

**Auf den Spuren der Befragungsoptimierung – Einflüsse der Befragungsgestaltung auf
verschiedene Gütekriterien der Befragung**

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

des Fachbereichs Humanwissenschaften

der Universität Osnabrück

vorgelegt von

Patrick Hofschröder, M.Sc.

aus

Mettingen

Osnabrück, 2022

BerichterstellerInnen:

Prof. Dr. Karsten Müller (Erstgutachter)

Prof. Dr. Nick Lin-Hi (Zweitgutachter)

Tag der mündlichen Prüfung:

27.09.2023

(genehmigte Kurzversion)

Danksagung

Allen beteiligten Personen, die mich bei der Anfertigung meiner Dissertation unterstützt haben, gilt meine Dankbarkeit.

Meinen besonderen Dank möchte ich Prof. Dr. Karsten Müller ausdrücken, der mich während der gesamten Zeit sehr unterstützt und die Durchführung dieser Arbeit erst ermöglicht hat. Durch sein Fingerspitzengefühl in den kniffligen Projekt-Phasen und wertvollen Impulse im wissenschaftlichen Diskurs war er eine große Hilfe.

Außerdem möchte ich mich bei Prof. Dr. Nick Lin-Hi bedanken, dass er zugestimmt hat meine Dissertation zu begutachten.

Darüber hinaus gilt mein Dank den Co-Autor:innen unserer Studien, die Teil dieser Dissertation sind: Dr. Svenja Schumacher, Philipp Hund, Franca Schoenemann und Thorsten Bauhaus. Dr. Svenja Schumacher möchte ich dabei besonders hervorheben, da wir den Weg der Promotion auch in schwierigen Zeiten gemeinsam gegangen sind und ihr Einsatz sowie ihre Leidenschaft für die Forschung maßgeblich zum Gelingen der Studien beigetragen haben. Ich danke Dir, Svenja! Auch Lisa Brockmann, Jana Britzwein, Gina Kalverkamp und Katja Oelher möchte ich für Ihre Unterstützung im Forschungsprozess danken.

Des Weiteren möchte ich mich bei meinem tollen Team der Arbeits- und Organisationspsychologie (Dr. Tammo Straatmann, Cosima Koßmann, Prof. Dr. Kai-Christoph Hamborg, Shirin Betzler, Jonas Brüggemann, Melina Cardena, Anna Engel, Nina Heidenescher, Christopher Klanke, Ulla Krimphoff, Prof. Dr. Regina Kempen, Dr. Alexander Meier, Thea Nieland, Lena Schmeiduch, Sarah Rethmann, Marina Schefer, Jan-Philip Schumacher, Sven-Oliver Spieß, Nina Tegeler, Tabea Weil) für Euren Rückhalt und Euren Einsatz bedanken. Ihr seid einmalig.

Auch Iris Gottschalk und Marion Meyer danke ich für ihre Unterstützung.

Zudem gilt großer Dank meinen Eltern: Reinhard Hofschröder und insbesondere Martina Hofschröder, die mir auch in anstrengenden Zeiten, zu jeder Zeit genau die Unterstützung gaben, die ich brauchte. Außerdem möchte ich meiner Schwester (Janina Hofschröder), meiner Freundin (Vanessa Opladen), sowie meinen Freund:innen danken, die als wichtiger Rückhalt in jeder Phase der Dissertation für mich da waren und mich bedingungslos unterstützt haben.

Danke!

Liste der Artikel

Diese Dissertation basiert auf den folgenden Artikeln:

Artikel 1: Hofschroerer*, P., Schumacher*, S. K., & Mueller, K. (under review). Does “very” make a difference? Effects of intensifiers in item stems of employee attitude surveys on response behavior.

Artikel 2: Schumacher, S. K., Hofschroerer, P., Hund, P. M., Schoenemann, F., & Mueller, K. (under review). Let’s make it an experience! Introducing SX as a respondent-centered criterion for successful survey design.

Artikel 3: Hofschroerer, P., Schumacher, S. K., Bauhaus, T. & Mueller, K. (under review). Gamified surveys in organizations: Does instant feedback enhance employees’ survey experience and data quality? - A Field Experiment.

*Die Autor:innen haben gleichermaßen beigetragen.

Inhaltsverzeichnis

Zusammenfassung	7
Abstract	8
Generelle Einführung	9
Trends in Employee Attitude Surveys.....	14
Qualifier als Interventionsmöglichkeit.....	14
Die Motivation als Herausforderung.....	17
Survey Design Features.....	18
Spezieller Survey Design-Lösungsansatz: Gamification	20
Die vorgestellte Forschung.....	22
Artikel 1: Does “very” make a difference? Effects of intensifiers in item stems of employee attitude surveys on response behavior.....	25
Artikel 2: Let’s make it an experience! Introducing SX as a respondent-centered criterion for successful survey design.....	27
Artikel 3: Gamified surveys in organizations: Does feedback with an aesthetic design enhance employees’ SX and data quality? - A Field Experiment.....	29
Diskussion	31
Generelle Implikationen für die Forschung.....	35
Generelle praktische Implikationen.....	38
Generelle Limitationen.....	41
Zukünftige Forschung	42
Generelle Schlussfolgerung.....	44
Referenzen.....	47

Abbildungsverzeichnis

Abbildung 1. Prozessmodell der Einflussmöglichkeiten von Befragungsgestaltung auf die Gütekriterien der Befragung.	12
Abbildung 2. Zusammenfassung der durchgeführten Studien	22

Zusammenfassung

Für langfristigen Unternehmenserfolg ist ein konstruktiver Austausch mit den Mitarbeitenden unter dynamischen Rahmenbedingungen wichtig. Dieser Austausch kann über Mitarbeitendenbefragungen (engl. employee attitude surveys) initiiert und abgebildet werden. Bei dem Einsatz dieser Befragung gibt es jedoch zwei wesentliche Herausforderungen. Zum einen zeigt sich eine hohe Zustimmungsrate im Antwortverhalten, die zu weniger Informationsgehalt in den Daten führt. Zum anderen begünstigt das Phänomen der auftretenden Befragungsmüdigkeit (engl. survey fatigue) inakkurate Antworten der Teilnehmenden. Beide Facetten erschweren so die Auswertung und Ableitung von Maßnahmen aus der Befragung. Entsprechend soll die vorliegende Forschungsarbeit mit Hilfe von verschiedenen Befragungsgestaltungen diese Herausforderungen adressieren. Zu Beginn untersuchen wir systematisch die Itemformulierung in Mitarbeiterbefragungen. Dabei nutzen wir Intensifier im Itemstamm, um dem Effekt der hohen Zustimmungsraten potenziell entgegenzuwirken. Für die Herausforderung der Befragungsmüdigkeit ist neben dem Antwortverhalten, die subjektive Wahrnehmung der Mitarbeitenden relevant. Umfragen erheben meist lediglich teilnahmeorientierte und qualitätsorientierte Evaluationskriterien. Die Erweiterung der Evaluationskriterien um die Teilnehmenden-Wahrnehmung (engl. survey experience) fördert ein besseres Verständnis der Befragungsqualität. Mit Hilfe der systematisch entwickelten und validierten SX-Kurzskala lässt sich diese effizient erfassen. Die Kurzskala dient in der anschließenden Studie der Validierung des genutzten Gamification-Elements namens Instant-Feedback und der Herausstellung des inkrementellen Nutzens für einzelne Facetten, wie das Antwortverhalten und die „pleasure“. Auf Basis der Ergebnisse werden Implikationen für die Praxis abgeleitet, die sowohl die Frageformulierung (z. B. spezifisch Überprüfungen der Vergleichbarkeit von Items) als auch die Befragungsgestaltung (z. B. Nutzung von Gamification-Elementen) betreffen. Darüber hinaus legt die Forschung die Basis für weiterführende Fragestellungen, die in dieser Arbeit diskutiert werden. Insgesamt zeigen die Ergebnisse das Potential einer durchdachten Befragungsoptimierung für die Befragungsgüte einer Befragung.

Schlüsselworte: Mitarbeiterbefragung, Antwortverhalten, Survey Experience, Gamification, Intensifier, Befragungsgüte

Betreuer: Prof. Dr. Karsten Müller

Abstract

Facing dynamic conditions, a constructive dialogue with employees is important for the long-term success of a company. This dialogue can be initiated and mapped by using employee attitude surveys. However, there are two main challenges in using these surveys. On the one hand, there is a high rate of agreement in the response behavior, which leads to bias in the data. On the other hand, the phenomenon of survey fatigue favors inaccurate answers by the participants. Both facets thus complicate the evaluation and derivation of measures from the survey. Accordingly, the present research aimed to address these challenges by using a variety of survey designs. To begin with, we systematically examined item formulation in employee surveys. For this, we used intensifiers in the item stem to counteract the potential effect of high agreement rates. In addition to response behavior, employees' subjective perceptions are relevant to the challenge of survey fatigue. Surveys usually only aim for participation-centered and quality-centered evaluation criteria. In our research, we extended the survey criteria to experience-centered criteria (survey experience, SX). To foster a better understanding of survey quality, we systematically developed and validated the SX short scale. In a further study, the SX short scale helped to validate the study on a used gamification element, namely instant feedback, and to highlight the incremental benefits for individual facets, such as response behavior and "pleasure". Based on the results, implications for employee surveys are derived, concerning both question formulation (e. g. specifically item comparability checks) and survey design (e. g. use of gamification elements). In addition, the research established the basis for further surveys, which are discussed in this paper. Overall, the results show the potential of a well thought-out survey optimization for the survey quality.

Keywords: employee survey, response behavior, survey experience, gamification, intensifier, survey quality

Supervisor: Prof. Dr. Karsten Müller

Generelle Einführung

Organisationen sehen sich dynamischen, komplexen und häufig kompetitiven Rahmenbedingungen ausgesetzt (Zhang et al., 2015). Um unter diesen dynamischen Rahmenbedingungen den langfristigen Unternehmenserfolg sicher zu stellen (Al-Haddad & Kotnour, 2015; Burke & Ng, 2006; von der Gracht et al., 2010), ist es daher umso wichtiger für Unternehmen den konstruktiven Austausch (Grant & Marshak, 2011; Heracleous, 2002) mit ihren Mitarbeitenden zu initiieren. Um diesen Austausch zu fördern, werden laut aktueller Feldforschung zunehmend „employee attitude surveys“ (EAS, dt. Mitarbeitenden-Befragungen) durchgeführt, sodass aktuell 88.5 % der größten Unternehmungen im deutschsprachigen Raum diese umsetzen (2008: 80% Frieg & Hossiep, 2020; Hossiep & Frieg, 2008). Für die USA zeigen Studien, dass über 75% der amerikanischen Unternehmen EAS einsetzen (Kraut, 2006). Das mit EAS verbundene klassische Ziel ist es, die Antworten von Mitarbeitenden zu verschiedenen Arbeitseinstellungen zu sammeln, analysieren, aggregieren und berichten, um so eine Evaluation und vielfältige Veränderungsprozesse zu initiieren (Müller et al., 2021).

Doch bereits während der Befragung zeigen sich im Antwortverhalten der Teilnehmenden Herausforderungen. Dies zeigen Studien (z. B. Braunscheidel et al., 2010; French-Lazovik & Gibson, 1984), in denen sich Trends zu links-schiefen Antwortverteilungen manifestieren, das heißt die Antwortverteilungen sind durch eine hohe Zustimmungsrates charakterisiert (Braunscheidel et al., 2010; French-Lazovik & Gibson, 1984; McCarty & Shrum, 2000; Peterson & Wilson, 1992; Stanton et al., 2002). Dies geht häufig mit einer fehlenden Variabilität in den Daten einher (Keppler-Seid et al., 1980; Lebow, 1982; McCarty & Shrum, 2000; Tanner & Stacy, 1985). Zusätzlich häufen sich Beobachtungen von Teilnehmenden mit möglicher Survey Fatigue (dt. Umfrage-Müdigkeit). Survey Fatigue ist charakterisiert durch eine geringere Bereitschaft und Motivation, an Befragungen teilzunehmen oder sie gewissenhaft zu beantworten (z. B. Porter et al., 2004). Die fehlende Motivation der Teilnehmenden führt folglich über sogenanntes *Satisficing* zu weniger oder inakkurateren Antworten (Krosnick, 1991). Dabei versteht sich *Satisficing* als lückenhafte oder flüchtige Beantwortung der Fragen (Krosnick, 1991). Die daraus resultierenden Antwortmuster oder die fehlende Variabilität verursachen beispielsweise Herausforderungen für Zusammenhangsanalysen (Vance, 2006). Letztlich bedeutet dies, dass inakkurate Daten und eingeschränkte Variabilität die statistischen Analysen erschweren und somit die Nützlichkeit von abgeleiteten Schlussfolgerungen in der Praxis und Forschung mindern (Lebow, 1982). Entsprechend wird deutlich, dass ohne eine geeignete Intervention

zur Erhöhung der Compliance beziehungsweise Minderung der Survey Fatigue-Effekte und des *Satisficing* die Güte der Befragung und die damit verbundene Aussagekraft von Ergebnissen maßgeblich beeinträchtigt sein kann.

Als Interventionen zur Gewährleistung der Güte einer Befragung, trotz zunehmend beeinträchtigter Compliance, bieten sich verschiedene Gestaltungsaspekte innerhalb von Befragungen an. Dazu lassen sich insbesondere die **Itemgestaltung**, die **Fragebogengestaltung** sowie der **Befragungsprozess** als Interventionsmöglichkeiten identifizieren.

Alle drei Gestaltungsebenen liefern Ansatzpunkte zur Verbesserung der Befragungsgüte. Die jeweilige Wirkung der Gestaltungsaspekte auf die Befragungsgüte lässt sich über verschiedene Kriterien überprüfen. Zur Differenzierung der Befragungsgüte werden in der Literatur üblicherweise zwei zentrale Arten von Gütekriterien herangezogen: „*quality-centered*“ Gütekriterien und „*participation-centered*“ Gütekriterien. Dabei beziehen sich *quality-centered* Gütekriterien auf die statistische Qualität der Messung und umfassen somit unter anderem Reliabilität, Validität (z. B. Schriesheim et al., 1989; Tourangeau et al., 2004), soziale Erwünschtheit (z. B. Heerwegh et al., 2005; Toepoel et al., 2008) und Messäquivalenz (Steenkamp & Baumgartner, 1998). Zusammengefasst erfassen *quality-centered* Gütekriterien unter anderem somit indirekt die Intensität der kognitiven Verarbeitung der Teilnehmenden. Demgegenüber stellen *participation-centered* Gütekriterien eher die individuelle Entscheidung an Befragungen teilzunehmen sowie das damit assoziierte Verhalten dar. Hierunter fallen vornehmlich Kriterien wie die Teilnahme-Quoten (z. B. Cobanoglu et al., 2001; Heberlein & Baumgartner, 1978; Manfreda et al., 2008) oder Item-Non-Response (dt. Nichtbeantwortung von Fragen; z. B. Funke, 2016; Peytchev et al., 2006). Neben diesen klassischen Bewertungskriterien von Gestaltungsaspekten von Befragungen zeigt sich, dass *experience-centered* Gütekriterien bislang wenig Aufmerksamkeit gewidmet wurde. Bei diesen Gütekriterien handelt es sich vornehmlich um die Wahrnehmung der Teilnehmenden im Zusammenhang mit der Befragung und die damit verbundenen Erlebnisse. Im Rahmen der Mensch-Computer-Interaktion ist die User-Experience als Teil der *experience-centered* Gütekriterien bereits ein häufig verwendetes und etabliertes Konstrukt zur Evaluation der Nutzerfreundlichkeit von Systemen (Hassenzahl, 2008; Law et al., 2009).

Positive Veränderungen der Gütekriterien durch Anpassungen in der Itemgestaltung, der Fragebogengestaltung oder der Befragungsprozessgestaltung zeigen sich in verschiedenen Studien (Cechanowicz et al., 2013; z. B. Dedrick et al., 2007; Iglesias & Torgerson, 2000). In Bezug auf die **Itemgestaltung** bestätigen Studien den exemplarischen Einfluss der

Itemformulierung auf die unterschiedlichen *quality-centered Gütekriterien* (Bishop et al., 1978; Dedrick et al., 2007; Dudycha & Carpenter, 1973). Zum Beispiel berichten McPherson und Mohr (2005) einen Unterschied von extrem formulierten im Vergleich zu moderat formulierten Items in Bezug auf die Zustimmungswahrscheinlichkeit. Genauer zeigt sich der Effekt in der Form, dass Teilnehmende extremen Aussagen (z. B. „Ich betrachte die Dinge immer von der positiven Seite.“) weniger oft zustimmten als moderaten (z. B. „Ich betrachte die Dinge meist von der positiven Seite.“). Demzufolge ließen sich die inhaltlichen Veränderungen im Itemstamm durch Adjektive oder adverbiale Bestimmungen als Gegenmaßnahme für links-schiefe Verteilungen und die hohe Zustimmungsrate in Befragungen identifizieren. Je nach Anwendungsfall werden diese auch als Qualifier (qualitätsbezogen), Intensifier (intensitätsbezogen) oder auch Quantifier (frequenzbezogen oder zeitbezogen) bezeichnet.

Auch in Bezug auf die **Fragebogengestaltung** zeigen sich Möglichkeiten, durch insbesondere Verkürzung der Umfrage, ansprechendes Erscheinungsbild, sorgfältig gewählte Themen und Auswahl des Befragungsmediums die Teilnahmebereitschaft der relevanten Stichprobe zu erhöhen. Im Detail lässt sich der Einfluss der Fragebogengestaltung insbesondere auf die *participation-centered Gütekriterien* in vielen Studien finden (z. B. Walston et al., 2006). Exemplarisch zeigen Iglesias & Torgerson (2000), dass längere Befragungen niedrigere Antwortraten bedingen. In ihrem quasi-randomisierten Experiment waren die Rücklaufquoten zu zugesandten Fragebögen in kürzeren Befragungen von circa 4 Seiten (49%) signifikant höher als in längeren Befragungen mit 7 Seiten (40%). Insofern kann auch die optimale Fragebogengestaltung zur Erhöhung der Befragungsgüte beisteuern.

