
Review: The effect of rating scale format on response styles: The number of response categories and response category labels

Alexander Büscher

Informationswirtschaft (B.Sc.) SPO 2009

Informationsdienste und elektronische Märkte

Karlsruher Institut für Technologie (KIT)

uxxxx@student.kit.edu

Abstract

Der Artikel von Weijters, Cabooter und Schillewaert [WCS10] beschäftigt sich mit dem Einfluss durch unterschiedliche Gestaltungen und Eigenschaften von Fragebögen auf deren Teilnehmer und ihre Antworten. Hierbei werden in Studie 1 unterschiedliche Antworttendenzen betrachtet, die durch die Anzahl der Antwortmöglichkeiten, die Art der Beschriftungen sowie der An- beziehungsweise Abwesenheit des Mittelpunktes bedingt werden. Studie 2 beschäftigt sich dann genauer mit dem expliziten Fall von Skalen mit fünf Antwortmöglichkeiten. Ein erstes Ergebnis gibt an, dass die gängige Befragungsart (7 Antwortmöglichkeiten) hinsichtlich MR nicht zwangsläufig die besten Ergebnisse liefert und stattdessen eher auf Befragungen mit 5 Antwortmöglichkeiten gesetzt werden sollte. Studie 2 stellt fest, dass die Validität von Kriterien eher gegeben ist, wenn man sich auf eine Beschriftung der äußeren Antwortmöglichkeiten beschränkt.

1 Einführung

Die Vermutung, dass die Anzahl der Antwortmöglichkeiten mitsamt ihrer Beschriftung überhaupt einen Einfluss auf das Antwortverhalten der Befragten haben könnte, wurde 1992 von Greenleaf aufgegriffen [Gre92]. Gleichzeitig zur Veröffentlichung dieses frühen Ansatzes rief er dazu auf, genauere Nachforschungen zur Thematik durchzuführen. Diesem Aufruf folgten vor allem Weathers, Sharma und Niedrich mit ihrer Ausarbeitung zu den Thematiken “status quo heuristic” und “Beeinflussung des Antwortverhaltens durch die Anzahl der Antwortmöglichkeiten” im Jahre 2005 [WSN05]. Wie das Autorenteam zeigt, beschäftigten sich andere Autoren (u.a. auch Greenleaf [Gre92]) oftmals mit einzelnen Befragungsarten (beispielsweise fünf Antwortmöglichkeiten), ohne dabei das Problem als Ganzes in den Fokus zu rücken. Eben diese Untersuchung wird durch die zwei Studien aufgegriffen und anhand von insgesamt 16 Hypothesen, die es zu verifizieren beziehungsweise falsifizieren galt, betrachtet. Diese Betrachtung ist Thema des vorliegenden Artikels, welcher im Umfeld des (theoretischen) analytischen CRM und somit im Bereich des Marketing anzusiedeln ist.

2 Inhaltliche Zusammenfassung

Der folgende Teil dieses Reviews stellt in einem knappen Rahmen die wichtigsten Inhalte der einzelnen Kapitel des Papers dar.

2.1 Introduction, Conceptional Background und Hypothesis Development

Neben der Motivation zur Erstellung des Papers werden als Einleitung in die Thematik die Abkürzungen der drei in diesem Artikel betrachteten Bias-Arten “misresponse to reversed items” (MR), “net acquiescence response style” (NARS) und “extreme response

style“ (ERS) erläutert. Auf den Grundlagen von Greenleaf [Gre92] werden im Kapitel *Conceptional background* die Eigenschaften und Gründe von NARS und ERS herausgearbeitet und von dem Phänomen MR abgegrenzt, dem sich auf Grundlage von Wang et al. [WRB03] genähert wird. Außerdem wird beschrieben, wie das Gesamtphänomen der Bias-Arten und nicht ein einzelner Fall betrachtet wird.

