

Antwortskalenrichtung und Umfragemodus

Dagmar Krebs

Universität Gießen (Deutschland)

Jan Karem Höhne

Universität Mannheim (Deutschland)

Zusammenfassung

Der Einfluss der Antwortskalenrichtung auf das Antwortverhalten ist ein bekanntes Phänomen in der quantitativen empirischen Sozialforschung. Es existieren verschiedene theoretische Erklärungsansätze für das Auftreten solcher Effekte, so dass die Literatur unterschiedliche empirische Befunde berichtet. Allerdings hat die Umfrageforschung gezeigt, dass das Antwortverhalten von Befragungspersonen insbesondere vom jeweiligen Umfragemodus abhängt. Aus diesem Grund vergleichen wir in dieser Studie das Auftreten von Effekten der Antwortskalenrichtung in zwei unterschiedlichen selbstadministrierten Umfragemodi: „Paper/Pencil (PaPi)“ und Web. Zu diesem Zweck haben wir ein Umfrageexperiment (n = 657) mit vier Gruppen durchgeführt indem sowohl die Antwortskalenrichtung (absteigend vs. aufsteigend) als auch der Umfragemodus (PaPi vs. Web) variiert wurde. Die Ergebnisse liefern Anhaltspunkte für Richtungseffekte im PaPi Modus. Für den Web Modus sind hingegen keinerlei Anhaltspunkte für Richtungseffekte zu finden. Darüber hinaus zeigen die Untersuchungen, dass bei gleichbleibender Skalenrichtung keine Messinvarianz zwischen den zwei Umfragemodi (PaPi und Web) besteht.

Keywords: Antwortverhalten, Pencil/Paper-Umfrage, Ratingskalen, Skalenrichtungseffekte, Umfragemodus, Web-Umfrage

Einleitung und Hintergrund

Zur Messung von Einstellungen und Meinungen werden für gewöhnlich geschlossene Fragen verwendet. Geschlossen bedeutet, dass die Befragungspersonen ihre Antwort aus vorgegebenen Antwortkategorien wählen müssen. Bei der Erstellung solcher Fragen sind einige grundlegende Designentscheidungen zu treffen. So ist zu entscheiden, ob die Antwortskala einen Mittelpunkt enthält, wie viele inhaltliche Ausprägungen die Antwortskala hat, ob die Antwortskala verbale Benennungen der einzelnen Kategorien beinhaltet, ob die Antwortskala uni- oder bipolar ist, ob die Antwortskala horizontal oder vertikal verläuft und ob die Antwortskala in einer ab- oder aufsteigenden Richtung verläuft (DeCastellarnau 2017; Höhne und Krebs 2017; Krosnick und Presser 2010).

This document is a preprint and thus it may differ from the final version: Krebs, D., & Höhne, J.K. (2020). Antwortskalenrichtung und Umfragemodus. In Mays, A., Dingelstedt, A., Hambauer, V., Schlosser, S., Berens, F., Leibold, J., & Höhne, J.K. (Eds.), *Grundlagen – Methoden – Anwendungen in den Sozialwissenschaften. Festschrift zu Ehren von Steffen-M. Kühnel*. Wiesbaden: VS Verlag. DOI: 10.1007/978-3-658-15629-9

Seit den frühen Arbeiten von Mathews (1929) sowie Rugg und Cantrill (1944) wurde die Auswirkung der Richtung von Antwortskalen auf das Antwortverhalten von Befragungspersonen vielfach untersucht. Sudman et al. (1996, S. 160) beschreiben die Situation wie folgt: „[...] the emergence of response order effects on opinion questions seems to follow the rule, sometimes you see them, sometimes you don‘t“. Richtungseffekte aufgrund der Antwortskala können in zwei verschiedenen Formen auftreten (Schwarz und Hippler 1991; Sudman et al. 1996): Einerseits können Antwortkategorien am Beginn einer Skala häufiger gewählt werden – in diesem Fall spricht man von sogenannten „primacy“ Effekten. Andererseits können auch Antwortkategorien am Ende einer Skala häufiger gewählt werden – in diesem Fall spricht man von sogenannten „recency“ Effekten. Im Spezialfall von Ratingskalen zur Messung von Einstellungen und Meinungen zeigt sich, dass die Antworten von Befragungspersonen im Allgemeinen eher zum Beginn der Skala hin verschoben sind (Höhne und Lenzner 2015; Höhne und Krebs 2017; Krosnick 1991; Sudman et al. 1996; Yan und Keusch 2015). Krebs und Bachner (2018) können zeigen, dass dieses Phänomen für unterschiedliche Leserichtungen – von links nach rechts sowie von rechts nach links – gleichermaßen zutrifft.

