

Swiss
Insights
News



#6

Aufmerksamkeitschecks in
Onlineumfragen

Datenqualität erhöhen: Was Fische mit Aufmerksamkeitschecks in Onlineumfragen zu tun haben



Raffael Meier,
Mitgründer/ CTO
onlineumfragen.com



Kathrin Staub,
Mitgründerin/Principal Consultant,
onlineumfragen.com



Nina Gwerder,
ehem. Lead Consultant,
onlineumfragen.com

Onlineumfragen werfen Fragen zur Datenqualität auf, etwa durch geringes Involvement der Teilnehmenden. Wir diskutieren Massnahmen, welche die Aufmerksamkeit identifizieren, fördern und die Datenqualität verbessern.

Mit webbasierten Befragungen lassen sich neue Zielgruppen realisieren. Neben der günstigen Zugriffsmöglichkeit, geografischer Unabhängigkeit und logistischen Vorteilen werden aber auch unseriöse Teilnahmen begünstigt. Der Anteil an verschmutzten Daten nimmt zu und die Aussagekraft der Ergebnisse wird vermindert. Weiter kann eine Website im Gegensatz zu einem handfesten Papier-Fragebogen auch das «Gefühl einer geringeren Verantwortlichkeit vermitteln» (Johnson, 2005, S. 108). Dies stellt eine ernsthafte Bedrohung für die Validität der Online-Forschung dar (Oppenheimer et al., 2009; Reips, 2002, 2009; Meier & Gwerder, 2022).

Im Gegenzug liefern Onlineumfragen meist auch Metadaten wie Ausfüllzeiten insgesamt, Ausfüllzeiten vom Laden der einzelnen Frage bis zu deren Beantwortung, Browserkennung, Betriebssystem, Bildschirmgrösse, IP-Adresse, Mausbewegungen und weitere Informationen, welche nach der Feldphase, wenn alle gesammelten Daten vorliegen, für ein Screening unaufmerksamer Antworten herangezogen werden können (Barge & Gehlbach, 2012; Bauermeister et al., 2012; Meade & Craig, 2012; Meier & Gwerder, 2022).

Mit unserem Artikel möchten wir einen Überblick über verschiedene Techniken zur Messung der Aufmerksamkeit in Onlineumfragen ermöglichen – Techniken, die durch akademische Erkenntnisse bestätigt oder überprüft wurden und/oder sich in der Praxis bewährt

haben, um die Anzahl «schlechter Antworten» zu begrenzen.

Aber was sind «schlechte Antworten»?

Generell werden als «schlecht» jene Antworten angesehen, die nicht die eigentliche Meinung oder das Wissen der Umfrageteilnehmenden wiedergeben, sondern irgendetwas anderes im Sinne eines Messfehlers. Die Gründe, weshalb Teilnehmende «schlechte» Antworten abgeben, sind vielfältig: Sie verweigern das korrekte Ausfüllen und geben stattdessen Fake-Antworten ab, sie sind schlichtweg abgelenkt, wissen die Antworten nicht oder nehmen sich nicht die nötige Zeit, ihre Antworten sorgfältig abzufüllen oder in die Antwortformate einzupassen. Dabei gibt es eine Auslegeordnung, was solche «schlechten» Daten angeht:

Als sogenannte «*sinnlose Antworten*» oder auch «*content responsive faking*» (Meade & Craig, 2012; Burns & Christiansen, 2011) werden in der Regel jene Antworten verstanden, bei denen Teilnehmende die Frage an sich aufnehmen und verstehen, aber bewusst keine gültige Antwort geben wollen («*intended faking*»). Dazu zählen betrügerisches Ausfüllen, z. B. in psychologischen Testverfahren wie dem MMPI2 (Rogers, 2003) oder in Job Assessments (Delgado, 2011) – aber auch bewusstes oder unbewusstes item-bezogenes, sozial erwünschtes Antwortverhalten (Paulhus, 1984). Demgegenüber stehen «*content nonresponsivity*»

Antworten, die in keinem Zusammenhang zum Inhalt der Fragen stehen (Nichols et al., 1989; Desimone et al., 2018). Sie werden auch als «random response» (Beach, 1989; Berry et al, 1992), «careless responding» (Curran, Kotrba, & Denison, 2010) oder «protocol invalidity» (Johnson, 2005) bezeichnet. Darunter fallen auch die Subkategorien «response sets» (Jandura, Peter, & Küchenhoff, 2012) und «response styles» (Van Vaerenbergh & Thomas, 2012) – alles formale Verschmutzungen von Daten (Meier & Gwerder, 2022), die mehr oder weniger unabhängig von der Fragestellung und nicht zentraler Gegenstand dieses Artikels sind.

Eine *Zwischenkategorie* bilden Teilnehmende, die an sich bereit wären, eine gültige Antwort abzugeben, dies aber nicht vollumfänglich tun. Beim Satisficing-Verhalten lesen Befragte die Fragestellung lediglich oberflächlich und geben jene Antwort ab, die ihnen zuerst in den Sinn kommt oder ihnen plausibel erscheint (Krosnick, 1991, 1999; Krosnick, Nayaran, & Smith, 1996;). «Pseudoopinions» (Bishop, Oldendick, Tuchfarber, & Bennett, 1980) und «nonattitudes» (Franzén, 2011; Schuman & Presser, 1980) sind Antworten von Befragten, die nicht über das nötige Wissen verfügen, um eine passende Antwort auszuwählen, die Frage falsch oder gar nicht verstehen oder die Fragen und Antworten nicht oder nicht akkurat genug lesen können. Gerade bei Fragebögen für Kinder ist solches «unintentional random responding» häufig. Auch nicht-item-bezogenes, mehr persönlichkeitsbedingtes, sozial erwünschtes Antwortverhalten fällt in diese Zwischenkategorie. Erwähnt sei auch das Under- and Overreporting als Abweichung hin zu mehr oder weniger wahren Antworten durch zu positiv oder zu negativ konnotierte Fragen.

Umfrageforschende stellen sich dieser Herausforderung verminderter Aufmerksamkeit schon, bevor Daten erhoben werden. Dabei gibt es zwei Grundstrategien: Sie integrieren diskret eingewobene Aufmerksamkeitstests in die Fragebögen, um unaufmerksame Teilnehmende zu entlarven und zugunsten der Datenqualität aus der Analyse auszuschliessen, oder sie integrieren explizite, gut sichtbare

Aufmerksamkeitsprompts, um Teilnehmende ganz offen darum zu bitten, aufmerksam zu antworten.

Zur Erinnerung: Aus Sicht des traditionellen kognitiven Modells der Umfragebeantwortung (CMSR, Cognitive Model of Survey Response) von Tourangeau, Rips und Rasinski (2000) durchlaufen Antwortende vier Schritte: erstens das Verständnis der Frage, zweitens das Abrufen relevanter Informationen aus dem Gedächtnis, drittens das Bilden eines Urteils aufgrund der abgerufenen Informationen und viertens das Auswählen einer passenden Antwort resp. das Einpassen oder Editieren der Antwort in das vorgegebene Antwortformat. Das nur oberflächliche oder unvollständige Durchlaufen dieser Schritte nennt Krosnick (1991; 1999) «non-optimal response behavior». Diese fehlende Aufmerksamkeit beeinflusst den kognitiven Beantwortungsablauf auf mindestens vier Arten:

- Teilnehmende verstehen die Frage nicht richtig, weil sie diese nicht oder nicht seriös gelesen haben. Dadurch werden nicht die richtigen Informationen abgerufen.
- Es werden nicht alle Informationen abgerufen, was die Urteilsbildung verzerrt.
- Ein Urteil wird heuristisch gebildet und weist eine mangelhafte Reliabilität auf.
- Teilnehmende können die passende Antwortkategorie nicht auswählen, weil sie die verfügbaren Optionen nicht genügend aufmerksam verarbeitet haben.

Zusammengefasst: Unaufmerksamkeit führt zu Mess- und Non Response-Fehlern.

Unaufmerksamkeit führt auch zu einer Situation, in der «das Rauschen, das durch Teilnehmer entsteht, die die Anweisungen nicht lesen, die Zuverlässigkeit der Daten verringert und die mit der Durchführung von Studien verbundenen Kosten erhöht, da die Anzahl der Teilnehmer, die für ein zuverlässiges Ergebnis erforderlich ist, künstlich erhöht wird» (Oppenheimer et al. 2009, S. 873).

Explizite Ernsthaftigkeitsprüfungen

Was kann nun aber konkret getan werden, um dieses komplexe Datenqualitätsproblem in Antwortdaten zu adressieren? Zunächst: Warum schwierig, wenn es auch einfach geht? Anstatt durch komplizierte Verfahren auf das Verhalten von Teilnehmenden zu schliessen, kann direkt gefragt werden, ob Aufmerksamkeit vorhanden war. Diesen Ansatz verfolgen Ernsthaftigkeitsprüfungen (engl. «seriousness checks»).