Der Hauptteil der ersten Studie startet mit dem Kapitel *Hypothesis development*. Es werden insgesamt 14 Hypothesen aufgestellt, die – in dieser Reihenfolge – den Einfluss auf NARS, ERS und MR in den Bereichen “Beschriftungsmöglichkeiten“, “An- /Abwesenheit des Mittelpunktes“, “Anzahl der Antwortmöglichkeiten“, “Beschriftung und Mittelpunkt“ sowie “Anzahl der Antwortmöglichkeiten und Beschriftungsmöglichkeiten“ untersuchen. Im Sinne der Lesbarkeit wurde auf die Einzelaufzählung aller Hypothesen verzichtet – diese finden sich in Anhang A.1.

2.2 Methodology und General discussion

Die Grundgesamtheiten der beiden Studien werden beschrieben. Des Weiteren werden die Grundlagen zur Überprüfung der Hypothesen beider Studien von Swain et al. aus dem Jahre 2008 [SWN08] und Bland & Altman aus dem Jahre 2000 [BA00] genannt. Sowohl die binäre Markierung im Zusammenhang mit dem gemessenen MR, als auch die Messungen bezüglich ERS und NARS, werden in diesem Teil durch eine Tabelle und ein Schaubild zusammengefasst. Diesem Teil folgt sowohl die Veranschaulichung, als auch die Interpretation der Ergebnisse, welche in den Tabellen 3 bis 5 zusammengefasst werden.

Als Vorgehensweise von Studie 2 wird eine einfache lineare Regression beschrieben, mit welcher der Einfluss durch Beschriftung der Antwortmöglichkeiten auf das NARS und ERS-Level gemessen wird (Hypothesen: siehe Anhang A.2). Abschließend werden die ermittelten Werte interpretiert und auf die beiden Hypothesen bezogen.

Im abschließenden Kapitel unterstreicht das Autorenteam, wie stark limitiert die Vergleichbarkeit zwischen unterschiedlich aufgebauten Umfragen ist. Ferner geben sie an, dass Umfragen mit 5 anstatt 7 Antwortmöglichkeiten hinsichtlich MR qualitativ bessere Ergebnisse liefern. Implikationen, wann welches Design in Umfragen angebracht ist, werden durch Schaubild 3 zusammengefasst. Ebenfalls wird der Bedarf nach weiteren Nachforschungen betreffend Studie 2 geäußert.

3 Diskussion

Hier beginnt der Hauptteil des Reviews, der eine kritische Stellungnahme sowohl zum Inhalt, Aufbau und Methodik, als auch zum Journal und Autorenteam beinhaltet.

3.1 Autoren

Im ersten Teil der Diskussion soll das dreiköpfige Autorenteam, das an der Publikation mitgearbeitet hat, untersucht werden.

Bert Weijters ist an der Universität in Gent angestellt. Laut seiner dortigen Homepage¹ hat er bereits an 16 veröffentlichten Artikeln mitgearbeitet, gut die Hälfte von diesen zum Thema “response styles“. Alle Artikel wurden dort auf einer Skala von 1 (gut) bis 5 (durchschnittlich) mit 1 bewertet. Zusätzlich hat er Veröffentlichungen zu zwei Konferenzen eingereicht, die beide mit 3 bewertet wurden. Dies zeigt, dass er auf dem Gebiet erfahren ist und sich bereits vorher mit der Thematik auseinander gesetzt hat.

Ebenfalls an der Universität Gent ist Niels Schillewaert². Bisher partizipierte er an der Veröffentlichung von 13 Artikeln, die im Bereich von 1 bis 4 bewertet wurden. Der Mittelwert der Bewertung beträgt genau 2. Auch bei ihm beschäftigen sich ca. die Hälfte

¹Uni Gent, Bert Weijters <https://biblio.ugent.be/person/802000878995> - Abruf: 22.05.2014

²Uni Gent, Niels Schillewaert https://biblio.ugent.be/publication?q=author\%3D\%22Schillewaert\%2C+Niels*\%22+or+\%28typeexact+bookEditor+and+editor\%3D\%22Schillewaert\%2C+Niels*\%22\%29 - Abruf: 22.05.2014

der Artikel mit der Thematik der “response styles“ – die letzten acht Veröffentlichungen tätigte er allesamt zusammen mit Bert Weijters.