In der Umfrageforschung existieren einige (teilweise) divergierende Erklärungsansätze für die Entstehung von Effekten aufgrund der Richtung der Antwortskala. So argumentieren einige Forscher, dass diese Effekte unmittelbar mit der Fragenschwierigkeit einhergehen (Bishop und Smith 2001). Schuman und Presser (1981) konnten jedoch Richtungseffekte auch in kurzen und vergleichsweise leichten Fragen feststellen, so dass die Schwierigkeit von Fragen nur bedingt als Erklärung plausibel erscheint. Gemäß der „opinion crystallization“ Hypothese (Rugg und Cantrill 1944) stellt die Einstellungsstärke seitens der Befragungspersonen einen zentralen Faktor dar, wobei Bishop (1990) zeigen konnte, dass Richtungseffekte nur geringfügig mit der Stärke von Einstellungen zusammenhängen. Krosnick (1991) hingegen argumentiert, dass Befragungspersonen entweder die am Beginn erscheinenden Kategorien kognitiv intensiver verarbeiten als die nachfolgenden Kategorien oder die erste plausibel erscheinende Antwortkategorie wählen ohne die nachfolgenden Kategorien zu berücksichtigen. Ein derartiges Antwortverhalten wird als „satisficing“ bezeichnet (Krosnick 1991; Krosnick und Alwin 1987; Krosnick et al. 1996) und impliziert einen oberflächlich durchgeführten Antwortprozess.

Die Literatur zur Umfrageforschung zeigt, dass auch der Umfragemodus einen Einfluss auf das Antwortverhalten von Befragungspersonen und deshalb auch auf das Auftreten unterschiedlicher Antworteffekte haben kann (de Leeuw und Hox 2011, 2014; Dillman und Messer 2010; Engel 2014; Engel et al. 2014). Der Großteil existierender Studien zu Effekten des Umfragemodus vergleicht primär auditive – bspw. Face-to-Face oder Telefon – und visuelle – bspw. postalische oder Web – Modi miteinander. Vergleiche zwischen verschiedenen visuellen Modi – bspw. „Paper/Pencil (PaPi)“ und Web – sind ausgesprochen selten (de Leeuw und Hox 2014).

Höhne und Krebs (2017) untersuchten das Auftreten von Effekten aufgrund der Richtung der Antwortskala in Zustimmungsfragen mit horizontalen Antwortskalen unter Verwendung eines PaPi Modus. Die verwendeten Fragen beschäftigten sich unter anderem mit Leistungsmotivation. Die Autoren konnten feststellen, dass die Antworten der

Befragungspersonen signifikant zum Beginn der Antwortskala hin verschoben waren. Höhne und Lenzner (2015) testeten identische Fragen in einer webbasierten Eyetrackingstudie. Hier konnten allerdings keinerlei Richtungseffekte beobachtet werden. Der Grund für die divergierenden Ergebnisse kann in den verschiedenen Modi (PaPi und Web) gesehen werden.

In diesem Beitrag wird das Auftreten von Effekten aufgrund der Antwortskalenrichtung in unterschiedlichen visuellen Umfragemodi (PaPi und Web) untersucht. Hierfür werden im Folgenden zunächst die Forschungshypothesen dargelegt. Daran anschließend werden das zugrundeliegende Forschungsdesign, die getesteten Fragen, die Stichprobe sowie die Vorgehensweise bei der Datenanalyse beschrieben. Sodann folgt die Darstellung der Ergebnisse sowie deren abschließende Diskussion.

Forschungshypothesen

Die beobachteten Ergebnisse der Eyetrackingstudie von Höhne und Lenzner (2015) für Fragen mit horizontal verlaufenden Antwortskalen haben gezeigt, dass Befragungspersonen in webbasierten Umfragen die Mittelkategorien besonders intensiv fixieren. Das bedeutet, dass diese Kategorien eine insgesamt höhere Fixationsanzahl und Fixationsdauer aufweisen als die Kategorien an den jeweiligen Enden der Antwortskala. Dieser Befund ist unabhängig von der Antwortskalenrichtung (absteigend vs. aufsteigend). In Übereinstimmung mit den Eyetrackingdaten konnte keine Verschiebung in den Antworten der Befragungspersonen zum Beginn der Antwortskala festgestellt werden. Wie von den Autoren angemerkt, weisen die empirischen Befunde vielmehr auf eine Tendenz zur Mitte hin (ebd., S. 373).

Im Hinblick auf das Fixationsverhalten von Befragungspersonen im PaPi Modus sind zum gegenwertigen Zeitpunkt keine vergleichbaren Untersuchungen und/oder empirische Befunde bekannt. Allerdings legen bisherige Untersuchungsbefunde nahe, dass die Wahl einer Antwortkategorie unmittelbar mit der Intensität ihrer Verarbeitung zusammenhängt (Krosnick 1991; Krosnick und Alwin 1987; Schwarz et al. 1992; Sudman et al. 1996). Demgemäß könnten die durch Höhne und Krebs (2017) beobachteten Richtungseffekte im PaPi Modus durch eine intensivere Verarbeitung der am Anfang stehenden Antwortkategorien erklärt werden. Das würde implizieren, dass Befragungspersonen im PaPi Modus – anders als im Web Modus – die Antwortkategorien am Beginn horizontaler Antwortskalen intensiver fixieren.

Basierend auf dieser Argumentation, postulieren wir die folgenden Forschungshypothesen: Im PaPi Modus erwarten wir einen signifikanten Richtungseffekt zwischen ab- und aufsteigender Antwortskala (Hypothese 1a), der aber wegen der Konzentration auf die mittleren Kategorien der Antwortskala im Web Modus nicht zu erwarten ist (Hypothese 1b). Weiterhin ist im PaPi Modus ein primacy Effekt zu erwarten (überzufällig häufige Wahl der Kategorien am Anfang der horizontalen Antwortskala), der im Web Modus nicht vermutet wird (Hypothese 2a und b). Als Folge des für den PaPi Modus postulierten Richtungseffekts sowie des primacy Effekts sind Unterschiede zwischen den latenten Mittelwerten der beiden Skalenrichtungen bei den Motivationskonstrukten im PaPi Modus zu erwarten (Hypothese 3a). Im Web Modus werden diese Unterschiede hingegen nicht erwartet (Hypothese 3b).

