

Einfluss des verwendeten Endgeräts auf das Nutzungsverhalten in Online-Befragungen

Helge Nissen, Monique Janneck

Fachbereich Elektrotechnik und Informatik, Technische Hochschule Lübeck

Zusammenfassung

Durch die steigende Nutzung von mobilen Endgeräten werden auch Online-Fragebogen zunehmend auf Smartphones bearbeitet. Ein Responsive Design von Fragebogen ermöglicht die Bearbeitung mit unterschiedlichen Geräteklassen. Dennoch stellt sich die Frage, inwiefern die erhobenen Daten insbesondere bei komplexen wissenschaftlichen Befragungen durch das verwendete Endgerät beeinflusst werden. Der vorliegende Beitrag untersucht diese Frage anhand einer umfangreichen Online-Befragung mit einem aktuellen responsiven Fragebogen-Template, woraus sich Hinweise für dessen Eignung für Forschungsarbeiten ergeben. 971 Personen nahmen an der Befragung teil. Es wurden die Bearbeitungszeit, die Abbruchquote und das Antwortverhalten zwischen auf kleinen Displays beantworteten Fragebogen und auf größeren Displays beantworteten Fragebogen verglichen. Die Ergebnisse zeigen eine statistisch signifikante Überlegenheit größerer Displays bei der Bearbeitungszeit und der Abbruchquote gegenüber kleineren Displays. Zudem zeigte sich ein gerätespezifisches Antwortverhalten.

1 Einleitung und Stand der Forschung

Online-Fragebogen werden in verschiedenen Forschungsdisziplinen eingesetzt, da sie für Wissenschaftler mit ihrer Flexibilität und als gute Grundlage zur Auswertung vergleichsweise viele Vorteile bieten. Durch den steigenden mobilen Zugriff steigt auch die Relevanz von für kleinere Displays optimierten Fragebogenoberflächen. So wird die Bearbeitung von webbasierten Fragebogen schon seit längerer Zeit nicht mehr nur auf Desktop-Computern erledigt (Callegaro, 2010). Das Bearbeiten von webbasierten Fragebogen mit dem Smartphone steigt stetig. Im Jahre 2011 wurden noch 4% der Teilnahmen über ein Smartphone abgeschlossen, während es im Jahre 2014 schon 18% waren (Sarraf, Brooks, & Cole, 2014). Ferner gibt es wenige Forschungsarbeiten, die eine freie Geräteauswahl als Grundlage haben, sodass die Erhebung aktueller Zahlen zur Gerätenutzung in Online-Umfragen durch den vorliegenden Beitrag zu leisten ist. Insbesondere bei jungen Menschen sind Smartphones

verbreitet (Lugtig, Toepoel, & Amin, 2016). Eine Optimierung würde also speziell die Reichweite in dieser Personengruppe erhöhen. Umgekehrt könnte eine mangelhafte Nutzbarkeit auf mobilen Geräten mit einer Unterrepräsentation dieser Gruppe einhergehen. Gerade in Situationen des Wartens werden Smartphones häufig verwendet (Thorsteinsson & Page, 2014), was großes Potential für die Bearbeitung von Fragebogen offenbart.

Auf den ersten Blick erscheint die Bearbeitung eines umfassenden wissenschaftlichen Fragebogens auf einem mobilen Gerät wenig sinnvoll. Bisherige Untersuchungen zeigen dies beispielsweise an der Abbruchquote. Demnach treten Abbrüche in der Bearbeitung eines Fragebogens vermehrt auf dem Smartphone auf (Mavletova, 2013; Sarraf et al., 2014). Ein weiterer häufig für Vergleiche zwischen den Geräten herangezogener Parameter ist die Bearbeitungszeit. Diesbezüglich zeigen verschiedene Studien, dass die benötigte Zeit zur Bearbeitung eines Fragebogens auf dem Smartphone signifikant höher ist (Horwitz, 2014; Lugtig & Toepoel, 2016; Mavletova, 2013). Gründe dafür werden bislang vor allem im unangepassten Layout gesehen, beispielsweise sind Tabellen-Layouts auf dem Smartphone eher problematisch (Couper & Peterson, 2016). Auch ein Zoom für die bessere Interaktion auf zu kleinen Flächen wurde beobachtet (Olmsted-Hawala, Nichols, Holland, & Gareau, 2016) und kann für eine längere Bearbeitungsdauer verantwortlich gemacht werden. Weiterhin wird untersucht, ob auf Smartphones erhobene Datensätze auffällige Effekte, beispielsweise Primacy-Effekte, sichtbar machen. Dies wurde jedoch mehrfach widerlegt (Mavletova, 2013; Wells, Bailey & Link, 2014). Dennoch gilt es stets zu beachten, ob hinsichtlich der Geräte signifikante Unterschiede im Antwortverhalten zu beobachten sind, um ein gerätespezifisches Antwortverhalten ausschließen zu können.