Durch Fragen wie «Gibt es Gründe, weshalb wir Ihre Antworten nicht in unsere Analyse einfließen lassen sollten?» oder «Es wäre sehr hilfreich, wenn Sie uns an dieser Stelle mitteilen könnten, ob Sie ernsthaft teilgenommen haben, sodass wir Ihre Antworten für unsere wissenschaftliche Analyse verwenden können, oder ob Sie sich nur durchgeklickt haben, um sich die Umfrage anzuschauen.», wird den Teilnehmenden die Möglichkeit gegeben, offen zu deklarieren, dass sie die Befragung nicht gewissenhaft ausgefüllt haben (Aust, 2013). Als Antwortmöglichkeiten bieten sich an: «Ich habe ernsthaft teilgenommen» und «Ich habe nur durchgeklickt, bitte verwenden Sie meine Daten nicht für die Analyse» (s. Abbildung 1).

FRAGEN

Gültigkeit der Angaben

Es wäre sehr hilfreich, wenn Sie uns an dieser Stelle mitteilen könnten, ob Sie ernsthaft teilgenommen haben, so dass Ihre Antworten vollumfänglich in die Ergebnisse einfließen, oder ob Sie sich nur durchgeklickt haben, um sich die Umfrage anzuschauen?

- Ich habe ernsthaft teilgenommen.
- Ich habe nur durchgeklickt, bitte verwenden Sie meine Daten nicht für die Analyse.

Speichern - nächste Frage

Abbildung 1: Expliziter Seriousness Check am Ende einer Befragung.

Die entsprechende Frage kann entweder zu Beginn (Reips, 2002, 2008, 2009) oder am Ende des Fragebogens platziert werden (Buchanan et al., 2010; Ihme et al., 2009). In der Praxis wird sehr oft die zweite Option bevorzugt, da die Teilnehmenden am Ende der Befragung ihr tatsächliches Verhalten während der Befragung zuverlässiger einschätzen können als ihr geplantes Handeln.

Ernsthaftigkeitsprüfungen wurden auch im Rahmen einer Befragung zu den Bundestagswahlen 2009

untersucht (Aust et al, 2013). Insgesamt gaben 112 (3.2 %) der 3'490 Teilnehmenden an, nicht seriöse Angaben gemacht zu haben. In der anschliessenden Analyse zeigte sich, dass die restlichen 3'378 Teilnehmenden konsistentere Angaben gemacht hatten und dass die Prognose des Wahlergebnisses besser mit der Realität übereinstimmte, wenn die 112 Fälle, die sich selbst als nicht ernsthaft deklariert hatten, ausgeschlossen wurden. Andere Autorinnen und Autoren fanden in unterschiedlichen Kontexten auch Werte in der Höhe von 5 bis 6 % oder 30 bis 50 % (Musch & Klauer, 2002; Reips, 2009). Grundsätzlich dürfte die Menge an unseriösen Teilnahmen stark vom Befragungskontext sowie von den Anreizen der Teilnehmenden abhängen. Beispielsweise ist bei Befragungen, bei welchen am Ende eine Belohnung winkt, mit sehr tiefen Selbstdeklarationsquoten zu rechnen, da die Teilnehmenden befürchten, auf eine Gewinnchance verzichten zu müssen, wenn sie zugeben, unseriös ausgefüllt zu haben. Dies hängt wohl stark davon ab, wie Ernsthaftigkeitsprüfung und «Gewinnseite» verquickt sind.

Die Herausforderung bei diesem direkten Ansatz liegt grundsätzlich in einer geschickten Formulierung, welche die Teilnehmenden zu einer ehrlichen Antwort motiviert, sie gleichzeitig aber nicht vor den Kopf stösst. Auch gibt es Befragungen, bei denen dieser Ansatz aufgrund der Beziehung zu den Teilnehmenden weniger geeignet ist. Beispielsweise besteht bei einer offenkundigen Ernsthaftigkeitsprüfung am Ende einer Kundenzufriedenheitsbefragung das Risiko, Kundinnen und Kunden, welche sich die Zeit für die Befragung genommen haben, zu verärgern oder zu irritieren, was bei diesem Anwendungsfeld von den Auftraggebenden weniger gut toleriert würde. Bei Panel-Befragungen und Studien hingegen eignet sich dieser Ansatz besser, solange eine ehrliche Antwort auf diese Frage den Erhalt der Incentivierung für die Umfrage nicht verhindert.

Fischers Fritz fischt rote Heringe

Eine weitere Methode zur Identifizierung von Aufmerksamkeit besteht darin, offenkundig falsche oder absurde Antworten in einen Fragebogen einzubauen. Was aber hat das mit Fischen zu tun? Im englischen Sprachgebrauch

bezeichnet ein Red Herring (roter Hering) ein Element, das in die Irre führt oder von einer relevanten oder wichtigen Frage ablenkt.

In der Umfrageforschung ist mit einem Red Herring eine Ablenkungsfrage als Massnahme zur Qualitätskontrolle gemeint. In eine Reihe von regulären Fragen werden ungewöhnliche Fragen eingefügt, um damit diejenigen Teilnehmenden zu identifizieren, welche die Umfrage vollständig gelesen und sich mit den Inhalten beschäftigt haben resp. jene, die dies nicht getan haben. Die bei Befragungen zu den sinnvollen, regulären Fragen gehörenden Ablenkungsfragen umrahmen dabei oft die dadurch nicht mehr saliente, «getarnte» Validitätsfrage. Als Hypothese gilt: Wer den roten Hering nicht sieht, ist nicht aufmerksam bei der Sache. Im Folgenden wird auf mehrere solche «Heringsarten» eingegangen, bei welchen Teilnehmende den Datenqualitätsfischern «ins Netz gehen» können.

Fiktive Antworten

In der NZZ-Leserbefragung 2005/2006 ergab sich ein kurioses Ergebnis: 189 von 1'883 Teilnehmenden (10 %) gaben an, dass ihnen das NZZ-Folio-Magazin mit dem Thema «Katastrophen» von allen Folios am besten gefallen habe. Der Clou? Dieses Heft gab es nie. Im Fall der NZZ scheint es plausibel, dass einige der 189 Probanden das Folio mit dem Thema «Katastrophen» wählten, weil sie sich nicht an alle NZZ-Folios erinnern konnten, dies aber nicht zugeben wollten und daher ein Heft mit einem spannend klingenden Titel wählten (Porst, 2014). Dabei ist anzumerken, dass zusätzliche Antwortvorgaben wie «*Ich kenne die NZZ-Folio-Magazine nicht*», «*keine Angabe/weiss nicht*» oder «*Ich kann mich nicht entscheiden*» hilfreiche Ausweichkategorien gewesen wären, die im NZZ-Szenario bei dieser Frage zur Steigerung der Datenqualität beigetragen, jedoch umgekehrt weniger Rückschlüsse auf verminderte Aufmerksamkeit zugelassen hätten.

Zu diesem sogenannten antizipierenden Antwortverhalten im Falle von Unwissen sind historisch zahlreiche Beispiele bekannt. Bishop

et al. (1986) beschreiben diesen «pressure to answer» ausführlich. So gaben 70 % der Befragten in einem Fragebogen eine klare Meinung zum «Metallic Metal Act» ab – einem völlig fiktiven Gesetz (Gill, 1947). Gleiches geschah mit 30.8 % zum «Agricultural Trade Act of 1978» (Schuman und Presser, 1981) und mit 26.4 % zum «Monetary Control Bill» (Schuman und Presser, 1981).

Auch Werner Wilken, ein aktuell nicht existierender Politiker, wird regelmässig in Umfragen gekannt, aber «man stimmt mit seiner Politik nicht ganz überein» (Porst, 2014). Dies gilt auch für weitere fiktive Politikerinnen und Politiker (EMNID, 1981; Reuband, 2000). Und je höher die formale Schulbildung der Befragten, desto bekannter sind die Fiktiven.

Aber warum ist das so? Die Kognitionspsychologie liefert eine Antwort: Befragungen wird Sinn unterstellt («die Ersteller des Fragebogens werden wohl seriös gearbeitet haben»): sinnlose Antworten widersprechen der Erwartungshaltung der Teilnehmenden. Dazu kommt eine Hemmung, zuzugeben, dass man etwas nicht kennt, das offenbar allgemein bekannt sein muss, da in einem Fragebogen ja kaum Spezialwissen abgefragt wird. Diese Annahmen treffen gehäuft Befragte, von denen viel formales Wissen erwartet wird. Stellt sich dann noch die Frage, ob man mit seiner Politik einverstanden ist, müssen die Flunkerer erneut lügen. Da sie ihn nicht kennen, liegt es nahe, dass die Politik nicht den eigenen Präferenzen entspricht. Fragen mit fiktiven Antworten sind also nur eingeschränkt geeignet, um nicht vorhandene Aufmerksamkeit zu erkennen. Sie widerspiegeln häufig nur die menschliche Eigenschaft, sich nicht blamieren zu wollen und deshalb möglichst passende oder konsistente Kommunikation zu selektieren.

Bogus-Items

Sogenannte «Bogus-Items» bergen das Risiko von heuristischen, spekulativen oder sozial erwünschten Ersatzantworten wie im vorherigen Abschnitt beschrieben deutlich weniger, weil die «richtige» Antwort sehr offensichtlich ist. Es handelt sich dabei zum Beispiel um Aussagen

wie «*Wasser ist nass*», bei denen auf einer Zustimmungsskala eigentlich nur «*stimme voll und ganz zu*» adäquat ist.