Elke Cabooter³ ist die dritte Person aus dem Autorenteam. Das vorliegende Paper ist ihr erster veröffentlichter Artikel. Neben diesem publizierte sie bereits zu drei Konferenzen, deren Veröffentlichungen im Bereich 1 bis 3 bewertet wurden. Sie ist damit zwar die unerfahrenste der drei Autoren, hat jedoch bereits zum Thema “response styles“ mit Bert Weijters zusammengearbeitet.

3.2 Journal

Der Artikel erschien im Jahre 2010 im International Journal of Research in Marketing. Zwischen der ersten Einsendung des Papers und der Veröffentlichung lagen 17 Monate, von denen $5\frac{1}{2}$ Monate Review-Zeit waren [WCS10]. Die wissenschaftliche Norm des Artikels ist durch bis zu vier Reviews gesichert, von denen zwei als peer-review stattfinden⁴. In einem der frühen Rankings wurde das International Journal of Research in Marketing 1997 in einer Studie von Hult, Neese, und Bashawin in einem Overall Rating mit Platz 26 von 41 beurteilt [HNB97]. In einem aktuelleren Ranking von 2012 können wir feststellen, dass sich das Journal bereits auf Platz 13 von 101 vorgearbeitet hat und somit im oberen Quartal der Liste der Marketing Journale zu finden ist⁵.

In einer genauen Auflistung der Jahre 2003 bis 2013 [Har] liegt das Journal regelmäßig im oberen Bereich der Bewertungen. Zur Zeit der Veröffentlichung, also im Jahre 2010, erhielt es auf einer Skala von 1 bis 4, wobei 1 schlecht und 4 gut ist, durch den *Association of Business Schools Academic Journal Quality Guide* eine Bewertung von 3. Insgesamt kann das Journal also als angemessene und seriöse Publikationsfläche angesehen werden.

3.3 Aufbau und Struktur

Der Artikel ist übersichtlich aufgebaut und sinnvoll strukturiert. In der Methodik wird erläutert, warum man mit der jeweiligen Methode vorgeht und warum sie auf die Problematik passt. Die Verweise sind angemessen platziert und stören die Übersicht nicht. Verschiedene Tabellen und Abbildungen ergänzen den Fließtext sinnvoll. Wünschenswert wäre, wenn im Hauptteil des Papers eine deutlichere Abgrenzung zwischen dem Ende von Studie 1 und dem Start von Studie 2 stattfinden würde. Auch ist fraglich, warum die durch Studie 2 untersuchten Hypothesen nicht formuliert und aufgelistet wurden, wie es noch mit den Hypothesen aus Studie 1 der Fall war.

3.4 Inhaltlich

Die Motivation des Artikels ist verständlich. Die Gründe, welche für eine Analyse in Abhängigkeit vom Zusammenspiel der Faktoren Beschriftung, Anzahl der Antwortmöglichkeiten und An-/ Abwesenheit des Mittelpunktes sprechen, sind nachvollziehbar. Dieser Artikel ist einer der wenigen, die sich nicht auf ein bestimmtes Format und dessen Auswirkungen spezifiziert haben. Verbesserungswürdig ist die Breite der Analyse. Auch wenn in Tabelle 1 die Ergebnisse der Analyse anderer Publikationen im International Journal of Research in Marketing sowie im Marketing Buch von Bruner et al. [BJH01] hinsichtlich der häufigsten Formate angegeben sind, wirken die Grenzen, nicht zuletzt auch wegen der Zahlen (Tabelle 1), eher willkürlich gewählt. Die Autoren erklären die Begrifflichkeiten ERS, NARS und MR gut, geben jedoch keinerlei Begründung an, warum ähnliche Verhaltensweisen nicht konsequenterweise ebenfalls untersucht wurden. Vorstellbar wären an dieser Stelle unter anderem “Disacquiescence response bias“ (DARS) oder “mid-point response style“ (MRS) gewesen. Die Erläuterungen im konzeptionellen

