

Markt- und Meinungsforschung

1 Wie gestalte ich gute Items und Interviewfragen?

Meinold T. Thielsch, Timo Lenzner & Torsten Melles

In diesem Kapitel geben wir anhand praktischer Beispiele Hinweise zur Gestaltung von Fragebogenitems und Interviewfragen. Neben grundsätzlichen Regeln stehen dabei besonders die Informationserhebung und die Verständlichkeit von Fragen im Fokus. Die dargestellten Grundsätze werden durch ein Praxisbeispiel zur Erfassung der Kundenzufriedenheit verdeutlicht.

1.1 Einleitung

Das Interview ist für Psycholog*innen in der Praxis das zentrale Werkzeug und ist die Methode, mit der am häufigsten Daten erhoben werden (Roth, Schmitt & Herzberg, 2010; Thielsch et al., 2012). So wird von ausgebildeten Psycholog*innen selbstverständlich erwartet, dass sie einen Interviewleitfaden oder einen Fragebogen selbst konstruieren, optimieren und anwenden können. Es erscheint zwar einfach, ist aber oft viel schwieriger als man zunächst vermutet: Die optimale Gestaltung der Fragen. Nicht alle Formulierungen und Fragetypen aus dem Alltagssprachgebrauch eignen sich für Interviews, Fragebögen oder diagnostische Verfahren. Dieses Kapitel soll daher Hilfestellung bei der Erstellung und Revision von Fragen oder Testitems geben und an praktischen Beispielen Probleme aufzeigen.

Erstellt man ein Instrument für eine einmalige Befragung sind besonders die grundsätzlichen Konstruktionsregeln für die Items zu beachten, eine genauere Itemanalyse wird sich in solchen Projekten praktisch nicht umsetzen lassen. Je stärker aber die Messung standardisiert erfolgen soll und je wichtiger die resultierenden Konsequenzen einer Befragung sind, desto mehr empfiehlt sich neben einem üblichen Fragebogenvortest (für eine Übersicht siehe Howard, 2018) eine nachfolgende statistische Itemanalyse und Itemrevision auf Basis empirischer Daten. Bei psychometrischen Testverfahren und anderen Fragebögen mit diagnostischer Zielsetzung sind diese unumgänglich – hier empfiehlt sich ein Blick in die einschlägige Literatur (bspw. Bühner, 2021; Moosbrugger & Kelava, 2020). Gut formulierte Items sind dabei die entscheidende Grundlage für objektive, reliable sowie valide Fragebögen und Interviews. Daher empfiehlt sich, wenn möglich immer der Rückgriff auf etablierte Instrumente; neu erstellte Fragebögen sollten unbedingt einem Vortest unterzogen werden (vgl. Krosnick & Presser, 2010 S. 294f.).

1.2 Grundsätzliche Regeln

Wenn man einen Interviewleitfaden oder einen Fragebogen entwickelt, so ist zunächst entscheidend, diesen so gut wie möglich an die gegebene Aufgabenstellung anzupassen. Hierbei können nicht nur umfangreiche Recherchen hinsichtlich schon vorhandener Verfahren oder relevanter Hintergrundinformationen von grundlegender Bedeutung sein, auch über Ziele und Adressaten sollte ausreichend nachgedacht werden. In manchen Situationen mag es sich anbieten, vor der eigentlichen Fragebogen- oder Leitfadenkonstruktion eine zusätzliche Anforderungsanalyse durchzuführen.

1.2.1 Ziele und Zielgruppe beachten

Von enormer Wichtigkeit ist es, sich klar zu werden, mit welchem Ziel und für welche Zielgruppe ein Fragebogen oder Interviewleitfaden erstellt wird. Hiervon hängt die gesamte Gestaltung der Fragen maßgeblich ab.

1. Fragen an das Verständnisniveau der Befragten anpassen

Beim Entwerfen der Fragen ist es wichtig, auf Sprache und Schwierigkeit zu achten. Man muss sie dem Sprach- und Bildungsniveau der Befragten anpassen. Diese sollten nicht mit Fragen konfrontiert werden, die sie aufgrund von Verständnisproblemen mit hoher Wahrscheinlichkeit nicht beantworten können (vgl. Bortz & Döring, 2015).

Beispiel: *„Bei welchem Provider haben Sie einen Postpaid-Vertrag?“* Besser: *„Bei welchem Mobilfunkanbieter haben sie einen Vertrag mit monatlicher Abrechnung?“*

2. An die Erfahrungen der Befragten anknüpfen

Soweit möglich empfiehlt es sich, an den Erfahrungshintergrund der Befragten anzuknüpfen. Beispielsweise ist es, anstatt die Führungsqualität des Vorgesetzten allgemein zu beschreiben, für den Interviewten einfacher zu beantworten, wie das Verhalten des Vorgesetzten in *konkreten* Situationen ausfiel, z. B. als ein Mitarbeiter wiederholt zu spät zur Arbeit kam (vgl. Fisseni, 2004). Zusätzlich wird empfohlen, bei der Formulierung von Fragen an Orte, Zeiten oder Zusammenhänge anzuknüpfen, die der oder die Befragte kennt. Besondere Möglichkeiten bieten hier Online-Fragebögen, die es erlauben, bestimmte Frageninhalte an die zuvor gegebenen Antworten der Befragten anzupassen.

Beispiel: *Bei einem Stromverbrauch von ca. 2.000 kWh: „Welchen der beiden folgenden Stromtarife bevorzugen Sie?“* (Dabei wird der Verbrauch an die genannte Zahl aus einer früheren Antwort im Fragebogen oder an Fragen zu Verbrauchern im Haushalt angepasst.)

3. Standardisierte Fragebögen: Itemformat an die Art der Testung anpassen
 Zusätzlich sollte man laut Bühner (2021) das Itemformat an die Art eines Tests anpassen. So erfordert z. B. ein Speedtest einfache Items, ein Persönlichkeitsfragebogen heterogene Fragen oder ein Leistungstest Aufgaben verschiedener Schwierigkeitsstufen.

