

Bewertung ausgewählter Studiendesigns zur Untersuchung persuasiver Selbstmonitoringsysteme

Katja HERRMANN¹, Sabine THEIS², Corinna A. CHRISTMANN³,
Gabriele BLESER³, Alexander MERTENS², Aysegül DOGANGÜN¹

¹ *Personal Analytics, Universität Duisburg-Essen, Forsthausweg 2, 47057 Duisburg*

² *Institut für Arbeitswissenschaft, Bergdriesch 27, 52062 Aachen*

³ *wearHEALTH, TU Kaiserslautern*

Gottlieb-Daimler-Straße Gebäude 48, D-67663 Kaiserslautern

Kurzfassung: Im Rahmen der Digitalisierung von Gesundheitsprozessen wird eine personalisierte Analyse des Gesundheitszustandes, des subjektiven Wohlbefindens und der Kontextparameter zunehmend durch mobile Tagebücher, Smartphone-gestützte Sensorik und am Körper getragene Computer (Wearables) ermöglicht. Für den erfolgreichen persuasiven Einsatz des persönlichen Monitorings ist jedoch eine arbeitswissenschaftliche Betrachtung nötig, die neue methodische Herangehensweisen, sowie Anpassungen bei gängigen Evaluationsmethoden erfordert. Dieser Beitrag wird daher auf Basis der Forschung dreier interdisziplinärer Projektteams eine Auswahl von gängigen (vergleichende Marktanalyse, Fokusgruppe, Mixed-Methods Design) und bisher weniger gebräuchlichen Studiendesigns (Felduntersuchung ohne Probandenkontakt) für arbeitswissenschaftliche Untersuchungen von persuasiven Selbst-Monitoring-Systemen vorstellen und diskutieren um damit eine Hilfestellung und Entscheidungsgrundlage für die Methodenauswahl zu liefern.

Schlüsselwörter: Mobile Gesundheit, Methodenauswahl, vergleichende Marktanalyse, Fokusgruppe, Felduntersuchung, Mixed-Methods Designs

1. Stand der Forschung

Persuasive Selbstmonitoringsysteme bezeichnen sensorbasierte Systeme, die eigene Verhaltens- und Körperdaten (z. B. Aktivität, Blutdruck) erfassen und zudem motivierend und unterstützend im Hinblick auf Verhaltensänderungen wirken sollen (Webb et al. 2010). Es besteht Einigkeit darüber, dass eine gründliche Evaluation essenziell für die effektive Gestaltung technischer Systeme ist (Consolvo & Walter, 2003). Dies impliziert nicht nur die summative Evaluation der persuasiven Wirkung des Systems, sondern auch einzelner persuasiver Elemente (Hamari et al. 2014), und Evaluationen im Designprozess (Consolvo & Walter 2003) bis hin zu partizipativem Design (Yetim, 2011). Solche Ansätze erlauben nicht nur Aussagen zur erzielten Verhaltensänderung, sondern auch die Gründe hierfür zu untersuchen, welche wiederum Rückschlüsse auf die Systemgestaltung erlauben (Klasnja et al., 2011). Ein Überblick zu Methoden und Studiendesigns, der die spezifischen Anforderungen der Evaluation persuasiver Selbstmonitoringsysteme aufgreift und dabei nach unterschiedlichen relevanten Fragestellungen differenziert, fehlt jedoch derzeit und soll mit diesem Beitrag adressiert werden.

2. Ausgewählte Studiendesigns für die Evaluation persuasiver Selbstmonitoringsysteme

Die nachfolgenden Studiendesigns für persuasive Selbstmonitoringsysteme (systematische Marktanalyse, Fokusgruppe, Felduntersuchung ohne Probandenkontakt, sowie Mixed-Methods-Design) werden jeweils in der Reihenfolge vorgestellt und diskutiert in der sie sich im Entwicklungsprozess bewährt haben.