³Uni Gent, Elke Cabooter <https://biblio.ugent.be/person/001999311759> - Abruf: 22.05.2014

⁴Guideline des International Journal of Research in Marketing zur Erstellung von Reviews <http://www.elsevier.com/reviewers/reviewer-guidelines> - Abruf: 23.05.2014

⁵Rankings 2012 <http://www.scimagojr.com/journalrank.php?category=1406> - Abruf: 22.05.2014

Hintergrund der Studien, welche in diesem Paper als Einleitung in die Ausarbeitung der 14 Hypothesen von Studie 1 fungieren, erscheinen durchdacht und zielführend. Ebenso wirken die Ausführungen, aufgrund derer die Hypothesen aufgestellt werden, schlüssig, sodass man auch ohne vorher angeeignetes Fachwissen nachvollziehen kann, wie die Autoren auf die 14 Hypothesen geschlossen haben.

Nun beginnt die Methodologie mit der Beschreibung des Aufbaus und der Grundgesamtheit der ersten Studie. Das beschriebene Design der Umfrage zu Studie 1 stimmt mit dem Fragebogen, der zur Verfügung gestellt wurde, überein. Unklar bleibt im Zusammenhang der Zuweisung der Teilnehmer, warum nicht auf eine komplett ausgewogene Verteilung geachtet wurde. Somit kommt es zu einer Differenz von 38 (175 - 137) Personen zwischen den beiden Umfragen mit 4 Antwortmöglichkeiten. Ebenfalls fragwürdig ist, dass in dieser ersten Studie nur Männer zur Teilnahme eingeladen wurden. Die spezifischen Gründe der Umfrage, mit denen im Text argumentiert wird, sind nach Einsicht in den Originalfragebogen nicht ersichtlich. Positiv bei der Beschreibung der Grundgesamtheit ist zu vermerken, dass sowohl ein breites Altersspektrum (15 bis 69 Jahre), als auch unterschiedliche Bildungsniveaus herangezogen wurden. Ein kleiner Kritikpunkt ist, dass lediglich der Median der Altersverteilung (49) angegeben ist, nicht jedoch die genaue Verteilung. Die Herkunft der Fragen zur Messung des MR-Levels ist aus dem Handbook of Marketing Scales [BN99] gegeben. Hierbei ist anzumerken, dass keine genauen Seitenzahlen vermerkt sind, sodass eine Überprüfung der Herkunft zwar möglich ist, jedoch mühselig sein kann. Die angegebene Methode von Swain aus dem Jahre 2008 [SWN08], auf deren Grundlage die Hypothesen zu MR geprüft werden, ist angemessen beschrieben und auch ein sinnvoller Indikator zur Messung von MR.

Auch im zweiten Teil der ersten Studie, der sich ausschließlich mit der Messung von ERS und NARS beschäftigt, wird die Herkunft der benutzten Fragen aus dem Handbook of Marketing Scales [BN99] und dem Marketing Scales Handbook [BJH01] geklärt - eine genaue Seitenzahl ist jedoch wiederholt nicht angegeben. Positiv anzumerken ist, dass zusätzlich die Reichweite der Korrelationskoeffizienten gegeben ist, um zu gewährleisten, dass es sich um Fragen mit geringer Abhängigkeit untereinander handelt. Als methodische Grundlage wird die Untersuchung zur "odds ratio" von Bland & Altman aus dem Jahre 2000 [BA00] genannt. In dieser wird beschrieben, dass die odds ratio ein probates Mittel ist, sowohl das Verhalten von 2 binären Variablen alleine, als auch ihren Einfluss untereinander aufzuzeigen [BA00]. In unserem Fall wären dies die beiden binären Variablen

$$\text{NARS tritt auf } (1 = ja \mid 0 = nein) \quad (1)$$

$$\text{ERS tritt auf } (1 = ja \mid 0 = nein) \quad (2)$$

Weiter unterscheiden Bland & Altman zwischen der odds ratio (s.o.) und der log odds ratio. Letztere wird ebenfalls im besprochenen Paper benutzt. Sie ist definiert als