Auch Layouts mit einem höheren Anpassungsgrad an Smartphones werden für Untersuchungen herangezogen. So wird bei einer vertikalen Darstellung einer Frage pro Seite und ausreichend großen Antwortbuttons auf dem Smartphone jedoch ebenfalls eine längere Bearbeitungszeit nachgewiesen (Andreadis, 2015).

Dass Websites heute auch für mobile Geräte optimiert werden, ist offenkundig. Wie verhält es sich jedoch mit dem *Responsive Design* für Fragebogenoberflächen? In der diesem Beitrag zugrundeliegenden Studie wird ein Standard-Template experimentell verwendet, um gerätespezifische Unterschiede herauszuarbeiten. Anders als in vorangegangenen Studien (Andreadis, 2015; Mavletova & Couper, 2014; Wells et al., 2014) wird hier ein einziger Fragebogen für mehrere Geräteklassen bereitgestellt, der über ein heute gebräuchliches Responsive Design auch am Smartphone zu betrachten ist. Der damit inhaltlich identische Fragebogen unterscheidet sich lediglich darin, dass die Anordnung der Antwortmöglichkeiten auf kleineren Displays vertikal dargestellt wird. Dieses Template erfährt keine besonderen Anpassungen für die mobile Nutzung. Um valide Aussagen zur Eignung herkömmlicher Templates treffen zu können, wird zwischen den Geräteklassen hinsichtlich der benötigten Bearbeitungszeit, der Abbruchquote und dem möglichen Aufkommen inhaltlicher Verzerrungen unterschieden. Weiterhin werden im Gegensatz zu anderen Studien (Andreadis, 2015; Buskirk & Andrus, 2014) keine inhaltlichen Anpassungen für kleinere Displays vorgenommen, sodass auch offene Fragen enthalten sind. Dieses Vorgehen zielt schließlich auf Ergebnisse ab, die eine relevante Entscheidungsgrundlage für Forschende bieten können und zudem die gerätevergleichende Online-Forschung weiter voranbringen sollen.

2 Methode

2.1 Studiendesign

Zur Untersuchung der o.g. Fragestellungen wurde im Dezember 2017 eine Online-Studie mit einer Laufzeit von zwei Wochen durchgeführt. Die Probanden wurden über einen Panel-Anbieter akquiriert und für die Teilnahme von diesem Anbieter entlohnt. Weiterhin wurde vom Anbieter stets auf die Möglichkeit der mobilen Teilnahme hingewiesen.

Um die Vergleichbarkeit der Ergebnisse zu gewährleisten, wurden unabhängige Messungen durchgeführt. Eine Gruppe wird nicht durch Messwerte einer anderen Gruppe beeinflusst.

Für die Untersuchung wurde das frei verfügbare und breit genutzte Tool Limesurvey verwendet. Die Anwendung wurde eigenständig betrieben, um vollen Zugriff auf die Datenbank (MySQL) zu erhalten und eigene Skripte (PHP, Javascript) zu integrieren. Für den vorliegenden Beitrag wurde anders als in bisherigen Studien keine spezielle Entwicklung für mobile Geräte betrieben, sondern das Standard-Template in Limesurvey verwendet. Als Darstellungsformen wurden Gruppierungen aus Fragen auf verschiedenen Seiten sowie eine Auflistung auf einer Seite eingesetzt. Die dadurch erzeugte Fragebogenoberfläche passt sich bis zu einem gewissen Maße an das verwendete Gerät an. So werden auf dem Smartphone beispielsweise die Antwortmöglichkeiten sowie der Fragentext vertikal angeordnet. Von weiteren technischen Anpassungen der Oberfläche wurde abgesehen, um ein realistisches Szenario für Forschende zu erzeugen, die typischerweise Standard-Tools verwenden und über keine Kapazitäten für die Entwicklung eigener Fragebogenanwendungen und -layouts verfügen.