Bei dem Vergleich der zwei selbstadministrierten Modi (PaPi und Web) sind zunächst keine Unterschiede in der Befragungssituation zu verzeichnen. Die Befragungspersonen

kontrollieren die Situation weitestgehend selbst, da es keine Interviewer gibt. Lassen sich jedoch die zuvor postulierten Unterschiede in den Antworten der Befragungspersonen nachweisen, dann stellt sich die Frage, ob und inwiefern zwischen den Befragungsmodi bei gleichbleibender Antwortskalenrichtung Messinvarianz besteht. Aufgrund der Tatsache, dass uns zum gegenwärtigen Zeitpunkt weder theoretische Argumente noch empirische Befunde über einen entsprechenden Vergleich von selbstadministrierten Modi (PaPi und Web) vorliegen, nehmen wir aufgrund der Ähnlichkeit der Befragungssituation an, dass Messinvarianz zwischen den zwei Umfragemodi besteht (Hypothese 4).

Methode

Forschungsdesign

Die Datenerhebung fand jeweils in der ersten Woche des Wintersemesters 2011 (PaPi) und 2013 (Web) an der Universität Bielefeld¹ statt. Befragt wurden Studierende einer Einführungsvorlesung in empirische Sozialforschung. Der Zeitpunkt der Befragung sollte sicherstellen, dass Personen ohne spezifische Kenntnisse der Methoden der empirischen Sozialforschung an der Befragung teilnehmen.

2011 wurden alle zum Erhebungszeitpunkt im Hörsaal anwesenden Studierenden eingeladen, an der Befragung teilzunehmen und erhielten einige Informationen zur Umfrage (bspw. das Thema der Befragung). Die Papierfragebögen wurden vor der Verteilung im Hörsaal systematisch sortiert, um die Randomisierung auf die zwei Split-Versionen (ab- und aufsteigende Skalenrichtung) zu gewährleisten.

2013 wurden die Studierenden über die Studienplattform Stud.IP eingeladen, an der Web Umfrage teilzunehmen. Hierzu wurde der Link zur Umfrage sowie ein kurzes Einladungsschreiben mit einigen Informationen (bspw. dem Thema der Befragung) an die Teilnehmenden der Veranstaltung verschickt. Die Randomisierung auf die zwei Split-Versionen erfolgte automatisch durch die Surveysoftware.²

Am Ende des jeweiligen Wintersemesters wurden die Studierenden darüber informiert, dass sie an einem Experiment teilgenommen haben. Zudem wurde über die Ergebnisse sowie die methodische Zielsetzung des Experiments informiert.

Tabelle 1. Experimentelles Untersuchungsdesign definiert durch Antwortskalenrichtung und Umfragemodus

Gruppe	Richtung der Antwortskala	Umfragemodus	Gruppengröße	Erhebungszeit
1	absteigend	PaPi	163	2011
2	aufsteigend	PaPi	165	
3	absteigend	Web	165	2013
4	aufsteigend	Web	164	

¹ Für die Kooperation bei der Datenerhebung danken wir Jost Reinecke und insbesondere Thomas Blank; für die Organisation der online Befragung danken wir Johannes Herrmann und für die finanzielle Unterstützung bei der Datenerhebung bedanken wir uns bei Forschung Raum und Gesellschaft (FRG) e.V.

² Es wurde kein adaptives Umfragedesign verwendet, so dass horizontales Scrolling nicht auszuschließen ist. Darüber hinaus kann nicht zwischen PCs und mobilen Endgeräten (bspw. Smartphones) unterschieden werden.

Testfragen

Zur Untersuchung der Effekte aufgrund der Antwortskalenrichtung im PaPi und Web Modus wurden insgesamt 12 Fragen aus dem Cross Cultural Survey for Work and Gender Attitudes (2010) verwendet – 4 Fragen beschäftigten sich mit Leistungsmotivation, 4 Fragen mit intrinsischer Jobmotivation und 4 Fragen mit extrinsischer Jobmotivation.

Leistungsmotivation bezieht sich auf das Konkurrenzdenken sowie persönliche Herausforderungen. Intrinsische Jobmotivation bezieht sich auf das Engagement im Job (bspw. die Realisierung eigener Ideen und Autonomie). Extrinsische Jobmotivation hingegen beschreibt die Wichtigkeit antizipierter Jobmerkmale (bspw. Karriereaussichten und Einkommen), die nicht ausschließlich der individuellen Kontrolle unterliegen.

Im Anschluss an die Motivationsfragen wurden einige weitere Fragen gestellt. Diese Fragen beschäftigten sich mit gesellschaftlichen und politischen Inhalten (bspw. Anomie und Autoritarismus) und sind nicht Bestandteil dieses Beitrags.