$$\ln \frac{\text{Event tritt auf}}{\text{Event tritt nicht auf}} \quad (3)$$

was in unserem Fall die beiden Formeln, die die Autoren aufstellen, erklärt:

$$NARS = \ln \frac{\text{agreements} + 1}{\text{disagreements} + 1} \quad \text{und} \quad ERS = \ln \frac{\text{extreme response} + 1}{\text{non-extreme response} + 1} \quad (4)$$

Als "extreme responses" definieren die Autoren Antworten im ersten oder letzten Punkt der Skala, was angemessen und intuitiv erscheint. Agreements und disagreements werden nicht näher definiert.

Für die Anwendung des Modells ist eine binäre Codierung nötig, die in Tabelle 2 und Figur 1 zusammengefasst wird. Während Tabelle 2 die Ergebnisse der Codierung sehr übersichtlich darstellt, wirkt Figur 1 etwas überfüllt und könnte den Leser des Papers daher vor Probleme stellen. Ebenfalls ist in dieser Figur fraglich, was mit den drei jeweils ausgehenden Kanten aus den Punkten NARS, ERS und MR gezeigt werden soll. Hier

wäre eine kurze Erläuterung angebracht gewesen.

Es folgen die Ergebnisse der durchgeführten Tests. War zur Einleitung dieses Kapitels die Erläuterung der Autoren zu eventuell unbekannten Größen beziehungsweise ihren Abkürzungen noch ein positiver Aspekt, versäumen sie es dieses Mal, die Begrifflichkeiten des durchgeführten Tests angemessen zu klären. Während man im χ^2 -Test die Werte von $\mathcal{X}$ und p noch interpretieren kann, können die Autoren von Lesern ohne spezielles Vorwissen nicht erwarten, mit den Abkürzungen CFI, TLI sowie RMSEA vertraut zu sein. Auch ein Nachschlagen im angegebenen Handbuch von Mplus [MM07] hilft hier wenig weiter. Der χ^2 -Test wird in diesem Fall angemessen benutzt. Dieser ist als

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (5)$$

definiert, wobei O_i die beobachteten (tatsächlichen) und E_i die erwarteten Werte darstellt. Der χ^2 -Test ist also genau dann dazu geeignet, die Nullhypothese zu verwerfen, wenn χ^2 entsprechend der Freiheitsgrade und der gewählten Irrtumswahrscheinlichkeit p groß genug ist.

Neben dem χ^2 -Test findet auch der t-Test zur Überprüfung der Hypothesen Anwendung. Dieser wird in den Hypothesen benutzt, in denen es nicht um die Beschriftung der Antwortmöglichkeiten in Kombination mit einem anderen Merkmal geht. Er prüft in unserem Fall die Signifikanz von unterschiedlichen beobachteten Mittelwerten, sowohl bei Anwesenheit als auch bei Abwesenheit der von der jeweiligen Hypothese untersuchten Merkmale (bspw. An-/ Abwesenheit des Mittelpunktes in H4-H6). Der t-Test ist genau dann dazu geeignet, die Nullhypothese zu verwerfen, wenn der **Betrag** der Teststatistik t entsprechend der Freiheitsgrade und der gewählten Irrtumswahrscheinlichkeit p groß genug ist.

$$t = \frac{\bar{x} - \bar{y}}{s} \cdot \sqrt{\frac{n_x \cdot n_y}{n_x + n_y}} \quad (6)$$

Hierbei sind $\bar{x}$ und $\bar{y}$ die beobachteten Mittelwerte der zu vergleichenden Stichproben, n_x der Stichprobenumfang von X, n_y der Stichprobenumfang von Y und s die Standardabweichung der beiden Stichproben. Wichtig ist, dass die Voraussetzung gleicher Varianzen beider betrachteter Verteilungen beim “two-sample t-Test“ nicht verletzt werden darf und im vorliegenden Fall auch nicht wird.