Der für die vorliegende Untersuchung verwendete Fragebogen enthält 133 Items und besteht u.a. aus standardisierten Fragebogen zur Untersuchung von Persönlichkeitsmerkmalen (Gosling, Rentfrow, & Swann, 2003) und der Motivation (Tondello et al., 2016). Die Bearbeitung wird aufgrund der Länge als verhältnismäßig aufwendig eingeschätzt. Neben Fragen, die mit einer Antwortoption beantwortet werden konnten, befanden sich auch Bewertungsskalen (1-7, 1-5) und offene Fragen im Fragebogen. Die Bildschirmbreite des verwendeten Geräts wurde zu Beginn der Umfrage eingelesen und in der Limesurvey-Datenbank gespeichert.

Für die Untersuchungen in diesem Beitrag wird zwischen kleinen Displays bis zu einer Bildschirmbreite von 768 px und großen Displays ab einer Bildschirmbreite von mehr als 1024 px unterschieden. Dazwischenliegende Bildschirmbreiten entsprechen denen auf Tablets, die jedoch nicht in den Untersuchungen zu diesem Beitrag betrachtet werden. Diese Einteilung orientiert sich an heute gängigen Media-Queries und daraus resultierenden Breakpoints im Responsive Design (Natda, 2013). Kleine Displays können dabei Smartphones und Tablets im Hochformat sein, während mit großen Displays überwiegend Desktop-Computer oder größere Laptops zusammengefasst werden. In den folgenden Ausführungen werden gemäß dieser Einteilung überwiegend die Begriffe *Smartphones* und *Desktop-Computer* verwendet.

2.2 Hypothesen

Um die Smartphone-Eignung von Fragebogen zu bewerten, orientieren wir uns im Rahmen dieses Beitrags an den Parametern der Bearbeitungszeit, der Abbruchquote und des Antwortverhaltens. Der Abbruch einer Teilnahme erzeugt einen unvollständigen Datensatz, welcher in vielen Fällen nicht für eine Auswertung herangezogen werden kann. Weiterhin ist eine Auswertung auf Skalen-Ebene bei einer Vielzahl fehlender Werte schwerer möglich und verliert dadurch häufig an Reliabilität. Darüber hinaus steht die Effizienz, hier hinreichend bestimmt durch die Bearbeitungszeit, bei der Nutzung von Anwendungen im unmittelbaren Zusammenhang mit ihrer Gebrauchstauglichkeit. Um gerätespezifische Verzerrungen der Daten ausschließen zu können, wird außerdem das Antwortverhalten untersucht. Es sind geräteabhängige Effekte vorstellbar, beispielsweise eine stärkere Tendenz der ersten Antwortoption durch eine einfachere Erreichbarkeit auf dem Smartphone. Zur Untersuchung des Antwortverhaltens wurde zunächst eine allgemeine Variable errechnet, indem alle Fragen, auf die mit einer Auswahl geantwortet werden konnte, codiert wurden. Jeweils die erste Antwort (auf dem Desktop-PC links, auf dem Smartphone oben) wurde mit einer 1 codiert, die letzte Antwort mit einer 5 bzw. einer 7. Daraus konnte schließlich ein Mittelwert berechnet werden, der inhaltlich keine Relevanz hat, für einen Vergleich zwischen den Geräten aber herangezogen werden kann. Zusätzlich wurden einzelne der in der Befragung enthaltenen standardisierten Fragebogen separat inhaltlich ausgewertet, um einen Bezug zur tatsächlichen Anwendung für Forschende herzustellen.