Alle 12 Fragen wurden im PaPi und im Web Modus in identischer Reihenfolge in Matrixform präsentiert. Die Antwortskalen waren vollverbalisiert und enthielten fünf Skalenpunkte ohne numerische Werte (siehe Anhang). Die Beschriftungen der Antwortkategorien für alle Fragen war wie folgt: Trifft voll und ganz zu, trifft eher zu, trifft in mittlerem Ausmaß zu, trifft eher nicht zu und trifft überhaupt nicht zu (absteigende Richtung).

Stichprobe

Insgesamt haben $n = 657$ Studierende an der Studie teilgenommen ($n = 328$ in 2011 und $n = 329$ in 2013). Es wurden alle Studierenden ausgeschlossen, die älter als 30 Jahre waren.³ In 2013 wurden zusätzlich diejenigen Studierenden ausgeschlossen, die bereits 2011 teilgenommen hatten. Die Identifizierung dieser Studierenden wurde aufgrund der Verwendung von persönlichen Codes zu beiden Erhebungszeitpunkten ermöglicht. Es stehen $n = 318$ Fälle in 2011 und $n = 325$ Fälle in 2013 für die statistischen Analysen zur Verfügung.

Tabelle 2. Stichprobenbeschreibung in 2011 und 2013

Erhebungszeit	Alter	Geschlecht (weiblich)	Studiengang (SoWi)	Erstsemester
2011	21.0 (2.3)	57%	85%	74%
2013	20.6 (1.9)	61%	87%	75%

Anmerkungen: Die Standardabweichung des Alters steht in Klammern. Die Abkürzung SoWi steht für Sozialwissenschaften.

Zur Beurteilung der der Stichprobenzusammensetzung in 2011 und 2013 wurde die Verteilung der folgenden Merkmale miteinander verglichen: Alter [$\chi^2(1) = 1.26$, $p = 0.15$], Geschlecht [$\chi^2(1) = 1.25$, $p = 0.15$] Studiengang [$\chi^2(1) = 0.45$, $p = 0.29$], und Semesterzahl [$\chi^2(1) = 0.06$, $p = 0.44$]. Hinsichtlich dieser vier Merkmale gibt es auch zwischen den vier Experimentalgruppen keine statistisch signifikanten Unterschiede. Bei diesen vier Merkmalen gibt es zwischen den beiden Skalenrichtungsgruppen innerhalb der beiden Präsentationsmodi

³ Als Kriterium wurde der Mittelwert plus/minus die vierfache Standardabweichung verwendet (siehe Höhne und Krebs 2017). Die in Tabelle 2 beschriebene Stichprobenzusammensetzung ändert sich durch den Ausschluss dieser Personen nicht.

(PaPi und Web) ebenfalls keine signifikanten Unterschiede. Dementsprechend können die Gruppen als vergleichbar angesehen werden.

Datenanalyse

Zum Zweck der Eindeutigkeit der statistischen Analysen sind alle Werte in absteigender Antwortskalenrichtung codiert. Zur Feststellung von primacy Effekten wurde zunächst eine Varianzanalyse (ANOVA) berechnet. Zur Feststellung der Messinvarianz für die zwei Skalenrichtungen (ab- und aufsteigend) in beiden Befragungsmodi wurde eine vergleichende konfirmatorische Faktorenanalyse (MGCFAs) mit drei latenten Variablen und schrittweise umfassenderen Gleichheitsrestriktionen für Faktorenladungen und Interzepte durchgeführt. Daran schließt sich ein Vergleich der Mittelwerte für die drei latenten Variablen zwischen ab- und aufsteigender Skalenrichtung im PaPi und Web Modus an. Abschließend erfolgt bei gleichbleibender Skalenrichtung die Gegenüberstellung der Befragungsmodi zur Feststellung der Messinvarianz zwischen PaPi und Web Modus. Hierzu wurde ebenfalls eine vergleichende konfirmatorische Faktorenanalyse mit zunehmenden Gleichheitsrestriktionen durchgeführt.

Da alle verwendeten Antwortskalen fünfstufig sind, wurde bei der Schätzung der Modellparameter die MLM Diskrepanzfunktion des Statistikprogramms Mplus (Version 6.12) verwendet. Die einfaktorielle Varianzanalyse wurde hingegen mit dem Statistikprogramm SPSS 24 berechnet.

Ergebnisse

Varianzanalyse zur Untersuchung von primacy Effekten

Um den in Hypothese 1a postulierten primacy Effekt im PaPi Modus zu überprüfen, haben wir zunächst die durchschnittliche Anzahl positiver Antworten über alle Fragen und alle Personen hinweg aggregiert. Entsprechend des 2×2 Forschungsdesigns wurden die Effekte der Skalenrichtung und des Umfragemodus in einer zweifaktoriellen Varianzanalyse überprüft. Zwischen ab- und aufsteigender Skalenrichtung ist in beiden Befragungsmodi kein signifikanter Unterschied zu beobachten. Damit ist Hypothese 1a für den PaPi Modus widerlegt, während dieses Ergebnis der Vermutung von Hypothese 1b entspricht. Ein signifikanter Unterschied besteht zwischen den Befragungsmodi. Im PaPi Modus die Anzahl der zustimmenden Antworten signifikant höher als im Web Modus. Dieses Ergebnis ist als Anhaltspunkt für einen primacy Effekt im PaPi Modus zu werten, was für das Zutreffen der Hypothese 2a spricht. Das gilt aber nur für die absteigende Skalenrichtung. Die bei aufsteigender Skalenrichtung im PaPi Modus im Vergleich zum Web Modus ebenfalls höhere Anzahl zustimmender Antworten widerspricht dagegen dem in Hypothese 2a postulierten primacy Effekt. Bezogen auf den PaPi Modus liefert dieses Ergebnis einen empirischen Beleg für die Annahme, dass die Antwortskala als Überredungskommunikation anzusehen ist und dass negative Überredung (wie sie in der aufsteigenden Skalenrichtung gegeben ist) zum Gegenteil, nämlich zu zustimmenden Antworten führt (siehe Sudman et al. 1996). Daraus folgt, dass die generelle Annahme eines primacy Effekts möglicherweise nur für die absteigende Skalenrichtung gilt (siehe Menold und Bogner 2015). Zur Erklärung des Antwortverhaltens bei aufsteigender Skalenrichtung scheinen daher zusätzliche Annahmen notwendig. Ein