1.2.2 Geschlossene versus offene Fragen

Während in Fragebogeninstrumenten geschlossene Fragen klar bevorzugt werden, hängt es im qualitativen Interview stark von der Situation und der Zielstellung ab, ob man offen oder geschlossen fragt. Nicht immer ist eine offene Frage der beste Weg zu interviewen, auch wenn viele dies zunächst annehmen würden, da sich dann der Befragte maximal frei äußern kann. Ein Beispiel: Einen Arbeitnehmer ohne Führungsaufgabe zu fragen, was er von der strategischen Ausrichtung des Unternehmens hält, kann wenig ergiebig sein – besser man fragt zunächst geschlossen, ob dieser überhaupt über die Kernpunkte der Firmenstrategie informiert ist/wurde, bevor man in die Tiefe geht. Ob offen oder geschlossen gefragt wird, hängt somit auch stark am Fokus der Frage ab. Ein weiteres Beispiel: Will man die Relevanz von Kaufkriterien erfassen, sind offene Fragen hierfür ebenfalls nur bedingt geeignet. Nicht immer ist das, was der oder die Befragte als erstes aus dem Gedächtnis abrufen bzw. woran sich er sich am leichtesten erinnert, das wirklich entscheidende Kaufkriterium (Beispielsweise kommt einem die Akkuleistung eines Handys nicht als erstes in den Sinn, wohl aber seine Funktionen). Die Relevanz kann er oder sie besser einschätzen, wenn ihm die Kaufkriterien zum Vergleich vorgegeben werden.

1. Offene Fragen

Im Interview bieten sich offene Fragen insbesondere in Explorationsphasen an. Dabei gibt der/die Interviewer*in keine Antwortmöglichkeiten vor, sondern formuliert die Frage offen. Insbesondere zu Beginn neuer Themenblöcke kann der oder die Interviewte zunächst einmal mit eigenen Worten erzählen, der oder die Interviewer*in kann später hieran anschließen.

Im Fragebogen können die Befragten ihre Antwort frei in einem Text formulieren. Offene Fragen einzubauen ist von Vorteil, wenn man ablehnendes oder sensibles Verhalten erfassen möchte (vgl. Converse & Presser, 1986, pp. 33-35) oder durch eine Vorgabe von Antwortmöglichkeiten das gesamte Spektrum an Äußerungen ggf. nicht erfasst wird. Hinsichtlich der Teilnahmemotivation der Befragten ist es jedoch meist vorteilhaft, derartige Items eher ans Ende einer Befragung zu stellen. Außerdem sollte die Anzahl von offenen Fragen in Fragebögen nicht zu groß sein, da deren Beantwortung relativ viel Zeit kostet (vgl. Borg 2003, S. 120).

2. Geschlossene Fragen

Geschlossene Fragen sollten verwendet werden, wenn man allen Befragten bei der Beantwortung denselben Referenzrahmen vorgeben möchte. Dies ist insbesondere in Konkretisierungsphasen eines Gesprächs angebracht. Bei einem Interview werden allerdings zu starre Antwortvorgaben (man denke nur an manches Interview in der Telefonmarktforschung) oft als ermüdend erlebt. Manche Befragte neigen zudem dazu, auf geschlossene Fragen dennoch offen zu antworten.

Bei einem geschlossenen Antwortformat in einem Fragebogen können nur Antworten ausgewählt werden, die zuvor von den Verfasser*innen des Fragebogens vorgegeben worden sind. Dies ist bei der Auswertung von Vorteil, da diese einheitlich und schnell ablaufen kann. Problematisch ist aber, wenn in diesem Format wichtige Antwortalternativen fehlen und dadurch die Angaben der Befragten verzerrt werden.

1.2.3 Die wichtigsten Formulierungsregeln

Sind Ziele und Zielgruppe der Befragung geklärt und die Fragen auf Basis von Anforderungsanalyse, Literatur oder anderen Informationen erstellt, so ist man noch nicht fertig. Mit der Beachtung einiger weniger Regeln lassen sich Fehler in der Konzeption von Interviewleitfäden und Fragebögen vermeiden (vgl. bspw. Krosnick & Presser, 2010; Porst, 2014). Im Folgenden sollen diese Grundregeln dargestellt werden:

1. Leicht verständliche Fragen verwenden

Allgemein wird empfohlen, leicht verständliche und grammatikalisch richtige Items zu verwenden. Diese sollten in Alltags- und nicht in Wissenschaftssprache sowie sachlich und neutral formuliert werden. Nur so kann man sicherstellen, dass alle Befragten die Fragen in gleicher Weise verstehen und in der Lage sind, sie richtig zu beantworten (vgl. Converse & Presser, 1986, pp. 10f). Wenn zwei oder mehr Befragte eine Frage unterschiedlich interpretieren, so unterscheiden sich ihre Antworten ggf. nicht aufgrund von unterschiedlichen Meinungen, sondern aufgrund des unterschiedlichen Fragenverständnisses. Hilfreich kann es daher sein, zu überprüfen, ob es für die zentralen Begriffe in den einzelnen Fragen Synonyme gibt, die geläufiger sind. Ist dies der Fall, so sollte man die weniger geläufigen Begriffe durch diese Synonyme ersetzen (z. B. statt „*Diskrepanz*“ besser „*Unterschied*“).

2. Kurze Items formulieren

Die Items an sich sollten möglichst kurz sein. Dies führt zu einem schnelleren Verständnis und Teilstücke werden nicht so schnell überhört oder überlesen. Man sollte Items nur verlängern, wenn Erklärungsbedarf besteht. Zum Beispiel dann, wenn es notwendig erscheint eine Definition zu nennen, damit eine bestimmte Frage von allen Interviewten beantwortet werden kann. Diese Definition sollte vor der eigentlichen Frage platziert werden. Dabei ist darauf zu achten, dass keine Wiederholungen vorkommen (vgl. Converse & Presser, 1986, pp.11f).

Beispiel: *„Ich lese Ihnen nun einige Einkommensgruppen vor. In welche Gruppe würden Sie Ihren Haushalt vom gesamten monatlichen Nettoeinkommen her einordnen. Ich meine die Summe aus Lohn, Gehalt, Einkommen aus selbständiger Tätigkeit, Rente oder Pension. Steuer und Sozialversicherungsbeiträge müssen abgezogen werden. Rechnen Sie bitte auch die Einkünfte aus öffentlichen Beihilfen, Einkommen aus Vermietung, Verpachtung, Wohngeld, Kindergeld und sonstige Einkünfte hinzu.“* Besser: *„Wie hoch ist das monatliche Nettoeinkommen Ihres Haushalts insgesamt? – Also von allen Personen zusammengerechnet nach Abzug von Steuern und Sozialabgaben.“*

3. Keine überflüssigen Fragen

Einen neu entwickelten Fragebogen oder Interviewleitfaden als Ganzes sollte daraufhin überprüft werden, ob wirklich jede seiner Fragen erforderlich ist. Überflüssige Items belasten den Befragten zusätzlich und verlängern den Fragebogen unnötig. Mit Fragen, die man nur eventuell auszuwerten gedenkt, sollte äußerst sparsam umgegangen werden (vgl. Bortz & Döring, 2015).