Die Autoren stellen alle Testergebnisse übersichtlich in Tabelle 5 dar und argumentieren anhand dieser nachvollziehbar über die Annahme bzw. Ablehnung ihrer Hypothesen. Positiv zu bewerten ist, dass die Autoren auch eigene Hypothesen aufgrund der Testergebnisse ablehnen. Dies deutet darauf hin, dass ein Publishing-Bias, welcher besagt, dass Autoren dazu tendieren, nur positive Ergebnisse zu veröffentlichen, hier nicht vorliegt. In der Beschreibung von Tabelle 7 findet sich eine Ungenauigkeit, da hier Angaben über die Variable “k“ gemacht werden, die nicht benutzt wird. Denkbar wäre, dass diese jedoch aufgrund der in der Tabelle erwähnten Freiheitsgrade ($df = n - k$) genannt wird. Den Autoren gelingt der Übergang zwischen Studie 1 und Studie 2, indem sie aufgrund ihrer Funde aus Studie 1 argumentieren, dass die gängige Art und Weise Fragebögen zu gestalten, nämlich 7 Antwortmöglichkeiten, bei denen die äußersten Punkte beschriftet sind, nicht notwendigerweise auch die besten Ergebnisse liefert. Vorgeschlagen wird ein Format mit 5 Antwortmöglichkeiten, welches in Studie 2 untersucht wird.

Die Definition des Ziels von Studie 2 wird angemessen dargestellt. Ebenfalls gut zu bewerten ist ein Rückverweis auf die gängige Beschriftung in diesem Format, die in Tabelle 1 dargestellt wurde. Erstmals gehen die Autoren auch auf das Thema Validität von Kriterien anhand des Wertes der “explained variance“ ein. Wenn wir auf die Beschaffenheit der Umfrage und der Grundgesamtheit zu sprechen kommen, stoßen wir auf ähnliche Probleme wie in Studie 1. Auch wenn die Beschreibung der Grundgesamtheit hinsichtlich Bildungsniveau, Geschlecht und Alter angegeben ist, wäre vor allem bei Letzterem eine genauere Auflistung wünschenswert gewesen. Wieder sind nur Grenzen und Median genannt. Des Weiteren fällt auf, dass in dieser Studie Frauen und Männer befragt wurden,

was in Studie 1 laut der Autoren nicht möglich war. Auch wenn es sich hierbei um eine andere Studie handelt, unterstreicht dies die Fragwürdigkeit, Frauen aus Studie 1 auszuschließen. Außerdem sind dieses Mal jeweils 113 Teilnehmer auf beide Studien verteilt. Nicht ersichtlich ist, warum die Autoren auf eine genaue Definition oder Benennung der beiden getesteten Hypothesen (siehe Anhang A.2) verzichten, während sie in Studie 1 noch übersichtlich alle auflisteten. Angemessen wurde wieder dargestellt, dass die verschiedenen Fragen mit einer Korrelation zwischen 0,18 und 0,21 verschiedenartig sind. Während die Ermittlung der Ergebnisse von Studie 2 gut beschrieben wird, finden sich in der Interpretation einige Unklarheiten und Ungenauigkeiten. Mit Bezug auf Tabelle 9 des Artikels werden der Reihe nach Aussagen über “regression weights“, “ R^2 “ und “variances“ gemacht. Wenn die Autoren in ihrem Text von einer Ausnahme bezüglich der regression weights sprechen, sind es tatsächlich zwei Ausnahmen. Betrachten wir die erwähnte Tabelle 9 des Artikels, erkennen wir, dass in der äußersten Spalte, in der das Verhältnis der Ergebnisse beider untersuchter Beschriftungsmöglichkeiten gebildet wird, ungenaue Rundungen auftreten. Wenn in Zeile zwei das Verhältnis $\frac{18}{18,1}$ auf 100% gerundet wird, damit dieser Wert diese These noch unterstützt und nicht verwirft, muss mindestens von Ungenauigkeiten gesprochen werden, da Thesen nur bei einem Verhältnis von $\geq 100\%$ akzeptiert werden. Ähnliches stellen wir beim Bestimmtheitsmaß und der Varianz fest. Bei der Varianz sprechen die Autoren von einem Verhältnis von sechs aus neun Fällen, in denen die Varianz im Falle der Beschriftung der äußeren Antwortmöglichkeiten größer ist. Betrachten wir nun die Tabelle genau, stellen wir fest, dass die Varianz in fünf von neun Fällen echt größer ist, in drei von neun Fällen echt kleiner und in einem Fall identisch. Auch wenn es sich hierbei um gerundete Werte handelt, ist dies ungenau vorgetragen und darf in der Auswertung der Ergebnisse nicht so dargestellt werden – schon gar nicht ohne Kommentar der Autoren. Durch eben dieses ungenaue Arbeiten erscheint das Ergebnis von Studie 2, welches beide Thesen der Autoren unterstützt, unglaubwürdiger. Das Paper wird mit einem Ausblick und einigen Empfehlungen der Autoren, basierend auf ihren Studien, angemessen abgeschlossen. Abbildung 3 fasst diese Empfehlungen, abhängig vom Einsatzfeld der Umfrage, anschaulich zusammen. Während der Hinweis, dass vor allem in Studie 2 noch deutlicher Forschungsbedarf besteht, begrüßenswert ist, sollte man kritisch betrachten, dass der kulturelle Unterschied der Grundgesamtheiten der beiden Studien lediglich mit einem (dem letzten) Satz angesprochen wird.