Interaktionseffekt zwischen Skalenrichtung und Umfragemodus und Skalenrichtung ist nicht zu beobachten.

Tabelle 3. Durchschnittliche Anzahl positiver (zustimmender) Antworten aggregiert über alle Fragen in beiden Skalenrichtungen und Umfragemodi

	absteigend	aufsteigend	F _{Richtung}
PaPi Modus	2.88	2.91	n.s.
Web Modus	2.70	2.80	
F _{Modus}	5.49*		

Anmerkungen: *p < 0.05. 4 Fragen zu Leistungsmotivation, 4 Fragen zu intrinsischer und 4 Fragen zu extrinsischer Jobmotivation.

Messinvarianz der Skalenrichtungen

Eine Voraussetzung für die Prüfung des in Hypothese 3 postulierten Unterschieds zwischen den latenten Mittelwerten der beiden Skalenrichtungen ist die Überprüfung der Invarianz der Messmodelle mit ab- und aufsteigender Skalenrichtung im PaPi und im Web Modus. Hierzu wurde ein multipler Gruppenvergleich (MGCFA) durchgeführt. Das Messmodell enthält drei latente Variablen mit jeweils vier Indikatoren: Leistungsmotivation sowie intrinsische und extrinsische Jobmotivation. Es ist ersichtlich (Tabelle 4), dass die globalen Fitparameter für die Übereinstimmung zwischen dem spezifizierten Modell und den Daten im Web Modus recht gut und im PaPi Modus akzeptabel sind (siehe Browne und Cudek, 1993). In beiden Umfragemodi sind die Tests der χ^2 -Differenzen zwischen den hierarchisch geschachtelten Modellen nicht signifikant, was die Folgerung einer Messinvarianz der Skalenrichtungen zulässt. Daher kann ein Vergleich der latenten Mittelwerte erfolgen.

Tabelle 4. Prüfung der Messinvarianz zwischen ab- und aufsteigender Antwortskalenrichtung im PaPi und im Web Modus

PaPi Modus	χ^2	df	χ^2 -Diff	RMSEA	CFI
Konfigural	155.11 (1.11)	100		0.059	0.941
Metrisch	157.90 (1.11)	112	2.79	0.051	0.951
Skalar	172.91 (1.10)	124	14.83	0.050	0.948
Web Modus					
Konfigural	132.63 (1.11)	100		0.045	0.977
Metrisch	147.57 (1.09)	112	14.76	0.044	0.975
Skalar	168.22 (1.09)	124	20.65	0.047	0.968

Anmerkungen: Alle Berechnungen wurden mittels MLM Diskrepanzfunktion durchgeführt. Die Skalierungsfaktoren zur Korrektur bei der Berechnung des χ^2 -Differenztests stehen in Klammern. In allen Modellen wurde eine Residuenkorrelation zwischen zwei – die Leistungsmotivation adressierenden – Indikatoren zugelassen.

Latente Mittelwertunterschiede

Die globalen Fitparameter der Modelle für den Vergleich der latenten Mittelwerte betragen für den PaPi Modus [$\chi^2(121) = 166.65 (1.10)$; RMSEA = 0.049; CFI = 0.951] und für den Web Modus [$\chi^2(121) = 164.48 (1.08)$; RMSEA = 0.047; CFI = 0.969]. Da alle Werte in absteigender Richtung codiert sind und die aufsteigende Skalenrichtung die Referenzgruppe bildet, geben die negativen Vorzeichen an, dass bei absteigender Skalenrichtung die Antworten positiver

ausfallen als bei aufsteigender Skalenrichtung. Allerdings ist dieser Effekt nur bei Leistungsmotivation ($p=0.015$) und bei extrinsischer Jobmotivation (mit $p=0.088$ knapp) signifikant. Aufgrund dieses Ergebnisses kann Hypothese 3a – in der für den PaPi Modus signifikante Unterschiede zwischen den beiden Skalenrichtungen postuliert wurden – nicht in dieser allgemeinen Form aufrechterhalten werden. Im Web Modus sind keine signifikanten Unterschiede zwischen ab- und aufsteigender Skalenrichtung zu beobachten, was die Vermutung in Hypothese 3b stützt.