4. Keine Suggestivfragen verwenden

Suggestivfragen sind weitgehend zu vermeiden (Beispiel: *„Sie sind sicher auch der Meinung, dass...“*). Der Stil der Fragen sollte die Befragten ermuntern, das zu sagen, was sie für richtig halten – nicht was Interviewer*in oder Auftraggeber*in der Befragung im Vorfeld vermuten. Items müssen so formuliert sein, dass sie keine bestimmten Antworten besonders nahelegen (vgl. Bortz & Döring, 2015).

5. Sinnvolle Struktur des gesamten Fragebogens oder Interviews

Die Komposition aller Items im Fragebogen oder Interview sollte einer sinnvollen Struktur folgen. Beispielsweise sollten Fragen zum gleichen Thema innerhalb eines Fragenblocks und nicht über das gesamte Interview oder den Fragebogen verteilt gestellt werden. Darüber hinaus sollten Fragebogenitems in der Regel vom Allgemeinen zum Speziellen sortiert sein, um den Befragten die Beantwortung zu erleichtern (vgl. Bradburn et al., 2004). Bei Leistungstests findet sich nach den Eisbrechern oft eine Itemreihung mit ansteigender Schwierigkeit, so können zumindest zu Beginn alle Getesteten erfolgreich ein Item lösen und die Itemreihung provoziert keine verzerrten Antwortmuster (bspw. durch Demotivation Einzelner). Die Position von offenen Fragen ist zu überdenken: Geht es um spontane Eindrücke gehören diese eher an den Anfang, summative Bewertungen eher ans Ende. Dies gilt so allerdings nicht in Online-Studien (vgl. Thielsch & Weltzin, 2021): In nicht rein qualitativen Online-Befragungen sollten offene Fragen eher sparsam verwendet und ans Ende gestellt werden, da viele offene Items zu Beginn erfahrungsgemäß die Abbruchquote erhöhen. In Online-Umfragen können auch Filterfragen die Effizienz der Studie erhöhen und die Belastung der Befragten verringern (indem Befragte beispielsweise nur Aspekte

bewerten, die sie vorher als bekannt genannt haben). Sollen in Online-Befragungen selektive Effekte bei Befragungsabbrechern erfasst werden, ist es sinnvoll, die Demografie an den Anfang der Befragung zu stellen, um eine Prüfung zu ermöglichen. Gleiches gilt auch für nicht-demografische Fragen, die selektive Effekte erklären können (z. B. Involvement für das Thema der Befragung). Grundsätzlich ist es sehr sinnvoll zu überlegen, welche Fragen sich zur Qualitätsprüfung der Angaben eignen. Insbesondere bei längeren schriftlichen Befragungen bieten sich hier explizite Kontrollfragen an: So empfehlen Meade und Craig (2012) ungefähr alle 50 bis 100 Fragen ein instruiertes Testitem (z.B. „Bitte antworten Sie bei dieser Frage mit stimme zu“) einzufügen, dabei sollte es in Summe nicht mehr als drei solcher Items im Fragebogen geben.

1.3 Verständlichkeit erhöhen

Mit der Beachtung einiger zusätzlicher Punkte lässt sich leicht die Verständlichkeit eines schriftlichen Fragebogens noch weiter erhöhen. Als Interviewer*in in einer face-to-face Situation sollte man versuchen, diese Regeln ebenso zu beachten, auch wenn es teilweise schwierig sein mag, diese „on the fly“ in einem Gespräch alle vor Augen zu haben.

1. Nur eine Aussage pro Frage!

Unabhängig vom Sprachniveau des Befragten ist es wichtig, komplexe Sachverhalte in Einzelfragen zu zerlegen und einfach darzustellen. Je Frage sollte nur eine einzige Aussage gemacht werden! Nach Fisseni (2004) sollte man Fragestellungen wie die Folgende tunlichst vermeiden: *„Schildern Sie mir bitte die einzelnen Bewältigungsstrategien je nach psychologischer Teildisziplin beziehungsweise je nach sozialer Situation. Also, Sie haben in Ihrem Studium unterschiedliche Fächer zu belegen: Statistik, Allgemeine Psychologie und so weiter. Bitte geben Sie an, welches Fach schwer und welches leicht für Sie ist und was Sie tun, wenn es Ihnen schwerfällt, je nach Fach.“*

2. Keine Doppelfragen nutzen

Mit Doppelfragen und insbesondere Oder-Fragen (Beispiel: *„Fiel es Ihnen leicht mit dem Rauchen aufzuhören oder hatten Sie sehr mit sich zu kämpfen?“*) ist vorsichtig umzugehen – diese führen zu zwei Problemen: Zum einen können diese einen suggestiven Charakter bekommen und die Befragten in eine bestimmte Richtung beeinflussen. Zum anderen sind die Befragten oft überfordert und wissen nicht, auf welchen Teil der Frage sie antworten sollen oder viel häufiger noch: vergessen alle erfragten Aspekte zu beantworten. Manchmal wird auf solche Fragen nur mit „Ja“ oder „Nein“ geantwortet und es ist unklar, auf welchen Fragenteil sich diese Antwort bezieht. Um solchen Missverständnissen aus dem Weg zu gehen, meidet man zumeist Oder-Fragen (Fisseni, 2004).

Ausnahme: Wenn die Frage alle möglichen Antwortalternativen adäquat abbildet und sich diese Alternativen klar ausschließen, sind derartige Fragen durchaus zulässig. Ein Beispiel: „*Meinen Sie mit ‚am Wochenende‘ den Samstag oder den Sonntag?*“

3. Keine doppelte Verneinung verwenden

Noch schwieriger zu verstehen und zu beantworten sind Fragen mit doppelter Verneinung (Beispiele: „*Könnten Sie nicht ausschließen, dass keine anderen Motive Sie beeinflusst haben?*“ oder auch „*Wie stark stimmen Sie der folgenden Aussage zu? – Beim Bausparen sind keine Einmalzahlungen möglich. Skala: „trifft überhaupt nicht zu“ bis „trifft voll und ganz zu“*). Hier muss sich der oder die Befragte erst Gedanken machen, wie die Frage zu verstehen ist, bevor er oder sie über die eigene Antwort nachdenken kann. Diese Verständnisschwierigkeiten können neben falschen Antworten zusätzlich zu einem weiteren unerwünschten Effekt führen: Einer generell geringeren Motivation weiter zu antworten (vgl. Bühner, 2021, S. 80f.).

