Ort haben Sie den Fragebogen ausgefüllt?

zuhause

an einem anderen Ort und zwar:

Waren Sie während des Ausfüllens des Fragebogens allein oder im Beisein anderer Personen? (Mehrfachauswahl möglich)

Falls Sie den Fragebogen in Anwesenheit anderer Personen ausgefüllt haben, geben Sie bitte Ihre Beziehung zu diesen Personen an.

Ich war allein.

Es waren mir unbekannte Personen anwesend.

Es waren mir bekannte Personen anwesend. Bitte geben Sie an, in welcher Beziehung sie zu der/n anwesenden Person/en sind:

In welcher Körperhaltung haben Sie die Befragung größtenteils ausgefüllt?

Sollte sich Ihre Körperhaltung während des Antwortprozesses verändert haben, wählen Sie die Haltung, die Sie die längste Zeit während der Beantwortung innehatten.

sitzend

stehend

liegend

in Bewegung

Sonstiges und zwar:

Bitte geben Sie das Ausmaß an, in dem Sie durch externe Reize (akustisch, visuell etc.) während der Befragung abgelenkt wurden.

Sehr geringe Ablenkung

1

2

3

4

5

6

Sehr starke Ablenkung

7

Auf welchem Gerät haben Sie die Umfrage beantwortet?

PC/Laptop

Smartphone

Tablet

Sonstiges und zwar

Bei Online-Befragungen, in denen Wissensfragen zu auf vorangehenden Seiten vermittelten Inhalten gestellt werden, machen sich die Befragten häufig ein Foto oder einen Screenshot dieser Inhalte oder starten die Befragung nochmal neu. War das bei Ihnen auch so?

Ja

Nein

Sind Sie während der Umfrage an Ihrem Ausfüllgerät anderen Aufgaben oder Tätigkeiten nachgegangen?

überhaupt nicht

vereinzelt und zwar folgenden Aktivitäten:

mehrfach und zwar folgenden Aktivitäten:

häufig und zwar folgenden Aktivitäten:

(fast) die gesamte Zeit über und zwar folgenden Aktivitäten:

Weiter



Vielen Dank für Ihre Teilnahme!

Zweck der Studie:

Mit Ihrer Teilnahme unterstützen Sie unsere Forschung. Dabei sind wir nicht nur an Ihren tatsächlichen Umwelteinstellungen und Ihrem Umweltbewusstsein interessiert, sondern haben noch einen anderen methodischen Schwerpunkt. Wir untersuchen, ob das Ausfüllgerät, mit dem an einer Befragung teilgenommen wird, einen Einfluss auf das Antwortverhalten der Befragten hat. Aus diesem Grund wurden verschiedene Prüfungen der Aufmerksamkeit in die Befragung integriert. Konkret wird betrachtet, ob sich die Aufmerksamkeit zwischen Personen, die die Befragung am Smartphone ausfüllen, und Proband*innen, die am PC oder Laptop teilnehmen, unterscheiden. Um Verzerrungen in den untersuchten Effekten zu vermeiden, möchten wir Sie bitten, bis zum Ende des Befragungszeitraums Ende November mit niemandem über das wahre Untersuchungsziel zu sprechen.

Teilnahme an der Verlosung:

Wenn Sie an der Verlosung der drei 25 Euro Best-Choice/WUNSCH-Gutscheine teilnehmen möchten, klicken Sie bitte auf [diesen Link](#) und geben Sie dort eine gültige E-Mail-Adresse an. Dadurch wird sichergestellt, dass Ihre Mailadresse separat von Ihren weiteren Angaben gespeichert wird.

Für Nutzer*innen von SurveyCircle (www.surveycircle.com): Der Survey Code lautet: 6HRR-WV5K-X3KF-QVDP

Herzlichen Dank!

Die Befragung ist hiermit beendet. Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

M.Sc. Mats Schumacher

Bergische Universität Wuppertal – 2020