Tabelle 5. Vergleich der latenten Mittelwerte für ab- und aufsteigende Skalenrichtung im PaPi und im Web Modus (nicht standardisierte Lösung)

PaPi Modus	Schätzung	Standardfehler	Kritischer Wert
Leistungsmotivation	-0.305	0.125	-2.444*
Intrinsische Jobmotivation	0.033	0.124	0.265
Extrinsische Jobmotivation	-0.206	0.120	-1.709
<hr/>			
Web Modus			
Leistungsmotivation	-0.120	0.117	-1.027
Intrinsische Jobmotivation	0.181	0.115	1.570
Extrinsische Jobmotivation	-0.049	0.117	-0.421

Anmerkungen: * $p < 0.01$. Antwortskalen wurden einheitlich codiert von 1 "positiv" bis 5 "negativ". Die Referenzgruppe ist die aufsteigende Skalenrichtung. 12 Fragen: 4 Fragen zu Leistungsmotivation, 4 Fragen zu intrinsischer und 4 zu extrinsischer Jobmotivation.

Messinvarianz der Umfragemodi

Zur Prüfung der Hypothese 4, die aufgrund der Ähnlichkeit der Befragungssituation von selbstadministrierten Befragungen im Papi und im Web Modus Messinvarianz zwischen den Befragungsmodi postuliert, wird erneut eine MGCFA mit den bekannten drei latenten Variablen berechnet. Die Vergleichsgruppen sind nun jedoch die Umfragemodi (PaPi und Web) für die ab- und die aufsteigende Skalenrichtung. Die Ergebnisse zeigen, dass zwischen den Umfragemodi keine Messinvarianz gegeben ist, was durch die signifikanten Tests der χ^2 -Differenzen zwischen den hierarchisch geschachtelten Modellen impliziert wird. Bei aufsteigender Skalenrichtung ist die der χ^2 -Differenz-Test zwischen der konfiguralen und der metrischen Invarianz nicht signifikant, was hier zumindest metrische, nicht aber skalare Invarianz vermuten lässt. Hypothese 4 ist auf Grund dieses Ergebnisses als widerlegt zu betrachten. Die Unterschiede zwischen den beiden Modi bei der Beantwortung von Fragen können durch die Ähnlichkeit der Befragungssituation nicht abgemildert werden.

Tabelle 6. Prüfung der Messinvarianz zwischen PaPi und Web Modus bei gleichbleibender (ab- und aufsteigender) Antwortskalenrichtung

Absteigend	χ^2	<i>df</i>	χ^2 -Diff	RMSEA	CFI
Konfigural	160.75 (1.13)	100		0.061	0.948
Metrisch	189.94 (1.12)	112	29.99*	0.066	0.933
Skalar	225.68 (1.10)	124	38.99***	0.071	0.912
Aufsteigend					
Konfigural	126.35 (1.09)	100		0.041	0.978
Metrisch	146.48 (1.08)	112	20.55	0.044	0.971
Skalar	177.24 (1.07)	124	32.00***	0.052	0.955

Anmerkungen: *p < 0.05; ***p < 0.01. Alle Berechnungen wurden mittels MLM Diskrepanzfunktion durchgeführt. Die Skalierungsfaktoren zur Korrektur bei der Berechnung des χ^2 -Differenztests stehen in Klammern. In allen Modellen wurde eine Residuenkorrelation zwischen zwei – die Leistungsmotivation adressierenden – Indikatoren zugelassen.

Diskussion und Fazit

Ziel dieser Studie war es, den Einfluss des Umfragemodus (PaPi vs Web) auf die Beantwortung von Fragen bei ab- und aufsteigender Antwortskalenrichtung zu untersuchen. Hierzu wurden Fragen verwendet, die sich mit Leistungsmotivation, intrinsischer und extrinsischer Jobmotivation beschäftigen. Die Formulierung der Forschungshypothesen stützt sich einerseits auf die satisficing Theorie (Krosnick 1991), der zufolge Befragungspersonen eher dazu neigen, Kategorien am Beginn einer Antwortskala zu wählen. Andererseits basieren die Forschungshypothesen auf empirischen Befunden einer webbasierten Eyetracking Studie (Höhne und Lenzner 2015), der zufolge Befragungspersonen bei der Beantwortung von horizontal angeordneten Ratingskalen eine Tendenz zur Mitte aufweisen. Für den PaPi Modus wurde ein signifikanter primacy Effekt erwartet, für den Web Modus hingegen wurde kein primacy Effekte erwartet. Interessanterweise zeigen die empirischen Ergebnisse, dass der postulierte primacy Effekt im PaPi Modus nur bei absteigender Antwortskalenrichtung auftritt (siehe Menold und Bogner 2015). In Einklang mit unseren Erwartungen ist im Web Modus kein primacy Effekt zu beobachten.

Wie im PaPi Modus treten zustimmende Antworten im Web Modus bei ab- und aufsteigender Skalenrichtung mit sehr ähnlichen Häufigkeiten auf. Allerdings sind diese signifikant niedriger als im Papi Modus. Dieses Ergebnis kann als Beleg für eine Tendenz zur Mitte bei der Beantwortung horizontal präsentierter Ratingskalen in webbasierten Umfragen angesehen werden. Dies korrespondiert mit den Ergebnissen von Höhne und Lenzner (2015).