4. Nur Begriffe mit eindeutiger Bedeutung einbringen

Man sollte es vermeiden, Begriffe mit mehreren Bedeutungen zu verwenden. Zum Beispiel wird bei der Aussage „*Ich bin in Gesprächen angriffslustig*“ nicht deutlich, ob sie positiv im Sinne von „*Ich vertrete meine Meinung offensiv*“ oder negativ in Form von „*Ich mache andere nieder*“ gemeint ist. Dies kann zu Verständnisproblemen führen. Gleichmaßen sollten abstrakte Begriffe vermieden werden, die per Definition vage sind. Abstrakte Begriffe (z.B. „*kulturelle Veranstaltungen*“) kann man sehr leicht daran erkennen, dass sie mehrere spezifischere und konkretere Begriffe einschließen (z.B. „*Theateraufführungen*“, „*Opern*“, aber auch „*Pop-Konzerte*“, „*Rock Festivals*“). Um die Verständlichkeit von Fragen zu erhöhen, sollte eher ein konkreter Begriff benutzt werden. Ein Beispiel: Es ist ungünstig zu fragen „*Wie häufig besuchen Sie in Ihrer Freizeit kulturelle Veranstaltungen?*“. Der /die Fragebogenentwickler*in sollte sich besser vor der Befragung Gedanken darüber machen, welche Arten von kulturellen Veranstaltungen hier interessieren und diese dann spezifisch abfragen, z. B.: „*Wie häufig besuchen Sie in Ihrer Freizeit Theateraufführungen?*“ (vgl. Lenzner et al., 2010). Alternativ können die relevanten Aspekte im Item aufgeführt werden. Beispiel: Statt „*Wie häufig informieren Sie sich über Finanzdienstleistungen?*“ ist es besser zu fragen „*Wie häufig informieren Sie sich über Spar- und Anlageprodukte, wie Tagesgeld, Festgeld oder Fonds?*“

5. Auf telegrafische Kürzel verzichten

Um die Verständlichkeit der Items weiter zu erhöhen, sollte man darauf achten, telegrafische Kürzel in Fragebögen zu unterlassen. Beispiel: „*U. U. ist es m. E. notwendig MA's am P.o.S. bes. zu schulen.*“.¹ Ausgenommen hiervon sind sehr geläufige Kürzel, die das Lesen vereinfachen wie „z. B.“.

6. Möglichst einfache syntaktische Strukturen verwenden

Die syntaktischen Strukturen der Fragen sollten möglichst wenig komplex sein, damit sie die Verarbeitungskapazität der Befragten nicht überlasten. Beispielsweise kann man die grammatikalische Komplexität von Fragen verringern, indem man Passivsätze in Aktivsätze überführt und Nominalisierungen vermeidet („*Mir steht ausreichend Arbeitszeit zur Verfügung, um neue Kenntnisse zu erwerben*“ vs. „*Mir steht ausreichend Arbeitszeit für den Erwerb neuer Kenntnisse zur Verfügung*“). Darüber hinaus sollte man sog. linksverzweigte Strukturen vermeiden, bei denen mehrere Satzteile ineinander geschachtelt sind und vor dem Hauptverb der Frage erscheinen (z. B. „*Auch wenn man manche Entscheidungen nicht für richtig hält, sollte man als Mitglied eines Teams dessen Entscheidungen im Allgemeinen selbst vertreten*“). Leichter verständlich sind rechtsverzweigte Strukturen: „*Als Mitglied eines Teams sollte man dessen Entscheidungen im Allgemeinen selbst vertreten, auch wenn...*“ (vgl. Lenzner et al., 2010).

7. Wichtige Inhalte hervorheben

Als hilfreich hat es sich erwiesen, wichtige Inhalte der Items in einem Fragebogen durch Fettdruck, Unterstreichen oder Ähnliches hervorzuheben. Man sollte hierbei allerdings darauf achten, dass diese Stilmittel nicht zu häufig angewendet werden, da sie sonst ihre Wirkung verlieren (vgl. Bühner, 2021, S. 81). In Online-Fragebögen wird optimalerweise hierfür die Schrift fett formatiert, da kursive Schriften am Bildschirm etwas schlechter lesbar sind und unterstrichene Wörter mit Hyperlinks verwechselt werden könnten.

¹ Für alle die sich wundern, was eigentlich mit diesem Item gefragt werden sollte, hier die ausformulierte Version: „Unter Umständen ist es meines Erachtens notwendig Mitarbeiter am Point-of-Sale besonders zu schulen.“

1.4 Sinnvoll Informationen erheben

Ist der Interviewleitfaden oder der Fragebogen schon entworfen oder fertig gestellt, so bringt eine Prüfung der folgenden Aspekte oft noch einen Zugewinn in der Qualität des Verfahrens. Auch auf neue Items lassen sich diese Aspekte selbstverständlich anwenden.

1. Ich-bezogene Formulierungen

Oftmals bietet es sich an, Items ich-bezogen zu formulieren. Ansonsten besteht die Gefahr, dass die/der Befragte nicht äußert, was sie/er selbst denkt, sondern was vermutlich die Mehrheit der Personen glauben könnte. Beispiel: Statt *„Die Projektdeadline war nur schwer einzuhalten“* sollte besser gefragt werden: *„Ich persönlich fand, die Projektdeadline war nur schwer einzuhalten“*.

Ausnahme: Fragen, die sich projektiver Techniken bedienen, um sensible Informationen zu erfassen. Beispiel: *„Wie stark stimmen Sie der folgenden Aussage zu? Ich kann verstehen, wenn mal unnütze Sachen im Einkaufswagen landen.“* statt *„Wie stark stimmen Sie der folgenden Aussage zu? In meinem Einkaufswagen landen gelegentlich unnütze Sachen.“* – Die erste Formulierung ist besser, da sich die/der Befragte nicht als irrationaler Konsument outen muss (= sozial unerwünscht). Die Zustimmung zur unpersönlichen Formulierung ist dennoch ein Indikator für das eigene Verhalten.