Über die Item- und Fragebogengestaltung hinaus ist auch innerhalb des **Befragungsprozesses** ein Effekt auf Verhaltensweisen und psychologische Zielvariablen möglich. Innerhalb des Prozesses können spezifische motivationale Anreize die Befragungsgüte verändern (Hamari et al., 2014). Dadurch werden auch direkt bzw. vermittelt Gütekriterien beeinflusst. Konkret zeigen Keusch und Zhang (2017) in ihrem Review bereits Zusammenhänge zwischen motivationalen Anreizen im Befragungsprozess in Form von spielerischen Elementen und *quality-centered* und *participation-centered Gütekriterien* zur Befragungsgüte auf. Diese Verwendung von spielerischen Elementen wird dabei auch als Gamification bezeichnet (Deterding et al., 2011). Exemplarisch zeigt die Studie von Cechanowicz und Kolleg:innen (2013) die positiven Effekte von Gamification auf die Teilnahmebereitschaft. Die Studie nutzt bewusste spielerische Elemente, wie beispielsweise Belohnungen, Herausforderungen und Feedback, um drei unterschiedlich stark gamifizierte

Befragungen (normale Befragung, teils spielerische Befragung, voll spielerische Befragung) zu erstellen. Der Vergleich der Bedingungen zeigt eine steigende Anzahl beantworteter Items durch die Hinzunahme von interaktiven spielerischen Elementen. Da es jedoch ebenfalls spielerische Elemente (z. B. Slogan Matching, d.h. eine spielerische Zuordnung von Slogans zu bestimmten Markenzeichen; Cechanowicz et al., 2013) gab, die zu keiner Verbesserung der korrekten Antworten führten, zeigt die Studie zeitgleich, dass spielerische Interaktion innerhalb des Befragungsprozesses die Datenqualität verbessern kann, aber nicht muss. In der Gesamtschau ist der Befragungsprozess bei adäquater Umsetzung dennoch ein guter Ansatzpunkt für die positive Beeinflussung der Befragungsgüte. Gleichzeitig sollte er sorgfältig untersucht werden.

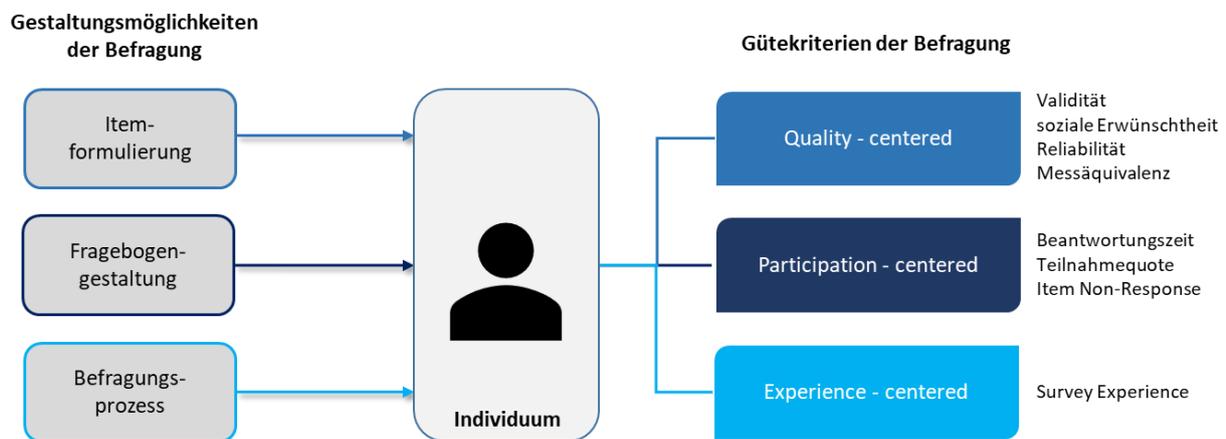


Abbildung 1. Prozessmodell der Einflussmöglichkeiten von Befragungsgestaltung auf die Gütekriterien der Befragung.

Um die Herausforderungen von EAS, die insbesondere in den zwei Facetten Antwortschiefe und Motivation liegen, zu adressieren, ergeben sich somit drei wesentliche Gestaltungsansätze: Item-Gestaltung, Fragebogengestaltung und der Befragungsprozess. Entsprechend behandelt die folgende Dissertation die Frage der optimalen Befragungsgestaltung zur Sicherung der Befragungsgüte in allen relevanten Facetten. Um die zwei Herausforderungen der schiefwinkligen Antwortverteilungen und der Motivationsverluste zu bewältigen, werden mögliche Ansatzpunkte und Strategien in den verschiedenen Befragungselementen identifiziert. Zunächst betrachtet die vorliegende Arbeit die Steigerung der Datenvariabilität durch die Anpassung von Itemformulierungen. Dabei kommen Intensifier im Itemstamm zum Einsatz.

Bis dato wurden Intensifier hauptsächlich im Kontext von Antwortskalen erforscht und inkonsistent verwendet (z. B. French-Lazovik & Gibson, 1984; Lam & Klockars, 1982;

Lam & Stevens, 1994; Weijters et al., 2013; Wildt & Mazis, 1978; Wyatt & Meyers, 1987). Daher stellt die vorliegende Forschungsarbeit einen wichtigen Beitrag zur Evaluation von unterschiedlichen Itemformulierungen sowie zur besseren retrospektiven Beurteilung der Vergleichbarkeit von Items im Rahmen des Benchmarkings dar. Gleichzeitig profitiert auch die Praxis, indem Intensifier zukünftige Fragebogenentwicklungen hinsichtlich der Befragungsgüte erleichtern. Im Anschluss werden die *quality-centered* und *participation-centered* Gütekriterien um ein *experience-centered* Gütekriterium erweitert. Dazu wird die „Survey-Experience“ (SX) - Skala aus theoretischen Grundlagen konzeptualisiert und empirisch validiert. Da bislang in diesem Bereich bis auf wenige Ausnahmen lediglich ad-hoc Skalen oder Einzelitems verwendet wurden (z. B. Couper et al., 2004; Walston et al., 2006; Yentes et al., 2012; Zukerberg et al., 1999), liefert die vorliegende Arbeit mit der SX - Skala ein dringend benötigtes einheitliches und validiertes Messinstrument für Forschung und Praxis. Im Kontext der EAS wird die SX verwendet, um psychologische Facetten, wie Freude und Benutzerfreundlichkeit, zu erheben. In diesem Rahmen wird die neuentwickelte SX - Skala in einer weiteren Studie genutzt, um Effekte von spielerischen Elementen (= Gamification) in EAS zu untersuchen. Damit folgen wir dem Aufruf von Keusch und Kolleg:innen (2020) zu einer genaueren Erforschung einzelner, spezifischer Gamification-Elemente. Spezifischer stellt sich in dieser Studie die Frage, ob sich neben der Survey Experience auch qualitätsbezogene Gütekriterien durch Gamification verändern lassen. Durch den erstmaligen Einsatz von spielerischen Elementen im Kontext einer Mitarbeiterbefragung in der Praxis trägt die experimentell-randomisierte Studie gleichzeitig zur Überprüfung der theoretischen Wirkmodelle bei. Insgesamt unterstützt diese Forschung das grundlegende Verständnis der Befragungsgüte und hilft mit möglichen Ansatzpunkten und Strategien Praktikern und Forschern in der Befragungsgestaltung.

Die folgenden Abschnitte werden sich jeweils den beiden dargestellten Herausforderungen - hohe Zustimmungstendenz und geringe Motivation während der Beantwortung - widmen. Dabei werden spezifische Gegenmaßnahmen zur Datenqualitätserhöhung identifiziert. Zur Exploration der potentiellen Gegenmaßnahmen erfolgt im darauffolgenden Abschnitt die Darstellung der durchgeführten Studien. Im Rahmen der anschließenden Diskussion werden die Ergebnisse der Studien aggregiert betrachtet und die theoretischen und praktischen Implikationen aus einer ganzheitlichen Perspektive bewertet. Zusätzlich werden mögliche Limitationen und Impulse für zukünftige Forschung diskutiert.

Trends in Employee Attitude Surveys

Wie oben ausgeführt zeigen Studien zu EAS einen zunehmenden Trend an links-schiefen Verteilungen (Braunscheidel et al., 2010; French-Lazovik & Gibson, 1984; McCarty & Shrum, 2000; Peterson & Wilson, 1992; Stanton et al., 2002). Dabei zeigt eine links-schiefe Verteilung eine hohe Zustimmung der Teilnehmenden in allen positiven Items (z. B. Braunscheidel et al., 2010; French-Lazovik & Gibson, 1984). Speziell die klassischen Themen der EAS, wie beispielsweise Arbeitszufriedenheit, Entwicklung und Beförderungen (Peterson & Wilson, 1992; Stanton et al., 2002), Empowerment, organisationales Commitment, Autonomie und Work-Life Balance (Church & Waclawski, 2017), subjektives Wohlbefinden (Diener, 1984), Gesundheit (Griffin et al., 2000) und Werte, werden dabei aufgrund der positiven Konstrukte weniger differenziert, häufig eher positiv wahrgenommen beziehungsweise links-schief beantwortet (McCarty & Shrum, 2000). Im konkreten Kontext der EAS konnten Eid und Rauber (2000) zeigen, dass 29% der Teilnehmenden eine Beantwortung mit extremer Antwortverzerrung aufwiesen. Daher stellen Tanner und Stacy (1985) die Allgemeingültigkeit von Umfrageergebnissen aufgrund von inflationären Werten in Frage. Stark positive Antwortmuster erschweren aufgrund der geringen Variabilität gleichzeitig für Unternehmen die Verarbeitung und Interpretation der Daten (Keppler-Seid et al., 1980; Lebow, 1982). Zusätzlich ist die Überprüfbarkeit der Reliabilität durch eine links-schiefe Verteilung limitiert (McCarty & Shrum, 2000; Tanner & Stacy, 1985). Andere statistische Analysen stoßen ebenfalls an ihre Grenzen, da auch sie auf eine ausreichende Varianz und Normalverteilung der Antworten angewiesen sind (Field, 2009; Howell, 2010). Dementsprechend lassen sich im Kontext der EAS bei ausschließlich positiven Bewertungen beispielsweise schwerlicher Verbindung oder Korrelationen zu anderen Leistungsindikatoren analysieren. Dadurch verringert sich ebenfalls der Nutzen der Daten und der Schlussfolgerungen (Lebow, 1982; Vance, 2006). Gerade für die Interpretation und Arbeit mit den Ergebnissen sowie die Ableitung von Maßnahmen im Folgeprozess der EAS kann dies zu Einschränkungen führen.

Qualifier als Interventionsmöglichkeit

Insofern gilt es Design-Elemente zu identifizieren, die den gegensätzlichen Fall von geringeren Schiefen und eine höhere Differenzierung in den Antworten begünstigen. In der Itemkonstruktion werden aufgrund ihrer Ökonomie in der Regel geschlossene Frageformate in der Auswertung häufiger verwendet als offene (Krosnick & Presser, 2010). Die geschlossenen Fragen in EAS werden typischerweise im Likert-Type (Likert, 1932) genutzt (Chyung et al., 2017; Joshi et al., 2015; Low, 1996; Purdey, 2013). Dabei fragen Likert-Type

Items gemeinhin den Grad der Zustimmung beziehungsweise der Ablehnung zu einer Aussage ab (Krosnick & Presser, 2010). Zum Einsatz kommt hierbei häufig eine 5-Punkte Skala (Borg & Mastrangelo, 2008; Joshi et al., 2015; Krosnick & Presser, 2010). Innerhalb der Items werden häufig verbale Anker benutzt, die meist im Antwortformat und teils zusätzlich im Itemstamm vorzufinden sind (z. B. 1 = stimme nicht zu, 5 = stimme zu; Rohrmann, 2007).

Eine Sonderform von verbalen Ankern sind *qualifier*. Diese verändern die Intensität eines Wortes oder Satzes, auf den sie sich beziehen (Cliff, 1959; Low, 1996). Diese Dissertation bezieht sich dabei auf das folgende Verständnis von Qualifiern: Qualifier sind in der Grammatik ein Wort oder eine Wortgruppe, das/die die Bedeutung eines anderen Wortes oder einer anderen Wortgruppe einschränkt oder verstärkt wie z. B. ein Adjektiv oder Adverb. Qualifier werden häufig in das Antwortformat von Befragungen integriert (z. B. *fully disagree* - *rather disagree* - *neutral* - *rather agree* - *fully agree*; Cools et al., 2006; Hofmans et al., 2007; Rohrmann, 2007) und können gleichzeitig auch im Itemstamm verwendet werden (z. B. "I am extremely satisfied with my job", C. A. O'Muirheartaigh et al., 1993; Wright et al., 1995). Dabei unterscheidet sich die Nutzung von Qualifiern sowohl innerhalb als auch zwischen Skalen beziehungsweise Befragungen. Beispielsweise können sie inkonsistent (vorhanden vs. nicht vorhanden) verwendet werden, wie in der Utrecht Work Engagement Scale (Schaufeli & Bakker, 2004): Dort werden in drei von 17 Items Qualifier verwendet (bspw. "I can continue working for *very* long periods at a time"; "At my job, I am *very* resilient, mentally"; "At my work I *always* persevere, even when things do not go well", p. 5). In den übrigen Items werden sie hingegen nicht eingesetzt.

Studien zeigen, dass bereits kleine Umformulierungen im Itemstamm das Antwortverhalten beeinflussen können (z. B. Bishop et al., 1978; Dedrick et al., 2007; Dudycha & Carpenter, 1973; Schuman & Presser, 1977; Schwarz, 1999; Sudman & Bradburn, 1982; Wright et al., 1995). Statt mehr oder weniger komplexe Umformulierungen vorzunehmen, können sogar durch kleinere Item-Ergänzungen Unterschiede erzielt werden. Beispielsweise können hinzugefügte Intensifier die Bedeutung eines Satzes verstärken. Intensifier bilden dabei eine Unterkategorie der Qualifier, die sich insbesondere auf die Intensität einer Beschreibung beziehen (z. B. sehr, extrem, ziemlich). Die dazugehörige Forschung hat bislang häufig Effekte von Intensifiern im Antwortformat untersucht (z. B. French-Lazovik & Gibson, 1984; Lam & Klockars, 1982; Lam & Stevens, 1994; Weijters et al., 2013; Wildt & Mazis, 1978; Wyatt & Meyers, 1987).

Die wenigen bereits existierenden Studien zur Nutzung von Intensifiern im Itemstamm zeigen vielversprechende Ergebnisse im Hinblick auf die Verhinderung von schiefen Antwortverteilungen. So finden Studien beispielsweise, dass die extreme Formulierung von Fragen einen Effekt auf das Antwortverhalten hat (McPherson & Mohr, 2005; Spector et al., 1997). Exemplarisch fanden McPherson und Mohr (2005) in ihrer Studie zu extremen Effekten von Likert-Type Skalen, dass Teilnehmende weniger wahrscheinlich zu extremen (positiven oder negativen) Items zustimmen. Ähnlich berichten auch Dedrick und seine Kolleg:innen (2007) in Studien den Einfluss von Intensifiern aufs Antwortverhalten. In ihrer experimentellen Erhebung zur Einstellung von Lehrenden zu Studierenden mit Behinderung wurden signifikante Mittelwertsunterschiede zwischen den milden (Studierende mit milden Behinderungen), neutralen (Studierende mit Behinderungen) und starken (Studierende mit starken Behinderungen) Bedingungen gefunden. Konkret zeigten sich aufgrund der negativ formulierten Frage höhere Mittelwerte für die neutrale Bedingung im Vergleich zur starken Bedingung. Ebenso hatte die neutrale Bedingung höhere Mittelwerte als die milde Bedingung. Eine weitere Studie von Wright und Kollegen (1995) unterstützt mit ihrer Forschung zur Lebenszufriedenheit die Gültigkeit dieser Ergebnisse. In ihr stimmen Teilnehmende der Frage nach der Lebenszufriedenheit signifikant weniger zu, wenn sie extrem formuliert ist (extrem glücklich) im Vergleich zu den schwächeren Formulierungen (glücklich/sehr glücklich).

Mögliche Erklärungsmodelle für die gefundenen Auswirkungen der Itemgestaltung auf das Antwortverhalten liefern Tourangeau und Rasinski (1988). Dabei postulieren Tourangeau and Rasinski (1988) einen vierschrittigen Prozess der Umfragebeantwortung. Im ersten Schritt bilden die Teilnehmenden ein Itemverständnis. Im anschließenden zweiten Schritt erinnern sie relevante Überzeugungen und Gefühle zum Inhalt des Items. Als Drittes bilden Sie ein Urteil auf der Grundlage der abgerufenen Inhalte und wählen schlussendlich im vierten Schritt eine passende Antwortoption aus. Durch die Verwendung von Intensifiern in Itemstämmen nehmen die Teilnehmenden das Item im Beantwortungsprozess vermutlich als extremer wahr (Schritt 1). Entsprechend passen sie ihre Beurteilung in Schritt 3 an, indem sie weniger wahrscheinlich zustimmen und im Folgenden (Schritt 4) eine weniger starke Antwortoption auswählen. Betrachtet man diesen Erklärungsansatz im Kontext der vorliegenden Studien, so sind Intensifier eine vielversprechende Gestaltungsmöglichkeit, um schiefe Antwortverteilungen zu verhindern.