Wer diese Frage mit «*stimme ganz und gar nicht zu*» beantwortet, hat höchstwahrscheinlich den Fragetext nicht aufmerksam genug gelesen (Gummer et al., 2021). Eine ähnliche Bogus-Frage ist «*Ich wurde am 30. Februar geboren*» (Beach, 1989) oder «*I am currently filling out a questionnaire*» (Hargittai, 2009; Meade & Craig, 2012). Bogus-Items werden oft in längere Abfolgen von Likert-Skalen, meist auf derselben Seite («Tabellenfragen») im Sandwich eingebettet.

Es wird also Aufmerksamkeit gemessen – und lediglich bei nicht offenkundig oder extrem genug formulierten Items auch die Tendenz, Nichtwissen mit Vermutungen zu kaschieren. Eine falsche Antwort lässt bei geeigneten Bogus-Items tatsächlich kaum Zweifel offen, dass Teilnehmende unaufmerksam oder absichtlich falsch geantwortet haben: Die Wahrscheinlichkeit, dass jemand falsch positiv als unaufmerksame Person klassifiziert wird, ist also geringer. Aber Vorsicht: In Item-Batterien mit einem untergemischten Bogus-Item, in welchen a) alle Items die gleiche Skalenrichtung aufweisen und b) das Bogus-Item auch die richtige Antwort in dieser Richtung anbietet, können falsch negative Antworten entstehen – Nichtaufmerksamkeit bleibt dann unentdeckt, weil bei allen Items und zum Beispiel auch beim Bogus-Item «*I am currently filling out a questionnaire*» die «*stimme voll und ganz zu*»-Antwort – ohne zu lesen oder nachzudenken – ausgewählt wurde.

Teilnehmende erwarten in den meisten Umfragekontexten keine «Trick»-Items und laufen Gefahr, mit einer Zustimmungstendenz zu antworten, sobald das Item nur ausreichend schwammig formuliert ist (Meade & Craig, 2012). So könnten einige Teilnehmende dem Item «*Meine Freunde vergleichen mich mit einem Pudel*» tatsächlich willentlich zustimmen, weil Pudel soziale, verträgliche Tiere sind und Teilnehmende annehmen könnten, genau dies solle latent mit diesem Item «gemessen» werden. Auch hier kommt das kognitionspsychologische Thema «Sinn und Sense-making» (Weick, 1995) zum Vorschein.

Das Ziel von fiktiven Fragen und Bogus-Items ist es also, unaufmerksame Teilnehmende mit einer

Antwort zu erwischen, die nicht möglich ist. Bei Bogus-Items (hier sind auch einfache Rechenaufgaben wie «*2+3=?*» beliebt) kann mit hoher Wahrscheinlichkeit davon ausgegangen werden, dass Teilnehmende, die diese nicht korrekt beantworten, zu wenig aufmerksam sind. Bei fiktiven Fragen/Items dagegen (wie jener zu den NZZ-Folios) ist es möglich, dass auch aufmerksame Personen den Check nicht bestehen. Dies kann jedoch aus Sicht der Datenqualität auch wünschenswert sein: Ist das Ziel der Befragung die Identifikation der besten Magazin-Ausgabe dieses Jahres, sollten optimalerweise nur Personen in der Stichprobe sein, die sich genügend mit dem Magazin auseinandersetzen, um dessen Ausgaben in einer Liste wiederzuerkennen. Es geht dann aber nicht nur um den Ausschluss von Unaufmerksamkeit, sondern auch von weiteren Störvariablen wie Inkompetenz, sozialer Erwünschtheit oder inhaltsunabhängiger Zustimmungstendenz/Akquieszenz (Bauer, 2000).

Der Wirkungsgrad des Einsatzes von Bogus-Items zur Erkennung von Nichtaufmerksamkeit wurde breiter diskutiert (Breitsohl and Steidelmüller, 2018; Curran, 2016; Goldsmith, 1989), es wurde aber nur eine spärliche qualitätsverbessernde Wirkung attestiert und es wurden andere Methoden, zum Beispiel das bloße Einbauen von «Weiss nicht»-Antworten, die einen veritablen Teil der unsicheren oder antwortunwilligen Personen abfangen.

Anweisungsprüfungen (IMC, IRI)

Ein weiterer Ansatz zur Überprüfung der Aufmerksamkeit von Teilnehmenden ist die Integration von Anweisungsprüfungen. Hier erhalten die Teilnehmenden eine klare Anweisung zur Beantwortung. Wer diese Anweisung nicht befolgt, weist einen ungenügenden Grad an Aufmerksamkeit auf.

Anweisungsprüfungen können in Form einer ganzen Frage («Instructional Manipulation Checks» IMC) umgesetzt werden. Diese werden auch «screener» genannt (Berinsky et al., 2014). Ein Beispiel ist das Hinzufügen eines zusätzlichen Satzes am Ende einer Frage – eine Bemerkung, die den Befragten anweist, die Frage zu ignorieren und eine bestimmte Antwort zu geben, zum Beispiel «*Bitte ignorieren*

Sie diese Frage und wählen Sie unten die vierte Antwort an.» (Alvarez & Li, 2021). Weitere Beispiele zeigen Abbildungen 2 und 3.

Workplace Facilities

Most modern theories of psychology recognize the fact that social perceptions do not take place in a social vacuum. Individual preferences and knowledge, along with situational variables can greatly impact the perception process. In order to facilitate our research on perceptions of workplace behaviors, we are interested in knowing certain factors about you, the perceiver. Specifically, we are interested in whether you actually take the time to read the directions; if not, then some of our manipulations that rely on changes in the instructions will be ineffective. So, in order to demonstrate that you have read the instructions, please ignore the facility items below. Instead, simply click the other option and in the corresponding box, enter the text: I read the instructions.

Which of these facilities are available at your workplace?
(Click on all that apply)

<input type="checkbox"/> Canteen/vending machine	<input type="checkbox"/> Washroom
<input type="checkbox"/> Lounge	<input type="checkbox"/> Windows
<input type="checkbox"/> Coffee maker	<input type="checkbox"/> Parking
<input type="checkbox"/> Air conditioning/heating	<input type="checkbox"/> Childcare Facilities
<input type="checkbox"/> Storeroom	<input type="checkbox"/> Other <input type="text"/>

Abbildung 2: Instructional Manipulation Check mit hohem Schwierigkeitsgrad (Kung, 2018)

FRAGEN

Frage 1

Bei dieser Frage geht es ausschliesslich darum, ob Sie aufmerksam sind. So helfen Sie uns, Datensätze, die unaufmerksam ausgefüllt wurden, zu identifizieren. Bitte wählen Sie unten keine Antworten aus und klicken Sie lediglich auf den Button "Speichern - nächste Frage". Klicken Sie also nichts an, ausser den Button. Vielen herzlichen Dank.

	nein	ja
Ich bin sehr konzentriert bei der Erledigung von Aufgaben.	<input type="radio"/>	<input type="radio"/>
Meine Fähigkeit, mich zu konzentrieren, variiert je nach Situation.	<input type="radio"/>	<input type="radio"/>
Wenn ich an meinem Computer arbeite, lasse ich mich oft durch E-Mails usw. ablenken.	<input type="radio"/>	<input type="radio"/>

Speichern - nächste Frage

Abbildung 3: Instructional Manipulation Checks mit mittlerem Schwierigkeitsgrad (eigenes Beispiel)

Als zweite Form kann ein einzelnes Item innerhalb einer ansonsten inhaltlich relevanten Tabellenfrage implementiert werden («Instructed Response Items» IRI), beispielsweise als fünftes Item von acht mit dem Wortlaut «Wählen Sie in dieser Zeile 'sehr zufrieden' an». (Gummer et al., 2021; DeSimone et al., 2015). Ein weiteres Beispiel zeigt Abbildung 4.

Frage 2

Bitte beantworten Sie nun einige Frage zu Ihrer täglichen Arbeit.

	Stimme sehr zu	Stimme etwas zu	Stimme kaum zu	Stimme gar nicht zu
Ich bin sehr konzentriert bei der Erledigung von Aufgaben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Meine Fähigkeit, mich zu konzentrieren, variiert je nach Situation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn ich an meinem Computer arbeite, lasse ich mich oft durch E-Mails usw. ablenken.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin vorläufig ganz zufrieden mit meiner Stelle.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Meine Arbeit gibt mir die Möglichkeit, Verantwortung zu übernehmen und Entscheidungen zu fällen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bitte wählen Sie hier 'stimme gar nicht zu', um Ihre Aufmerksamkeit zu bestätigen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann hier meine Ideen verwirklichen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin stolz, für diese Firma zu arbeiten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Speichern - nächste Frage

Abbildung 4: Instructed Response Item (IRI) in Zeile 6 (angelehnt an Podsakoff et al., 1990)

IMC und IRI finden in der Umfrageforschung unter anderem aufgrund der relativ einfachen Umsetzung grossen Anklang.