2. Keine zu allgemeinen Formulierungen benutzen

Fragebogenitems sollten nicht zu allgemein formuliert werden, Informationen lassen sich mit konkreten Fragen besser erheben. Hierauf ist besonders zu achten, wenn man mit dem Fragebogen zwischen Gefühlen, Wissen, Einstellungen und Verhalten differenzieren will (vgl. Bortz & Döring, 2015). Verallgemeinerungen wie *„Alle Kinder machen Lärm“* sollte man nicht als Item in einen Fragebogen aufnehmen. Es wird wahrscheinlich passieren, dass die Befragten solche pauschalen Aussagen unreflektiert ablehnen oder zustimmen (vgl. Bühner, 2021). Dies führt direkt zum nächsten Punkt:

3. Items vermeiden, denen immer zugestimmt wird oder die immer abgelehnt werden

Items, bei denen die Befragten nahezu immer zustimmen oder die sie fast immer ablehnen, sind zu vermeiden. Beispiel: *„Der Staat sollte dafür sorgen, dass die Renten sicher sind“*. Derartige Fragen liefern keine diagnostischen Informationen und sind überflüssig, da ihre Beantwortung keine Hinweise auf die Einstellung oder das Verhalten der untersuchten Teilnehmer*innen offenbaren kann.

4. Bei Fragen zu Einstellungen keine Sachverhalte darstellen

Für die Ermittlung von Einstellungen sind Itemformulierungen ungeeignet, mit denen wahre Sachverhalte dargestellt werden. Beispiel: „*Beim Bausparen gibt es hohe Guthabenzinsen.*“ Eine Äußerung zu diesem Item würde keine Meinung, sondern allenfalls Fachkenntnisse über Anlageprodukte erfassen. Für eine Einstellungsmessung besser geeignet wäre z. B. die Formulierung: „*Ich finde die Guthabenzinsen beim Bausparen attraktiv.*“ (Bortz & Döring, 2015).

5. Nur Fragen stellen, die Befragte auch beantworten können

Generell ist zu beachten, dass in einem Interview oder einem Fragebogen nur solche Fragen gestellt werden, die von den Befragten auch beantwortet werden können, d. h. für deren Beantwortung sie genügend Informationen besitzen (vgl. Fowler & Cosenza, 2008). Beispiel: Einen Mobilfunknutzer zum Betriebssystem eines Smartphones zu befragen, macht nur dann Sinn, wenn ihm die Bedeutung/Funktion des Betriebssystems bekannt sind.

Zudem empfiehlt sich, keine Kompetenzen vorauszusetzen, die nicht selbst Gegenstand der Befragung sind. Beispiel: Verlangt man von den Befragten Antwortwerte auf eine Frage im Kopf zu berechnen, führt dies möglicherweise zu Fehlern und Motivationsproblemen. Hier ist es besser die Einzelwerte abzufragen und nachlaufend automatisch zu verrechnen – es sei denn die Studie befasst sich mit dem Thema Rechenfähigkeit.

6. Zeitspannen genau spezifizieren

Fragen sollten möglichst konkret und ereignisbezogen formuliert sein. Wenn in einem Item auf einen Zeitpunkt oder eine Zeitspanne verwiesen wird, sollte dieser/diese eindeutig definiert sein. Nur so kann man den Probanden die Chance geben, die Items genau zu beantworten, und reliable sowie valide Ergebnisse erhalten.

Ein Beispiel für eine schlechte Zeitangabe: „*In der letzten Zeit konnte ich Weiterbildungsangebote nutzen, um mein Wissen und Können zu verbessern.*“ Aus Zustimmung oder Ablehnung dieser Frage lässt sich wenig ablesen, da der Zeitraum, auf den sich die Frage bezieht, vollkommen vage bleibt. Manche Mitarbeiter*innen werden die Zeitspanne relativ eng („in den letzten vier Wochen“), andere eher weit interpretieren („in den letzten 6 Monaten“ oder gar „im letzten Jahr“). Eine bessere Umsetzung der Frage wäre: „*In den letzten 6 Monaten konnte ich Weiterbildungsangebote nutzen, um mein Wissen und Können zu verbessern.*“

Auch bei Fragen, die in die Zukunft gerichtet sind, empfiehlt sich ein Zeit- und Ereignisbezug („*Wenn Ihnen beim Abschluss Ihres nächsten Mobilfunkvertrages [oder in den nächsten 6 Monaten], folgendes Handy angeboten wird: [Einblendung von Handy und Preis] Wie wahrscheinlich würden Sie das Handy kaufen?*“).

1.5 Im Fragebogen: Das Design der Antwortskalen

Bei der Erstellung der Antwortskalen ist zu beachten, dass Items und Antwortskalen zusammenpassen. Die Antwortvorgaben sollten das gesamte mögliche Spektrum der Antworten abdecken und dabei inhaltlich disjunkt, d. h. überschneidungsfrei, sein. In manchen Fällen empfiehlt sich statt der Erhebung mittels einer Skala eine direkte, offene Abfrage. Dies gilt insbesondere für metrische Abfragen wie zum Beispiel das Alter: Hier ist es besser offen zu fragen und das metrische Skalenniveau zu erhalten, als Kategorien wie „unter 18“, „18-25“, „26-35“ usw. vorzugeben und das Skalenniveau auf ordinal zu reduzieren (es sei den aus Datenschutzgründen soll das Alter nur grob erfasst werden). Kategorien bilden kann man mit metrisch erhobenen Daten hinterher immer noch – im gegebenen Beispiel kann man bei einer ordinalen Abfrage nicht einmal mehr berechnen wie das Alter im Durchschnitt war. Nur mit der offenen Abfrage können im Nachhinein ein arithmetischer Mittelwert oder bestimmte neue Kategorien gebildet werden.

1. Die Likertskala

Wenn es um das Erfassen von Meinungen oder Einstellungen geht, ist es häufig sinnvoll, eine Likertskala zu verwenden. Bei diesem Skalentyp werden die Fragen als Feststellungen formuliert und die Befragten sollen angeben, inwieweit sie den Aussagen zustimmen oder sie ablehnen. Ein Vorteil von Likertskalen besteht darin, dass Befragte mehrere Fragen anhand der gleichen Skala beantworten können, was die Beantwortung der Fragen erleichtert und sie weniger zeitintensiv gestaltet. Die Breite einer Likertskala sollte generell zwischen vier und maximal neun Stufen liegen, um die Befragten nicht zu überfordern. Zu beachten ist die jeweils vermutete Fähigkeit der Zielgruppe zur Abstraktion des erfragten Sachverhalts. Bei verbalisierten Skalen (bei denen jeder Skalenpunkt verbal benannt ist) liegt die Reichweite optimalerweise zwischen vier und sieben Skalenpunkten (vgl. Porst, 2014; Krosnick & Fabrigar, 1997). Bei mehr als sieben Abstufungen fällt hier das Formulieren der Anker schwer.