2.1 Systematische Marktanalyse

Der Markt für persuasive Selbstmonitoringsysteme ist gekennzeichnet von einer großen Fülle verfügbarer und erschwinglicher Anwendungen. Um festzustellen, welches dieser Geräte sich für die zu untersuchende Aufgabe eignet oder welche Anforderungen an ein zu entwickelndes Gerät gestellt werden, sollte im Vorfeld zu eigenen Entwicklungen oder Evaluationen eine umfassende Marktanalyse hinsichtlich der Stärken und Schwächen erhältlicher Systeme durchgeführt werden. Hierbei können verschiedene Aspekte im Fokus liegen, z. B. technische Anforderungen (zusätzliche Wearables, Betriebssystem, Mindestvoraussetzungen der Hardware), Benutzerfreundlichkeit, Ästhetik, Datenschutz, die Integration verhaltenstheoretischer Konstrukte oder die Qualität und Quantität der vermittelten Informationen.

Die Analyse sollte grundsätzlich auf Basis eines vorab definierten Kriterienkatalogs erfolgen. Hierbei bietet es sich an, auf bereits vorhandene Kriterienkataloge (z. B. Abraham & Michie, 2008 für Verhaltensänderungstechniken oder Stoyanov et al., 2015 für Usability, Ästhetik und Informationsgehalt) zurückzugreifen, insbesondere wenn diese bereits validiert und regelmäßig verwendet wurden, da die Untersuchungsergebnisse so leichter in die bisherige Literatur eingebettet werden können. Der Aspekt des Datenschutzes wurde bislang jedoch nur in wenigen Studien thematisiert. Hier ist es z. B. denkbar den Kriterienkatalog an aktuellen Empfehlungen des Gesetzgebers oder der Europäischen Union (z.B. Data Protection Working Party, 2013) zu orientieren.

Um die Inhaltsvalidität der zu untersuchenden Aspekte zu gewährleisten, sollte nicht nur auf Vollständigkeit geachtet werden, was eine umfassende Recherche voraussetzt, sondern sollten auch eindeutige Definitionen gewählt werden, was durch das zusätzliche Bereitstellen von Positiv- und Negativbeispielen erleichtert wird. Die Datenerhebung erfolgt durch mindestens zwei trainierte Beurteiler, deren Übereinstimmung Hinweise auf die Objektivität und Reliabilität der Daten gibt (Cohen, 1960). Die Gütekriterien der Marktanalyse hängen somit sowohl von der Qualität des verwendeten Kriterienkatalogs als auch vom Training der Beurteiler ab. Eine Einschätzung durch den Nutzer selbst erfolgt nicht. Die Ergebnisse spiegeln daher eher die Expertenperspektive wider. Um die Meinung des späteren Nutzers zu berücksichtigen kann deshalb z. B. im Anschluss eine Fokusgruppe gebildet werden.

2.2 Fokusgruppen

Die allgemeinen Vorteile von Fokusgruppen zur partizipativen Systementwicklung, die u. A. in der Möglichkeit der Erfassung tiefgehender qualitativer Daten und der Diskussion von Gestaltungsaspekten bestehen, sind aus anderen Kontexten hinreichend bekannt. Ein Problem stellen Meinungsführerschaft und Anpassung der Meinung Einzelner an die der Mehrheit oder besonders dominanter Teilnehmer dar. Genau diese Gruppendynamik kann bei einigen Fragestellungen hinsichtlich Selbst-