Hierzu werden die folgenden Hypothesen aufgestellt, die auf dem in Abschnitt 1 dargestellten Stand der Forschung aufbauen:

- H1: Die benötigten Bearbeitungszeiten der Gruppen „Smartphone“ (G1) und „Desktop“ (G2) unterscheiden sich signifikant.
- H2: Die Abbruchquoten der Gruppen „Smartphone“ (G1) und „Desktop“ (G2) unterscheiden sich signifikant.
- H3: Das Antwortverhalten unterscheidet sich nicht zwischen den Gruppen „Smartphone“ (G1) und „Desktop“ (G2).

3 Ergebnisse

Insgesamt nahmen an der Studie 971 Personen (Frauen: 489, Männer: 474, andere: 1) im Alter zwischen 18 und 79 Jahren teil. Davon bearbeiteten 327 (33,7 %) den Fragebogen auf einem Smartphone und 644 (66,3%) auf einem Desktop-Rechner.

3.1 Bearbeitungszeit

Die Bearbeitungszeit wurde in Limesurvey erfasst und für jeden Fall in die Datenbank geschrieben. Im Datensatz befinden sich bei der Betrachtung der deskriptiven Statistik der

Bearbeitungszeit unabhängig vom verwendeten Gerät einige Ausreißer. Bei besonders extremen Werten von annähernd 60 000 Sekunden ist davon auszugehen, dass der Fragebogen pausiert und am nächsten Tag fortgesetzt wurde. Um eine Aussage über die tatsächliche Bearbeitungszeit zu erhalten, werden für die Auswertung nur noch Fälle bis zu einer Dauer von 2500 Sekunden betrachtet. Für diesen Zeitrahmen ist für den überwiegenden Teil der Fälle von einer pausenlosen Bearbeitung auszugehen. Außerdem stellt diese festgelegte Grenze etwa das Doppelte der mittleren Bearbeitungszeit dar. Weiterhin werden für die Untersuchung der benötigten Bearbeitungszeit naturgemäß nur abgeschlossene Fälle einbezogen. Die zu untersuchende Stichprobe ist nach Auswahl dieser Fälle noch $N = 826$.

Beide Gruppen sind gemäß Shapiro-Wilk-Test nicht normalverteilt ($p < .001$). Da die Stichprobe für die Gruppen mit $N_{G1} = 272$ bzw. $N_{G2} = 554$ verhältnismäßig groß ist, kann der t-Test jedoch verwendet werden, um systematische Unterschiede zwischen den benötigten Bearbeitungszeiten in den beiden Gruppen zu finden. Die abhängige Variable ist dabei die Zeit, während als unabhängige Variable die Gruppenzugehörigkeit (Smartphone, G1 bzw. Desktop, G2) verwendet wird.

Die benötigte Bearbeitungszeit auf kleineren Displays ($M = 1205$ sek., $SD = 468$ sek.) ist höher als auf dem Desktop-PC ($M = 1089$, $SD = 478$). Eine Übersicht über die ermittelten Werte zeigt Tabelle 1.

Display	Anzahl	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
w < 769 px (G1)	272	1204.7192	468.23549	28.39095
w > 1024 px (G2)	554	1089.2442	478.40920	20.32566

Tabelle 1: Vergleich der benötigten Zeit

Der Test auf Homogenität der Varianzen (Levene-Test) für die benötigte Bearbeitungszeit wird mit $p = .974$ statistisch nicht signifikant. Es kann also von einer Homogenität der Varianzen ausgegangen werden. Es zeigt sich ein statistisch signifikanter Unterschied zwischen den benötigten Bearbeitungszeiten der beiden Gruppen Smartphone (G1) und Desktop (G2). Die Bearbeitungszeit liegt beim Smartphone durchschnittlich 115.47 Sekunden höher (95%, CI[46.4333, 184.517], $t(824) = 3.283$, $p < .001$, Cohen's $d = .244$).

Hypothese 1 kann somit angenommen werden: Auf Smartphones ist das Bearbeiten von Fragebogen mit einem höheren Zeitaufwand verbunden als auf Desktop-Computern.