Ob eine gerade oder ungerade Skala gewählt wird, hängt an der Zielsetzung der Items. Einerseits möchte man die Möglichkeit, in eine Mittelkategorie auszuweichen (der Antworttendenz zur Mitte) vielleicht ausschließen. Andererseits kann genau diese mittlere Kategorie eine sinnvolle Antwortoption sein (vgl. Krosnick & Presser, 2010). Ein Beispiel: *„Die Häufigkeit von Mitarbeitergesprächen sollte...verringert werden, so bleiben wie sie ist, erhöht werden.“* Falls, wie im Beispiel, eine Mittelkategorie konzeptuell von der Frage verlangt wird, so sollte sie immer angeboten werden. Ein alternatives Beispiel: *„Wie beurteilen Sie die Freundlichkeit der Mitarbeiter von Unternehmen X im Vergleich zu Unternehmen Y? eindeutig besser, eher besser, [gleich,] eher schlechter, eindeutig schlechter.“* Es sollte dann ohne Mittelkategorie gefragt werden, wenn es wichtig erscheint Mildeeffekte auszuschließen und auch leichte Tendenzen in der Sicht des Befragten identifizieren zu können.

Beim Erstellen von Antwortskalen kann es zudem von Vorteil sein, alle Stufen der Skala verbal zu benennen. Wenn nur einige Stufen verbal verankert werden, die Anker schlecht formuliert sind oder man für die Bezeichnung Zahlen nutzt, ist die Skala womöglich für die Befragten uneindeutig – und damit weniger reliabel und valide. Die verbalen Label für die Punkte der Antwortskalen sollten für die Befragten präzise Bedeutungen haben. Durch sie sollte die Skala außerdem in äquivalente Intervalle aufgeteilt werden (vgl. Krosnick & Fabrigar, 1997, pp. 149-152).

Ausnahme: Wenn aus früheren Untersuchungen bereits bekannt ist, dass die meisten Befragten sehr positiv oder negativ antworten, kann man auch eine unausgeglichene Antwortskala verwenden. Beispiel für eine derartige Skala: „*unzufrieden – weniger zufrieden – zufrieden – sehr zufrieden – vollkommen zufrieden*“. Dies erfordert zwar fundiertes Vorwissen über die zu erfassenden Einstellungen, kann aber zu größerer Streuung und damit differenzierterer Messung führen (vgl. Brown, 2004).

Typischerweise werden folgende verbale Anker für fünfstufige Likertskalen genutzt: „*trifft gar nicht zu – trifft wenig zu – teils-teils – trifft ziemlich zu – trifft völlig zu*“ (vgl. Rohrmann, 1978, S. 239f). Denkbar ist hier auch: „*stimme gar nicht zu – stimme eher nicht zu – neutral – stimme eher zu – stimme vollkommen zu*“. Geht man davon aus, dass manche Befragte zu einer Frage *keine* Meinung haben können, so kann man eine zusätzliche Verweigerungsoption anbieten (z. B. „weiß nicht“, „keine Angabe“ oder „nicht sinnvoll beantwortbar“). Allerdings sollte man mit diesen Kategorien vorsichtig sein, da sie von manchen Befragten als Fluchtkategorie missbraucht werden, um die Beantwortung der Fragen zu beschleunigen (Krosnick & Fabrigar, 1997). Alternativ können, zumindest in Telefon- oder Online-Umfragen, derartige Ausweichmöglichkeiten weggelassen werden, dabei sollte allerdings den Befragten die Möglichkeit gegeben werden, Items ohne eine Antwort zu überspringen.

2. Weitere Antwortskalen

Es gibt aber auch Fragen bei denen man andere Antwortskalen als die Likertskala verwenden sollte. Ein Beispiel hierfür wäre das Item *„Der Projektleiter hat zu Beginn des Projekts den Meilensteinplan ausreichend dargestellt“*. In diesem Fall wäre es schlecht, wenn die Befragten dieses Item anhand einer Skala mit den Punkten *„trifft gar nicht zu – trifft wenig zu – teils-teils – trifft ziemlich zu – trifft völlig zu“* bewerten müssten. Eine derartig differenzierte Antwortskala passt hier nicht – entweder wurde der Meilensteinplan besprochen oder nicht. In diesem Fall würde es vollkommen ausreichen, wenn man als Antwortmöglichkeiten *„Ja“* oder *„Nein“* anbieten würde.

Manchmal ist es auch nötig, nach einer bestimmten Anzahl zu fragen, wie zum Beispiel in dem Item *„Wie oft ist der Mitarbeiter zu spät zur Arbeit erschienen?“*. Als Antwortmöglichkeit könnte man hier die Skala *„nie – selten – gelegentlich – oft – immer“* nutzen. In diesem Fall ist es aber sinnvoller, sich die genaue Anzahl der Verspätungen angeben zu lassen, hierdurch erhält man eine qualitativ hochwertigere Information.

Für die drei Urteilsdimensionen Häufigkeit, Intensität und Wahrscheinlichkeit sind für fünfstufige Skalen häufig folgende Antworten in Gebrauch:

- Bei Häufigkeitsskalen bieten sich folgende Anker an: *„nie – selten – gelegentlich – oft – immer“*.
- Bei Intensitätsskalen bieten sich folgende Anker an *„gar nicht – wenig – mittelmäßig – überwiegend – völlig“*.
- Die Antwortskala zur Wahrscheinlichkeit von z. B. einem Ereignis beinhaltet die Punkte *„bestimmt nicht [alternativ: auf keinen Fall] – wahrscheinlich nicht – vielleicht – wahrscheinlich – bestimmt [alternativ: auf jeden Fall]“*.

Es gibt auch Fragen von der Art *„War der Umfang des Lernmaterials, das im Weiterbildungsseminar angeboten wurde zu wenig, genau richtig oder zu viel?“*. Als Antwortmöglichkeiten sollten hier diejenigen aufgegriffen werden, die schon in der Frage genannt werden. Also *„zu wenig – genau richtig – zu viel“* (vgl. Lyons, 1998, pp. 79f).