Die Exklusion von Teilnehmenden, welche IMCs nicht bestehen, führt dann zu höherer Konsistenz in der Datenanalyse (Oppenheimer et al., 2009). Das Exkludieren von IRI-Fällen aus dem Datenmaterial wird allerdings kontrovers diskutiert und hat nicht in allen Fällen eine höhere Datenqualität zur Folge (Gummer et al., 2021; Grezki et al., 2015; Anduiza & Galais, 2016). IRIs sind aber wie herkömmliche Methoden definitiv genauso oder teilweise sogar etwas zuverlässiger in der Lage, nichtseriöse Teilnehmende zu identifizieren (Jones et al., 2015; Gummer et al., 2021), beispielsweise mittels Messung der Bearbeitungsdauer (Speeder) oder der Identifikation von Null-Varianz-Antwortverhalten, sogenanntem Straightlining (Meier & Gwerder, 2022). Moderne Systeme für Onlineumfragen können bei falsch beantworteten IMC und IRI die Frage auch nochmals stellen mit dem Hinweis, dass eine unplausible Antwort gegeben wurde, und der Bitte, die Fragen und Antworten genau zu lesen. Mit dieser alternativen Strategie sollen die fehlbaren Teilnehmenden nicht mehr ausgeschlossen werden. Sie sollen ihre Antworten korrigieren. Und genau dieses erneute Stellen einer IMC, so lange, bis die Teilnehmenden diese bestehen, führt zu erhöhter Aufmerksamkeit in den Folgefragen. Daher sollte diese spezifische Variante von IMC und IRI dann auch in einer der ersten Fragen des Fragebogens umgesetzt oder alternativ vor besonders wichtigen Fragen platziert sein, deren Beantwortung besonders gewissenhaft erfolgen soll. Diese Erkenntnis wurde in weiteren Studien besonders auch für komplexe Folgefragen bestätigt (Miller & Baker-Prewitt, 2009; Hauser & Schwarz, 2015), was den Wert von Anweisungsprüfungen als «moral changer» unterstreicht.

Wie bei den Ernsthaftigkeitsprüfungen liegt die Herausforderung der Anweisungsprüfungen bei der Akzeptanz der Befragten. Eine oder in langen Fragebögen zwei IRIs oder IMCs werden von den meisten Teilnehmenden akzeptiert, eine zu hohe Zahl kann Befragte jedoch verärgern und deren Motivation negativ beeinflussen.

Pseudo-Fragen

Eine Alternative zu Anweisungsprüfungen stellen sogenannte «Mock Vignettes» dar, eine kurze Aufmerksamkeitsprüfung («Mock Vignette Check», MVC), die vor der eigentlichen Befragung eingefügt wird. Die Teilnehmenden werden dabei aufgefordert, einen kurzen informativen Text («Vignette») zu lesen und anschliessend einige Fragen dazu zu beantworten. Mit diesen Fragen wird geprüft, ob die Vignette aufmerksam gelesen und korrekt verstanden wurde. Teilnehmende, welche einen eingangs gestellten MVC bestehen, weisen im weiteren Fragebogenverlauf (gemessen an typischen Aufmerksamkeitsmetriken) höhere Aufmerksamkeit aus (Kane et al., 2023). Ausserdem zeigte sich, dass die zu Beginn des Fragebogens gestellten Pseudo-Fragen keinen negativen Effekt auf das Antwortverhalten im restlichen Fragebogen aufweisen, diesen also abgesehen von den Auswirkungen höherer Aufmerksamkeit nicht beeinflussten. Eine solche «Mock Vignette» zu Beginn des Fragebogens könnte mit einer Formatübung kombiniert werden, um die Teilnehmenden auf die Fragetypen vorzubereiten. Zum Beispiel: Diese erste Frage ist inhaltlich nicht relevant. Wir möchten Sie mit dem Ausfüllen vertraut machen und bitten Sie, den folgenden Text trotzdem genau zu lesen und die vier Fragen dazu auf der folgenden Seite korrekt zu beantworten.

Unserer praktischen Erfahrung nach sind solche experimentell anmutenden Checks für Kunden- und Mitarbeitendenbefragungen leider nur eingeschränkt geeignet und werden von Auftraggebenden selten akzeptiert, da Teilnehmende das Gefühl haben könnten, unter einen Generalverdacht mangelnder Aufmerksamkeit gestellt worden zu sein. Gummer et al. (2021) stellen in ihrer Studie denn auch fest, dass ein nicht unerheblicher Teil der Befragten Aufmerksamkeitskontrollen als lästig (16.8 %), verwirrend (10.1 %), manipulierend (10.4 %) oder als Kontrolle empfanden (25.0 %) resp. nicht belehrt werden wollen (24.3 %). Immerhin befand ein Drittel der Stichprobe (31.4 %) Aufmerksamkeitskontrollen als motivierend.

Widersprüchliche Antwort-Paare/-Sets

Widersprüchliche Antworten über zwei oder mehr inhaltlich gleiche Fragen hinweg sind eine weitere Methode, um Aufmerksamkeit zu erkennen. Und sie können in der Regel bei der Betrachtung einzelner Fragebogen relativ einfach identifiziert werden. Geben Teilnehmende bei zwei Fragen, die dasselbe Konstrukt erfassen, komplett unterschiedliche Antworten, oder ist eine bestimmte Antwort-Kombination schlichtweg unmöglich (beispielsweise ein 21-jähriger Student mit 15 Jahren Berufserfahrung), kann dies ein Hinweis darauf sein, dass Befragte dem Fragebogen nicht genügend Aufmerksamkeit schenken. Wenn eine Befragung zahlreiche Tabellenfragen enthält, kann bei der ersten und bei der letzten Tabellenfrage je einmal dasselbe Item eingebaut werden (z. B. «*Meine Arbeitsstelle gefällt mir sehr.*» und «*Ich bin mit meiner Arbeitsstelle sehr zufrieden.*»). Es kann dann überprüft werden, ob die Antworten diametral abweichen. Solche Fälle können dann für die Analysen exkludiert werden.

Aber Achtung: Bei heiklen Themen sind auch die aufmerksamen Teilnehmenden nicht a priori willig, wahre Antworten zu geben. In einer aktuellen Online-Studie zu riskantem Cannabisgebrauch fanden sich ähnliche Werte – 45.3 % der Teilnehmenden wiesen inkonsistente Antworten auf (Schell et al., 2022).

Wie aber sollte nun mit widersprüchlichen Antworten als Indikator für fehlende Aufmerksamkeit verfahren werden? Geeignete Methodiken für die jeweilige Befragung könnten sein (Bauer & Johnson, 2000):

- Nichts tun: Auch widersprüchliche Antworten werden für die Analyse verwendet. Die Anzahl gibt lediglich einen Hinweis auf die Datenqualität und relativiert die Ergebnisse und deren Interpretation im Sinne eines Konfidenzintervalls.
- Torhüter («Gatekeeper»): Nur die erste Antwort eines Falles wird für die Analyse verwendet. Darauf folgende einzelne Antworten, welche der ersten widersprechen, werden als «missing items» umcodiert. Der

Datensatz an sich fließt mit ein.

- **Limitierter Ansatz:** Es wird eine Liste von besonders wichtigen Fragen definiert. Widerspricht sich ein/e Teilnehmer/in innerhalb dieser wichtigen Fragen, werden alle Antworten der Teilnehmerin/des Teilnehmers bei diesen wichtigen Fragen als «missing items» in die Ergebnisse aufgenommen.
- **Globaler Ansatz:** Sämtliche Fälle mit inkonsistenten Antworten werden aus der Analyse ausgeschlossen.
- **Ansatz der inhaltlich überwiegenden Beweislage.** Der Fragebogen und die besonders wichtigen Fragen aller Befragten, welche widersprüchliche Antworten gaben, werden sorgfältig geprüft und der Fallstatus wird auf der Grundlage der «überwiegenden Beweislage» zugewiesen, die durch die Auswertung der Antworten ermittelt wird (z. B. wenn es darum geht, Fragebögen zu typisieren und einer Kategorie zuzuordnen). Alle Antworten, die dem zugewiesenen Fallstatus widersprechen, werden als «missing items» betrachtet.

Je nach verwendetem Ansatz zum Umgang mit inkonsistenten Antworten ergibt sich eine unterschiedliche Zahl von exkludierten Fällen – 33 bei «*nichts tun*» bis zu 1374 beim «globalen Ansatz» (Bauer & Johnson, 2000) – und es ergeben sich damit signifikant unterschiedliche Analyseergebnisse. Es ist daher von grosser Bedeutung, die gewählte Methode kritisch zu bewerten und im Auswertungsreport genauestens zu deklarieren.

Umgepolte Items

Eine Spezialform widersprüchlicher Angaben sind doppelte Items, von denen eines rotiert ist. Solche «Fallen» (wir sind wieder bei roten Heringen) müssen dann zweimal genau entgegengesetzt beantwortet werden, ansonsten wäre eine Antwort widersprüchlich. Zum Beispiel kann ein Item im ersten Teil des Fragebogens untergebracht werden, ein zweites Item dann gegen Ende, wobei typischerweise dieselbe Likert-Skala verwendet wird. Das zweite Item erfragt denselben Inhalt wie

das erste in leicht veränderter Formulierung, wird aber negiert (= rotiert) formuliert. Beispielsweise kann gefragt werden: «*Das Produkt gefällt mir im Allgemeinen sehr gut.*» (5-er-Skala von «*trifft sehr zu*» bis «*trifft gar nicht zu*»). Im weiteren Fragebogenverlauf könnte mit ausreichend grossem Abstand gefragt werden: «*Ich finde die Produkte generell mangelhaft.*» (5-er-Skala von «*trifft sehr zu*» bis «*trifft gar nicht zu*»). Konsistentes, aufmerksames Ausfüllverhalten müsste hier mindestens einigermassen diametrale Ergebnisse erzeugen. Die beiden Skalenwerte innerhalb desselben Falles dürften nach Umpolung eines der beiden Items (Gleichausrichtung) nur eine geringe Varianz aufweisen. Fälle, bei denen diese Bedingung verletzt wurde, könnten wegen Nichtaufmerksamkeit für die Datenanalyse exkludiert werden. Ein konkretes Beispiel zeigt Abbildung 5.