Die genauere mögliche Einflussnahme mittels Intensifiern hat in den einzelnen Prozessschritten der Fragebogenbeantwortung (Tourangeau & Rasinski, 1988) jedoch auch Einschränkungen, da Intensifier ihr Potential für die Datenqualität nur dann entfalten können,

wenn sie sowohl in Schritt 1 (Wahrnehmung) erkannt als auch in Schritt 3 (Urteilsbildung) nicht vernachlässigt werden. Studien zeigen jedoch, dass in einigen Fällen genau dies passiert (Low, 1996) - nämlich, dass Intensifier im ersten Schritt erkannt, aber für die Beurteilung nicht berücksichtigt werden. Ursächlich für derartige Schwierigkeiten in der Item-Beantwortung kann ein zu hoher kognitiver Aufwand sein, der für das Verstehen und die Beurteilung erforderlich ist (Cannell et al., 1981). Besonders die schwierigen und gleichzeitig komplexen Fragen überfordern die Verarbeitungsmöglichkeiten Einzelner (Lenzner, 2012), begünstigen dadurch Satisficing (Krosnick, 1991) und führen so zu schlechterer Antwortqualität (Lenzner, 2011). Zusätzlich können aber auch Zeitdruck oder motivationale Probleme ursächlich für die fehlende Intensifier-Verarbeitung sein (Krosnick, 1991). Auch das klassische Überlesen aufgrund der geringeren Relevanz für die inhaltliche Fragestellung könnte ein Grund sein und zeigte sich in Studienergebnissen (z. B. Gaskell. et al., 1993).

Zusammenfassend zeigen Qualifier in der Itemgestaltung das Potenzial um links-schiefen Antwortverteilungen entgegenzuwirken und so die Datenqualität zu erhöhen. Zur Entfaltung ihres vollen Potenzials bedarf es weiterer begünstigender Faktoren, die eine adäquate kognitive Verarbeitung der Intensifier im Befragungsprozess ermöglichen, wie insbesondere eine ausreichend hohe Antwortmotivation.

Die Motivation als Herausforderung

Doch wie kann insbesondere die - für die Datenqualität wichtige - Motivation gewährleistet werden. Simple Umfragen sind mittlerweile selbst für Laien einfach und schnell durchführbar (Porter et al., 2004), was ihre Beliebtheit als Instrumente für einen schnellen Einblick in die Meinung der Mitarbeitenden noch erhöht. Der Trend einer stetig zunehmenden Anzahl an Befragungen bestätigt dieses Phänomen (Adams & Umbach, 2012; Karlberg, 2015; Porter et al., 2004). Entsprechend erscheinen Mitarbeiterbefragungen omnipräsent (Adams & Umbach, 2012), sodass steigende Befragungsanzahlen meist mit einer Abnahme von Antwortraten einhergehen (Stedman et al., 2019). Begründet wird dies durch den Stress in Teilnehmenden, den jede Befragung auslöst (z. B. Antwortdruck, soziale Erwünschtheit). Dieser Stress könnte eine Hauptursache für den Trend sinkender Antwortraten und den damit verbundenen Motivationsverlust sein. Studien unterstützen diese Annahme, da sie zeigen, dass die Anzahl an Befragungsanfragen einen Einfluss auf die sogenannte Survey Fatigue hat (z. B. Desinteresse bezüglich der Befragung). Als Folge zeigt sich, teilweise begünstigt durch die Survey Fatigue und mögliche Motivationsverluste, ein Satisficing Verhalten (Zhang & Conrad, 2014). Dabei beschreibt Satisficing eine Art inakkuraten Antwortverhaltens, das durch eine weniger sorgfältige Verarbeitung der vier Schritte zur Fragebeantwortung (d.h.

Verstehen der Aufgabe, Erinnern relevanter Informationen, Beurteilen und Antworten; Tourangeau & Rasinski, 1988) charakterisiert ist. Infolgedessen kann derartiges Antwortverhalten zu unvollständigen, ungenauen oder fehlenden Daten führen (Krosnick, 1991). Da die Anzahl der Befragungsanfragen für ein Individuum schwer beziehungsweise nicht kontrollierbar ist, sind entsprechende Gegenmaßnahmen zur Survey Fatigue wichtig. Da die fehlende Motivation - neben individuellen Fähigkeiten und kognitiven Anforderungen - dabei auch von Krosnick (1991) als potentielle Ursache für inadäquates Antwortverhalten gesehen wird (z. B. Satisficing), stellt sie einen Ansatzpunkt zur Erhöhung der Datenqualität dar. Entsprechend schlussfolgern auch Turner und Kolleginnen (2014), dass Befragungen mehr auf motivationale Aspekte fokussieren sollen. Zielstellung ist dabei, Teilnehmenden eine Befragungsteilnahme mit mehr Freude zu ermöglichen und dadurch dafür Sorge zu tragen, dass sie als teilnahmewürdiger wahrgenommen wird (Turner et al., 2014).

Survey Design Features

Zur Erhöhung der wahrgenommenen Motivation von Teilnehmenden ist die Fragebogengestaltung mittels verschiedener Survey Design Features ein wesentlicher Ansatzpunkt. Das Fragebogendesign ist dabei ein bereits stark erforschter Kontext, der verschiedene Strategien und Indikatoren für erfolgreiche Fragebogengestaltung berücksichtigt. Die vorliegende Dissertation orientiert sich dabei an den existierenden Vorschlägen einer Taxonomie von Rogelberg und Kolleg:innen (2001). Da in der vorliegenden Dissertation insbesondere die Motivation der Teilnehmenden im Fokus steht, ist die Erweiterung um weitere emotionale und einstellungsbezogene Gütekriterien substantiell, um mögliche Ansatzpunkte und Strategien zur wirksamen Fragebogengestaltung zu identifizieren. Neben den typischen Gütekriterien-Kategorien, *participation-centered* und *quality-centered*, soll daher auch die häufig vernachlässigte Kategorie der *experience-centered Gütekriterien* betrachtet werden. Dabei orientiert sich die erste Kategorie der *participation-centered Gütekriterien* meist an der Teilnahmebereitschaft der potentiellen Stichprobe. Hierfür werden in der Literatur meist Indikatoren wie Antwortraten (z. B. Cobanoglu et al., 2001; Heberlein & Baumgartner, 1978; Manfreda et al., 2008), Unit-Non-Response (z. B. Deutskens et al., 2006; Yu & Cooper, 1983), Abbrüche (z. B. Deutskens et al., 2004; Lenzner, 2011; Silber et al., 2013) und Item-Non-Response (z. B. Funke, 2016; Peytchev et al., 2006) herangezogen. Die zweite Kategorie der *quality-centered Gütekriterien* umfasst Parameter, die die Güte der Messung evaluieren. Darunter fallen beispielsweise Validität und Reliabilität (z. B. Schriesheim et al., 1989; Tourangeau et al., 2004), soziale Erwünschtheit (z. B. Heerwegh et al., 2005; Toepoel et al., 2008), Insufficient Effort

Responding (z. B. Ran et al., 2015; Ward & Pond, 2015; Weijters & Baumgartner, 2012), die Bearbeitungszeit (z. B. Buskirk, 2015; Fricker et al., 2005; Toepoel et al., 2009), die Messäquivalenz (z. B. Christian & Dillman, 2004; Fricker et al., 2005; Redline & Lankford, 2001) oder auch korrekt ausgeführte Aufgaben beziehungsweise die Aufmerksamkeit der Teilnehmenden (z. B. Couper et al., 2001; Jones & Lang, 1980). Die neu integrierte dritte Kategorie, die in der Forschung bislang wenig beachtet wurde, beinhaltet *experience-centered Gütekriterien*, die sich auf das Erleben der Teilnehmenden während der Befragung konzentrieren. Frühere Studien zeigen, dass Survey Design Features Teilnehmende und ihre gemachten Erfahrungen beeinflussen können (Christian & Dillman, 2004; Couper et al., 2001; Kaczmirek, 2008, 2018; Manfreda et al., 2002). Den Mehrwert solcher Survey Design Features zeigen bereits einige Studien, die auf die generelle (z. B. Goyder, 1986; Helgeson & Ursic, 1994; Rogelberg et al., 2001; Schleifer, 1986; Sjoberg, 1954; Walker & Cook, 2013) oder spezifische Einstellung (z. B. Couper et al., 2004; Manfreda et al., 2002; Peytchev et al., 2006; Walston et al., 2006; Yentes et al., 2012; Zuckerberg et al., 1999) gegenüber Befragungen abzielen. Bei der Betrachtung der bisherigen Messinstrumente zur Erhebung dieser Einstellungen lässt sich exemplarisch der Fragebogen von Rogelberg und Kolleg:innen (2001) nennen mit den darin enthaltenen Faktoren *survey enjoyment* (d.h. „Ich fülle gerne Umfragen aus“) und *survey value* (d.h. „Umfragen sind ein nützliches Mittel zur Sammlung von Informationen“).

Bei der Betrachtung dieser exemplarischen Faktoren wird deutlich, dass es sich lediglich um persönliche Einstellungen zu Befragungen im Allgemeinen handelt. Entsprechende direkte Rückschlüsse auf die erfolgreiche Gestaltung spezifischer Umfragen lassen sie dabei jedoch nicht zu. Es bedarf daher unbedingt eines weiteren Messinstruments, das die Überprüfung des Erlebens der Teilnehmenden in Bezug auf eine spezielle Befragung - statt der allgemeinen Einstellung – ermöglicht.

Die bisherigen *experience-centered Gütekriterien* in Befragungen waren geprägt von Einzelfragen zur Zufriedenheit mit der Umfrage (Couper et al., 2004; Peytchev et al., 2006), offenen Kommentaren (Manfreda et al., 2002; Peytchev et al., 2006) und semantischen Differentialen (Walston et al., 2006; Zuckerberg et al., 1999). Natürlich haben derartige ad-hoc Messmethoden mit fehlender theoretischer Basis in gewisser Weise ihre Legitimation. Dennoch erfordert zukünftige Forschung eine validierte und theoretisch fundierte Skalenentwicklung. Entsprechend lieferten Yentes und Kolleg:innen (2012) in ihrer Studie mit Hilfe von Faktorenanalyse erste Ansatzpunkte für eine validierte Messung der SX. Als Resultat konnten die Faktoren „*survey enjoyment*“ mittels 12 Fragen und „*survey focus*“ mit

3 Fragen erhoben werden. Trotz des Erfolgs, diese Skala auch mit weiteren Indikatoren wie Teilnehmerleistung in Verbindung zu setzen, fehlte jedoch eine im Vorfeld definierte theoretische Basis oder Hypothese zur Struktur dieser Faktoren. Um diese Forschungslücke zu schließen, gilt es in einem deduktiven Ansatz das Konstrukt der SX zu finden.

Mit Hilfe dieses entwickelten Konstrukts lassen sich anschließend verschiedene spezifische Design-Features evaluieren, indem sich ihre Auswirkungen auf SX abbilden lassen. Dabei lassen sich auf Basis der bestehenden Literatur verschiedene Oberkategorien für Design-Features aggregieren (Kommunikation, Umfragelänge, Item-Konstruktion und Erscheinungsbild). Die Effekte auf die SX lassen sich aufgrund bestehender Forschungsarbeiten annehmen (z. B. Cook et al., 2000; Couper et al., 2001; Deutskens et al., 2004; Dillman, 1991, 2000; Dillman et al., 1993; Funke, 2016; Funke et al., 2011; Garcia et al., 2004; Heerwegh et al., 2005; Jones & Lang, 1980; Krosnick & Presser, 2010; Lietz, 2010; Manfreda et al., 2002; O’Cathain & Thomas, 2004; C. O’Muircheartaigh et al., 2000; Peytchev et al., 2006, 2006; Rogelberg et al., 2001; Rolstad et al., 2011; Tourangeau et al., 2004; Tourangeau & Rasinski, 1988; Villar et al., 2013; Walston et al., 2006, 2006; Whelan, 2015; Yentes et al., 2012; Zuckerberg et al., 1999). Eine zusätzliche neue Kategorie, die sich in der letzten Zeit herauskristallisiert, sind die verschiedenen Gestaltungsmöglichkeiten durch Gamification.

Insbesondere da SX ein Konstrukt aus dem Bereich der *experience-centered* Gütekriterien ist, welches auch mit *quality-centered* und *participation-centered* Gütekriterien zusammenhängen kann (z. B. Hassenzahl et al., 2003; Hassenzahl & Tractinsky, 2006; Yentes et al., 2012), bietet es sich an, auch im Kontext der Gamification die SX einzusetzen, um die positiven Einflüsse der Gamification nachvollziehen zu können.

Spezieller Survey Design-Lösungsansatz: Gamification

Das Review von Keusch und Zhang (2017) zeigt die Veränderung von auf psychologische und verhaltensbezogene Konstrukte innerhalb der Person durch die Verwendung von Gamification-Elementen in Befragungen. Diese psychologischen und verhaltensbezogenen Konstrukte weisen eine Ähnlichkeit zu *experience-*, *quality-* und *participation-centered Gütekriterien* auf. Generell umfassen Gamification-Elemente unter anderem Regeln, Zeitchallenges, Bildmaterial, Frageformulierungen, Fortschrittsbalken, Avatare, Punkte, Abzeichen, Geschichten, Ziele, Herausforderungen, Belohnungen und Feedback (Hamari et al., 2014; Keusch & Zhang, 2017). Bei genauerer Betrachtung der genannten Elemente zeigt sich, dass sich diese in der Komplexität und der Schwierigkeit der Befragungsimplementierung unterscheiden. Avatare und Geschichten benötigen

beispielsweise eine eher komplexe Integration im Vergleich zu einfach umzusetzenden Elementen wie Punkten oder Feedback. Das dazugehörige Verständnis über die Nutzung mehrerer Gamification-Elemente zur Schaffung einer ansprechenden Teilnehmenden-Erfahrung schaffen bereits einige Studien (z. B. Downes-LeGuin et al., 2012; Mekler et al., 2013). Im Detail zeigen sich neben Freude, Interesse und Zufriedenheit mit der Umfrage bessere Effekte für das selbstberichtete Engagement und die Motivation, wenn mehrere Gamification-Elemente verwendet werden (Downes-LeGuin et al., 2012; Mekler et al., 2013). Dadurch verbessert die entsprechend positivere Befragungserfahrung nachweislich die Datenqualität (Downes-LeGuin et al., 2012; Mekler et al., 2013; Puleston & Sleep, 2011). Speziell die Datenqualität als Schlüsselfaktor für den Erfolg von EAS sollte durch die Gamification nicht verschlechtert werden.

Die verwendeten Gamification-Elemente oder „motivational affordances“, wie Hamari und Kolleg:innen (2014) sie in ihren Studien nennen, können bei Teilnehmenden zu Motivationssteigerungen führen. Ein Indiz dafür können beispielsweise mehr beziehungsweise längere Antworten und damit eine verbesserte Auseinandersetzung mit der Befragung sein (Harms, Biegler, et al., 2015). Zusätzlich berichten Studien über erhöhte Zufriedenheit und Freude an der Befragung (Harms, Biegler, et al., 2015; Mekler et al., 2013; Triantoro et al., 2019). Des Weiteren beeinflussen sie weitere psychologische Konstrukte und haben dadurch auch verhaltensbezogene Auswirkungen (Hamari et al., 2014). Erfasste psychologische Konstrukte sind dabei häufig: Vergnügen, Engagement, Spaß und Motivation. Diese psychologischen Konstrukte vermitteln eine mögliche Veränderung des Verhaltens, das sich in diesem Modell meist auf die Qualität und Quantität der Antwort und die Teilnahmebereitschaft bezieht. Daher sollten nicht nur Verbesserungen in der Datenqualität genau betrachtet werden. Auch mögliche unerwünschte Effekte im Antwortverhalten (z. B. Assimilierung oder Kontrast-Effekte Couper et al., 2007) sollten sorgfältig überprüft werden, um eine künstliche Manipulation durch Gamification-Elemente auszuschließen. Trotz der Eingängigkeit des Modells bedarf es der Berücksichtigung zweier Punkte: Als Erstes sind die analysierten Studien nicht ausschließlich aus dem Bereich der Fragebogenforschung, sondern aus einem breiteren Kontext. Zweitens sind ungeachtet der positiven Auswirkungen auf den Befragungsprozess (z. B. Harms et al., 2014) Vergleiche der Mittelwerte nicht ohne Weiteres anwendbar, da Gamification parallel auch positive Effekte induzieren kann.

Viele Fragen bleiben noch unbeantwortet, da die Studien im Kontext der Gamification entsprechend häufig lediglich deskriptive Statistiken berichten oder aber Effekte der Gamification-Elemente untereinander konfundiert sind (z. B. Link et al., 2014; Malinoff &

Puleston, 2011; Puleston & Sleep, 2011; Triantoro et al., 2019). Folgerichtig fordern Keusch und seine Kolleg:innen (2020) daher mehr experimentelle Studien, die spezifische Gamification-Elemente und ihre Effekte untersuchen. Aus der holistischen Perspektive ist für die Bewertung der Befragungsgüte von gamifizierten Befragungen ein multidimensionaler Ansatz wichtig. Zwar sind *quality-centered* und *participation-centered Gütekriterien* wichtige Gütekriterien – sie haben jedoch nicht das Potenzial, das explizite Erleben der Teilnehmenden während der Befragung abzubilden. Um ein ganzheitliches Bild zu erhalten, ist daher die Berücksichtigung von *experience-centered Gütekriterien* erforderlich.

Die vorgestellte Forschung

Diese Dissertation besteht insgesamt aus drei Studien, die jeweils einen wichtigen Beitrag zur aktuellen Forschung und Praxis im Rahmen der Umfrageforschung leisten. Ausgehend von der Bedeutung akkurater Daten im Kontext der EAS für die Ergebnisinterpretation sowie der resultierenden Maßnahmenableitung adressieren die vorliegenden Studien die Gestaltungsmöglichkeiten innerhalb von Befragungen zur Verbesserung verschiedener Qualitätskriterien. Dabei sollen Veränderungen – insbesondere durch Itemformulierungen und Gamification – das Antwortverhalten und das Erleben von Teilnehmenden während der Befragung verbessern. Dabei fokussiert jede Studie einen relevanten Teilaspekt des EAS-Beantwortungsprozesses. In der Zusammenschau der Studien ergibt sich so ein breites Portfolio an einfachen Interventionen, um Befragungsdaten akkurater und besser zu gestalten. Figur 2 zeigt die verschiedenen Ansatzpunkte der Studien innerhalb des Gesamtkonzepts von Befragungen und deren Datenqualität.