monitoringsystemen jedoch von besonderem Interesse sein. So beinhalten viele gängige persuasive Selbstmonitoringsysteme Strukturen sozialer Netzwerke. Die Nutzung oder Nichtnutzung und somit die persuasive Wirkung der Anwendung bzw. ihrer sozialen Funktionalitäten hängt zu einem großen Teil von der Verbreitung und Akzeptanz in der Peer-Group der Nutzer ab. Besonders Fokusgruppen, bei denen sich die Teilnehmer kennen, haben sich in unserer Arbeit als geeignet erwiesen, um Effekte einer sozial abhängigen Meinungsbildung zu berücksichtigen. Auch für ethische und datenschutzbezogene Aspekte, die oft von der Meinung des Umfeldes mitgeprägt werden, haben sich Fokusgruppen als sinnvoll erwiesen. Fokusgruppen sind somit insbesondere dann zu empfehlen, wenn die Verbreitung der Anwendung in der Peer-Group relevant bzw. die Interaktion zwischen den Nutzern ein wesentlicher Bestandteil der Anwendung ist (z. B. Wettbewerb oder Kooperation in Activity-Tracking-Anwendungen) oder Aspekte fokussiert werden, die stark von der Meinung der Peer-Group abhängig sind.

2.3 Felduntersuchung ohne Probandenkontakt

Studiendesigns zur Untersuchung des Nutzungsverhaltens und der persuasiven Wirkung von Systemen zur Gesundheitsförderung erfordern typischerweise direkten Kontakt zwischen Teilnehmern und Versuchsleiter bei Rekrutierung, Briefing und Übergabe erforderlicher Hardware sowie (abschließender) Befragungen. Jedoch kann der direkte Kontakt Effekte der sozialen Erwünschtheit beim Nutzungsverhalten hervorrufen - insbesondere da Gesundheitsverhalten ein sozial relevantes Verhalten ist. Bei Untersuchungen der Nutzungsdauer oder Drop-Out-Raten können weitere Verzerrungen durch Angabe eines Endzeitpunktes der Studie entstehen indem die Nutzer trotz nachlassender Motivation bis dahin „durchhalten“. Eine weitere Problematik ist der Hawthorne-Effekt (Rice 1982), d. h. eine Verhaltensänderung allein durch das Bewusstsein an einer Studie teilzunehmen. Die Salienz der Studiensituation wird durch persönlichen Kontakt, die Nutzung studienspezifischer Geräte, das Wissen um einen Endzeitpunkt und zusätzliche Befragungen noch erhöht.

Als Lösungsansatz stellen wir ein Studiendesign vor, das diese Problematiken weitestgehend umgeht und dabei die Spezifika von persuasiven Selbstmonitoringsystemen nutzt: Da die Erfassung von Daten bereits integraler Bestandteil des Monitoringsystems ist, können diese zu Studienzwecken ausgewertet werden. Eine weite Verbreitung von Smartphones ermöglicht den Verzicht auf ungewohnte, studienspezifische Geräte und deren persönliche Übergabe. Der Kern des vorgeschlagenen Designs ist daher das Vereinen von Untersuchungsinstrument und Untersuchungsgegenstand in einer App, die im App-Store zum freien Download zur Verfügung gestellt wird. Die Aufklärung über Studienteilnahme, Freiwilligkeit und Datenerfassung erfolgt im Rahmen des Installationsprozesses. Für vergleichende Studien kann bei der Installation eine von mehreren App-Versionen randomisiert zugewiesen werden. Relevante Daten lassen sich weitestgehend durch die Nutzung erfassen. Die Nutzungsdauer sollte bei einem solchen Studiendesign aus o.g. Gründen zeitlich nicht begrenzt sein.

Durch die Form der Rekrutierung und den Verzicht auf direkten Kontakt wird eine Distanz zwischen Teilnehmern und Versuchsleitern geschaffen und absolute Anonymität der Teilnehmer gewährleistet. Diese ist eine wichtige Voraussetzung dafür, ein Gefühl der Beobachtung und Bewertung des persönlichen Verhaltens und somit eine Reaktion mit sozial erwünschtem Verhalten zu vermeiden. Die Verwendung des eigenen Handys, der bekannte Distributionsweg sowie der Verzicht auf rein studien-

bedingte Aktivitäten und auf eine zeitliche Begrenzung erhöhen die Natürlichkeit der Nutzungssituation. Auch ermöglicht die Bereitstellung in einem App Store eine valide Beantwortung der Frage, ob die Teilnehmer die Anwendung real nutzen (würden), was in bisherigen Studien meist fiktiv erfragt wurde. Die Zusammensetzung der Nutzergruppe hinsichtlich der erfassten Daten kann ebenfalls ohne Stichprobenbias ermittelt werden, wobei aufgrund des Studiendesigns wenige Informationen über die Nutzer vorliegen.