Frage 3

Bitte beurteilen Sie folgende Aussagen zu Ihrer Arbeit.

	Stimme sehr zu	Stimme etwas zu	Stimme kaum zu	Stimme gar nicht zu
Meine Arbeit macht mir wenig Spaß, aber man sollte nicht zu viel erwarten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann meine Arbeit selbst einteilen und planen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Meine Arbeit läuft immer im gleichen Trott; daran kann man nichts machen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Insgesamt gesehen: Meine Arbeit ist wirklich interessant und befriedigend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Meine Arbeit gibt mir die Möglichkeit, Verantwortung zu übernehmen und Entscheidungen zu fällen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe nichtige Freude an der Arbeit.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Speichern - nächste Frage

Abbildung 5: Erstes und drittes Item rotiert (angelehnt an Podsakoff et al., 1990)

Es ist allerdings zu bedenken, dass reverse Items unterschiedliche Bedingungen für gute und schwächere Leserinnen und Leser schaffen, da negativ formulierte Items eine höhere Sprachkompetenz und höheren kognitiven Aufwand erfordern und damit einen Bias in die Ergebnisse einsteuern können (Suárez-Alvarez et al., 2018), vor allem bei sprachlich schwächeren Zielgruppen (z. B. bei Kindern und Jugendlichen).

Offensichtliche Fragen

Eine weitere Technik zur Aufmerksamkeitsprüfung sind Fragen, die eine einzige plausible Antwort zulassen, nämlich «*trifft zu*». Werden solche Fragen auf einer Likert-Skala mit «*trifft nicht zu*» beantwortet, kann auf mangelnde

Aufmerksamkeit geschlossen werden. Beispiel: «Es gibt Menschen, die eine andere Meinung haben als ich» oder «E ist ein Buchstabe». Allerdings ist eine Abstützung auf einzig diese Art von Aufmerksamkeitscheck kaum empfehlenswert und solche Fragen laufen Gefahr, als deutliche Fremdkörper wahrgenommen zu werden, weil deren Funktion von einer Vielzahl an Befragten nicht verstanden wird und sie dadurch irritierend wirken.

A Priori Fakes mit Bots

In den letzten Jahren wurde ein neues Problem erkannt, welches auf den ersten Blick aussieht wie aufmerksamkeitsloses Ausfüllen: das Einsetzen von Botnetzen (automatisierten Scripts, die Formulare ausfüllen) und dedizierten Umfrage-Bots zur Beantwortung kompletter Fragebögen, auch mit offenen Textantworten, bei welchen qualitätsmässig komplett sinnlose, willkürliche Antworten technisch abgefüllt werden. Dabei erzielen Teilnehmende im besten Fall Vergütungen für komplettierte Fragebögen, ohne dass sie selbst tatsächlich Fragen beantworten (Dupuis, 2018; Fullerton und McCullough, 2022). Gegen diese Techniken gibt es wenige wirkungsvolle Massnahmen. Das Abfragen einer anonymen «uniform identifier», einer «einzigartigen Identifikationskennung der Teilnehmenden». So kann beispielsweise in einer Single-Choice-Frage mit einigen Antwortvorgaben das Domizilland der Teilnehmenden abgefragt werden, wobei Bots dann über alle Datensätze hinweg in der Regel eine breite Varianz erzeugen, echte Teilnehmende aber nur das effektive Land auswählen, in dem sie wohnen und in dem die Studie auch durchgeführt wird oder nach dem sie rekrutiert werden – schlichtweg deshalb, weil sie die Frage verstehen (Fullerton & McCullough, 2022). Panels verwenden auch einen Test-Retest-Mechanismus, bei dem gleiche Fragen mit zeitlich stabilen Antwortwartungen über mehrere Fragebögen hinweggestreut sind. Wenn die Antworten der einzelnen Teilnehmenden auf diese Fragen nicht genau oder akzeptabel nahe übereinstimmen, wird der entsprechende

Befragte aus dem Datensatz entfernt (Fullerton et al., 2009). Auch Google ReCAPTCHA können zu Beginn eines Online-Fragebogens eingesetzt werden; diese für Menschen einfach zu lösenden kleinen Aufgaben verwenden «eine fortschrittliche Risikoanalyse-Engine und adaptive Herausforderungen, um bösartige Software von missbräuchlichen Aktivitäten auf Ihrer Website abzuhalten» (Google, 2024).

Item Response Theory

Ein elaboriertes Verfahren zur Aufmerksamkeitserkennung post hoc, also nach bereits erfolgter Datenerhebung, stellen auch Rasch-Personen-Fit-Indices dar. Sie bieten einen methodisch fortschrittlichen Ansatz zur Erkennung abweichender Antworten und identifizieren atypische Antwortmuster auf Personenebene, die zum Beispiel als Folge von Betrug oder fehlender Aufmerksamkeit auftreten können (Beck et al., 2019; Li & Olejnik, 1997). Dieses Verfahren ist allerdings nur anwendbar in Skalen, die nach der Item Response Theory (IRT) erstellt wurden (van den Wittenboer et al., 1997) und die eine ausreichende Länge und eine volle Range von Itemschwierigkeiten aufweisen. Für die Marktforschungspraxis sind diese Voraussetzungen im Unterschied zu psychologischen Testverfahren oder Kompetenztests oft nicht gegeben.

Guttman Errors

Guttman-Fehler, welche die Grundlage vieler nichtparametrischer Person-Fit-Statistiken bilden, eignen sich auch direkt für eine Post-hoc-Erkennung abweichender Antwortmuster: Diese treten dann auf, wenn ein Befragter ein schwieriges Item zu einem Thema richtig und danach ein leichtes Item zum selben Thema falsch beantwortet. Voraussetzung dazu sind auch hier Skalen, die Konstrukte mit mehreren Items unterschiedlicher Schwierigkeit (item difficulty) messen. Guttman-Fehler können dann gezählt oder in Relation zur Gesamtzahl der Items gesetzt werden. So können abweichende Antwortmuster in Daten mit dichotomen

Antwortskalen genau identifiziert werden (Emons, 2008; Karabatsos, 2003; Meijer, 1994; Meijer, Egberink, Emons, & Sijtsma, 2008). Beck et al. (2019) zeigen auch weitere Person-Fit-Methoden wie U3 personfit statistic, HT coefficient und standardized log likelihood auf.

Fazit

Eine proaktive Behandlung des Themas «**Aufmerksamkeit in Umfragen**» kann mit zahlreichen Methoden umgesetzt werden, womit in der Regel die Datenqualität gesteigert werden kann. Es gibt aber keine eigentlichen Standardmethoden oder klare Empfehlungen. Viel wichtiger ist es, die Methoden den Erfordernissen der einzelnen Befragung und der Zielgruppe anzupassen und – gerade wenn Meinungsforschung für Auftraggebende durchgeführt wird – die Methodenwahl transparent zu begründen und zu dokumentieren. Umfrageverzerrungen werden somit vermindert und die Validität von Studien wird erhöht. Eher seltener führt das Exkludieren von nicht aufmerksamen Teilnehmenden jedoch zu einer thematisch relevanten Verzerrung der Stichprobe – dieses Risiko muss von Fall zu Fall geprüft werden.

Aufmerksamkeitsprüfungen haben auch sekundäre Effekte und können aus verschiedenen Gründen die Beteiligung der Befragten erhöhen:

- Sie können die Umfragen für Teilnehmende interessanter machen.
- Rote Heringe oder Instructional Manipulation Checks unterbrechen die Monotonie einer Umfrage.
- Sie halten die Aufmerksamkeit der Befragten aufrecht.
- Aufmerksamkeitsprüfungen machen zwar den Fragebogen etwas länger, erhöhen jedoch die Wahrscheinlichkeit, dass die Befragten jede Frage vollständig lesen und verstehen.
- Einzelne Teilnehmende ärgern sich über «Fremdkörper» oder «Fallen».
- Partizipanten erkennen die Checks und setzen Vertrauen in die Wichtigkeit und Professionalität der Befragung.

- Die Datenanalyse kann nach Exklusion betroffener Datensätze zuverlässiger und ressourcenschonender sein.
- Ein sehr hoher Anteil an unaufmerksamen Teilnehmenden weist im Pretesting auf einen langen, komplizierten, unverständlichen oder monotonen Fragebogen hin oder darauf, dass die Teilnahmemotivation in der Stichprobe nicht ausreicht – ein Thema, dem allenfalls mit der richtigen Ansprache der Teilnehmenden und mit der Kommunikation des persönlichen immateriellen Profits für die Teilnehmenden begegnet werden könnte. Eine Incentivierung mit Geld kann gefährlich sein und Aufmerksamkeitsprobleme sogar im Sinne einer Jagd auf abgeschlossene Fragebögen verschärfen.