3.5 Ergebnisse

Die Ergebnisse der Studie sind in Verhältnis zu den Ergebnissen anderer Studien zu stellen. Wie von den Autoren eingangs des Papers richtig erwähnt, beschäftigen sich die meisten alternativen Paper mit den Eigenschaften eines speziellen Aufbaus. Als anderes, großes Feld wird hauptsächlich nach interkulturellen Unterschieden geforscht – auch hier sind Ergebnisse also nur bedingt vergleichbar. Unterstützung in der These, dass Antwortmöglichkeiten in einem gesunden Verhältnis möglichst gering gehalten werden sollten (in unserem Fall also die Variante mit 5 Antwortmöglichkeiten), erhalten die Autoren aus der Studie von Weathers et al. aus dem Jahre 2005 [WSN05] bezüglich der status quo heuristic (SQH).

Dem gegenüber stehen die Erkenntnisse von Preston und Colman aus dem Jahr 2000 [PC00], die besagen, dass vor allem in den Bereichen Validität und (Test–Retest–)Verlässlichkeit die Tests mit 7 bis 9 Antwortmöglichkeiten besser abschneide. Dies wird von den Autoren nicht erwähnt, obwohl diese Studien zum Erscheinungstermin des Artikels bereits 10 Jahre alt waren.

4 Fazit und Anwendbarkeit

Inhalt und Thema des Papers entsprechen den Ansprüchen des International Journal of Research in Marketing. Verschiedene Frameworks und Tabellen ergänzen den Text sinnvoll und illustrieren die Thematik angemessen. Die jeweils benutzten Grundgesamtheiten

sowie Methoden sind mit Quellen vermerkt, die Rohdaten waren ohne größere Probleme vom Hauptautor (Bert Weijters) zu erhalten. Leider schwächt das Paper diesen guten Eindruck durch Ungenauigkeiten. Vor allem die Darstellung und Interpretation der Ergebnisse von Studie 2 sind ungenau dargestellt. Die hieraus resultierenden Implikationen wirken dadurch weniger einleuchtend. Ebenfalls verbesserungswürdig ist die Genauigkeit der Angabe der Grundgesamtheiten, speziell im Bereich der Altersverteilung. Der Übersicht hätte eine genaue Beschreibung der zu prüfenden Hypothesen in Studie 2 gut getan.