3.2 Abbruchquote

Da im vorliegenden Fragebogen alle Fragen verpflichtend beantwortet werden mussten, wird als Abbruch jeder Datensatz gewertet, in dem mindestens ein Wert fehlt. Der Zeitpunkt des Abbruchs und die damit verbundene Anzahl unbeantworteter Fragen werden in diesem Beitrag nicht betrachtet. Die Abbruchquote wurde berechnet, indem in Abhängigkeit der letzten

Variable eine 1 (für Abbruch) bzw. eine 0 (bei vollständiger Bearbeitung) vergeben wurde. Die Quote errechnet sich daraufhin als Mittelwert dieser Variable über alle Fälle einer Gruppe (G1, Smartphone vs. G2, Desktop). Der Test auf Homogenität der Varianzen (Levene-Test) für das Antwortverhalten wird statistisch signifikant mit $p < .001$. Es kann also nicht von einer Homogenität der Varianzen ausgegangen werden. Aus diesem Grund wird im Folgenden der Welch-Test interpretiert. Beide Gruppen sind gemäß Shapiro-Wilk-Test nicht normalverteilt. ($p < .001$).

Display	Anzahl	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
w < 769 px (G1)	327	.097859	.297580	.016456
w > 1024 px (G2)	644	.060559	.238705	.009406

Tabelle 2: Vergleich der Abbruchquote

In der Gruppe G1 werden von 327 Teilnahmen 32 Abbrüche (9,8 %) festgestellt, während in der Gruppe G2 von 644 Teilnahmen 39 Abbrüche (6,1 %) zu verzeichnen sind.

Es zeigt sich ein signifikanter Unterschied zwischen den errechneten Mittelwerten der beiden Gruppen (G1) und (G2). Der Mittelwert ist in der Gruppe G2 um .037 höher (95%, CI [.001, .075], $t(544.361) = 1.968$, $p = .049592$, Cohen's $d = .138$). Eine Übersicht über die Auswertung der Abbruchquote zeigt Tabelle 2.

Hypothese H2 kann folglich angenommen werden: Auf Geräten mit kleineren Displays, hier mit einer Bildschirmbreite bis 768 px, wird die Bearbeitung des Fragebogens im Vergleich zu Geräten mit größeren Displays, hier mit einer Bildschirmbreite ab 1024 px, vermehrt abgebrochen.

3.3 Antwortverhalten

Die Mittelwerte zwischen den Gruppen Smartphone (G1) und Desktop (G2) wurden mit einem ungepaarten t-Test verglichen. Die abhängige Variable ist dabei der Mittelwert, während die unabhängige Variable die Gruppe darstellt.

Display	Anzahl	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
w < 769 px (G1)	323	4.2534	.56904	.03166
w > 1024 px (G2)	641	4.3555	.44324	.01751

Tabelle 3: Vergleich des Antwortverhaltens

Der Test auf Homogenität der Varianzen (Levene-Test) für das Antwortverhalten wird statistisch nicht signifikant mit $p = .095$. Es kann also von einer Homogenität der Varianzen ausgegangen werden. Beide Gruppen sind gemäß Shapiro-Wilk-Test nicht normalverteilt ($p < .001$).

Es zeigt sich ein signifikanter Unterschied zwischen den errechneten Mittelwerten der beiden Gruppen Smartphone (G1) und Desktop (G2). Der Mittelwert ist in der Gruppe G2 um $.1021$ höher (95%, CI[-.16754, -.03659], $t(962) = -3.059$, $p = .002$, Cohen's $d = -.200$). Die Standardabweichung ist für die Gruppe G1 um $.1258$ höher. Eine Übersicht über die Auswertung des Antwortverhaltens ist in Tabelle 3 dargestellt.

Die Untersuchung auf gerätespezifische Unterschiede des Fragebogens zur Motivation (Tondello et al., 2016) ergibt über alle sechs Skalen einen geringeren Mittelwert für die Gruppe G1, für zwei der sechs Skalen, „Free Spirit“ $t(898) = -2.476$, $p = .013$ und „Achiever“ $t(898) = -2.177$, $p = .030$ sogar einen signifikanten Unterschied der Mittelwerte. Die Auswertung des in der Studie enthaltenen Fragebogens zu Persönlichkeitsmerkmalen (Gosling et al., 2003) ergab bis auf die Skala „Extraversion“ stets einen niedrigeren Mittelwert für die Gruppe G1. Der t-Test für den Unterschied der Mittelwerte wurde für die Skalen „Conscientiousness“ mit $t(905) = -2.775$, $p = .006$ und „Emotional Stability“ mit $t(905) = -2.105$, $p = .036$ sogar signifikant.