Leiner (2019) empfiehlt generell, als wichtigstes Merkmal von Aufmerksamkeit eine angemessene Ausfülldauer des Fragebogens zu Rate zu ziehen, vor allem, wenn im Fragebogen keine Informationen nachgeschlagen werden müssen oder andere komplexe Fragen gestellt werden, die Unterbrüche generieren.

Möglicherweise können mehrere Methoden und Fragen zur Aufmerksamkeitsprüfung eingesetzt werden, die sich methodologisch ergänzen. Auch die Position von Aufmerksamkeitsfragen innerhalb des Fragebogens sollte bedacht werden. Während die Positionierung am Anfang des Fragebogens auf die Teilnehmenden eher präventive Wirkung auf die weitere Aufmerksamkeit im Fragebogen zeigt und für die Auswertungen ein Signal für eine von Anfang an mangelnde Aufmerksamkeit sein kann (z. B. bei teilweise Teilnehmenden, Incentivierungsjägern, Teilnahmen aus Neugier), motivieren zwischendurch gestellte Aufmerksamkeitschecks die Teilnehmenden und «halten sie wach». Gegen Ende der Befragung gestellte Checks dienen meist primär der Identifikation von Teilnehmenden, deren Interesse oder deren Konzentration im Fragebogenverlauf verloren gegangen ist.

Um Teilnehmende nicht zu verärgern oder als Versuchsobjekte zu deklassieren, sollten solche Items jedoch insgesamt vor allem bei

fragilen Zielgruppen und allenfalls bei Personen, die nur einmal teilnehmen (z. B. Kunden, Messebesucher etc.), zurückhaltend oder sogar nur in Pretests eingesetzt werden, um eine Kenngrösse für eine allfällige Aufmerksamkeitsproblematik zu erhalten und diese für Konfidenzintervalle der Ergebnisse der Hauptbefragung im Hinterkopf zu behalten.

Die Autoren und Kontaktangaben

Raffael Meier

Mitgründer/CTO von onlineumfragen.com und Pionier der deutschsprachigen Onlineumfragetechnologie. Er befasst sich mit gesellschaftlichen, technischen und methodologischen Aspekten von Daten.
Raffael Meier, +41 44 500 5137
raffael.meier@onlineumfragen.com

Kathrin Staub

Mitgründerin von onlineumfragen.com und Principal Consultant. Spezialisiert auf die Beratung namhafter nationaler und internationaler Unternehmen in den Bereichen Methodik und Fragebogenkonstruktion.
Kathrin Staub, +41 44 500 5136
kathrin.staub@onlineumfragen.com

Nina Gwerder

ehemalige Lead Consultant bei onlineumfragen.com mit Schwerpunkten in den Themen Statistik, Datenqualität und effektive Auswertung von Onlineumfragen.
Nina Gwerder, +41 44 500 5140
nina.gwerder@onlineumfragen.com

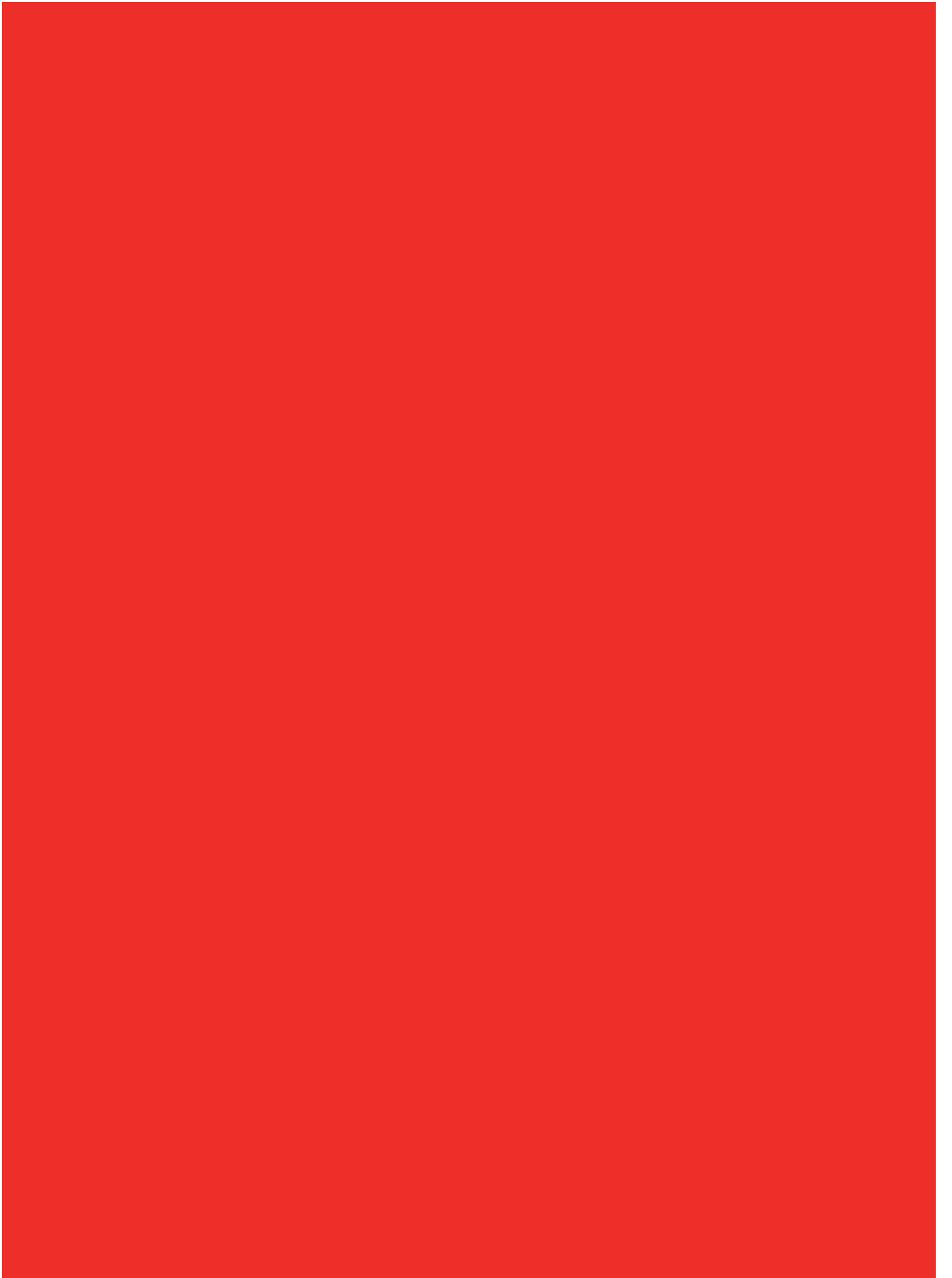
Literatur

- [1] Alvarez, R. M., & Li, Y. (2022). Survey Attention and Self-Reported Political Behavior, *Public Opinion Quarterly*, Volume 86, Issue 4, Winter 2022, Pages 793–811, <https://doi.org/10.1093/poq/nfac048>
- [2] Anduiza, Eva, & Carol Galais (2016). Answering Without Reading: IMCs and Strong Satisficing in Online Surveys. *International Journal of Public Opinion Research*. doi:10.1093/ijpor/edw007.
- [3] Aust, F., Diedenhofen, B., Ullrich, S. & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavioral Research Methods*, 45, S. 527-535.
- [4] Barge, S. & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, 53(2), 182–200. doi:10.1007/s11162-011-9251-2
- [5] Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a biethnic population. *European Journal of Public Health*. 20 (5): 543–548. doi:10.1093/eurpub/ckq052.
- [6] Bauer, E., & Johnson, T. (2000). Editing Data: What Difference Do Consistency Checks Make? *American Journal of Epidemiology*, 151(9), S. 921-926.
- [7] Bauermeister, J. A., Pingel, E., Zimmerman, M., Couper, M., Carballo-Diequez, A., & Strecher, V. J. (2012). Data quality in HIV/AIDS webbased surveys: Handling invalid and suspicious data. *Field Methods*, 24(3), 272– 291. doi:10.1177/1525822X12443097
- [8] Beach, D. A. (1989). Identifying the random responder. *Journal of Psychology: Interdisciplinary and Applied*, 123(1), 101-103.
- [9] Beck, M. F., Albano, A. D., & Smith, W. M. (2019). Person-Fit as an Index of Inattentive Responding: A Comparison of Methods Using Polytomous Survey Data. *Applied psychological measurement*, 43(5), 374–387. <https://doi.org/10.1177/0146621618798666>
- [10] Berinsky, A.J., Margolis, M.F., & Sances, M.W. (2014), Separating the Shirkers from the Workers? Making Sure Respondents Pay