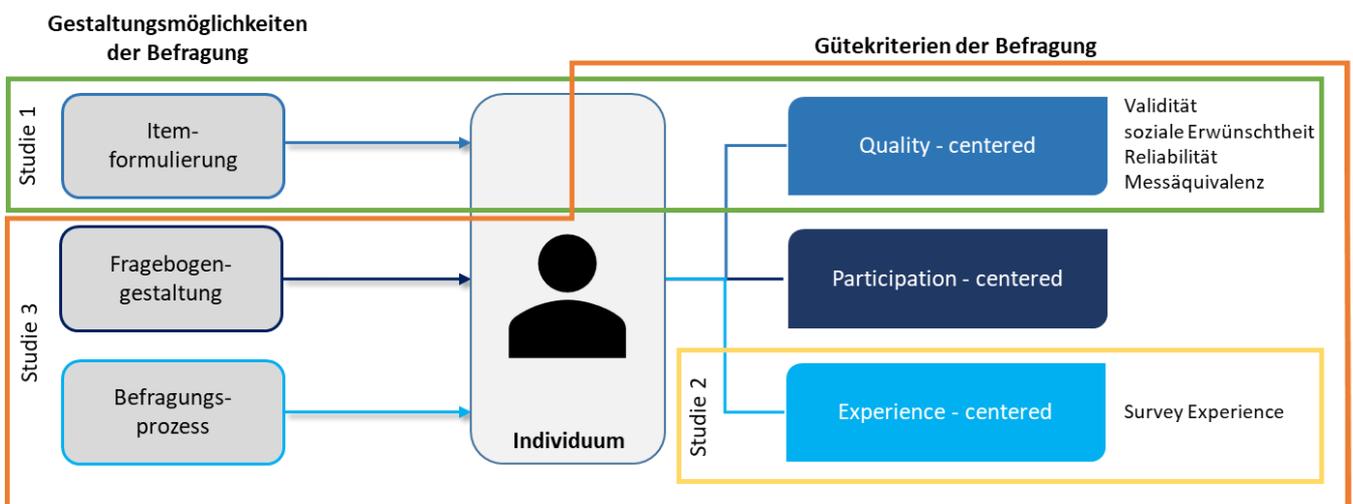


Abbildung 2. Zusammenfassung der durchgeführten Studien

Die **Studie 1** erforscht die Effekte der Itemgestaltung im Kontext der EAS. Herausfordernd sind in diesem Kontext häufig die links-schiefen Verteilungen, die mit einer verringerten Variabilität in den Daten einhergehen. Diese verringerte Variabilität bedingt auch, dass potentielle Informationsquellen nicht vollständig ausgeschöpft werden können (z. B. Keppler-Seid et al., 1980; Lebow, 1982; Tanner & Stacy, 1985). Daher werden die konkreten Einflussmöglichkeiten von Intensifiern im Itemstamm über drei verschiedene Experimente untersucht. Innerhalb der Experimente wurde dabei sowohl die Intensität als auch die Darstellung manipuliert. Wenn starke und saliente Intensifier den Ausdruck verstärken und wahrgenommen werden, sollte die Zustimmung zu den entsprechenden Items eher sinken (Nye et al., 2010; C. A. O’Muircheartaigh et al., 1993). Dies sollte sich ebenfalls in Schiefe, aber auch in Mittelwert, Varianz und Korrelationen mit anderen Kriterien widerspiegeln. Die Ergebnisse der Untersuchung liefern wichtige Erkenntnisse für die zukünftige (konsistente) Nutzung von Intensifiern in der Forschung und Praxis und wird als möglicher Ansatzpunkt gesehen, um die links-schiefe Antwortverteilung bei EAS zu reduzieren.

Die **zweite Studie** schafft eine theoretische Grundlage für die Arbeit an Motivationsverlusten, indem ein valides Messinstrument zur Erfassung von SX entwickelt wird. Die bisherige Forschung konzentrierte sich hauptsächlich auf participation- und quality-centered Gütekriterien von Befragungen, vernachlässigte aber sehr häufig die subjektive Wahrnehmung der Teilnehmenden. Falls diese doch erfasst wurde, erfolgte die Erhebung lediglich durch ad-hoc Skalen zur Zufriedenheit mit der Befragung (z. B. Couper et al., 2004; Walston et al., 2006; Yentes et al., 2012; Zuckerberg et al., 1999) ohne wissenschaftliche Fundierung. Entsprechende Validierungsversuche sind eher selten (Yentes et al., 2012). Daher ist die theoretisch-fundierte Entwicklung der SX-Kurzskala ein wichtiger Schritt um zukünftig Befragungen aus der Wahrnehmung von Teilnehmenden ökonomisch bewerten zu können. Zusätzlich wurde die Wirkung weiterer Design-Features überprüft (z. B. Ansprache, Grafiken, Fortschrittsanzeigen). Die resultierenden Validitätsgewinne und Erkenntnisse erleichtern Forschern und Praktikern den Einfluss von Fragebögen auf die Wahrnehmung der Teilnehmenden zu erfassen und zu beeinflussen. Der Einsatz der SX-Skala ist insbesondere von praktischer Bedeutung, da mit ihrer Hilfe Ansatzpunkte für die Befragungskonzeption zur Erhöhung der Teilnehmendenmotivation identifizierbar sind.

In **Studie 3** nutzen wir die empirisch validierte SX-Kurzskala aus Studie 2, um den Einfluss von Gamification auf die Teilnehmenden zu erfassen, und erheben zugleich *quality-centered* Gütekriterien (z. B. Ankreuzverhalten, Item-Non-Response). Hier wird in einem

randomisierten Feldexperiment innerhalb einer Mitarbeiterbefragung Instant Feedback als leicht einzubauendes Gamification-Element erprobt. Da in der Forschung häufig mehrere Gamification-Elemente parallel untersucht wurden (Bailey et al., 2015; Cechanowicz et al., 2013; Downes-LeGuin et al., 2012; Harms, Biegler, et al., 2015), kann durch die Einzelbetrachtung eine Konfundierung von Effekten ausgeschlossen werden. Auf dieser Basis lassen sich in Forschung und Praxis die Datenqualität verbessern und Impulse für neue Forschung setzen.

Zusammengefasst adressieren diese drei Studien direkt und indirekt die Steigerung der Datenqualität von Befragungen. Dabei schließen sie existierende Forschungslücken durch die Entwicklung von leicht nutzbaren Interventionen. Sie zeigen mögliche Einflussfaktoren, wie Qualifier (Studie 1) und Gamification (Studie 3), und den Bedarf an validen Messinstrumenten für die Wahrnehmung von Teilnehmenden (Studie 2). Dadurch helfen sie zukünftig, bewusstere Entscheidungen in der Fragebogenforschung zu treffen und dadurch die Qualität wissenschaftlicher Forschung und Praxis in allen Bereichen zu unterstützen.

Artikel 1: Does “very” make a difference? Effects of intensifiers in item stems of employee attitude surveys on response behavior.

Hofschroer*, P.¹, Schumacher*, S. K.¹, & Mueller, K.¹. (under review). Does “very” make a difference? Effects of intensifiers in item stems of employee attitude surveys on response behavior.

¹Department of Industrial and Organizational Psychology, Osnabrueck University, Osnabrueck, Germany

* These two authors contributed equally.

Abstract

Employee attitude surveys are important tools for organizational development. To gain insights into employees' attitudes, surveys most often use Likert-type items. Measures assessing these attitudes frequently use *intensifiers* (e.g., *extremely*, *very*) in item stems. To date little is known about the effects of intensifiers in the item stem on response behavior. They are frequently used inconsistently, which potentially has implications for the comparability of results in the context of benchmarking. Also, results often suffer from left-skewed distributions limiting data quality for which the use of intensifiers potentially offers a remedy. Therefore, we systematically examine the effects of intensifiers on response behavior in employee attitude surveys and their potential to remedy the issue of left-skewed distributions. In three studies, we assess effects on level, skewness and nomological structure. experiment 1 examines the effects of *intensifier strength* in the item stem, while experiment 2 and 3 assess whether *intensifier salience* would increase these effects further. Interestingly, results did not show systematic effects. Future research ideas in regards to item design and processing as well as practical implications for the design of employee attitude surveys are discussed.

Key Words: employee attitude survey, survey design, intensifier, item stem, ceiling effect, left-skewed distribution, intensifier salience, intensifier strength

Artikel 2: Let's make it an experience! Introducing SX as a respondent-centered criterion for successful survey design.

Schumacher, S. K.¹, Hofschroer, P.¹, Hund, P. M.¹, Schoenemann, F.¹, & Mueller, K.¹ (under review). Let's make it an experience! Introducing SX as a respondent-centered criterion for successful survey design.

¹Department of Industrial and Organizational Psychology, Osnabrueck University, Osnabrueck, Germany

Abstract

Surveys play a critical role in research and business practice. Most survey designers use survey evaluation criteria of participation and quality to evaluate survey design. A third important criterion is often neglected: the experience of taking the survey. This study introduces the construct of *survey experience* as an experience-centered measure of successful survey design. We systematically developed a short-scale of survey experience on the theoretical foundation of user experience theory and validated it in two studies. Overall, the short-scale showed good reliability and was associated with other experience-centered survey evaluation criteria but distinct from participation- or quality-centered survey evaluation criteria. Further, we tested differences in survey experience when manipulating survey design features. In the preliminary study, we found significant differences for two design features impacting survey experience namely the number of response categories and the use of graphics in the survey. As the results indicated that survey experience might be influenced by a configuration of design features rather than individual features, we examined the effects of configurations of survey design features on survey experience in the main study. We found enhancement effects for pleasure and overall survey-experience in the corresponding conditions compared to the control conditions. This study contributes to survey design research by drawing on user experience theory to develop the construct and respective measure of survey experience as a third category of survey evaluation criteria for successful survey design. As SX is associated with the willingness to participate in similar future surveys, practitioners may use SX as a possible lever to counter potential adverse effects of survey fatigue. The short-scale provides an economic and reliable way to assess participants' SX and derive design implications and evaluations for future surveys tailored to specific target groups.

Keywords: survey design, user experience, survey experience, attitudes towards surveys, satisfaction with surveys

Artikel 3: Gamified surveys in organizations: Does feedback with an aesthetic design enhance employees' SX and data quality? - A Field Experiment

Hofschroerer, P. ¹, Schumacher, S. K. ¹, Bauhaus, T. ², & Mueller, K. ¹ (under review).
Gamified surveys in organizations: Does feedback with an aesthetic design enhance employees' SX and data quality? - A Field Experiment.

¹ Department of Industrial and Organizational Psychology, Osnabrueck University, Osnabruck, Germany

² Department of Industrial and Organizational Psychology, University of Wuppertal, Wuppertal, Germany

Abstract

To face the challenge of survey participant's decreasing motivation, research considers the improvement of survey experience through survey design. A specific design strategy is the implementation of gamification elements to enhance participants' survey experience as a potential lever for motivation and data quality. Thus, we examined the effects of a gamified survey version in the context of an employee attitude survey of a food retail organization in a randomized field experiment ($N = 735$). Participants were randomly assigned to either a standard or a gamified (i.e., instant feedback in aesthetics design) survey version and their perceived survey experience (i.e., pleasure and usability) as well as resulting data quality (i.e., straightlining and item-nonresponse) were assessed. Results indicate that especially the pleasure of taking a survey improved in the gamified survey version while usability of the survey remained stable. Further, data quality, in terms of less straightlining and item-nonresponse, was enhanced by the implemented gamification elements. The findings are in line with previous research. This study addresses the need identified in previous research to examine the effects of specific gamification elements on participants' survey experience and related outcome variables. The study further looks at the specific context of employee attitude surveys and derives important practical implications for their design. In particular, results show the promise of using instant feedback as a potential remedy for decreasing survey participant motivation in this applied context.

Key Words: gamification, gamified survey, survey experience, field study, experimental design, data quality, respondent behavior.

Diskussion

Der konstruktive Austausch innerhalb von Organisationen ist für den langfristigen Unternehmenserfolg essentiell (Al-Haddad & Kotnour, 2015; Burke & Ng, 2006; von der Gracht et al., 2010). Zum Einsatz kommen dabei neben der persönlichen Kommunikation auch Befragungsinstrumente wie beispielsweise EAS (Frieg & Hossiep, 2020). Das Ziel vieler Befragungsinstrumente ist, das gegenseitige Verständnis zu verbessern. Die Zielerreichung birgt jedoch verschiedene Herausforderungen, die insbesondere eine Interpretation der Befragungsdaten erschweren. Beispielsweise sind bestimmte Antwortmuster - ähnlich wie beispielsweise die Ja-Sage-Tendenz - auch in Befragungen enthalten. Die resultierenden Einschränkungen der Datenqualität beeinflussen die Maßnahmenableitung und stellen so ein Gefahrenpotential für die strategische Ausrichtung und Positionierung von Unternehmen dar. Deshalb soll die vorliegende Forschungsarbeit einen Beitrag zum Verständnis von Befragungsdaten leisten, indem das Konzept der Befragungsgüte umfassend erfasst und mögliche Survey Design Features als Einflussfaktoren identifiziert werden.

Basierend auf der aktuellen Forschung wurden zwei relevante Phänomene ausgewählt, die im Kontext von Befragungen zunehmen: Links-schiefe Antwortverteilungen und fehlende Antwortmotivation. Mit Hilfe der drei vorliegenden Studien sollen diese adressiert werden: Zwei der drei Studien untersuchen konkrete Interventionsmöglichkeiten (d.h. Qualifier und Gamification), um den zwei Herausforderungen gerecht zu werden. In diesem Zusammenhang untersucht die erste Studie Veränderungen beziehungsweise Anpassungen in der Frageformulierung, die die inhaltliche Stärke von Aussagen variieren. Demgegenüber fokussiert die zweite Studie ganzheitliche Gestaltungsmöglichkeiten in Form von Survey Design Features (Gamification) innerhalb von EAS. Um den Erfolg der Interventionen messbar zu machen, stellt die Erweiterung des Befragungsgütemodells um den Aspekt der *experience-centered Gütekriterien* in Studie 2 einen wichtigen Beitrag dar.

Zusammenfassend lassen sich aus den drei Studien Hinweise auf die Wirksamkeit angemessener Interventionen zur Verbesserung der ganzheitlichen Befragungsgüte ableiten. Durch die Adressierung verschiedener Gestaltungs- und Evaluationsaspekte innerhalb der EAS bieten die Studien neue Erkenntnisse zur erfolgreichen Umsetzung von zukünftigen Befragungen. Dadurch liefern sie einen wichtigen Beitrag für die theoretische und praktische Fragebogengestaltung im Anwendungsfeld der EAS. Dieser Beitrag wird im Folgenden diskutiert.

Zusammenfassung der studienspezifischen Ergebnisse und Implikationen

Die zentrale Zielstellung einer jeden Befragung sind verlässliche Ergebnisse. Als Interventionen zur Gewährleistung der Güte einer Befragung - trotz zunehmend beeinträchtigter Compliance - bieten sich verschiedene Gestaltungsaspekte innerhalb von Befragungen an: **Itemgestaltung, Fragebogengestaltung** und der **Befragungsprozess**.

Studie 1 betrachtet zunächst die genauen Effekte von Veränderungen im Itemstamm von häufig genutzten Likert-Type Items auf das Antwortverhalten. Welche Bedeutung dies für die Vergleichbarkeit von inkonsistent verwendeten Intensifiern in Items hat, ist bislang unerforscht. Ähnlich ist der Einfluss auf links-schiefe Verteilungen bislang weitgehend unklar. Daher werden in dieser Studie mit Hilfe von Intensifiern – einer Sonderform der Qualifier – konkrete Aussagen innerhalb von Items in ihrer Intensität systematisch manipuliert und die Auswirkungen auf das Antwortverhalten untersucht. In der Untersuchung sollen Unterschiede in statistischen Kennzahlen wie beispielsweise Mittelwert, Schiefe und Varianz in drei ähnlichen experimentellen Settings sichtbar werden. Inhaltlich wäre somit eine Zustimmung zu positiven extremformulierten Items im Vergleich zu positiven normalformulierten Items weniger wahrscheinlich. Wider Erwarten zeigten die Ergebnisse aus der ersten Untersuchung keine konsistenten Mittelwerts-, Varianz- und Schiefe-Unterschiede durch die unterschiedliche Verwendung von Intensifiern (kein Intensifier, moderater Intensifier und starker Intensifier). Zur Überprüfung der korrekten Verarbeitung der Intensifier auf Basis der Theorien von Krosnick (1991) und Tourangeau und Rasinski (1988) wurden die Items weiter angepasst. Das Antwort-Prozessmodell (Tourangeau & Rasinski, 1988) geht von einer schrittweisen Verarbeitung der Intensifier aus: Zunächst die Wahrnehmung des Intensifiers, gefolgt von der Verarbeitung des Intensifiers, der entsprechend resultierenden Urteilsbildung und schließlich der Übertragung ins Antwortformat. Die Bedenken, Intensifier könnten übersehen oder in der Urteilsbildung vergessen worden sein, sollten durch optische Hervorhebung (Studie 1.2 & 1.3), verkürzte Darstellung (Studie 1.3) und die Fokussierung auf den Vergleich von starken versus keinen Intensifiern (Studie 2 & 3) überprüft werden. Auch in den folgenden Untersuchungsergebnissen ließen sich jedoch keine systematischen Unterschiede in Bezug auf Mittelwert, Varianz oder Schiefe identifizieren. Intensifier scheinen damit, entgegen der Erwartungen, keinen systematischen Einfluss auf das Antwortverhalten der Teilnehmenden zu haben. Dies könnte ein Anzeichen dafür sein, dass Intensifier aufgrund der erhöhten Salienz wahrgenommen werden, jedoch aufgrund fehlender Motivation durch Satisficing nicht in den weiteren Beantwortungsprozess (Schritt 2-4) mit einbezogen werden (Krosnick, 1991).