Die Hypothese H3 kann demnach nicht angenommen werden: Es zeigen sich Unterschiede bezüglich des Antwortverhaltens zwischen den Geräten.

4 Diskussion und Ausblick

Für den vorliegenden Beitrag wurde eine vergleichende Studie zwischen der Nutzung von Smartphones und Desktop-Computern bei der Bearbeitung von wissenschaftlichen Fragebogen durchgeführt. Als konkrete Untersuchungsparameter dienten die benötigte Bearbeitungszeit, die Abbruchquote und das Antwortverhalten. Anders als in vorangegangenen Studien wurde ein Standard-Template einer frei verfügbaren Software verwendet, von der heute eine automatische Anpassung an die Bildschirmbreite erwartet werden kann. Zusammenfassend zeigt sich, dass die untersuchten Parameter auf Smartphones tendenziell schlechter ausfallen bzw. sich sogar Unterschiede im Antwortverhalten zeigen, was bei der Bewertung der Ergebnisse zu berücksichtigen ist. Die durch diese Studie erlangten Ergebnisse können als Entscheidungsgrundlage für die praktische Planung zukünftiger wissenschaftlicher Untersuchungen dienen.

Bei einer durchschnittlichen *Bearbeitungszeit* von 1089 bzw. 1205 Sekunden beträgt die benötigte Zeit auf Smartphones im Durchschnitt 116 Sekunden länger. In Relation zu einer für die vorliegende Studie benötigten Gesamtbearbeitungszeit von etwa 20 Minuten erscheint eine um etwa zwei Minuten längere Dauer jedoch moderat. Die praktische Relevanz dieses Wertes, insbesondere bei der Frage, ob eine mobile Version des Fragebogens bereitgestellt werden sollte, gilt es individuell zu bewerten. Zukünftige Studien sollten untersuchen, ob

auch bei kürzeren Fragebogen ein Unterschied auftritt und ab welcher Länge ein Unterschied überhaupt feststellbar ist.

Bei der Betrachtung der *Abbruchquoten* gilt es zu berücksichtigen, dass die Teilnehmenden über einen externen Panel-Anbieter akquiriert und von diesem auch für eine abgeschlossene Teilnahme entschädigt wurden. Somit ist von einer geringeren Abbruchquote im Vergleich zu Befragungen ohne Incentivierung auszugehen. Weiterhin sind durch die Wahl der Darstellungsformen, insbesondere durch die Auflistung der Fragen auf einer Seite, nicht alle tatsächlichen Abbrüche registrierbar. Die Vergleichbarkeit zwischen den beiden Gruppen kann dennoch als gegeben erachtet werden, da dies gruppenübergreifend galt. Weiterhin ist zu berücksichtigen, dass neben Fragen mit einfacher Auswahl einer Antwort auch Fragen mit einer freien Textantwort im Fragebogen enthalten waren. Diese Fragen sind unabhängig vom Grad der Anpassung der Fragebogenoberfläche an mobile Geräte allein wegen der Texteingabe über die virtuelle Tastatur auf dem Smartphone vergleichsweise mit mehr Schwierigkeiten verbunden (Page, 2013). Vor diesem Hintergrund scheint die etwas höhere Abbruchquote auf dem Smartphone zwar nicht gänzlich relativierbar zu sein, weitere Untersuchungen mit reinen Auswahlfragen könnten jedoch zu einer Annäherung der beiden Quoten führen.