Unsere Erfahrungen weisen darauf hin, dass mit dem gewählten Studiendesign der Hawthorne-Effekt und sozial erwünschtes Verhalten weitestgehend vermieden werden können. Während Untersuchungen mit klassischem Studiendesign nur vereinzelte Abbrüche berichten, entspricht die in unseren Studien ermittelte Abbruchrate von 76% nach einer Woche annähernd den Ergebnissen aus Erhebungen zum Mediennutzungsverhalten, die zeigen, dass durchschnittlich 80% aller installierten Apps nicht genutzt werden (van Eimeren und Frees 2012).

Eine Limitation des vorgestellten Studiendesigns besteht darin, dass die Stichprobe nur Teilnehmer mit bereits vorhandenem Interesse und passendem Endgerät umfasst. Es ist daher nur zu empfehlen, wenn dies auch der Zielgruppe der Erhebung entspricht. Die Teilnehmerzahl ist schwer kalkulierbar, weshalb ausreichend Zeit für die Verbreitung vorgesehen und diese z. B. durch Pressearbeit unterstützt werden sollte. Für aussagekräftige Daten sollte nicht nur die passive, sondern auch die aktive Nutzung erfasst werden. Dabei ist mit hohen Ausschlussraten zu rechnen, die die Datenqualität erhöhen. Der Umsetzungsaufwand ist nicht zu unterschätzen, da die zu untersuchende bzw. als Untersuchungsinstrument dienende App marktreif entwickelt werden muss. Auch erwarten Nutzer technischen Support sowie die Behebung von Fehlern und Erweiterung der App um gewünschte Funktionalitäten. Bei der Entscheidung für oder gegen das vorgestellte Studiendesign ist insbesondere abzuwägen, welche Daten benötigt werden. Es ist nicht geeignet für Untersuchungen, die neben den vom Monitoringsystem erfassten Daten ergänzende Informationen benötigen. Gut geeignet ist es hingegen, wenn die Anwendungsnutzung und die Wirkung unterschiedlicher Systemgestaltungen auf das erfasste Verhalten valide untersucht werden sollen.

2.4 Mixed-Methods-Studien

Mixed-Methods-Studien (MMS) vereinen qualitative und quantitative Ansätzen, um Vorteile beider zu nutzen und Nachteile zu minimieren. Beide Ansätze können in unterschiedlicher Reihenfolge und Gewichtung Teil einer Studie werden. Erhebungen und Auswertung können sowohl hypothesentestend als auch -generierend gestaltet werden. Beim eingebetteten Vorgehen wird eines der beiden Verfahren verstärkt durchgeführt und lediglich durch das andere ergänzt. Hierzu zählen das explanative und das explorative Design. Ersteres betrachtet quantitative Ergebnisse genauer und sichert die quantitativen Ergebnisse qualitativ ab. Beim explorativen Design kehrt sich die Reihenfolge um: erst qualitativ, dann quantitativ. Dadurch können qualitative Ergebnisse statistisch verallgemeinert und die Anwendungsnähe quantitativer Untersuchungen erhöht werden. MMS eignen sich besonders zur Untersuchung persuasiver Selbstmonitoringsysteme, da deren sozio-technischer Kontext bei der Gestaltung und der Evaluation berücksichtigt werden muss. Ergebnisse aus kontaktlosen Felduntersuchungen können, zum Beispiel durch qualitative Befragung von Probanden eines Subsamples der gleichen Stichprobe Zusammenhänge beleuchten oder Hintergründe erklären. Ebenso können MMS zur Bedarfsanalyse herangezogen werden.