3. Richtung der Skala

Hierbei geht es um die Ausrichtung der Skala, also ob den Befragungspersonen eine Skala angeboten wird, die von links nach rechts ansteigt oder vice versa. Beides ist denkbar und hier gelten vor allem Gewohnheitsregeln. So scheint im angloamerikanisch- und europäisch geprägten Kulturraum eine Skala *„von links nach rechts“* als intuitiver und damit sinnvoller (vgl. Porst, 2014, S. 88f.).

1.6 Projekt- und Praxisbeispiel: Erfassung von Kundenzufriedenheit

Im Folgenden wird in Ausschnitten die Entwicklung eines Fragebogens zur Messung von Kundenzufriedenheit dargestellt, um daran die Umsetzung der zuvor dargestellten Regeln und Empfehlungen zu verdeutlichen.

Kundenzufriedenheit ist definiert als (zeitlich überdauernde) Einstellung gegenüber einem Objekt, die eine affektive und eine kognitive Komponente umfasst. Die kognitive Komponente besteht in der Bildung einer Meinung, die affektive Komponente im Auftreten von Gefühlen beim Bewerten der jeweiligen Objekte (vgl. Nerdinger et al., 2015).

Soll die Zufriedenheit von Kunden mit Leistungen eines Unternehmens umfassend eruiert werden, müssen in einem ersten Schritt die (möglichen) Kontaktpunkte des Kunden mit dem Anbieter ermittelt werden. Dies kann beispielsweise im Vorfeld durch eine Vorabbefragung von Kunden, durch eine Befragung von Mitarbeiter*innen mit Kundenkontakt oder auch durch die Analyse von Dokumenten, CRM-Systemen (Systeme für die Pflege von Kundenbeziehungen, in welchen bspw. Kundeninformationen dokumentiert werden) etc. erfolgen. In einem Blueprint (ein spezielles Analysediagramm) können die Kundenkontaktpunkte festgehalten und visualisiert werden.

In der anschließenden Zufriedenheitsbefragung ist es von Vorteil, den Kunden zunächst nach der Gesamtzufriedenheit mit dem Unternehmen zu befragen.

Beispiel:

Wie zufrieden sind Sie mit der XY-Versicherung insgesamt?

- *unzufrieden*
- *weniger zufrieden*
- *zufrieden*
- *sehr zufrieden*
- *vollkommen zufrieden*

Es wird gezielt eine asymmetrische Skala eingesetzt, um typische Deckeneffekte bei Zufriedenheitsskalen zu vermeiden (vgl. Abschnitt 5 Punkt 1). Alternativ ließe sich die Frage auch folgendermaßen operationalisieren:

Wie beurteilen Sie die XY-Versicherung insgesamt?

- *schlecht*
- *mittelmäßig*
- *gut*
- *sehr gut*
- *ausgezeichnet*

Diese Skala weist eine noch etwas stärkere Asymmetrie auf als die Zufriedenheitsskala. Sie zielt allerdings primär auf die kognitive Komponente der Kundenzufriedenheit (Bilden einer Meinung) ab und weniger auf die affektive Komponente (Auftreten von Gefühlen).

Durch die Erhebung der Gesamtzufriedenheit am Anfang der Befragung wird gewährleistet, dass der Kunde oder die Kundin ein unbefangenes Urteil abgibt. Würden stattdessen zunächst einzelne Leistungsmerkmale des Unternehmens beurteilt (z. B. Erreichbarkeit des Versicherungsvermittlers), wird das Gesamturteil hierdurch beeinflusst. Bestimmte Aspekte werden möglicherweise erst durch die Befragung salient und bekommen durch eine bestimmte Positionierung im Fragebogen eine größere oder auch eine geringere Bedeutung.

Ergänzend zum Gesamturteil kann eine offene Frage eingesetzt werden, in der das Urteil begründet wird. Zum einen haben unzufriedene Kunden damit die Möglichkeit, Ihren Ärger direkt zu äußern, zum anderen wird vermieden, dass durch den Fragebogen evtl. relevante Leistungsaspekte übersehen werden (vgl. Abschnitt 2 Punkt 1).

Damit Kunden nur zur Zufriedenheit mit Sachverhalten befragt werden, die sie beurteilen können, wird im nächsten Schritt die Nutzung von bzw. Erfahrung mit Kontaktpunkten ermittelt (siehe Abschnitt 2 Punkt 2 sowie Abschnitt 4 Punkt 5):

*Wann hatten Sie das letzte Mal Kontakt mit einem/einer Versicherungsvertreter*in der XY-Versicherung?*

- *in den letzten 6 Monaten*
- *vor mehr als 6 Monaten, aber innerhalb der letzten 12 Monate*
- *vor mehr als 12 Monaten, aber innerhalb der letzten 24 Monate*
- *vor mehr als 24 Monaten*
- *noch nie*

Bestand noch nie ein Kontakt oder ist der Kontakt länger als 24 Monate her, werden in diesem Fall keine weiteren Fragen zur Zufriedenheit mit dem Versicherungsvermittler gestellt. Ein Kontakt, der sehr weit in der Vergangenheit liegt, wird wahrscheinlich nur unzureichend erinnert, so dass die Zufriedenheitseinschätzungen dann kaum noch auf tatsächlichen Erlebnissen basieren.

Wir geben bei dieser Frage Kategorien vor, da eine offene Frage mit einer Benennung des Zeitpunktes die meisten Befragten überfordern dürfte und in Online-Befragungen die Wahrscheinlichkeit eines Abbruchs erhöht. Die Zeitspanne wird in den vorgegebenen Kategorien genau spezifiziert (vgl. Abschnitt 4 Punkt 6).

Im nächsten Schritt wird die Gesamtzufriedenheit mit dem Vermittler erfasst: *Wie zufrieden sind Sie mit dem/der Versicherungsvertreter*in der XY-Versicherung insgesamt?*

- *unzufrieden*
- *weniger zufrieden*
- *zufrieden*
- *sehr zufrieden*
- *vollkommen zufrieden*

Wir verwenden die Begriffe „Versicherungsvertreter*in“ und „Versicherung“, da diese der Zielgruppe geläufiger sind als die Begriffe „Versicherungsvermittler“ und „Versicherer“ (vgl. Abschnitt 2 Punkt 1).