Ergänzend zur Abbruchquote erscheint für zukünftige Studien die Analyse fehlender Werte relevant. Da in der vorliegenden Untersuchung alle Fragen verpflichtend beantwortet werden mussten, konnte diese Variable nicht betrachtet werden. Online-Fragebogen verwenden jedoch häufig auch optionale Fragen, um zu verhindern, dass das Nichtbeantworten einer Frage zum vollständigen Abbruch führt, gleichzeitig vermindert eine hohe Anzahl fehlender Werte die Datenqualität. Sollten sich hierbei Unterschiede zwischen verschiedenen Geräteklassen zeigen, wäre das für die Planung und Interpretation von Online-Studien von erheblicher Relevanz.

Der allgemeine für den Vergleich des *Antwortverhaltens* errechnete Wert hat keine inhaltliche Bedeutung und wurde lediglich für die Vergleichbarkeit bestimmt. Um dieses Ergebnis mit einem höheren Maß an inhaltlicher Relevanz zu bestätigen, wurden zudem zwei der enthaltenen standardisierten Fragebogen ausgewertet. Alle drei Einzeluntersuchungen zeigten einen Effekt der Geräteklasse. Da die Mittelwerte auf dem kleineren Gerät nahezu bei jeder Variable und teilweise sogar signifikant niedriger sind, liegt die Interpretation nahe, dass auf diesen Geräten tendenziell eher eine der oberen Antwortmöglichkeiten ausgewählt wurde, z.B. weil die Teilnehmenden auf dem Smartphone eher bestrebt waren, die Umfrage schnell und flüchtig zu beenden. Zwar legt die Systematik der beobachteten Unterschiede (stets geringerer Mittelwert bei Beantwortung auf dem Smartphone) nahe, dass diese auf die Art der Bearbeitung des Fragebogens auf dem Endgerät zurückzuführen ist, jedoch kann nicht ausgeschlossen werden, dass ein tatsächlicher inhaltlicher Zusammenhang zugrunde liegt, z.B. weil sich die Probanden der Gruppen „Smartphone“ bzw. „Desktop“ systematisch unterscheiden. In zukünftigen Untersuchungen planen wir daher, den Einfluss von Variablen wie Alter und Geschlecht auf das Antwortverhalten zu analysieren.

Insgesamt gilt es, die möglichen Vorteile des Einsatzes von Smartphones bei der Beantwortung von Fragebogen zu analysieren und diese den hier aufgezeigten Nachteilen gegenüberzustellen. Im Hinblick auf Bearbeitungszeit und Abbruchquote erscheinen die negativen Effekte vor dem Hintergrund der verbreiteten Smartphone-Nutzung und der damit verbunde-

nen Reichweite eher gering. Der festgestellte Effekt auf das Antwortverhalten ist hingegen als kritischer zu betrachten, da dies die Qualität der erhobenen Daten unmittelbar berührt. Allerdings fanden frühere Studien keine Unterschiede im Antwortverhalten (Mavletova, 2013; Wells, Bailey & Link, 2014). In diesen Studien wurde jedoch lediglich ein Unterschied bezüglich des Primacy-Effekts mithilfe einer vertauschten Reihenfolge der Antwortmöglichkeiten untersucht. Im vorliegenden Beitrag hingegen stand der Vergleich zwischen den Geräten und die Frage, ob es Unterschiede in der Beantwortung gibt, im Mittelpunkt. Wir planen weitere Untersuchungen, um zu prüfen, ob sich die hier beobachteten Effekte replizieren lassen. Dabei sollen auch mögliche Stichprobeneffekte (z.B. der Einfluss von Alter, Geschlecht, Erfahrungen im Umgang mit Technologie...) untersucht werden.

Weitere Studien sollten neue, besser auf Smartphones angepasste Fragebogenoberflächen und deren Potential für Verbesserungen untersuchen. Beispielsweise ließe sich der Zeitwert verbessern, wenn die Bedienung auf dem Smartphone effizienter möglich wäre. Fragebogen-Templates, die vom hier verwendeten Standard abweichen und eine andere Einteilung von Fragen und Seiten verwenden (vgl. Andreadis, 2015; de Bruijne & Wijnant, 2013), sollten mit den hier erzielten Ergebnissen verglichen werden. Weiterhin wären Fragebogen ohne freie Textantworten bezüglich der Abbruchquote genauer zu untersuchen, da hier die Schwächen des Smartphones in der Texteingabe weniger ausgeprägt sein sollten.