Verwendung einzelner Methoden sollte sich alleine durch entsprechende Fragestellungen begründen. Die Integration quantitativer und qualitativer Daten in Form von MMS besitzt das Potential Genauigkeit und Reichhaltigkeit von Anforderungs- und Evaluationsstudien persuasiver Selbstmonitoringsysteme zu steigern. Ein Mixed-Methods-Design zu wählen, das der Forschungsfrage am besten entspricht und dabei zeitlich und personelle Ressourcen berücksichtigt, ermöglicht tiefere, bedeutsamere Erkenntnisse.

3. Fazit

Während die Relevanz der Evaluierung persuasiver Selbstmonitoringsysteme außer Frage steht, ist die Entscheidung für die Untersuchungsmethode schwieriger.

Tabelle 1: Stärken, Schwächen und zugrundeliegende Fragestellungen von Studiendesigns zur Untersuchung persuasiver Selbstmonitoringsysteme

		Vorteile	Nachteile	Anwendungsgebiet
Studiendesign	Systematische Marktanalyse	z.T. standardisierte Kriterienkataloge vorhanden, Perspektive des Experten, bereits vor Entwicklungsprozess	Intensives Training der Beurteiler notwendig, Validität hängt von der Qualität des Kriterienkatalogs ab, keine Beurteilung durch den Nutzer	Verhaltensänderungstechniken, Gamification, Usability, Datenschutz, Ethik, technische Eigenschaften, Preis etc.
	Fokusgruppe	Sozial abhängige Meinungsbildung wird bereits im Entwicklungsprozess berücksichtigt	Sozial erwünschte oder durch einige Teilnehmern dominierte Antworten	Interfacegestaltung, Datenschutz, Ethik, Akzeptanz
	Felduntersuchung ohne Probandenkontakt	Große, repräsentative Stichprobe möglich, hohe externe Validität, Reduktion des Hawthorne-Effekts und sozialer Erwünschtheit, valide Untersuchung von Nutzungsverhalten und Drop-Out-Raten, Kosten für Ausstattung der Teilnehmer mit Hardware entfallen, langzeitliche Erhebung, Zwischenauswertung der Daten jederzeit möglich	Nur Teilnehmer mit Interesse und passendem Endgerät, verschiedene Drop-Out-Raten pro Bedingungen möglich, hohe Ausschlussraten, marktreflektierte Entwicklung und technischer Support nötig, kaum weitere Informationen neben Monitoringdaten, Gründe für Nutzerverhalten kaum erfassbar	Allgemeine Wirksamkeitsuntersuchung, Langzeitliche motivationale Wirkung verschiedener Gestaltungselemente, Zusammensetzung der Nutzergruppe, unverzerrtes Nutzungsverhalten, ggf. Aufdecken nicht intendierten Verhaltens, Gewohnheiten, Dauer, Häufigkeit der Nutzung, genutzte Funktionen
	Mixed-Methods Studien	Erfasst Komplexität realer Zusammenhänge; Vorteile qual./quant. Ansätze vereint, Triangulation kann die Ergebnisvalidität stärken; fördert interdisziplinäre Forschung	Erfordert breite Methodenkenntnis; große Datenmengen; hoher Zeitaufwand; Methodenpräferenz des Versuchsleiters kann einen Einfluss darstellen	Kontext- und Anforderungsanalyse, Hypothesenfindung, komplexe Zusammenhänge, nutzerzentrierte Fragestellungen; Studien mit speziellen Zielgruppen,

Das gewählte Studiendesign muss dem Stand des Entwicklungsprozesses, den verwendeten Geräten, der Zielpopulation und nicht zuletzt zur untersuchten Fragestellung anpasst werden.