Der Umstand, dass die zustimmenden Antworten in beiden Umfragemodi bei ab- und aufsteigender Skalenrichtung mit ähnlichen (wenn auch Modus-spezifisch unterschiedlichen) Häufigkeiten auftreten, gibt Anlass zu der Vermutung, dass die auf der satisficing Theorie basierende Erklärung für primacy Effekte vornehmlich für die absteigende Skalenrichtung zutrifft. Zur Erklärung der Häufigkeit zustimmender Antworten bei aufsteigender Skalenrichtung erscheint es sinnvoll die „cognitive elaboration“ Theorie (Sudman et al. 1996) heranzuziehen, die die Antwortskala als eine Art Überredungskommunikation versteht. Dieser Erklärungsansatz nimmt nicht nur die Antwortskala, sondern die Frage-Antwort-Kombination in den Blick. Dadurch wird es möglich, Effekte der Verknüpfung von Skalenrichtung und Frageinhalt zu berücksichtigen. Bei den hier verwendeten Fragen zur Leistungs- und

Jobmotivation impliziert eine negative Antwort eine wenig positive Selbstbeschreibung. Im Sinne der Impression-Management Theorie (siehe Mummendey und Bolten 1993) scheinen Befragungspersonen negative Antworten zu vermeiden, so dass bei aufsteigenden Antwortskalen nicht die erste negativ konnotierte Kategorie, sondern die erste positiv selbstdarstellende Kategorie gewählt wird.

Die Messinvarianz der Skalenrichtungen (ab- und aufsteigend) wurde bereits bei Höhne und Krebs (2017) dargelegt und konnte in dieser Studie für beide Umfragemodi (PaPi und Web) repliziert werden. Der Vergleich der latenten Mittelwerte von Leistungsmotivation sowie intrinsischer und extrinsischer Jobmotivation zeigen im Web Modus keinerlei signifikante Richtungsunterschiede. Dieser Befund entspricht unseren Erwartungen. Im PaPi Modus zeigt sich entgegen unserer Erwartungen ein etwas anderes Bild, da lediglich für Leistungsmotivation ein signifikanter Unterschied zwischen den latenten Mittelwerten zu verzeichnen ist. Dennoch ist unseres Erachtens selbst bei bestehender Messinvarianz der Skalenrichtungen Vorsicht geboten, da zumindest im PaPi Modus die Berücksichtigung der Antwortskalenrichtung essentiell erscheinen kann.

Ausgehend von der vermeintlichen Ähnlichkeit der Befragungssituation wurde bei gleichbleibender Skalenrichtung Messinvarianz zwischen den Modi (PaPi und Web) postuliert und widerlegt. Zwischen PaPi und Web Modus besteht keine Invarianz der Messungen und deshalb ist beim Vergleich von Daten aus Befragungen im PaPi und Web Modus zunächst einmal Vorsicht geboten. Dieses Ergebnis ist deshalb von besonderer Bedeutung, da bedeutende nationale und internationale Umfragen aus Zeit- und Kostengründen zunehmend „mixed-mode“ Designs verwenden. Die unterschiedliche Art des Beantwortens im PaPi und Web Modus scheint die Kombination dieser zwei Modi – außer zu methodologischem Erkenntnisgewinn – in Frage zu stellen.

Abschließend sind noch einige Einschränkungen zu erwähnen, die mit der vorliegenden Studie verbunden sind. Zu nennen ist vor allem die Stichprobe von Studierenden, auf deren Basis die vorgestellten Ergebnisse entstanden sind. Da es hier jedoch um die Prüfung von Hypothesen geht und nicht um eine Verallgemeinerung der Ergebnisse, sehen wir die interne Validität der Studie nur geringfügig eingeschränkt. Ein weiterer wichtiger Aspekt ist, dass einerseits kein adaptives Umfragedesign im Web Modus verwendet wurde und andererseits nicht zwischen Gerätetypen (bspw. PC und Smartphones) unterschieden werden kann. Es wäre methodisch wünschenswert, dass folgende Studien beide Aspekte berücksichtigen. Des Weiteren wäre es wünschenswert weitere Frageinhalte mit unterschiedlichen Personengruppen zu testen.

Literaturverzeichnis

- Bishop, G.F. 1990. Issue involvement and response effects in public opinion surveys. *Public Opinion Quarterly* 54: 209–218.
- Bishop, G.F., und A. Smith. 2001. Response-order effects and the early Gallup split-ballots. *Public Opinion Quarterly* 65: 479–505.
- Browne, M.W., und R. Cudek. 1993. Alternative ways of assessing model fit. In *Testing Structural Equation Models*, Hrsg. K.A. Bollen und J.S. Long, 136–162. San Francisco: Sage.