Durch die Hinzunahme der subjektiven Teilnehmendenperspektive (z.B. Motivation, Freude an der Befragung) lieferte **Studie 2** daher einen wichtigen Beitrag, indem sie die Evaluationsmöglichkeiten von Befragungen um ein validiertes *experience-centered* Gütekriterium ergänzt: SX. Erstaunlicherweise hat die bisherige Forschung die Wahrnehmung von Teilnehmenden kaum systematisch betrachtet. In der Forschungslandschaft existieren fast ausschließlich ad-hoc Instrumente ohne theoretische Fundierung, die lediglich die Zufriedenheit mit der Befragung erheben (z. B. Couper et al., 2004; Walston et al., 2006; Yentes et al., 2012; Zukerberg et al., 1999). Differenzierte Validierungsversuche wie die von Yentes und Kollegen (2012) sind äußerst rar. Um diese Forschungslücke zu schließen, war das Ziel der zweiten Studie die Entwicklung eines theoriebasierten, validierten Messinstruments für SX. Entwickelt wurde SX dabei zu großen Teilen auf der Basis von User Experience (Hassenzahl et al., 2000, 2003). Entsprechend besteht sie - ähnlich wie auch die User Experience - aus den zwei Dimensionen *Usability* und *Pleasure*. Im Detail beinhaltet der erste Faktor *Usability* Facetten, die dem Konzept des Pragmatismus (Hassenzahl et al., 2003) ähneln und Effektivität und Effizienz des Survey Designs fokussieren. Exemplarisch geht es um Eindrücke wie die technische Umsetzung und Handhabbarkeit. Der zweite Faktor *Pleasure* ist demgegenüber durch emotionale oder hedonistische Vorzüge (Jordan, 1998) charakterisiert, die durch die Umfrageteilnahme entstehen. In der Praxis dienen beispielsweise Challenges, Interaktionen oder visuelle Effekte der Erreichung von *Pleasure* (Hassenzahl et al., 2003; Hassenzahl & Tractinsky, 2006). Wissenschaftlich überprüft wurde die SX-Kurzskala im Rahmen einer Vorstudie, indem sie einer Marktforschungsumfrage ($N = 799$) hinzugefügt wurde. Aus den durchgeführten Faktoranalysen ergab sich eine zweifaktorielle Lösung mit insgesamt drei Items für *Usability* (*comprehensible – incomprehensible, clear – confusing, satisfying – frustrating*) und drei Items für *Pleasure* (*inventive – fanciless, attractive – unattractive and enjoyable – unenjoyable*). Gleichzeitig luden erwartungsgemäß nicht alle Items aus dem Itempool auf die a priori festgelegten Faktoren. Die letztlich für die SX-Kurzskala ausgewählten Items sind über die verschiedenen Anwendungskontexte im Ladungsmuster jedoch stabil. Die Reliabilitäten für die Subskalen waren gut (George & Mallery, 2003, as cited in Gliem & Gliem, 2003) - besonders im Vergleich zu sonstigen Kurzskalen (Rammstedt & Beierlein, 2014). Auch die diskriminante Validität - gemessen über die Korrelation zu anderen participation centered- und quality centered Kriterien - war gegeben.

Auf den Ergebnissen aufbauend wurde in der Hauptstudie die Skala mit Hilfe verschiedener Survey-Design Konfigurationen in einem anderen Anwendungskontext erneut untersucht und weiterentwickelt. Die resultierende zweifaktorielle Lösung bestätigte sich erneut. Die entsprechende Itemauswahl der Hauptstudie - unter Berücksichtigung der Vorstudie - ergab für den Faktor *Usability* zwei Items (d.h., *comprehensible* – *incomprehensible*; *clear* – *confusing*) und für den Faktor *Pleasure* drei Items (d.h., *enjoyable* – *unenjoyable*, *inventive* – *fanciless*, *attractive* – *unattractive*). Die resultierende Kurzskala zeigte - ähnlich den Ergebnissen der Vorstudie - gute Reliabilitäten sowie gute konvergente Validität durch signifikante Zusammenhänge zu weiteren Konstrukten der experience-centered Gütekriterien (z.B. Bereitschaft zur Teilnahme an ähnlichen zukünftigen Befragungen). Zugleich bestätigten die ausbleibenden Korrelationen zu anderen Konstrukten der participation-centered und quality-centered Gütekriterien eine gute diskriminante Validität. Dass die SX-Kurzskala zusätzlich ein Indikator zur perspektivischen Verhinderung von Survey Fatigue (Porter et al., 2004) ist, zeigt sich insbesondere durch den gefundenen Zusammenhang zwischen SX und der zukünftigen Teilnahmebereitschaft. Durch die Hinzunahme der SX-Kurzskala in organisationalen Befragungen (Saari & Scherbaum, 2011) können die Erfahrungen der Teilnehmenden sehr ökonomisch erfasst werden. Die Ergebnisse der Studie zeigen zudem, dass bestimmte Konfigurationen von Survey-Design Features (z. B. Bilder und Farben) speziell den Faktor *Pleasure* und in Kombination mit weiteren Features (z.B. Mittelkategorie) auch die gesamte *SX* erhöhen. Gleichzeitig ließen sich keine Unterschiede durch die exklusive Nutzung von Radio Buttons und einer Mittelkategorie für die *Usability* nachweisen. Jedoch scheint dies aufgrund der hohen Grundfunktionalität des Fragebogens durchaus plausibel. In der Summe wurde die SX-Skala zufriedenstellend validiert und darüber hinaus wurden noch erste Implikationen zu Auswirkungen von Survey Design Features abgeleitet.

Folglich greift auch **Studie 3** den Punkt der Survey Design Features auf. Um dem Effekt fehlender Motivation und dem daraus resultierenden Datenqualitätsverlust vorzubeugen, wurde Instant Feedback als Gamification Element - beziehungsweise generisches Survey Design Feature - in Teile des EAS eines Lebensmitteleinzelhändlers integriert. Innerhalb des randomisierten Feldexperiments wurden die 735 Teilnehmenden zufällig der spielerischen Bedingung (mit Instant-Feedback) versus der Standard-Bedingung (ohne Instant-Feedback) zugeordnet. Dabei misst die SX-Skala aus Studie 2 als experience-centered Gütekriterium die Auswirkungen der Gamification auf die Erfahrung der Teilnehmenden. Zusätzlich werden weitere Gütekriterien für die Datenqualität (z.B.

Straightlining) erfasst. Die Ergebnisse zeigten dabei, dass die spielerische Bedingung mit Instant Feedback im Vergleich zur Standard-Bedingung signifikant bessere *experience-centered* Gütekriterien (z.B. *Pleasure*) aufwies. Weitere positive Ergebnisse zeigten sich in Bezug auf die klassische Datenqualität (z.B., Straightlining und Item-Non-Response). Zeitgleich blieb die Facette *Usability*, wie auch schon in Studie 2 zu beobachten, unverändert. Da kleine Anpassungen innerhalb der Befragung die Möglichkeit schaffen, die Motivation und Datenqualität zu modulieren, bieten die Studienergebnisse ökonomische Interventionsmöglichkeiten mit großem Skalierungspotential insbesondere für daraus resultierende Maßnahmenableitungen.

Generelle Implikationen für die Forschung

Die drei dargestellten Studien liefern neue Erkenntnisse in Bezug auf die Beeinflussung der Befragungsgüte im Rahmen von EAS. Einzelne Einflussfaktoren und Gütekriterien werden dabei im Gesamtkontext auf mögliche Verbesserungspotentiale untersucht.

Zunächst lassen sich durch **Studie 1** die verschiedenen Antworttheorien von Tourangeau und Rasinski (1988) und Krosnick (1991) überprüfen. Trotz der Annahme der vier Schritte zur Beantwortung von Befragungen (Tourangeau & Rasinski, 1988) ließen sich die aus der Manipulation der Items abgeleiteten systematischen Effekte im Antwortverhalten der Teilnehmenden nicht finden. Die ausbleibenden Effekte stimmen mit Ergebnissen von Gaskell, Wright, und O’Muircheartaigh (1993) überein, die entsprechend die Annahme vertreten starke Intensifier würden überwiegend übersehen. Die erfolgten Anpassungen der Studie zur Erhöhung der Intensifier-Salienz – wie beispielsweise in Studien von Low (1995, 1996) und O’Muircheartaigh und Kollegen (1993) vorgeschlagen – zeigten jedoch – trotz der literaturkonformen Variation von Salienz und Itemkomplexität – wider Erwarten keine systematischen Effekte im Antwortverhalten (z. B. French-Lazovik & Gibson, 1984; Lam & Stevens, 1994; Weijters et al., 2013; Wildt & Mazis, 1978; Wyatt & Meyers, 1987). Für das Ausbleiben signifikanter Ergebnisse können zwei mögliche Ursachen in Betracht gezogen werden:

Die fehlende Berücksichtigung der Intensifier im Bewertungsprozess (d.h. Schritt 2 und Schritt 3: Abrufen der relevanten Informationen und Urteilsbildung) aufgrund einer zu hohen kognitiven Anstrengung ist eine erste plausible Erklärungsmöglichkeit. Diese Erklärung würde bedeuten, dass beispielsweise neue Themen und die damit verbundenen Vorstellungen mehr kognitive Ressourcen beanspruchen. Infolgedessen verhindern die limitierten kognitiven Ressourcen eine Fokussierung der Intensifier beziehungsweise eine differenzierte Urteilsbildung (Krosnick, 1991; Wright et al., 1995). In der Konsequenz steigt

die Wahrscheinlichkeit für *Satisficing* Verhalten. Unterstützung findet die Annahme durch ein Experiment von Low (1996). Seine think-aloud Task zeigt trotz Lesens der Intensifier eine fehlende Integration in die Urteilsbildung.

Als zweite mögliche Ursache kann die häufig gleichzeitige Verwendung von Intensifiern sowohl im Itemstamm als auch im Antwortformat (Cools et al., 2006; Hofmans et al., 2007; C. A. O’Muircheartaigh et al., 1993; Rohrman, 2007; Wright et al., 1995) diskutiert werden. Die möglicherweise dominantere Form der Intensifier im Antwortformat könnte die Wirkung der Intensifier im Itemstamm beeinflussen und so auf den Beantwortungsprozess einwirken. Zugleich könnte auch hier die kognitive Anstrengung, beide Intensifier parallel zu verarbeiten und zu bewerten, höher sein. Zur Vermeidung dieser höheren Anstrengung könnten Teilnehmende die Intensifier im Itemstamm bewusst vernachlässigen (Tourangeau et al., 2000).

Dieses Vermeidungsverhalten lässt sich jedoch vermutlich nicht allein durch die Varianz von Item-Formulierungen verändern, sondern ist auch bedingt durch eine gute – die Motivation der Teilnehmenden fördernde – Befragungsgestaltung. Um das Motivationspotenzial einer spezifischen Befragung zu erfassen und damit den Erfolg von Survey Design Features konsequent zu evaluieren bedarf es validierter Messinstrumente, die über die Erfassung der generellen Einstellung der Teilnehmenden zu Befragungen hinaus gehen.

Dabei ist die Konzeptionalisierung von SX ein wichtiger Schritt zur Erfassung und Entwicklung möglicher Motivationsgewinne. SX ermöglicht, den Einfluss der Fragebogengestaltung im Kontext des experience-centered Gütekriteriums zu evaluieren. Auch wenn das anfängliche Ladungsmuster der SX Skala sich teils von den bekannten Mustern aus dem Bereich der User Experience unterschied, ist die letztliche Itemauswahl der SX-Kurzskala mit den Faktoren *Pleasure* und *Usability* mit der bisherigen Forschung zur User Experience konform. Folglich bestätigen die Ergebnisse die theoretischen Überlegungen zur Anwendbarkeit von User Experience im Kontext der SX. Da das Ursprungs-konstrukt User Experience häufig im Kontext der Rahmenbedingungen betrachtet wird (Law et al., 2009), könnte die Struktur von SX in Teilen von Inhalten, Zielgruppen und Survey Settings abhängig sein. Diese Facetten variieren in den verschiedenen Anwendungsbereichen wie beispielsweise sozialen Befragungen (Vehovar et al., 2000), Marktforschung (Evans & Mathur, 2018) oder Organisationsentwicklung (Smither et al., 2016). Trotz der unterschiedlichen Rahmenbedingungen der Anwendungsbereiche innerhalb der Studien 2.1. und 2.2 zeigt sich hier im Vergleich der SX-Kurzskala-Items eine sehr hohe Konsistenz. Diese

Übereinstimmung wiederum demonstriert eine sehr gute und vielversprechende Robustheit der entwickelten SX-Kurzskala. Nichtsdestotrotz sollte sie in weiteren Anwendungsbereichen wissenschaftlich überprüft werden.

In Studie 3 wird diese robuste SX-Kurzskala nun verwendet, um nachzuweisen, dass Instant Feedback – als Gamification Element – im Einklang mit der bisherigen Forschung (z. B. Keusch & Zhang, 2017) die *Pleasure* und Antwortqualität beeinflusst. In Bezug auf diese vielversprechenden Gestaltungsergebnisse ist eine Konfundierung von Instant Feedback mit anderen Survey Design Features weitgehend ausgeschlossen. Da die Vielfalt möglicher Survey Design Features nur partiell erforscht ist, brauchte es weitere Forschung zu spezifischen Survey Design Features (Keusch et al., 2020), die leicht einzusetzen sind. Die ersten Forschungsansätze zu Badges (Harms, Seitz, et al., 2015) lassen sich durch die vorliegende Studie bestätigen: Kleine Interventionen reichen aus, um Teilnehmende zu beeinflussen. In Übereinstimmung mit der bisherigen Forschung (Hamari et al., 2014) zeigen gamifizierte Umfragen eine motivationsfördernde Charakteristik. Drei mögliche Mechanismen, die von Jung und Kollegen (2010) vorgeschlagen wurden, könnten die auftretenden Effekte erklären: Erstens, da das Feedback direkt mit dem Antwortprozess der Teilnehmer verknüpft ist, funktioniert das Feedback als Belohnung, um wesentliche Informationen über die Gruppenleistung zu erhalten. Zweitens kann diese Gruppeninformation zusätzlich das Urteil oder die Item-Verarbeitung des Teilnehmers bewerten, indem er seine Antwort mit der Antwort seiner Peergruppe vergleicht. Drittens werden die Teilnehmer, die sehen, dass ihr Feedback übereinstimmt, eine Art von Verstärkung spüren, während unbedachte Antworten möglicherweise zu einer negativen Verstärkung führen, da sich die Ergebnisse stark unterscheiden. Zeitgleich konnten keine Unterschiede in der *Usability* gefunden werden. Dies kann zum einen an der hohen Grundfunktionalität der heutigen Befragungen liegen. Zum anderen scheinen so keine negativen Auswirkungen durch Gamification-Elemente (Koenig-Loewis et al., 2013 zitiert nach Keusch & Zhang, 2017) zu entstehen. Dieses Phänomen ist daher als positiv zu bewerten.

In der Zusammenschau der Studien wird deutlich, dass SX ein wichtiges Konstrukt der *experience-centered* Kriterien ist, das die *participation-centered* und *quality-centered* Kriterien sinn- und wertvoll ergänzt. Insbesondere die Ergebnisse aus Studie 2 und 3 zeigen, dass es sich um ein reliables Messinstrument handelt, das durch Gestaltungsmöglichkeiten innerhalb der Befragung beeinflusst wird. Da sich innerhalb der Befragungen sowohl die SX als auch die Item-Non-Response verbessert hat, wird deutlich, dass *experience-centered*

Gütekriterien potentiell auch im Zusammenhang zu den anderen (participation-centered und quality-centered) Gütekriterien stehen können. Bezüglich des beobachtbaren Effekts zwischen der gamifizierten und nicht-gamifizierten Bedingung ist weitere Forschungsarbeit in Bezug auf mögliche Moderatorvariablen innerhalb und außerhalb des Befragungsdesigns (z. B. Charakteristik der Teilnehmenden, Itemformulierungen, Rahmenbedingungen der Befragung) nötig.

Gleichzeitig bieten auch die ausbleibenden systematischen Effekte in Bezug auf die quality-centered Gütekriterien wichtige Erkenntnisse: Gamification scheint hier die Interpretation der Items - gemessen an den Mittelwerten der Gruppen - nicht zu beeinflussen. Die Beeinflussung von quality-centered Gütekriterien ist in unseren Studien bislang nur durch die Itemumformulierungen in Studie 1 (unsystematisch) gelungen. Perspektivisch bietet aber auch Gamification die Möglichkeit, Items umzuformulieren und so gegebenenfalls quality-centered Kriterien zu beeinflussen. Inwieweit die Kombination aus Itemformulierung und Gamification einen Einfluss auf die drei Gütekriterien hat, könnte somit eine interessante Forschungsfrage sein. Bei kumulierter Nutzung diverser Gestaltungsmöglichkeiten der Befragungen könnte deren relativer Einfluss auf die drei Gütekriterien der Befragung mit Hilfe von Regressionsanalysen untersucht werden. Bei gezielter Adressierung aller Qualitätskriterien könnten zudem auch die Abhängigkeiten zwischen den Gütekriterien untersucht werden.

Generelle praktische Implikationen

Gerade in komplexen Umgebungen und unter dynamischen Randbedingungen gelingt der langfristige Erfolg von Unternehmen durch den konstruktiven Austausch mit Mitarbeitenden (Grant & Marshak, 2011; Heracleous, 2002). Entsprechend wird der Austausch mit Hilfe von Befragungen häufig unterstützt und katalysiert, wobei eine Zunahme der EAS zu erkennen ist (Frieg & Hossiep, 2020). In der Praxis sind gleichzeitig Herausforderungen in der Interpretation und entsprechenden Entwicklung von adäquaten Maßnahmen erkennbar: Linksschiefe Verteilungen und Satisficing Verhalten sind Trends, die den konstruktiven Austausch behindern. Ziel dieser Arbeit ist es, entsprechende Gestaltungsmechanismen zu identifizieren, die einen positiven Einfluss auf die Gütekriterien der Befragung - wie beispielsweise Datenqualität und SX - haben.