Ziel von Kundenzufriedenheitsstudien ist meistens, nicht nur den Zufriedenheitsgrad der Kunden mit Merkmalen des Anbieters zu erheben, sondern auch konkrete Hinweise für Verbesserungsmöglichkeiten zu gewinnen. Daher wird beispielsweise im Anschluss an die Gesamtbeurteilung des Vermittlers dessen Verhalten eingeschätzt. In der Praxis erfolgt dies entweder über Urteils- oder Zufriedenheitsskalen – analog zur Gesamtbeurteilung – oder über Zustimmungsskalen oder Faktfragen. Soll etwa die Einschätzung der Erreichbarkeit erfasst werden, sind folgende Fragen möglich:

Zufriedenheitsskala: Wie zufrieden sind Sie mit der Erreichbarkeit des Versicherungsververtreters/der Versicherungsvertreterin?

- *unzufrieden*
- *weniger zufrieden*
- *zufrieden*
- *sehr zufrieden*
- *vollkommen zufrieden*

*Zustimmungsskala (vgl. Abschnitt 5 Punkt 1): Wie stark stimmen Sie der folgenden Aussage zu? Ich kann den/die Versicherungsvertreter*in gut erreichen.*

- *trifft gar nicht zu*
- *trifft wenig zu*
- *teils-teils*
- *trifft ziemlich zu*
- *trifft völlig zu*

Faktfrage: Als Sie das letzte Mal mit Ihrem Versicherungsvertreter/Ihrer Versicherungsvertreterin in Kontakt treten wollten: Wie viele Versuche haben Sie unternommen, bis Sie ihn/sie erreicht haben?

- *einen (habe ihn auf Anhieb erreicht)*
- *zwei*
- *drei*
- *vier*
- *mehr als vier*

Die Zustimmungsskala und die Faktfrage gelten als weniger anfällig für Deckeneffekte und Ausstrahlungseffekte (hier: Ausstrahlung der Gesamtzufriedenheit mit dem/der Berater*in auf die Beurteilung der Leistungsmerkmale). Die Antwortkategorien der Faktfrage sind darüber hinaus an konkreten Leistungsmerkmalen verankert, so dass sich hieraus am besten Maßnahmen (z. B. Definition von Servicelevels) ableiten lassen. Die Zufriedenheits- und die Zustimmungsskala weisen gegenüber der Faktfrage allerdings den Vorteil auf, dass sich die gleichen Antwortkategorien für verschiedene Fragen (z. B. Einhalten von Zusagen, Eingehen auf die persönliche Situation des Befrag-

ten) verwenden lassen. Somit lassen sich in Zusammenhangsanalysen (hier: mit der Gesamtbeurteilung des Vermittlers) verschiedene Leistungsmerkmale direkt miteinander vergleichen. Weist dann beispielsweise die Erreichbarkeit einen höheren Zusammenhang mit der Beurteilung des Vermittlers auf als das Eingehen auf die persönliche Situation, wäre dies ein Indikator dafür, dass die Verbesserung der Erreichbarkeit die effektivere Maßnahme wäre.

Viele weitere praktische Beispiele und hilfreiche Tipps zur Fragebogenerstellung finden sich bei Porst (2014).

Literaturverzeichnis

- Borg, I. (2003). *Führungsinstrument Mitarbeiterbefragung* (3. Überarb. Aufl.). Göttingen: Hogrefe-Verlag.
- Bortz, J. & Döring, N. (2015). *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler* (5. überarb. Aufl.). Heidelberg: Springer Medizin Verlag.
- Bradburn, N., Sudman, S. & Wansink, B. (2004). *Asking Questions. The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco, CA: Jossey-Bass.
- Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports*, 94, 1015-1024. doi: <https://doi.org/10.2466/pr0.94.3.1015-1024>
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4. Aufl.). München: Pearson Studium.
- Converse, J. M. & Presser, S. (1986). *Survey Questions Handcrafting the standardized questionnaire*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-063. Beverly Hills: Sage Pubns.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik* (3. erw. und überarb. Aufl.). Göttingen: Hogrefe.
- Fowler, F. J. & Cosenza, C. (2008). Writing effective questions. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 136-160). New York, NY: Taylor & Francis.
- Howard, M. C. (2018). *Scale Pretesting. Practical Assessment, Research & Evaluation*, 23(5). <https://doi.org/10.7275/hwpz-jx61>
- Krosnick, J. A. & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 141-164). New York, NY: John Wiley & Sons, Inc.
- Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In P. V. Marsden & J. D. Wright (Eds.) *Handbook of Survey Research* (pp. 263-312). Bingley (UK): Emerald.

- Lenzner, T., Kaczmarek, L. & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, 24(7), 1003-1020. <https://doi.org/10.1002/acp.1602>
- Lyons, W. (1998). Beyond agreement and disagreement: the inappropriate use of Likert items in the applied research culture. *Int. J. Social Research Methodology*, 1(1), 75-83. <https://doi.org/10.1080/13645579.1998.10846864>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437-455.
- Nerdinger, F. W., Neumann, C. & Curth, S. (2015). Kundenzufriedenheit und Kundenbindung. In K. Moser (Hrsg.), *Wirtschaftspsychologie* (S. 119-138). Heidelberg: Springer
- Moosbrugger, H. & Kelava, A. (Hrsg.) (2020). *Testtheorie und Fragebogenkonstruktion* (3. Aufl.). Heidelberg: Springer.
- Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9, 222-245.
- Porst, R. (2014). *Fragebogen - Ein Arbeitsbuch* (4. Aufl.). Wiesbaden: Springer VS.
- Roth, M., Schmitt, V. & Herzberg, P. Y. (2010). Psychologische Diagnostik in der Praxis: Ergebnisse einer Befragung unter BDP-Mitgliedern. *Report Psychologie*, 35(3), 118-128.
- Thielsch, M. T., Brandenburg, T. & Kanning, U. P. (2012). Diagnostische Verfahren im Praxiseinsatz. In R. Reinhardt (Hrsg.), *Wirtschaftspsychologie und Organisationserfolg* (S. 57-64). Lengerich: Pabst Science Publishers.
- Thielsch, M. T. & Weltzin, S. (2021). Online-Umfragen und Online-Mitarbeiterbefragungen. In T. Brandenburg, P. Mehlich & M. T. Thielsch (Hrsg.), *Praxis der Wirtschaftspsychologie II: Fragen, Führen und Verändern*. (S. 31-48). Lengerich: Pabst Science Publisher.