- DeCastellarnau, A. 2017. A classification of response scale characteristics that affect data quality: A literature review. *Quality and Quantity*. Doi: 10.1007/s11135-017-0533-4
- de Leeuw, E.D., und J.J. Hox. 2014. Survey mode and mode effects. In *Improving Survey Methods: Lessons from Recent Research*, Hrsg. U. Engel, B. Jann, P. Lynn, A. Scherpenzeel und P. Sturgis, 22–34. New York: Routledge.
- de Leeuw, E.D., und J.J. Hox. 2011. Internet surveys as part of a mixed-mode design. In *Social and Behavioral Research and the Internet. Advances in Applied Methods and Research Strategies*, Hrsg. M. Das, P. Ester, und L. Kaczmirek, 45–76. New York: Routledge.
- Dillman, D.A., und B.L. Messer. 2010. Mixed-mode surveys. In *Handbook of Survey Research*, Hrsg. P.V. Marsden und J.D. Wright, 551–574. Bingley: Emerald.
- Engel, U. 2014. Survey modes and response effects: Overview and introduction. In *Improving Survey Methods: Lessons from Recent Research*, Hrsg. U. Engel, B. Jann, P. Lynn, A. Scherpenzeel und P. Sturgis, 15–21. New York: Routledge.
- Engel, U., B. Jann, P. Lynn, A. Scherpenzeel, und P. Sturgis. 2014. Improving survey methods: General introduction. In *Improving Survey Methods: Lessons from Recent Research*, Hrsg. U. Engel, B. Jann, P. Lynn, A. Scherpenzeel und P. Sturgis, 1–11. New York: Routledge.
- Höhne, J.K., und D. Krebs. 2018. Scale direction effects in agree/disagree and item-specific questions: A comparison of question formats. *International Journal of Social Research Methodology* 21: 91–103.
- Höhne, J.K., und L. Lenzner. 2015. Investigating response order effects in web surveys using eye tracking *Psihologija* 48: 361–377.
- Krebs, D. und Y.G. Bachner, Y.G. 2017. Effects of rating scale direction under the condition of different reading conditions. *Methods, Data, Analyses*, 1-21; doi: 10.12758/mda.2017.08.
- Krosnick, J.A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5: 213–236.
- Krosnick, J.A., und D.F. Alwin. 1987. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly* 51: 201–219.
- Krosnick, J.A., S. Narayan, und W.R. Smith. 1996. Satisficing in surveys: Initial evidence. In *New Directions for Evaluation: Advances in Survey Research*, Hrsg. M.T. Braverman und J.K. Slater, 29–44. San Francisco: Jossey-Bass Publisher.
- Krosnick, J.A., und S. Presser. 2010. Question and questionnaire design. In *Handbook of Survey Research*, Hrsg. P.V. Marsden und J.D. Wright, 263–313. Bingley: Emerald.
- Mathews, C.O. 1929. The effect of the order of printed response words on an interest questionnaire. *Journal of Educational Psychology* 30: 128–134.
- Menold, N., und K. Bogner. 2015. Gestaltung von Ratingskalen in Fragebögen. http://www.ssoar.info/ssoar/bitstream/handle/document/39366/ssoar-2014-menold_et_al-Gestaltung_von_Ratingskalen_in_Fragebogen.pdf?sequence=1. Zugegriffen: 13. Januar 2018.
- Mummendey, H.D., und H.G. Bolten. 1993. Die Impression-Management Theorie. In *Theorien der Sozialpsychologie: Motivations- und Informationsverarbeitungstheorien*, Hrsg. D. Frey, 57–77. Bern: Huber.

- Rugg, D., und H. Cantril. 1944. The wording of questions. In *Gauging Public Opinion*, Hrsg. H. Cantril, 23–50. Princeton: Princeton University Press.
- Schuman, H., und S. Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks: Sage.
- Schwarz, N., und H.J. Hippler. 1991. Response alternatives: The impact of their choice and presentation order. In *Measurement Errors in Surveys*, Hrsg. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, und S. Sudman, 41–56. Hoboken: John Wiley and Sons.
- Schwarz, N., H.J. Hippler, und E. Noelle-Neumann. 1992. A cognitive model of response-order effects in survey measurement. In *Context Effects in Social and Psychological Research*, Hrsg. N. Schwarz und S. Sudman, 187–202. New York: Springer.
- Sudman, S., N.M. Bradburn, und N. Schwarz. 1996. *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass Publishers.
- Yan, T., und F. Keusch. 2015. The effects of the direction of rating scales on survey responses in a telephone survey. *Public Opinion Quarterly* 79: 145–165.

Anhang

Formulierung der Fragen mit Antwortskalen in absinkender Richtung.

Es macht mir Spaß mit anderen im Wettbewerb zu stehen. (LM)

Es befriedigt mich, wenn meine Leistungen (im Studium/Beruf) besser sind als die von anderen. (LM)

Bei allem was ich tue, bemühe ich mich, besser zu sein als andere. (LM)

Ich streng mich mehr an, wenn ich mit anderen im Wettbewerb stehe. (LM)

Selbstständig arbeiten zu können ist mir wichtig. (IJM)

Eigene Talente und Fertigkeiten einzusetzen ist mir wichtig. (IJM)

Aufgaben, die Verantwortungsbewusstsein erfordern, sind mir wichtig. (IJM)

Eigene Ideen verwirklichen zu können ist mir wichtig. (IJM)

Ein hohes Einkommen ist mir wichtig. (EJM)

Gute Aufstiegsmöglichkeiten sind mir wichtig. (EJM)

Klare Karriereperspektiven sind mir wichtig. (EJM)

Ein Beruf, der anerkannt und geachtet wird, ist mir wichtig. (EJM)

trifft voll und ganz zu, trifft eher zu, trifft in mittlerem Ausmaß zu, trifft eher nicht zu, trifft überhaupt nicht zu

Anmerkungen: Die 12 Testfragen wurden in Matrixform präsentiert. LM = Leistungsmotivation; IJM = Intrinsische Jobmotivation; EJM = Extrinsische Jobmotivation