Auf Basis der Erfahrung dreier interdisziplinärer Forschergruppen zu persuasiven Selbstmonitoringsystemen wurden ausgewählte Studiendesigns vorgestellt, die sich in diesem Kontext bewährt haben. Tabelle 1 fasst Stärken und Schwächen der einzelnen Designs sowie zugrundeliegende Fragestellungen zusammen. Die vorliegende Betrachtung einzelner Methoden kann sowohl Praktikern als auch Forschern mit Interesse an der nutzerzentrierten Gestaltung und Evaluation persuasiver Selbstmonitoringsysteme als Hilfestellung und Entscheidungsgrundlage dienen.

4. Literatur

- Abraham C, Michie S (2008) A taxonomy of behavior change techniques used in interventions. *Health Psychology* 27(3):379-387.
- Article 29 Data protection working party (2013) Opinion 02/2013 on apps on smart devices. Accessed September 13, 2016. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1:37-46.
- Consolvo S, Walker M (2003) Using the Experience Sampling Method to Evaluate Ubicomp Applications. *Pervasive Computing* 2(2):24-31.
- Creswell JW, Clark, VLP (2007) *Designing and conducting mixed methods research*. London: SAGE Publications
- Hamari J, Koivisto J, Pakkanen, T (2014) Do persuasive technologies persuade? - A review of empirical studies. In Spagnolli A, Chittaro L, Gamberini L (Hrsg.): *Persuasive Technology*. Springer International Publishing, 118-136.
- Klasnja P, Consolvo S, Pratt W (2011) How to evaluate technologies for health behavior change in HCI research. In: CHI '11. New York: ACM, 3063-3072.
- Rice B (1982) The Hawthorne defect: Persistence of a flawed theory. *Psychology Today* 16(2):70-74.
- Stoyanov SR, Hides L, Kavanagh D, Zelenko O, Tjondronegoro D, Mani M (2015) Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth* 3(1):e27.
- Trochim WM, Donnelly JP (2001) *Research methods knowledge base*. Cincinnati: Atomic Dog Publishing.
- Van Eimeren B, Frees B (2012) 76 Prozent der Deutschen online–neue Nutzungssituationen durch mobile Endgeräte: Ergebnisse der ARD/ZDF-Onlinestudie 2012. *Media Perspektiven* 7–8:362-379.
- Webb TL, Joseph J, Yardley L, Michie S (2010) Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research* 12(1):e4.
- Yetim, F (2011) A set of critical heuristics for value sensitive designers and users of persuasive systems. In *ECIS Proceedings 2011*. Paper 185.

Danksagung: Die vorgestellte Arbeit wurde vom Bundesministerium für Bildung und Forschung gefördert und ist im Rahmen der Arbeit der Nachwuchsforschergruppen PAnalytics (Ref.Nr: 16SV7110), Tech4age (Ref.Nr: 16SV7111) und wearHEALTH (Ref.Nr: 16SV7115) entstanden.



Gesellschaft für
Arbeitswissenschaft e.V.

Soziotechnische Gestaltung des digitalen Wandels – kreativ, innovativ, sinnhaft

63. Kongress der
Gesellschaft für Arbeitswissenschaft

FHNW Brugg-Windisch, Schweiz

15. – 17. Februar 2017

GfA Press

Bericht zum 63. Arbeitswissenschaftlichen Kongress vom 15. – 17. Februar 2017

FHNW Brugg-Windisch, Schweiz

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.

Dortmund: GfA-Press, 2017

ISBN 978-3-936804-22-5

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Alle Rechte vorbehalten.

© **GfA-Press, Dortmund**

Schriftleitung: Matthias Jäger

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet, den Kongressband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen.

USB-Print: Dr. Philipp Baumann, Olten

Screen design und Umsetzung

© 2017 fröse multimedia, Frank Fröse

office@internetkundenservice.de · www.internetkundenservice.de