In Bezug auf die Unterstützung des Beantwortungsprozesses der Teilnehmenden liefert diese Studie vielversprechende Erkenntnisse für die Umfragepraxis. Da keine systematischen Effekte von Intensifiern in Bezug auf linksschiefe Verteilungen und die Variabilität gefunden wurden, führt die bislang häufig inkonsistente Verwendung von

Intensifiern im Itemstamm damit wider Erwarten nicht zu einer verzerrten Vergleichbarkeit der Einzelitems. Mehr noch wäre es den Ergebnissen zufolge möglich, zukünftig Intensifier bewusst wegzulassen oder strategische Intensifier zu platzieren, ohne Verluste von Benchmarking-Möglichkeiten zu befürchten. Insgesamt könnte dies für die Benchmarking-Praxis also eine Lockerungsmöglichkeit der Vergleichbarkeitsprüfung bedeuten, die jedoch einer weiteren statistischen Validierung bedarf. Aufgrund der aktuellen Forschungslage scheinen Intensifier somit ungeeignet, um trennschärfere Informationen aus den Daten zu extrahieren und in Form von beispielsweise Zusammenhangsanalysen auszuwerten.

Neben der spezifischen Itemformulierung zeigt die Gesamtheit der Fragebogengestaltung Verbesserungsmöglichkeiten für die Befragungsgüte. Das vorgeschlagene auf die Teilnehmendenerfahrung ausgerichtete Konstrukt der SX dient dabei als wichtige Grundlage. Die theoretisch-fundierte SX-Kurzskala ist ein effizientes und ökonomisches Instrument, um Praktikern zeitnahes und wertvolles Feedback zu durchgeführten Fragebogenanpassungen zu vermitteln. Als Kurzskala mit lediglich 5 Items ist sie ideal geeignet, als begleitendes Erhebungsinstrument in alle Arten von Befragungen integriert zu werden. So verhindert sie als Frühindikator für *Usability*- und *Pleasure*-Schwierigkeiten die Entstehung von Survey Fatigue. Darüber hinaus bietet die SX-Kurzskala als validiertes Evaluationsinstrument der Befragungsgüte die Möglichkeit, Survey Design Features einheitlich in Bezug auf SX zu überprüfen. Beispielsweise zeigen erste Erkenntnisse eine positivere *Pleasure*- und SX-Bewertung bei einer Kombination von Grafiken und farblichen Survey Design Elementen innerhalb einer Befragung. Im Gegensatz dazu erzielten die verschiedenen Konfigurationen von Survey Design Elementen keine signifikanten Effekte für die *Usability*. In der Praxis scheint die *Usability* in vielen etablierten Befragungen bereits gegeben zu sein. Als Ansatzpunkt zur Befragungsverbesserung bleiben somit insbesondere *Pleasure*-bezogene Komponenten wie zum Beispiel Grafiken und Farben. Gleichzeitig beschränken sich die Gestaltungsmöglichkeiten nicht nur auf die Verwendung dieser Komponenten, sondern können um weitere Survey Design Elemente erweitert werden. Schließlich zeigen Keusch und Zhang (2017), dass speziell Gamification-Elemente geeignet sind, das Engagement, die Zufriedenheit und Motivation der Befragten zu erhöhen. Es ist zu erwarten, dass sich dies entsprechend auch in der SX beziehungsweise ihren Subskalen *Pleasure* und *Usability* widerspiegelt.

Folglich setzt genau an dieser Stelle Studie 3 an und untersucht mit Hilfe der entwickelten SX-Kurzskala den Einfluss der Gamification auf das experience-centered Gütekriterium. Parallel werden ebenfalls qualitätsbezogene Facetten (z.B. Straightlining,

Item-Non-Response) der Befragungsgüte untersucht. Nach dem Beispiel von Harms und Kolleg:innen (2015), die lediglich Badges in Befragungen einbauten, wurden unter Verwendung eines ökonomischen und leicht zu implementierenden Gamification Elements psychologische und verhaltensbezogene Komponenten der Teilnehmenden beeinflusst: In unseren Studien konnte Instant Feedback als Gamification Element die *Pleasure* Komponente der Befragung signifikant verbessern. Zusätzlich ließen sich mit Instant Feedback im Bereich der Datenqualität positive Ergebnisse erzielen: Instant Feedback führte zu signifikant weniger fehlenden Antworten. Beim Satisficing Verhalten – geprüft über Straightlining – gibt es darüber hinaus einen deskriptiven Trend zu einer verbesserten Datenqualität. Zusammenfassend lassen sich durch einfache gamifizierte Ergänzungen zum Fragebogen experience-centered und participation-centered Gütekriterien verbessern. Gegenüber der häufig von komplexen Gamification Ansätzen geprägten Praxis zeigen diese Ergebnisse, dass bereits gut durchdachte minimale Ansätze zu einer Verbesserung der Befragungsgüte führen und dadurch im Vergleich zu aufwendigen Gamification-Ansätzen Kosten sparen können.

Aus diesen Erkenntnissen ergibt sich ein mögliches Vorgehen für die Weiterentwicklung von Befragungen in der Praxis. Durch die Diversifizierung der Gütekriterien von Befragungen bietet sich nun die Möglichkeit, eine spezifische Befragung sowohl aus Sicht der quality-centered, participation-centered und experience-centered Gütekriterien zu evaluieren. Insbesondere die SX-Skala ist ein neues Instrument, das die Wahrnehmung der Mitarbeitenden während der Befragung fokussiert und so als zusätzlicher Indikator für einen möglichen Handlungsbedarf dienen kann. Sollte dieses Instrument kritische Werte identifizieren, ließe sich im Folgeprozess qualitativ mit Hilfe offener Fragen eruieren, welche Aspekte der Fragebogengestaltung diese Wahrnehmung begründet haben. Zugleich bietet der Einsatz der SX-Skala die Möglichkeit, den Mitarbeitenden zu vermitteln, dass ihr Erleben während der Befragung für das Unternehmen relevant ist und der Zeiteinsatz wertgeschätzt wird.

Gemäß den Ergebnissen aus unseren Studien kann kritischen Werten in den *experience-centered* oder *participation-centered* Kriterien insbesondere durch das spezifische Gamification-Element des Instant Feedbacks begegnet werden. Gerade im Kontext von sich wiederholenden Befragungen (wie beispielsweise EAS und ihren Follow-Up-Befragungen), in denen standardisierte Fragen und Instruktionen immer wieder verwendet werden, empfiehlt sich der Einsatz von Instant-Feedback zur Abfederung der Befragungsmüdigkeit. Durch die Ergänzung der Befragung um Gamification-Elemente sind positive Einflüsse auf die *experience-centered* und *participation-centered* Gütekriterien zu erwarten. Zugleich sollten in

der Praxis auch die *quality-centered* Gütekriterien auf Veränderungen hin überprüft werden, da ein Einfluss nicht gänzlich ausgeschlossen werden kann.

Treten in der Evaluation von Befragungen hingegen kritische Werte im Bereich der *quality-centered* Gütekriterien auf - wie beispielsweise stark links-schiefe Verteilungen – ist die Nutzung von Intensifiern in der Itemformulierung anstelle von Gamification-Elementen im Befragungsprozess ein erster möglicher Ansatzpunkt. Die Anwendung von Intensifiern könnte trotz der vorliegenden Ergebnisse erprobt werden, da sich hier zumindest unsystematische Effekte zeigten. Eine sorgfältige Evaluation aller Anpassungen in der Gestaltung von Befragungen in Bezug auf die drei Bereiche der Befragungsgüte ist auch in der Praxis in jedem Fall zu empfehlen.

Generelle Limitationen

Neben den vielversprechenden Ergebnissen sind in den jeweiligen Studien bestimmte Besonderheiten zu beachten, die sich auf die Interpretation der Ergebnisse und ihre Generalisierung auswirken.

Der Chronologie der Durchführung folgend stellt zunächst die erhobene **Stichprobe** eine mögliche Limitation dar. Während in Studie 3 der EAS - im praxisnahen Kontext - in einem wirtschaftlichen Unternehmen erhoben wurde, erreichten die Studien 1 und 2 ihre Praxisnähe über Zugangsvoraussetzungen. Konkret wurde in Studie 1 auf Arbeitnehmer fokussiert, doch die Probandengewinnung wurde über Social-Media und persönliche Kontakte abgewickelt, sodass die Teilnehmenden mitunter aus unterschiedlichen Organisationen, teils sogar unterschiedlichen Industriebereichen, kamen. Ähnlich wurden in Studie 2 Panel-Daten von Arbeitnehmern mit vorgegebenen Voraussetzungskriterien erhoben. Dadurch variiert der demographische und organisationale Hintergrund der Teilnehmenden. Entsprechend könnten durch die Erhebungsmethoden Teilnehmende mit einer Affinität zu Befragungen in den Stichproben enthalten sein. Dies könnte neben der zusätzlichen Bezahlung, die in den anderen Studien entfiel, die Effekte auf das Antwortverhalten beeinflusst haben.

Während der **Erhebung** ist die Überprüfung der *experience-centered* Kriterien nicht immer trivial. Da die Erhebung der SX-Kurzskala nur am Ende der Befragung sinnvoll einzusetzen ist, stellt der Dropout von Teilnehmenden während der Befragung eine Herausforderung in Studie 2 und Studie 3 dar. Durch Abbrüche begründete fehlende Antworten erschweren hier die Interpretation der Zusammenhänge auf Individual-Niveau, da die Ursachen für den Dropout nicht endgültig erklärbar sind und letztlich keine Daten der SX vorliegen. Infolgedessen sind einige Analysen nur auf aggregiertem Niveau möglich.

Dennoch tragen die Studien dazu bei, wichtige Erkenntnisse über mögliche Zusammenhänge mit SX und anderen Survey Design Features zu gewinnen.

Für die Auswertung und Interpretation der Ergebnisse im Kontext des Studiendesigns ist die erreichte **Power** ein zentraler Maßstab zur Bewertung der Adäquatheit. Über alle Studien hinweg konnten jeweils eine ausreichende Power für mittlere Effekte erreicht werden. In **Studie 1** war die Power ausreichend hoch (.99 – 1.00), um potentiell mittlere Effekte ($d = 0.50$) zu identifizieren, die wiederum einer Veränderung von 0.36 bis 0.73 auf der Antwortskala entsprechen würden. Trotz der hohen Power ließen sich jedoch keine signifikanten Effekte in der Studie finden. Die Untersuchung auf kleine Effekte ($d = 0.20$), die abhängig von der Standardabweichung einen Unterschied von circa 0.14 bis 0.29 auf der Antwortskala ausmachen würden, hat wahrscheinlich keine praxisrelevanten Implikationen im Kontext der EAS und wurde daher nicht weiter verfolgt. Auch **Studie 2** zeigt im Vergleich zu Studie 1 etwas geringere aber ausreichende Power von 0.84, um mittlere Effekte zu finden ($d = 0.50$). Dabei konnten - im Gegensatz zur Untersuchung des Intensifiereinflusses in Studie 1 - für die Facetten *Pleasure* und Gesamt-SX signifikante Unterschiede gefunden werden. Die potentiellen Effekte für die *Usability* wiesen vergleichsweise minimale Effektstärken ($d = 0.07$) auf und hätten eine Stichprobengröße von 4518 benötigt, um signifikant zu werden. Zudem wären diese Effekte zu klein, um als sinnvolle Unterschiede in Bezug auf SX interpretiert zu werden. Insgesamt sind die erreichten Stichprobengrößen und die jeweilige Power zufriedenstellend. In der praxisnahen **Studie 3** konnten wir mit den vorangegangenen Power-Analysen aus Studie 1 und Studie 2 für mittlere Effekte sehr gut mithalten und diesmal sogar für kleine Effekte ($d = 0.20$) eine mit 0.77 ausreichende Power gewährleisten. Gerade in Bezug auf die Kontrolle des ähnlichen Antwortverhaltens für Arbeitszufriedenheit, Commitment und Engagement war dies ein wichtiger Parameter, da Gamification keine Veränderungen in der Bewertung der arbeitsbezogenen Einstellungen erzielen sollte. Zusammenfassend wird ersichtlich, dass die Power in allen Studien zur Entdeckung mittlerer Effekte ausreichend hoch war.

Zukünftige Forschung

Mit Hilfe dieser Forschungsarbeit wurden Survey Design Kriterien validiert, die einen positiven Einfluss auf die SX - als *experience-centered* Gütekriterium von Umfragen - haben. Dennoch bedarf es weiterer Forschung: Besonders die fehlenden Unterschiede im Antwortverhalten in Studie 1 werfen die Frage auf, warum die Manipulation durch Intensifier trotz der erhöhten Salienz nicht erfolgreich war. Laut den Modellannahmen von Tourangeau und Rasinski (1988) werden diese vermutlich wahrgenommen aber nicht im

Bewertungsprozess berücksichtigt. Hypothetisch könnten angepasste Rahmenbedingungen in Form von adaptierten Instruktionen die Verarbeitung im Beantwortungsprozess unterstützen. Da in unserer Studie 1 keine Bezahlung erfolgte, die Teilnahme freiwillig war und keine direkten Implikationen für die Teilnehmenden ersichtlich waren, wäre eine erneute Durchführung der Studie im organisationalen Setting interessant. Dieser Wechsel aus dem Kontext der „Laborbefragungen“ hin zu realen Befragungen mit persönlicher Relevanz begünstigt wahrscheinlich gewissenhafteres Antwortverhalten - insbesondere da beispielsweise Mitarbeiterbefragungen durch ihren direkten und indirekten Einfluss auf die Mitarbeitenden die intrinsische Motivation der entsprechenden Teilnehmenden erhöhen (z. B. Church & Waclawski, 2001; Mueller et al., 2011). Im Einklang mit Krosnick's (1991) Idee führt die erhöhte Motivation dann wahrscheinlicher zum gewünschten Optimizing Verhalten an Stelle des - die Datenqualität mindernden - Satisficing Verhaltens. Dabei ist anzunehmen, dass Optimizing Verhalten eine Verarbeitung und Einbeziehung der Intensifier begünstigt.

Zusätzlich ist die Überprüfung, ob Intensifier tatsächlich gelesen und wahrgenommen werden, mit Hilfe von Eye-tracking oder neuro-image Technologie sinnvoll. Erste Studien mit der Methode des Lauten Denkens von Low (1996) legen nahe, dass Intensifier gesehen und wahrgenommen werden, aber nicht in die Verarbeitung (Schritt 3) einbezogen werden. Durch die neuen Methoden könnte überprüft werden, wie lange Intensifier betrachtet werden und ob dies zu Veränderungen in der Gehirnaktivität führt. Die Studie von Lenzner und Kolleg:innen (2011) konnte durch diesen Ansatz bereits erste Erkenntnisse zur Verwendung von bestimmten Textelementen liefern (z. B. geläufige Begrifflichkeiten, verständliche Syntax, wenig Mehrdeutigkeiten). Eine Anpassung des methodischen Vorgehens auf die Intensifier wäre leicht umsetzbar. Auch die Anwendung von EEG-Technologie (z. B. Freunberger & Nieuwland, 2016) oder fMRI-Technologie (z. B. Zhan et al., 2017) kann zur Beobachtung von mentalen Prozessen während der Verarbeitung und Interpretation von Intensifiern genutzt werden. Durch das bessere Verständnis der kognitiven Prozesse könnten weitere Implikationen hinsichtlich der Verwendung von Intensifiern für die erfolgreiche Gestaltung von EAS abgeleitet werden. Gleichzeitig sind derartige Methoden nicht auf den Anwendungsbereich Intensifier beschränkt, sondern können auch für die Gamification wichtige Erklärungsmodelle liefern, indem sie auf der physiologischen Ebene demonstrieren, welche Gamification-Elemente besonders ansprechend wirken und ob dadurch positive Effekte auf die Antwortqualität vermittelt werden.

Die langfristigen Auswirkungen von Gamification-Elementen zu untersuchen ist ebenfalls relevant: Gamification-Elemente könnten zum Beispiel in ihrer Wirkung durch

Gewöhnungseffekte langfristig abnehmen. Ähnliches könnte auch für SX und ihren Einfluss auf die Datenqualität relevant sein. Am Beispiel von Gamification zeigt eine Langzeitstudie (Mavletova, 2015), dass sich die zunächst positiven Effekte auf Straightlining und Item-Non-Response über die Zeit wieder verringern. Auch hier können Gewöhnungseffekte offensichtlich eintreten. Zukünftige Experimente sollten daher häufiger zwei oder mehrere aufeinanderfolgende Messzeitpunkte beinhalten, um so mögliche Veränderungen feststellen zu können.

Neben der Gestaltung der Studien sollten auch die Prozesse, die im Individuum vorgehen, näher beleuchtet werden. Der häufige Unterschied zwischen Verhalten und Kognition zeigt die Notwendigkeit, den Beantwortungsprozess auf der Verhaltens- und Kognitionsebene zu betrachten. Während wir in den vergangenen Experimenten häufig von einer Art Blackbox ausgegangen sind, sollten die Komponenten während der Beantwortung genauer untersucht werden. Dabei zeigen sozialpsychologische Experimente, dass neben der kognitiven Verarbeitung und dem letztlich gezeigten Verhalten (z.B. Dropout, Item-Non-Response) auch die affektive/motivationale Ebene berücksichtigt werden sollte.