Zur Bewertung der Ergebnisse sollte jedoch herangezogen werden, dass durch das Setting dieser Studie mit einem hohen Fragenumfang und zahlreichen Freitexteingaben die Nachteile von Smartphones besonders zur Ausprägung gekommen sein können. Zukünftige Studien sollten die hier betrachteten Parameter mit einer Variation in der Fragebogendarstellung untersuchen. Denkbar wäre eine veränderte Anzahl der pro Seite dargestellten Fragen. Auch die Weiterleitung zur nächsten Seite kann ein zum Abbruch oder schlechter Datenqualität führender Faktor sein, den es zu untersuchen gilt. Daneben bieten Usability-Fragen, insbesondere zur Erreichbarkeit der Antworten auf kleineren Displays, Forschungspotential.

Literaturverzeichnis

- Andreadis, I. (2015). Web Surveys Optimized for Smartphones. *Methods, Data, Analyses*, 9(2), 213–228. <https://doi.org/10.12758/MDA.2015.012>
- Buskirk, T. D., & Andrus, C. (2014). Smart Surveys for Smart Phones : Exploring Various Approaches for Conducting Online Mobile Surveys via Smartphones - Overview of Current Smartphone Landscape. *Survey Practice*, 5(1), 1–11.
- Callegaro, M. (2010). Do you know which device your respondent has used to take your online survey? *Survey Practice*, 3(6), 1–12.
- Couper, M. P., & Peterson, G. J. (2016). Why do web surveys take longer on smartphones? *Social Science Computer Review*, 1–21. <https://doi.org/10.1177/0894439316629932>
- de Bruijne, M., & Wijnant, A. (2013). Can mobile web surveys be taken on computers ? A discussion on a multi-device survey design. *Survey Practice*, 6(4), 1–8.

- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Horwitz, R. (2014). Usability of the ACS Internet Instrument on Mobile Devices 1. In *Proceedings of Statistics Canada Symposium 2014*.
- Lutig, P., & Toepoel, V. (2016). The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error. *Social Science Computer Review*, 34(1), 78–94. <https://doi.org/10.1177/0894439315574248>
- Lutig, P., Toepoel, V., & Amin, A. (2016). Mobile-Only Web Survey Respondents. *Survey Practice*, 9(4), 1–8.
- Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review*, 31(6), 725–743. <https://doi.org/10.1177/0894439313485201>
- Mavletova, A., & Couper, M. P. (2014). Mobile web survey design: scrolling versus paging, SMS versus e-mail invitations. *Journal of Survey Statistics and Methodology*, 2(4), 498–518. <https://doi.org/10.1093/jssam/smu015>
- Natda, K. V. (2013). Responsive Web Design. *Eduvantage*, 1(1). <https://doi.org/10.11635/2319-9954/1/1/18>
- Olmsted-Hawala, E. L., Nichols, E. M., Holland, T., & Gareau, M. (2016). *Results of Usability Testing of the 2014 American Community Survey on Smartphones and Tablets Phase I: Before Optimization for Mobile Devices*. *Survey Methodology*.
- Page, T. (2013). Usability of text input interfaces in smartphones. *J. of Design Research*, 11(1), 39. <https://doi.org/10.1504/JDR.2013.054065>
- Sarraf, S., Brooks, J., & Cole, J. S. (2014). Taking Surveys with Smartphones: A Look at Usage Among College Students. In *Aapor* (pp. 1–16). <https://doi.org/10.1017/CBO9781107415324.004>
- Thorsteinsson, G., & Page, T. (2014). User attachment to smartphones and design guidelines. *International Journal of Mobile Learning and Organisation*, 8(3), 201–215. <https://doi.org/10.1504/IJMLO.2014.067020>
- Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The Gamification User Types Hexad Scale. *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, 229–243. <https://doi.org/10.1145/2967934.2968082>
- Wells, T., Bailey, J. T., & Link, M. W. (2014). Comparison of Smartphone and Online Computer Survey Administration. *Social Science Computer Review*, 32(2), 238–255. <https://doi.org/10.1177/0894439313505829>