- Attention on Self-Administered Surveys. *American Journal of Political Science*, 58: 739-753. <https://doi.org/10.1111/ajps.12081>
- [11] Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*, 4(3), 340-345. doi:10.1037/1040-3590.4.3.340
- [12] Bishop, G. F., Oldendick, R. W., Tuchfarber, A. J., & Bennett, S. E. (1980). Pseudopinions on public affairs. *Public Opinion Quarterly*, 44(2), 198–209.
- [13] Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1986). Opinions on Fictitious Issues: The Pressure to Answer Survey Questions. *The Public Opinion Quarterly*, 50(2), 240–250. <http://www.jstor.org/stable/2748887>
- [14] Breitsohl, H. & Steidelmüller, C. (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. doi:10.1111/apps.12121
- [15] Buchanan, T., Heffernan, T. M., Parrott, A. C., Ling, J., Rodgers, J., & Scholey, A. B. (2010). A short self-report measure of problems with executive function suitable for administration via the internet. *Behavior Research Methods*, 42, S. 709-714.
- [16] Burns, G. N., & Christiansen, N. D. (2011). Methods of Measuring Faking Behavior, *Human Performance*, 24:4, 358-372, DOI: 10.1080/08959285.2011.597473
- [17] Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4–19.
- [18] Delgado, K, Tristan, E., Kung, M (Mei-Chuan) & O'Connell, M. (2011). Mean score and validity differences among bogus item endorsement groups.
- [19] DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, 36(2), 171–181. doi:10.1002/job.1962
- [20] Dupuis, M., Meier, E., & Cuneo, F. (2018). Detecting computergenerated random responding in questionnairebased data: A comparison of seven indices. *Behavior Research Methods*, 51(5), 2228–2237. <https://doi.org/10.3758/s13428-018-1103-y>
- [21] EMNID (1981): Die Popularität von Bundesministern – eine methodenkritische Befragung. In:EMNID-Informationen 8: 14-16
- [22] Emons W. H. M. (2008). Nonparametric personfit analysis of polytomous item scores. *Applied Psychological Measurement*, 32, 224-247
- [23] Farrell, A., Danish, S., & Howard, C. (1991) Evaluation of data screening methods in surveys of adolescents' drug use. *Psychol Assess*, 1991(3), S. 295-298.
- [24] Franzén, M. (2011). Nonattitudes / pseudopinions: Definitional problems, critical variables, cognitive components and solutions. (C/D Extended Essay No. 14). Retrieved from <http://www.diva-portal.org/smash/get/diva2:1032161/FULLTEXT01.pdf>
- [25] Fullerton, S., & McCullough, T. (2023). Using quality control checks to overcome pitfalls in the collection of primary data via online platforms. *Journal of Marketing Analytics*, 11(4), 602–612. doi:10.1057/s41270-023-00249-z
- [26] Fullerton, S., D. Taylor, and J. Watson. 2009. Webbased instruction and online delivery of classes: Where are we now? *American Journal of Business Education* 2 (1): 91–100.
- [27] Gill, S. (1947). How do you stand on sin? *Tide*, 74(March), 72
- [28] Goffin, R. D., & Christinasen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment*, 11: 340–344.
- [29] Goldsmith, R. E. (1989). Reducing spurious response in a field survey. *The Journal of Social Psychology*, 129(2), 201–212. doi:10.1080/00224545.1989.9711721
- [30] Google (2023). reCAPTCHA protects your website from fraud and abuse without friction <https://www.google.com/recaptcha/about/>. Accessed 2 June 2024.
- [31] Greszki, R., Meyer, M., & Schoen, H. (2015).

- Exploring the Effects of Removing “Too Fast” Responses and Respondents from Web Surveys. *Public Opinion Quarterly*, 79(2), 471–503. <https://doi.org/10.1093/poq/nfu058>
- [32] Gummer, T., Rossmann, J., & Silber, H. (2021). Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation. *Social Methods & Research*, 50(1), S. 238-264.
- [33] Haghghat, R. (2007). The Development of the Brief Social Desirability Scale (BSDS). *Europe’s Journal of Psychology*, 3(4). <https://doi.org/10.5964/ejop.v3i4.417>
- [34] Hargittai, E. (2009). An update on survey measures of weboriented digital literacy. *Social Science Computer Review*, 27(1), 130–137. doi:10.1177/0894439308318213
- [35] Hauser, D. & Schwarz, N. (2015). It’s a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on ‘Tricky’ Tasks. *Sage Open*, 5(2), S. 1-6.
- [36] Ihme, J. M., Lemke, F., Lieder, K., Martin, F., Müller, J. C., & Schmidt, S. (2009). Comparison of ability tests administered online and in the laboratory. *Behavior Research Methods*, 41, S. 1183-1189.
- [37] Jones, M., House, L., & Gao, Z. (2015). Respondent Screening and Revealed Preference Axioms: Testing Quarantining Methods for Enhanced Data Quality in Web Panel Surveys. *Public Opinion Quarterly*, 79(3), S. 687-709.
- [38] Johnson, J. A. (2005). Ascertaining the validity of individual protocols from webbased personality inventories. *Journal of Research in Personality*, 39(1), 103–129. doi:10.1016/j.jrp.2004.09.009
- [39] Kane, J., Velez, Y., Barbas, J. (2023). Analyze the attentive and bypass bias: Mock vignette checks in survey experiments. *Political Science Research and Methods*, 11(2), S.293-310.
- [40] Karabatsos G. (2003). Comparing the aberrant response detection performance of thirty-six personfit statistics. *Applied Measurement in Education*, 16, 277-298
- [41] Kemper, C. J., Beierlein, C., Bensch, D., Kovaleva, A., & Rammstedt, B. (2012). Eine Kurzskala zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens: Die Kurzskala Soziale Erwünschtheit-Gamma (KSE-G) (GESIS Working Papers 2012|25). Köln: GESIS.
- [42] Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236. doi:10.1002/acp.2350050305
- [43] Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537–567. doi:10.1146/annurev.psych.50.1.537
- [44] Krosnick, J. A., Narayan, S. S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman, & J. K. Slater (Eds.), *Advances in Survey Research* (pp. 29-44). San Francisco: Jossey-Bass.
- [45] Kung, F.Y.H., Kwok, N. and Brown, D.J. (2018), Are Attention Check Questions a Threat to Scale Validity? *Applied Psychology*, 67: 264-283. <https://doi.org/10.1111/apps.12108>
- [46] Li, M. F., & Olejnik, S. (1997). The power of rasch personfit statistics in detecting unusual response patterns. *Applied Psychological Measurement*, 21, 215–231.
- [47] Meade, A. W. & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. doi:10.1037/a0028085
- [48] Meier, R., & Gwerder, N. (2023). Dirty on the go? Data quality in online surveys between PC and mobile devices. In *Swiss Insights (Hrsg.), Swiss Insights Annual Report 2023*.
- [49] Meier, R., & Gwerder, N. (2022). Dirty Data in Online Surveys. How to improve data quality before and after the field phase. In *Swiss Insights (Hrsg.), Swiss Insights Annual Report 2022* (S. 51-62).
- [50] Meijer R. R. (1994). The number of Guttman errors as a simple and powerful personfit statistic. *Applied Psychological Measurement*, 18, 311-314
- [51] Meijer R. R., Egberink I. J. L., Emons W. H. M., Sijtsma K. (2008). Detection and validation of unscalable item score patterns using item response theory: An illustration with Harter’s Self-Perception Profile for children. *Journal of Personality Assessment*, 90, 227-238

- [52] Miller, J. & Baker-Prewitt, J. (2009). Beyond 'Trapping' the Undesirable Panelist: The Use of Red Herrings to Reduce Satisficing. Präsentiert an der CASRO Panel Quality Conference, 2. Februar 2009. New Orleans, USA.
- [53] Nichols, D. S., Greene, R. L., & Schmolck, P. (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. doi:10.1002/1097-4679(198903)45:2<239::AID-JCLP2270450210>3.0.CO;2-1
- [54] Nießen, D., Partsch, M. V., Kemper, C. J., & Rammstedt, B. (2019). Social Desirability-Gamma Short Scale (KSE-G). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis271_exz
- [55] Oppenheimer, D., Meyvis, T., & Davidenko, N. (2009). Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power. *Journal of Experimental Social Psychology*, 45, S. 867-72.
- [56] Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598-609.
- [57] Podsakoff, P.M., MacKenzie, S.B., Moorman, R.H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *Leadership Quarterly*, 1(2), 107-142. [http://doi.org/10.1016/1048-9843\(90\)90009-7](http://doi.org/10.1016/1048-9843(90)90009-7)
- [58] Porst, R. (2014). Fragebogen - Ein Arbeitsbuch. 4. Auflage, Wiesbaden: Springer Fachmedien.
- [59] Reips, U.-D. (2002). Standards for internetbased experimenting. *Experimental Psychology*, 49, S. 243-256.
- [60] Reips, U.-D. (2008). How Internetmediated research changes science (S. 268-294). Cambridge, MA: Cambridge University Press.
- [61] Reips, U.-D. (2009). Internet experiments: Methods, guidelines, metadata. *Human Vision and Electronic Imaging XIV, Proceedings of SPIE*, 7240, 724008.
- [62] Reuband, Karl-Heinz (2000): „Pseudo-Opinions“ in Bevölkerungsumfragen. Wie die Bürger fiktive Politiker beurteilen. In: ZA-Informationen 46: 26-38
- [63] Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, 10(2), 160-177. doi:10.1177/1073191103010002007
- [64] Schell, C., Godinho, A., & Cunningham, J. (2022). Using a consistency check during data collection to identify invalid responding in an online cannabis screening survey. *BMC Medical Research methodology*, 22(67).
- [65] Schuman, H. & Presser, S. (1980). Public opinion and public ignorance: The fine line between attitudes and nonattitudes. *American Journal of Sociology*, 85(5), 1214- 1225.
- [66] Schuman, H., & Presser, S. (1981). *Questions and Answers in Attitude Surveys*. New York: Academic Press.
- [67] Stevens, H. (2017, November 19). Catching a red herring. *Chicago Tribune*. <https://www.chicagotribune.com/lifestyles/ct-tribu-words-work-herring-20110202-story.html>
- [68] Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñoz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158.
- [69] Ross, R.S. (2008). Popularization of 'red herring' by English political agitator William Cobbett. In: *Comments on Etymology* 38:1-2, 2008, S. 62-69.
- [70] Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A reply to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt. *Personnel Psychology*, 60: 267-293.
- [71] Tourangeau, R., Rips, L. J., & Rasinski, K. (Eds.). (2000). *The psychology of survey response*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511819322>
- [72] Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks [u.a.] : Sage.
- [73] van den Wittenboer, G., Hox, J., & de Leeuw, E. (1997). Aberrant Response Patterns in Elderly Respondents: Latent Class Analysis of Respondent Scalability (pp. 155-162). Münster, Germany: Waxman.