Die zukünftige Forschung zur Gamification sollte im Kontext der EAS weitere konkrete ökonomische Maßnahmen (beispielsweise Avatare, Fortschrittsbalken oder Punktesysteme) und ihre Langzeitwirkung untersuchen. Zusätzlich könnte die Interaktion mit Intensifiern überprüft werden, um die Dynamik unterschiedlicher Interventionsmethoden besser zu verstehen. Für die Theorie und Praxis lassen sich so verlässlichere Aussagen und damit auch Implikationen ableiten, die weitere konstruktive Forschungen anstoßen.

In der Gesamtbetrachtung der vorliegenden Forschungsarbeiten wird zugleich deutlich, dass sich die zukünftige Forschung nicht wie bislang hauptsächlich auf *quality-centered* und *participation-centered* Gütekriterien konzentrieren sollte, sondern eine holistische Evaluation durch die Hinzunahme von *experience-centered* Gütekriterien das Ziel sein sollte. Nur durch diese Gesamtbetrachtung lassen sich die Gestaltungsmöglichkeiten von Befragungen in ihrer Gesamtheit begreifen und optimieren. So könnten neue Hypothesen zur Verarbeitung von Intensifiern und die Wahrnehmung von Gamification-Elementen unter Berücksichtigung des spezifischen Erlebens der Befragungsteilnehmenden aufgestellt werden, die sich nicht nur von den eher mechanischen *quality-centered* und *participation-centered* Gütekriterien inspirieren lassen.

Generelle Schlussfolgerung

Umfragen stellen ein wichtiges und beliebtes Instrument in der Marktforschung und organisationalen Praxis dar (z. B. Evans & Mathur, 2018; Smither et al., 2016). Durch die

Zunahme der Befragungen sehen sich die potentiellen Teilnehmenden mit immer mehr Befragungen konfrontiert (Adams & Umbach, 2012; Saari & Scherbaum, 2011). Daher ist es umso wichtiger, die Umfragen so zu gestalten, dass Teilnehmende motiviert sind, diese gewissenhaft auszufüllen und gleichzeitig sicherzustellen, dass die Daten sinnvolle Ableitungen zulassen. Dabei betrachtet die vorliegende Arbeit sowohl den Beantwortungsprozess (Krosnick, 1991; Tourangeau & Rasinski, 1988), die SX als auch Elemente der Gamification (Keusch & Zhang, 2017), um Maßnahmen für eine gute Befragungsgüte zu identifizieren. Konkret helfen die Erkenntnisse im Kontext der Umfrageforschung, den Beantwortungsprozess zu verstehen und dadurch zu verändern sowie die Motivation durch eine ansprechende Umfragegestaltung zu erhöhen und durch ergänzende spielerische Elemente nochmals zu verstärken - alles mit dem Ziel, die Befragungsgüte zu verbessern. Durch die drei vorliegenden Studien innerhalb dieses Forschungsprojektes lassen sich drei wichtige Ansatzpunkte identifizieren.

Zunächst lassen sich verschiedene Intensifier im Itemstamm in unterschiedlichen Befragungsinstrumenten wiederfinden. Die aus der Theorie abgeleiteten Gefahren und Befürchtungen bezüglich der validen Interpretation und Vergleichbarkeit innerhalb und zwischen Items können mit der vorliegenden Datengrundlage abgeschwächt werden. Da die systematische Untersuchung der Intensifier zu keinen systematischen Abweichungen in Bezug auf Mittelwerte, Varianzen, Schiefe und nomologischer Struktur führen, können die bisherigen Benchmarks in der Praxis liberaler und gleichzeitig flexibler angewendet werden. Um dennoch die Befragungsgüte zu gewährleisten, sollte der Motivation und Freude an der Beantwortung besondere Aufmerksamkeit geschenkt werden, da diese die adäquate Verarbeitung (Tourangeau & Rasinski, 1988) in Form von Optimizing (Krosnick, 1991) begünstigt.

Dabei kann SX eben diese Freude und Nützlichkeit als wichtiges Kriterium zur *experience-centered* Befragungsgüte erheben und somit einer Survey Fatigue effektiv entgegenwirken (Manfreda et al., 2008; Porter et al., 2004). Die entwickelte und validierte SX-Kurzskala stellt ein zuverlässiges, wissenschaftlich fundiertes und ökonomisches Instrument dar. Es bietet die Möglichkeit, viele Studiendesigns auf ihre motivationalen beziehungsweise *Pleasure*-bezogenen und anwenderfreundlichen Facetten zu überprüfen und entsprechende (Gegen-)Maßnahmen für die Praxis zu evaluieren.

Eine solche Maßnahme zur Erhöhung der Motivation gelingt durch den Einsatz des direkten Feedbacks als Gamification-Element. Die konkreten motivationalen Anreize (Hamari et al., 2014) beeinflussen sowohl *Pleasure*-bezogene Facetten der wahrgenommenen SX als

auch die Datenqualität. Organisationale Befragungen profitieren durch diese kleinen ökonomischen Anpassungen im Fragebogendesign und machen so aufwendige Überarbeitungen obsolet.

Die vorliegen Studien zeigen damit eindrucksvoll, dass die bisher vorherrschende Praxis, die Güte von Befragung vor allem anhand von *quality-centered* und *participation-centered* Gütekriterien festzumachen, zu kurz greift und dass die Erhebung von *experience-centered* Gütekriterien einen vielversprechenden Ansatz für die zukünftige Forschung darstellt. Für die Praxis zeigt sich, dass Veränderungen im Bereich der Itemformulierungen nicht immer die gewünschte Wirkung erzielen und genau evaluiert werden sollten. Deutlicher fällt hingegen der Einfluss von selbst minimalen Gamification-Elementen auf die Wahrnehmung der Teilnehmenden bezüglich der spezifischen Befragung aus. Die resultierenden, qualitativ besseren Daten dienen als wichtige Grundlage für eine strategische Ableitung von Maßnahmen in Unternehmen und können so zum Unternehmenserfolg maßgeblich beitragen.

Referenzen

- Adams, M. J. D., & Umbach, P. D. (2012). Nonresponse and Online Student Evaluations of Teaching: Understanding the Influence of Salience, Fatigue, and Academic Environments. *Research in Higher Education*, 53(5), 576–591. <https://doi.org/10.1007/s11162-011-9240-5>
- Al-Haddad, S., & Kotnour, T. (2015). Integrating the organizational change literature: A model for successful change. *Journal of Organizational Change Management*, 28(2), 234–262.
- Bailey, P., Pritchard, G., & Kernohan, H. (2015). Gamification in market research: Increasing enjoyment, participant engagement and richness of data, but what of data validity? *International Journal of Market Research*, 57(1), 17–28.
- Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1978). Change in the Structure of American Political Attitudes: The Nagging Question of Question Wording. *American Journal of Political Science*, 22(2), 250. <https://doi.org/10.2307/2110616>
- Borg, I., & Mastrangelo, P. M. (2008). *Employee surveys in management: Theories, tools, and practical applications*. Hogrefe.
- Braunscheidel, M. J., Suresh, N. C., & Boisnier, A. D. (2010). Investigating the impact of organizational culture on supply chain integration. *Human Resource Management*, 49(5), 883–911. <https://doi.org/10.1002/hrm.20381>
- Burke, R. J., & Ng, E. (2006). The changing nature of work and organizations: Implications for human resource management. *Human Resource Management Review*, 16(2), 86–94. <https://doi.org/10.1016/j.hrmr.2006.03.006>
- Buskirk, T. D. (2015). Are Sliders Too Slick for Surveys? An Experiment Comparing Slider and Radio Button Scales for Smartphone, Tablet and Computer Based Surveys. *Methods, data*, No 2 (2015)-. <https://doi.org/10.12758/mda.2015.013>
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on Interviewing Techniques. *Sociological Methodology*, 12, 389–437. <https://doi.org/10.2307/270748>
- Cechanowicz, J., Gutwin, C., Brownell, B., & Goodfellow, L. (2013). Effects of gamification on participation and data quality in a real-world market research domain. *Proceedings of the First International Conference on Gameful Design, Research, and Applications - Gamification '13*, 58–65. <https://doi.org/10.1145/2583008.2583016>
- Christian, L. M., & Dillman, D. A. (2004). The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions. *Public Opinion Quarterly*, 68(1), 57–80. <https://doi.org/10.1093/poq/nfh004>

- Church, A. H., & Waclawski, J. (2001). *Designing and using organizational surveys: A seven-step process* / (1st Aufl.). Jossey-Bass.
- Church, A. H., & Waclawski, J. (2017). *Designing and Using Organizational Surveys*. Routledge. <https://doi.org/10.4324/9781315258119>
- Chyung, S. Y. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement*, 56(10), 15–23. <https://doi.org/10.1002/pfi.21727>
- Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, 66(1), 27–44. <https://doi.org/10.1037/h0045660>
- Cobanoglu, C., Moreo, P. J., & Warde, B. (2001). A Comparison of Mail, Fax and Web-Based Survey Methods. *International Journal of Market Research*, 43(4), 1–15. <https://doi.org/10.1177/147078530104300401>
- Cook, C., Heath, F., & Thompson, R. L. (2000). A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement*, 60(6), 821–836. <https://doi.org/10.1177/00131640021970934>
- Cools, W., Hofmans, J., & Theuns, P. (2006). Context in category scales: Is “fully agree” equal to twice agree? *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 56(4), 223–229. <https://doi.org/10.1016/j.erap.2005.09.007>
- Couper, M. P., Conrad, F. G., & Tourangeau, R. (2007). Visual Context Effects in Web Surveys. *Public Opinion Quarterly*, 71(4), 623–634. <https://doi.org/10.1093/poq/nfm044>
- Couper, M. P., Tourangeau, R., Conrad, F. G., & Crawford, S. D. (2004). What They See Is What We Get: Response Options for Web Surveys. *Social Science Computer Review*, 22(1), 111–127. <https://doi.org/10.1177/0894439303256555>
- Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web Survey Design and Administration. *Public Opinion Quarterly*, 65(2), 230–253. <https://doi.org/10.1086/322199>
- Dedrick, R. F., Marfo, K., & Harris, D. M. (2007). Experimental Analysis of Question Wording in an Instrument Measuring Teachers’ Attitudes Toward Inclusive Education. *Educational and Psychological Measurement*, 67(1), 116–131. <https://doi.org/10.1177/0013164406292034>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining „gamification“. *Proceedings of the 15th International Academic*

- MindTrek Conference on Envisioning Future Media Environments - MindTrek '11, Tampere, Finland, 9–15.* <https://doi.org/10.1145/2181037.2181040>
- Deutskens, E., de Ruyter, K., & Wetzels, M. (2006). An Assessment of Equivalence Between Online and Mail Surveys in Service Research. *Journal of Service Research, 8*(4), 346–355. <https://doi.org/10.1177/1094670506286323>
- Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters, 15*(1), 21–36. <https://doi.org/10.1023/B:MARK.0000021968.86465.00>
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542–575.
- Dillman, D. A. (1991). The design and administration of mail surveys. *Annual review of sociology, 17*(1), 225–249.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2. Aufl.). Wiley.
- Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys. *Public Opinion Quarterly, 57*(3), 289. <https://doi.org/10.1086/269376>
- Downes-LeGuin, T. D.-L., Baker, R., Mechling, J., & Ruyle, E. (2012). Myths and Realities of Respondent Engagement in Online Surveys. *International Journal of Market Research, 54*(5), 613–633. <https://doi.org/10.2501/IJMR-54-5-613-633>
- Dudycha, A. L., & Carpenter, J. B. (1973). Effects of item format on item discrimination and difficulty. *Journal of Applied Psychology, 58*(1), 116–121. <https://doi.org/10.1037/h0035197>
- Eid, M., & Rauber, M. (2000). *Detecting Measurement Invariance in Organizational Surveys*. 11.
- Evans, J. R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. *Internet Research, 28*(4), 854–887. <https://doi.org/10.1108/IntR-03-2018-0089>
- Field, A. P. (2009). *Discovering statistics using SPSS: (And sex and drugs and rock „n“ roll)*. SAGE.
- French-Lazovik, G., & Gibson, C. L. (1984). Effects of Verbally Labeled Anchor Points on the Distributional Parameters of Rating Measures. *Applied Psychological Measurement, 8*(1), 49–57. <https://doi.org/10.1177/014662168400800106>
- Freunberger, D., & Nieuwland, M. S. (2016). Incremental comprehension of spoken quantifier sentences: Evidence from brain potentials. *Brain Research, 1646*, 475–481. <https://doi.org/10.1016/j.brainres.2016.06.035>

- Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly*, *69*(3), 370–392.
<https://doi.org/10.1093/poq/nfi027>
- Frieg, Dr. P., & Hossiep, Dr. G. R. (2020, Juni 20). Mitarbeiterbefragungen—Ein etablierter Klassiker bei den Unternehmen. *WIRTSCHAFTSPSYCHOLOGIE HEUTE*.
<https://www.wirtschaftspsychologie-heute.de/mitarbeiterbefragungen-ein-etablierter-klassiker-bei-den-unternehmen/>
- Funke, F. (2016). A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales. *Social Science Computer Review*, *34*(2), 244–254. <https://doi.org/10.1177/0894439315575477>
- Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the Smart: Type of Rating Scale on the Web Interacts With Educational Level. *Social Science Computer Review*, *29*(2), 221–231. <https://doi.org/10.1177/0894439310376896>
- Garcia, J., Evans, J., & Reshaw, M. (2004). Is There Anything Else You Would Like to Tell Us – Methodological Issues in the Use of Free-Text Comments from Postal Surveys. *Quality & Quantity*, *38*(2), 113–125.
<https://doi.org/10.1023/B:QUQU.0000019394.78970.df>
- Gaskell., G., Wright, D., & O’Muircheartaigh, C. (1993). Reliability of surveys. *The Psychologist*, *6*(11), 500–503.
- Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach’s alpha reliability coefficient for Likert-type scales*.
- Goyder, J. (1986). Surveys on surveys: Limitations and potentialities. *Public Opinion Quarterly*, *50*(1), 27–41.
- Grant, D., & Marshak, R. J. (2011). Toward a Discourse-Centered Understanding of Organizational Change. *The Journal of Applied Behavioral Science*, *47*(2), 204–235.
<https://doi.org/10.1177/0021886310397612>
- Griffin, M. A., Hart, P. M., & Wilson-Evered, E. (2000). Using employee opinion surveys to improve organizational health. In L. R. Murphy & C. L. Cooper (Hrsg.), *Healthy and productive work: An international perspective* (S. 15–36). Taylor & Francis.
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? -- A Literature Review of Empirical Studies on Gamification. *2014 47th Hawaii International Conference on System Sciences*, 3025–3034. <https://doi.org/10.1109/HICSS.2014.377>
- Harms, J., Biegler, S., Wimmer, C., Kappel, K., & Grechenig, T. (2015). Gamification of Online Surveys: Design Process, Case Study, and Evaluation. In J. Abascal, S. Barbosa,

- M. Fetter, T. Gross, P. Palanque, & M. Winckler (Hrsg.), *Human-Computer Interaction – INTERACT 2015* (S. 219–236). Springer International Publishing.
- Harms, J., Seitz, D., Wimmer, C., Kappel, K., & Grechenig, T. (2015). Low-Cost Gamification of Online Surveys: Improving the User Experience through Achievement Badges. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 109–113. <https://doi.org/10.1145/2793107.2793146>
- Harms, J., Wimmer, C., Kappel, K., & Grechenig, T. (2014). Gamification of online surveys: Conceptual foundations and a design process based on the MDA framework. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction Fun, Fast, Foundational - NordiCHI '14*, 565–568. <https://doi.org/10.1145/2639189.2639230>
- Hassenzahl, M. (2008). User experience (UX): Towards an experiential perspective on product quality. *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine on - IHM '08*, 11–15. <https://doi.org/10.1145/1512714.1512717>
- Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler (Hrsg.), *Mensch & Computer 2003* (Bd. 57, S. 187–196). Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-80058-9_19
- Hassenzahl, M., Platz, A., Burmester, M., & Lehner, K. (2000). Hedonic and ergonomic quality aspects determine a software's appeal. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '00*, 201–208. <https://doi.org/10.1145/332040.332432>
- Hassenzahl, M., & Tractinsky, N. (2006). User experience—A research agenda. *Behaviour & Information Technology*, 25(2), 91–97. <https://doi.org/10.1080/01449290500330331>
- Heberlein, T. A., & Baumgartner, R. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. *American Sociological Review*, 43(4), 447. <https://doi.org/10.2307/2094771>
- Heerwegh, D., Vanhove, T., Matthijs, K., & Loosveldt, G. (2005). The effect of personalization on response rates and data quality in web surveys. *International Journal of Social Research Methodology*, 8(2), 85–99. <https://doi.org/10.1080/1364557042000203107>
- Helgeson, J. G., & Ursic, M. L. (1994). The role of affective and cognitive decision-making processes during questionnaire completion. *Psychology and Marketing*, 11(5), 493–510. <https://doi.org/10.1002/mar.4220110506>