Die Autoren verweisen auf ihre Vorschläge, unter welchen Bedingungen welcher Umfragetyp anzuwenden ist und gewährleisten somit im Zusammenhang mit Figur 3 ebenfalls die Anwendbarkeit der Ergebnisse der untersuchten Hypothesen. Die Relevanz des Artikels ist gegeben – laut Google Scholar wurde dieser 71 Mal zitiert (Stand: 16.07.2014).

Literatur

- [BA00] Bland, J.M. und Altman, D.G. (2000) Statistic notes: The odds ratio. *British Medical Journal*, 7247(320), S. 1468.
- [BJH01] Bruner, G., James, K. und Hensel, P. (2001) *Marketing Scales Handbook*. American Marketing Association, Chicago.
- [BN99] Bearden, W.O. und Netemeyer, R.G. (1999) *Handbook of Marketing Scales*. Sage: Los Angeles.
- [Gre92] Greenleaf, E.A. (1992) Improving Rating Scale Measures by Detecting and Correctiong Bias Components in Some Response Styles. *Journal of Marketing Research*, 29(2), S. 176–188.
- [Har] Harzing Journal Rankings. http://www.harzing.com/download/jql_issn.pdf, accessed: 2014-05-22.
- [HNB97] Hult, G.T.M., Neese, W.T. und Bashaw, R.E. (1997) Faculty Perceptions of Marketing Journals. *Journal of Marketing Education*, 19(1), S. 37–52.
- [MM07] Muthén, L. und Muthén, B. (2007) *Mplus User's Guide*. Muthén and Muthén, Los Angeles, CA.
- [PC00] Preston, C.C. und Colman, A.M. (2000) Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), S. 1–15.
- [SWN08] Swain, S., Weathers, D. und Niedrich, R. (2008) Assessing three sources of misresponse to reversed likert items. *Journal of Marketing Research*, 45(1), S. 116–131.
- [WCS10] Weijters, B., Cabooter, E. und Schillewart, N. (2010) The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), S. 236–247.
- [WRB03] Wong, N., Rindfleisch, A. und Burrouhs, J. (2003) Do Reverse-Worded Items Confound Measures in Cross-Cultural Consumer Research? The Case of the Material Values Scale. *Journal of Consumer Research*, 30(1), S. 72–91.
- [WSN05] Weathers, D., Sharma, S. und Niedrich, R.W. (2005) The impact of the number of scale points, dispositional factors, and the status quo decision heuristic on scale reliability and response accuracy. *Journal of Business Research*, 58(11), S. 1516–1524.

A Anhang

A.1 Hypothesen Studie 1

H1: *Labeling all response categories leads to higher levels of NARS.*

H2: *Labeling all response categories leads to lower levels of ERS.*

H3: *Labeling all response categories leads to lower levels of MR.*

H4: *NARS increases when adding a midpoint.*

H5: *ERS decreases when adding a midpoint.*

H6: *MR decreases when adding a midpoint.*

H7: *ERS decreases when more gradations of (dis)agreement are offered.*

H8: *The full labeling of the response categories strengthens the positive effect of offering a midpoint on NARS.*

H9: *Full labeling of the response categories strengthens the negative effect of offering a midpoint on ERS.*

H10: *The full labeling of the response categories strengthens the negative effect of inclusion of a midpoint on MR.*

H11: *The presence of a midpoint offsets the negative effect of adding more gradations of (dis)agreement on ERS.*

H12: *Adding more gradations reduces the negative effect of offering a midpoint on MR.*

H13: *Fully labeling scales weakens the negative effect of adding more gradations of (dis)agreement on ERS.*

H14: *In formats with labels for the endpoints only, adding more gradations of (dis)agreement leads to an increase in MR.*

A.2 Hypothesen Studie 2⁶

H1: *Labeling all response categories in a 5-point scale format leads to higher NARS.*

H2: *Labeling all response categories in a 5-point scale format leads to lower levels of ERS.*

⁶Die beiden genannten Hypothesen wurden nicht vom Autorenteam genau definiert, sondern nur sinngemäß im Text angegeben.