Swiss Insights

Swiss Insights ist der Verband und die Interessensvertretung aller Unternehmen, die Daten und prädiktive Modelle im Rahmen von Marketing, Innovationsprozessen, Kundenservice, Angebotsgestaltung, Kommunikation und Zielgruppendefinitionen erheben, analysieren, einsetzen und daraus Handlungsempfehlungen ableiten.

Die Corporate Mitglieder von Swiss Insights sind Unternehmen, die sich für den fairen Umgang mit Auskunftspersonen sowie Auftraggeber und den Schutz der Privatsphäre engagieren. Corporate Mitglieder sind berechtigt, das Logo "SWISS INSIGHTS - Corporate Member" zu tragen. Corporate Member haben zudem die Möglichkeit, das Label "Market and Social Research by Swiss Insights" und/oder das "Label Data Fairness by Swiss Insights" zu beantragen. Unternehmen, die eines der beiden Label tragen, garantieren, dass sie sich an die strengen Richtlinien und Reglemente von Swiss Insights und Esomar halten.

Das Label "Market and Social Research by Swiss Insights" kann von Unternehmen mit Sitz/Niederlassung in der Schweiz beantragt werden. Sie unterliegen einem strengen Regelwerk von schweizerischen und internationalen Normen und garantieren, dass keine Interviews mit Werbe- oder Verkaufsabsichten durchgeführt, wissenschaftlich abgesicherte Methoden angewendet und die Datenschutzrichtlinien eingehalten werden.

Das Label "Data Fairness by Swiss Insights" kann sowohl von Schweizer wie auch von ausländischen Unternehmen beantragt werden. Mit dem Label zeigen sie, dass sich ihr Unternehmen oder ihre Abteilung für den transparenten und verantwortungsvollen Umgang bei der Bearbeitung und Analyse von grossen Datenmengen verpflichtet.



Herausgeber und Kontakt

Swiss Insights, Swiss Data Insights Association, Gruebengasse 10, 6055 Alpnach, Switzerland
+41 44 3501960, info@swiss-insights.ch, www.swiss-insights.ch

SWISS INSIGHTS

Corporate Member

amPuls Market Research

Hirschengraben 49, 6000 Luzern 7
+41 41 612 14 14 / info@ampuls.ch
www.ampuls.ch



gfs.bern. Menschen. Meinungen. Märkte.

Effingerstrasse 14, Postfach, 3001 Bern
+41 31 311 08 06 / info@gfsbern.ch
www.gfsbern.ch



amrein+heller MarktforschungTreuhand AG

Südweid 7, 6274 Eschenbach
+41 748 63 70 / contact@ah-feedback.ch
www.ah-feedback.ch



gfs-zürich, Markt- & Sozialforschung

Riedtlistrasse 9, 8006 Zürich
+41 44 360 40 20 / gfs@gfs-zh.ch
www.gfs-zh.ch



Bilendi Schweiz AG

Reinhardstrasse 19, 8008 Zürich
+41 79 801 88 80 / contact.ch@bilendi.com
www.bilendi.ch



gfs-befragungsdienst

Schaffhauserstrasse 491, 8052 Zürich
+41 44 360 26 40 / info@gfs-bd.ch
www.gfs-bd.ch



Boomerang Ideas AG

Sihlquai 131, 8005 Zürich
+41 44 500 88 60 / raphael@boomerangideas.com
www.boomerangideas.com



GIM Suisse AG

General-Wille-Strasse 10, 8002 Zürich
+41 44 283 18 18 / info@g-i-m.ch
www.g-i-m.ch



BSI Business Systems Integration AG

Täferweg 1, 5405 Baden
+41 58 255 90 00, info@bsi-software.com
www.bsi-software.com



Human8

Staufacherstrasse 101, 8048 Zürich
+41 44 204 16 26 / christian@wearehuman8.com
www.wearehuman8.com



Constant Dialog

Alte Steinhäuserstrasse 33, 6330 Cham
+41 41 310 05 40 / info@constant-dialog.ch
www.constant-dialog.ch



INNOFACT (Schweiz) AG Research & Consulting

Flurstrasse 50, 8048 Zürich
+41 43 931 77 82, Info@innofact.ch
www.innofact.ch



DemoSCOPE Data + Research

Klusenstrasse 18, 6043 Adligenswil
+41 41 375 40 00 / demoscope@demoscope.ch
www.demoscope.ch



Insight Institute AG

Bergstrasse 138, 8032 Zürich
+41 44 387 90 90 / info@insightinstitute.ch
www.insightinstitute.ch



Die Schweizerische Post

Wankdorffallee 4, 3030 Bern
www.post.ch



intervista

Optingenstrasse 5, 3013 Bern
+41 31 511 39 00 / anfragen@intervista.ch
www.intervista.ch



Gallup AG

Reinhardstrasse 19, 8008 Zürich
+41 78 891 31 15 / office@gallup.swiss
www.gallup.swiss



IPSOS Suisse SA

11, Chemin du Château-Bloch, 1219 Le Lignon
+41 22 591 06 00 / Contact_Switzerland@ipsos.com
www.ipsos.com/de-ch



gff Swiss Research Services

Baarerstrasse 25, 6300 Zug
+41 41 560 01 60 / gut@gff.ag
www.gff.ag



just-medical!

Blegistrasse 5, 6340 Baar
+41 41 766 11 55 / info@just-medical.com
www.pharmaagentur.ch



SWISS INSIGHTS

Corporate Member

Kantar Media Switzerland AG

Bahnhofstrasse 4, 3073 Gümligen
+41 31 537 79 00 / ch.panel@kantarmedia.com
www.kantarmedia.com



super computing Systems AG

Technopark, Technoparkstr. 1, 8005 Zürich
info@scs.ch
www.scs.ch

Marketagent.com Schweiz AG

Feldstrasse 40, 8004 Zürich
+41 43 555 06 50 / schweiz@marketagent.com
www.marketagent.com



TALK Online Panel

Lindenmoosstrasse 4, 8910 Zürich
+41 43 550 14 62 / rfq@talkonlinepanel.com
www.talk-group.com



M.I.S. Trend SA

Pont Bessières 3, 1005 Lausanne
+41 21 320 95 03 / info@mistrend.ch
www.mistrend.ch



TransferPlus AG Market Research

Haldenstrasse 11, 6006 Luzern
+41 41 618 33 11 / transfer@transferplus.ch
www.transferplus.ch



NielsenIQ / GfK (Switzerland) GmbH

Park 6/8, 6039 Root D4
nielsen-ch@nielsen.com / info.ch@gfk.com
www.NielsenIQ.com / www.gfk.com



YouGov

Baslerstrasse 60, 8048 Zürich
+41 41 367 73 73 / zurich@yougov.ch
www.business.yougov.com/de



onlineumfragen.com

Kernserstrasse 15, 6056 Kägiswil
+41 44 500 50 54 / info@onlineumfragen.com
www.onlineumfragen.com



POLYQUEST AG

Flurstrasse 26, 3014 Bern
+41 31 335 64 00 / info@polyquest.ch
www.polyquest.ch



Publicom AG

Winterthurerstrasse 92, 8006 Zürich
+41 44 716 55 11 / publicom@publicom.ch
www.publicom.ch



QUALINSIGHT Sàrl

Route des Mosses 34, 1613 Maraçon
+41 79 911 62 19 / e.seve@qualinsight.ch
www.qualinsight.ch



Qualitest AG, Institut für Marketing- und Sozialforschung

Rosenberghöhe 3, 6004 Luzern
+41 41 712 12 21 / qualitest@qualitestag.ch
www.qualitestag.ch



SensoPLUS

Industriestrasse 16, 6300 Zug
+41 41 710 71 61 / info@sensoplus.ch
www.sensoplus.ch