- Heracleous, L. (2002). The contribution of a discursive view to understanding and managing organizational change. *Strategic Change*, 11(5), 253–261.
<https://doi.org/10.1002/jsc.601>
- Hofmans, J., Theuns, P., Baekelandt, S., Mairesse, O., Schillewaert, N., & Cools, W. (2007). Bias and changes in perceived intensity of verbal qualifiers effected by scale orientation. *Survey Research Methods*, 1, 97–108.
- Hossiep, R., & Frieg, P. (2008). Der Einsatz von Mitarbeiterbefragungen in Deutschland, Österreich und der Schweiz. *Planung & Analyse*, 6, 55–59.
- Howell, D. C. (2010). *Statistical methods for psychology* (7th ed). Thomson Wadsworth.
- Iglesias, C., & Torgerson, D. (2000). Does Length of Questionnaire Matter? A Randomised Trial of Response Rates to a Mailed Questionnaire. *Journal of Health Services Research & Policy*, 5(4), 219–221. <https://doi.org/10.1177/135581960000500406>
- Jones, W. H., & Lang, J. R. (1980). Sample Composition Bias and Response Bias in a Mail Survey: A Comparison of Inducement Methods. *Journal of Marketing Research*, 17(1), 69. <https://doi.org/10.2307/3151119>
- Jordan, P. W. (1998). Human factors for pleasure in product use. *Applied Ergonomics*, 29(1), 25–33. [https://doi.org/10.1016/S0003-6870\(97\)00022-7](https://doi.org/10.1016/S0003-6870(97)00022-7)
- Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*, 7(4), 396–403.
<https://doi.org/10.9734/BJAST/2015/14975>
- Jung, J. H., Schneider, C., & Valacich, J. (2010). Enhancing the Motivational Affordance of Information Systems: The Effects of Real-Time Performance Feedback and Goal Setting in Group Collaboration Environments. *Management Science*, 56(4), 724–742.
<https://doi.org/10.1287/mnsc.1090.1129>
- Kaczmirek, L. (2008). *Human-Survey Interaction Usability and Nonresponse in Online Surveys* (Mannheim, Deutschland). 202.
- Kaczmirek, L. (2018). Attention and Usability in Internet Surveys: Effects of Visual Feedback in Grid Questions. In *Social and Behavioral Research and the Internet* (S. 191–214). Routledge.
- Karlberg, C. (2015). *The Survey Fatigue Challenge: Understanding young People's Motivation to participate in Survey Research Studies* [Lunds University, Department of Psychology]. <https://lup.lub.lu.se/student-papers/search/publication/7792773>

- Keppeler-Seid, H., Windle, C., & Woy, J. R. (1980). Performance measures for mental health programs: Something better, something worse, or more of the same? *Community Mental Health Journal*, 16(3), 217–234. <https://doi.org/10.1007/BF00835726>
- Keusch, F., Atkinson, P., Delamont, S., Cernat, A., Sakshaug, J. W., & Williams, R. A. (2020). *Gamification in web surveys*. <https://methods.sagepub.com/foundations/gamification-in-web-surveys>
- Keusch, F., & Zhang, C. (2017). A review of issues in gamified surveys. *Social Science Computer Review*, 35(2), 147–166. <https://doi.org/10.1177/0894439315608451>
- Kraut, A. I. (2006). *Getting action from organizational surveys: New concepts, technologies and applications*. Jossey Bass.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236. <https://doi.org/10.1002/acp.2350050305>
- Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In P. V. Marsden & J. D. Wright (Hrsg.), *Handbook of survey research* (S. 263–313). Emerald.
- Lam, T. C. M., & Klockars, A. J. (1982). Anchor Point Effects on the Equivalence of Questionnaire Items. *Journal of Educational Measurement*, 19(4), 317–322.
- Lam, T. C. M., & Stevens, J. J. (1994). Effects of content polarization, item wording and rating scale width on rating response. *Applied Measurement in Education*, 7(2), 141–158.
- Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., & Kort, J. (2009). *Understanding, Scoping and Defining User Experience: A Survey Approach*. 10.
- Lebow, J. (1982). Consumer satisfaction with mental health treatment. *Psychological Bulletin*, 91(2), 244–259. <https://doi.org/10.1037/0033-2909.91.2.244>
- Lenzner, T. (2011). *A psycholinguistic look at survey question design and response quality*. *Dissertation*. https://madoc.bib.uni-mannheim.de/29478/1/Lenzner_Dissertation.pdf
- Lenzner, T. (2012). Effects of Survey Question Comprehensibility on Response Quality. *Field Methods*, 24(4), 409–428. <https://doi.org/10.1177/1525822X12448166>
- Lenzner, T., Kaczmirek, L., & Galesic, M. (2011). Seeing Through the Eyes of the Respondent: An Eye-tracking Study on Survey Question Comprehension. *International Journal of Public Opinion Research*, 23(3), 361–373. <https://doi.org/10.1093/ijpor/edq053>
- Lietz, P. (2010). Research into Questionnaire Design: A Summary of the Literature. *International Journal of Market Research*, 52(2), 249–272. <https://doi.org/10.2501/S147078530920120X>

- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 5–55.
- Link, M. W., Lai, J., & Bristol, K. (2014). Not so fun? The challenges of applying gamification to smartphone measurement. *International Conference of Design, User Experience, and Usability*, 319–327.
- Low, G. (1995). ‘Hallelujah, very!’ responding to ‘very’ in questionnaire items. *Evaluation & Research in Education*, 9(1), 15–27. <https://doi.org/10.1080/09500799509533369>
- Low, G. (1996). Intensifiers and Hedges in Questionnaire Items and the Lexical Invisibility Hypothesis1. *Applied Linguistics*, 17(1), 1–37. <https://doi.org/10.1093/applin/17.1.1>
- Malinoff, B., & Puleston, J. (2011). How far is too far. *Traditional, flash and gamification interfaces, and implications for the future of market research online survey design (ESOMAR Publication Series Volume S250)*. Amsterdam, The Netherlands: ESOMAR.
- Manfreda, K. L., Batagelj, Z., & Vehovar, V. (2002). Design of Web Survey Questionnaires: Three Basic Experiments. *Journal of Computer-Mediated Communication*, 7(3). <https://doi.org/10.1111/j.1083-6101.2002.tb00149.x>
- Manfreda, K. L., Berzelak, J., Vehovar, V., Bosnjak, M., & Haas, I. (2008). Web Surveys versus other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research*, 50(1), 79–104. <https://doi.org/10.1177/147078530805000107>
- Mavletova, A. (2015). Web Surveys Among Children and Adolescents: Is There a Gamification Effect? *Social Science Computer Review*, 33(3), 372–398. <https://doi.org/10.1177/0894439314545316>
- McCarty, J. A., & Shrum, L. J. (2000). The Measurement of Personal Values in Survey Research. *Public Opinion Quarterly*, 64(3), 271–298. <https://doi.org/10.1086/317989>
- McPherson, J., & Mohr, P. (2005). The Role of Item Extremity in the Emergence of Keying-Related Factors: An Exploration With the Life Orientation Test. *Psychological Methods*, 10(1), 120–131. <https://doi.org/10.1037/1082-989X.10.1.120>
- Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Disassembling gamification: The effects of points and meaning on user motivation and performance. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, 1137. <https://doi.org/10.1145/2468356.2468559>
- Mueller, K., Voelkle, M. C., & Hatstrup, K. (2011). On the relationship between job satisfaction and non-response in employee attitude surveys: A longitudinal field study: Non-

- response in employee attitude surveys. *Journal of Occupational and Organizational Psychology*, 84(4), 780–798. <https://doi.org/10.1348/096317910X526777>
- Müller, K., Kempen, R., & Straatmann, T. (2021). *Mitarbeiterbefragung: Organisationales Feedback wirksam gestalten* (1. Auflage). Hogrefe.
- Nye, C. D., Newman, D. A., & Joseph, D. L. (2010). Never Say “Always”? Extreme Item Wording Effects on Scalar Invariance and Item Response Curves. *Organizational Research Methods*, 13(4), 806–830. <https://doi.org/10.1177/1094428109349512>
- O’Cathain, A., & Thomas, K. J. (2004). „Any other comments?“ Open questions on questionnaires – a bane or a bonus to research? *BMC Medical Research Methodology*, 4(1). <https://doi.org/10.1186/1471-2288-4-25>
- O’Muircheartaigh, C. A., Gaskell, G. D., & Wright, D. B. (1993). Intensifiers in Behavioral Frequency Questions. *Public Opinion Quarterly*, 57(4), 552. <https://doi.org/10.1086/269395>
- O’Muircheartaigh, C., Krosnick, J. A., & Helic, A. (2000). Middle alternatives, acquiescence, and the quality of questionnaire data. *American Association for Public Opinion Research Annual Meeting, St. Petersburg, FL*.
- Peterson, R. A., & Wilson, W. R. (1992). *Measuring customer satisfaction: Fact and artifact*. 11.
- Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. D. (2006). Web Survey Design: Paging versus Scrolling. *Public Opinion Quarterly*, 70(4), 596–607. <https://doi.org/10.1093/poq/nfl028>
- Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 2004(121), 63–73. <https://doi.org/10.1002/ir.101>
- Puleston, J., & Sleep, D. (2011). The game experiments. *Proceedings of the ESOMAR (European Society for Opinion and Marketing Research) Congress*, 10, 13.
- Purdey, B. (2013). Occupant stimulus response workplace productivity and the vexed question of measurement. *Facilities*, 31(11/12), 505–520. <https://doi.org/10.1108/F-03-2012-0021>
- Rammstedt, B., & Beierlein, C. (2014). Can’t We Make It Any Shorter?: The Limits of Personality Assessment and Ways to Overcome Them. *Journal of Individual Differences*, 35(4), 212–220. <https://doi.org/10.1027/1614-0001/a000141>

- Ran, S., Liu, M., Marchiondo, L. A., & Huang, J. L. (2015). Difference in Response Effort Across Sample Types: Perception or Reality? *Industrial and Organizational Psychology*, 8(02), 202–208. <https://doi.org/10.1017/iop.2015.26>
- Redline, C. D., & Lankford, C. P. (2001). Eye Movement Analysis: A New Tool for Evaluating the Design of Visually Administered Instruments (Paper and Web). *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 8.
- Rogelberg, S. G., Fisher, G. G., Maynard, D. C., Hakel, M. D., & Horvath, M. (2001). Attitudes toward Surveys: Development of a Measure and Its Relationship to Respondent Behavior. *Organizational Research Methods*, 4(1), 3–25. <https://doi.org/10.1177/109442810141001>
- Rohrman, B. (2007). *Verbal qualifiers for rating scales*: 28.
- Rolstad, S., Adler, J., & Rydén, A. (2011). Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health*, 14(8), 1101–1108. <https://doi.org/10.1016/j.jval.2011.06.003>
- Saari, L. M., & Scherbaum, C. A. (2011). *Identified employee surveys: Potential promise, perils, and a professional practice guidelines*. 4, 435–448. <https://doi.org/1754-9426/11>
- Schaufeli, W. B., & Bakker, A. B. (2004). *Utrecht work engagement scale: Preliminary manual*. Occupational Health Psychology, Utrecht University. https://www.wilmarschaufeli.nl/publications/Schaufeli/Test%20Manuals/Test_manual_UWES_English.pdf
- Schleifer, S. (1986). Trends in attitudes toward and participation in survey research. *Public Opinion Quarterly*, 50(1), 17–26.
- Schriesheim, C. A., Kopelman, R. E., & Solomon, E. (1989). The effect of grouped versus randomized questionnaire format on scale reliability and validity: A three-study investigation. *Educational and Psychological Measurement*, 49(3), 487–508.
- Schuman, H., & Presser, S. (1977). Question Wording as an Independent Variable in Survey Analysis. *Sociological Methods & Research*, 6(2), 151–170. <https://doi.org/10.1177/004912417700600202>
- Schwarz, N. (1999). How the Questions Shape the Answers. *American Psychologist*, 13.
- Silber, H., Lischewski, J., & Leibold, J. (2013). Comparing Different Types of Web Surveys: Examining Drop-Outs, Non-Response and Social Desirability. *Advances in Methodology and Statistics*, 10(2), 121–143.
- Sjoberg, G. (1954). A Questionnaire on Questionnaires. *Public Opinion Quarterly*, 18(4), 423. <https://doi.org/10.1086/266535>

- Smither, R. D., Houston, J. M., & McIntire, S. A. (2016). *Organization development: Strategies for changing environments* (Second edition). Routledge.
- Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When Two Factors Don't Reflect Two Constructs: How Item Characteristics Can Produce Artifactual Factors. *Journal of Management*, *23*(5), 659–677.
<https://doi.org/10.1177/014920639702300503>
- Stanton, J. M., Sinar, E. F., Balzer, W. K., Julian, A. L., Thoresen, P., Aziz, S., Fisher, G. G., & Smith, P. C. (2002). Development of a Compact Measure of Job Satisfaction: The Abridged Job Descriptive Index. *Educational and Psychological Measurement*, *62*(1), 173–191. <https://doi.org/10.1177/001316440206200112>
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, *25*(1), 78–107.
<https://doi.org/10.1086/209528>
- Sudman, S., & Bradburn, N. M. (1982). *Asking Questions*. Josey-Bass Inc. Publishers.
- Tanner, B. A., & Stacy, W. (1985). A validity scale for the sharp consumer satisfaction scales. *Evaluation and Program Planning*, *8*(2), 147–153. [https://doi.org/10.1016/0149-7189\(85\)90009-6](https://doi.org/10.1016/0149-7189(85)90009-6)
- Toepoel, V., Das, M., & Van Soest, A. (2008). Effects of Design in Web Surveys. *Public Opinion Quarterly*, *72*(5), 985–1007. <https://doi.org/10.1093/poq/nfn060>
- Toepoel, V., Das, M., & Van Soest, A. (2009). Design of Web Questionnaires: The Effects of the Number of Items per Screen. *Field Methods*, *21*(2), 200–213.
<https://doi.org/10.1177/1525822X08330261>
- Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly*, *68*(3), 368–393. <https://doi.org/10.1093/poq/nfh035>
- Tourangeau, R., & Rasinski, K. A. (1988). *Cognitive Processes Underlying Context Effects in Attitude Measurement*. 16.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511819322>
- Triantoro, T., Gopal, R., Benbunan-Fich, R., & Lang, G. (2019). Would you like to play? A comparison of a gamified survey with a traditional online survey method. *International Journal of Information Management*, *49*, 242–252.
<https://doi.org/10.1016/j.ijinfomgt.2019.06.001>

- Turner, G., van Zoonen, L., & Adamou, B. (2014). *Research through Gaming: Public Perceptions of (the Future of) Identity Management*. SAGE Publications, Ltd. <https://doi.org/10.4135/978144627305013496519>
- Vance, R. J. (2006). *Employee engagement and commitment: A guide to understanding, measuring and increasing engagement in your organization*. 53.
- Vehovar, V., Manfreda, K. L., Batagelj, Z., & Center, C. (2000). Design issues in web surveys. *Proceedings of the American Statistical Association, Survey Research Methods Section*, 983–988.
- Villar, A., Callegaro, M., & Yang, Y. (2013). Where Am I? A Meta-Analysis of Experiments on the Effects of Progress Indicators for Web Surveys. *Social Science Computer Review*, 31(6), 744–762. <https://doi.org/10.1177/0894439313497468>
- von der Gracht, H. A., Vennemann, C. R., & Darkow, I.-L. (2010). Corporate foresight and innovation management: A portfolio-approach in evaluating organizational development. *Futures*, 42(4), 380–393. <https://doi.org/10.1016/j.futures.2009.11.023>
- Walker, R. W., & Cook, W. A. (2013). You Can't Put a Price Tag on A Survey Participant's Enjoyment: The Latest Findings from The ARF's "Foundations of Quality" Research. *Journal of Advertising Research*, 53(3), 254–257. <https://doi.org/10.2501/JAR-53-3-254-257>
- Walston, J. T., Lissitz, R. W., & Rudner, L. M. (2006). The Influence of Web-based Questionnaire Presentation Variations on Survey Cooperation and Perceptions of Survey Quality. *Journal of Official Statistics*, 21.
- Ward, M. K., & Pond, S. B. (2015). Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys. *Computers in Human Behavior*, 48, 554–568. <https://doi.org/10.1016/j.chb.2015.01.070>
- Weijters, B., & Baumgartner, H. (2012). Misresponse to Reversed and Negated Items in Surveys: A Review. *Journal of Marketing Research*, 49(5), 737–747. <https://doi.org/10.1509/jmr.11.0368>
- Weijters, B., Geuens, M., & Baumgartner, H. (2013). The Effect of Familiarity with the Response Category Labels on Item Response to Likert Scales. *Journal of Consumer Research*, 40(2), 368–381. <https://doi.org/10.1086/670394>
- Whelan, T. J. (2015). *Response Rates in 21st Century Organizational Survey Research: A Conceptual Model and Meta-analysis* [Dissertation, North Carolina State University]. <https://repository.lib.ncsu.edu/bitstream/handle/1840.16/10384/etd.pdf?sequence=1>

- Wildt, A. R., & Mazis, M. B. (1978). Determinants of Scale Response: Label versus Position. *Journal of Marketing Research*, *15*(2), 261–267.
<https://doi.org/10.1177/002224377801500209>
- Wright, D. B., Gaskell, G. D., & O’Muircheartaigh, C. A. (1995). Testing the Multiplicative Hypothesis of Intensifiers. *Applied Cognitive Psychology*, *9*, 167–177.
- Wyatt, R. C., & Meyers, L. S. (1987). Psychometric Properties of Four 5-Point Likert Type Response Scales. *Educational and Psychological Measurement*, *47*(1), 27–35.
<https://doi.org/10.1177/0013164487471003>
- Yentes, R. D., Toaddy, S. R., Thompson, L. F., Gissel, A. L., & Stoughton, J. W. (2012). *Effects of Survey Progress Bars on Data Quality and Enjoyment*. [Data set]. American Psychological Association. <https://doi.org/10.1037/e518332013-801>
- Yu, J., & Cooper, H. (1983). A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. *Journal of Marketing Research*, *20*(1), 36.
<https://doi.org/10.2307/3151410>
- Zhan, J., Jiang, X., Politzer-Ahles, S., & Zhou, X. (2017). Neural correlates of fine-grained meaning distinctions: An fMRI investigation of scalar quantifiers: Neural Correlates of Fine-Grained Meaning Distinctions. *Human Brain Mapping*, *38*(8), 3848–3864.
<https://doi.org/10.1002/hbm.23633>
- Zhang, C., & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey research methods*, *8*(2), 127–135.
- Zhang, Y., Waldman, D. A., Han, Y.-L., & Li, X.-B. (2015). Paradoxical Leader Behaviors in People Management: Antecedents and Consequences. *Academy of Management Journal*, *58*(2), 538–566. <https://doi.org/10.5465/amj.2012.0995>
- Zukerberg, A., Nichols, E., & Tedesco, H. (1999). Designing surveys for the next millennium: Internet questionnaire design issues. *National Center for Education Statistics*, 79.