



Toolbox zur Website-Evaluation: Erfassung der User Experience von Onlinegesundheits- informationen

Einleitung

Laut der ARD/ZDF-Onlinestudie 2019 nutzen 90 % der Personen ab 14 Jahren in Deutschland das Internet zumindest gelegentlich, 71 % täglich [1] – digitale Medien sind heutzutage eine zentrale Informationsquelle. Sehr schnell und spontan treffen die NutzerInnen (User) dabei eine Auswahl aus der Vielzahl an unterschiedlichen verfügbaren Websites. Entscheidend ist hier das persönliche Erleben einer Website, die sogenannte User Experience. Dieser Begriff umfasst alle Erfahrungen der NutzerInnen bei der Interaktion mit einer Website, was auch die Erwartungen hinsichtlich der zukünftigen Benutzung einschließt [2]. Inhalt, Usability (Gebrauchstauglichkeit), ästhetische Wahrnehmungen sowie das Vertrauen in eine Website und deren Anbieter sind zentrale Dimensionen der Erfahrungen der NutzerInnen [3–6]. Jeder dieser Aspekte hat entscheidenden Einfluss auf Akzeptanz, Wertschätzung, Wiederbesuch und Weiterempfehlung einer Website durch die Nutzenden. Wenn diese beispielsweise den Inhalt nicht verstehen oder dem Anbieter misstrauen, werden sie sich neu orientieren und nach einer anderen Website suchen. Daher ist es notwendig, Websites zu evaluieren, um die Wahrnehmung und die Eindrücke der NutzerInnen zu verstehen. Evaluation in Form von systematischen Bewertungen durch die NutzerInnen kann hierbei den Websiteverantwortlichen helfen, ihre Angebote zu verbessern. Aufgrund

der weitreichenden Folgen des Themas ist dies für E-Health-Angebote von besonderer Bedeutung, zumal schätzungsweise zwei von drei deutschen Internetnutzenden online nach Gesundheitsinformationen suchen [7].

Die Erstellung der im Folgenden dargestellten Evaluationstoolbox für Websites [8]¹ war Teil des Projektes CAHPOT (Community Access to Health Promotion via Online Tools) der Bundeszentrale für gesundheitliche Aufklärung (BZgA) im Forschungsverbund „Gesund Aufwachsen“. Gefördert wurde das Projekt vom Bundesministerium für Bildung und Forschung (BMBF, Förderkennzeichen: 01EL1426B). CAHPOT beabsichtigte, Vermögen und Grenzen der Vernetzung durch Onlinecommunitys im Themenbereich der BZgA zu erforschen sowie Einflussfaktoren auf die Vernetzung festzustellen [9]. Die hier geschaffene Toolbox stellt umfassend verschiedene bewährte Instrumente der Website-Evaluation zusammen. Mögliche Zielgruppen der Toolbox sind insbesondere betreibende Institutionen und Projektverantwortliche im Bereich von Onlinegesundheitsinformationen, aber auch Experten in User Experience und Usability sowie Personen mit Tätigkeitsschwerpunkten im Bereich

Gestaltung/Webdesign, Marktforschung oder Grundlagenforschung zu E-Health.

Theoretischer Hintergrund

Komponenten der Web User Experience

In der Forschung erweisen sich drei Gestaltungseigenschaften von Websites als zentral für die Wahrnehmung der NutzerInnen: Inhalt, Usability und Ästhetik [5, 10, 11]. Website-Inhalt wird in der DIN-EN-ISO-Norm 9241-151 [2] definiert als die Zusammenstellung von Informationsobjekten, die in Form von Ton, Text oder Video präsentiert werden können. Der Inhalt zeichnet sich dabei zum einen aus durch objektivierbare Eigenschaften, wie syntaktische Struktur, Anzahl und Muster von Wörtern oder Text-Bild-Verhältnis, zum anderen durch die individuell-subjektive Wahrnehmung der NutzerInnen, zum Beispiel hinsichtlich der Aktualität, Glaubwürdigkeit oder Verständlichkeit der Texte [6].

Usability ist in der DIN EN ISO 9241-11 [12] beschrieben als das Ausmaß, in dem ein interaktives System, ein Produkt oder eine Dienstleistung von bestimmten Stakeholdern (wie bspw. Websiteusern) genutzt werden kann, um bestimmte Ziele mit Effektivität, Effizienz und Zufriedenstellung in einem bestimmten Nutzungskontext zu erreichen. Wendet man diese ISO-Norm auf Websites an, bedeutet dies beispielsweise, dass sie sich leicht bedienen lassen und den NutzerInnen

¹ Frei verfügbar via <https://doi.org/10.17623/BZGA:224-2.1> bzw. via <https://www.bzga.de/forschung/methoden-und-verfahren/> (Toolbox inkl. Anwendungshilfen).

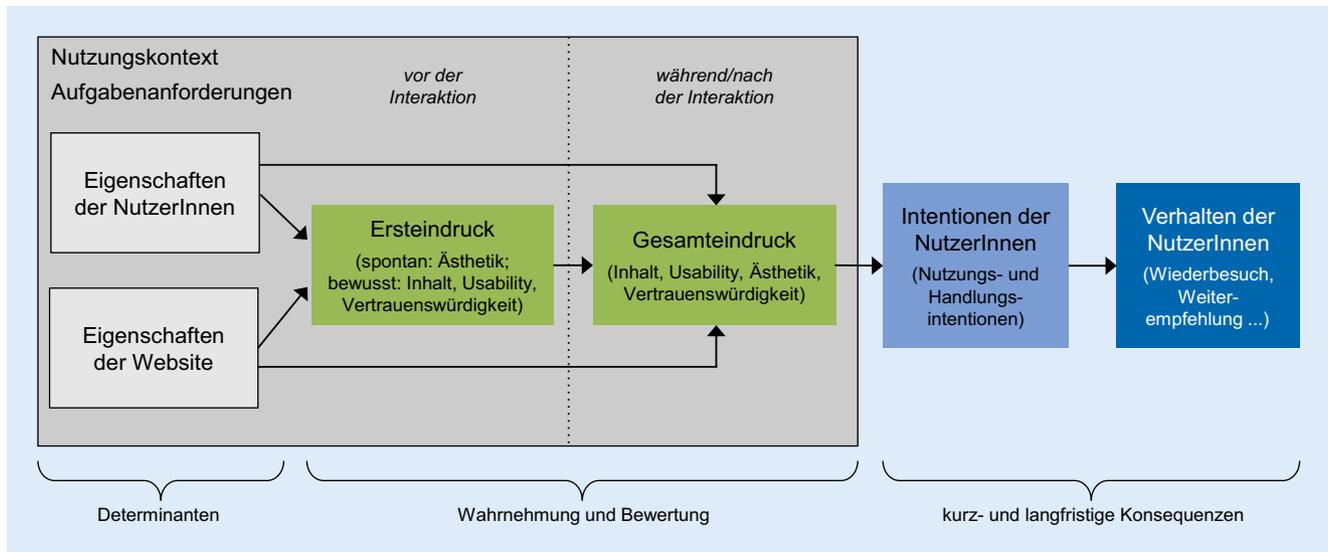


Abb. 1 ▲ Veranschaulichung von Prozessen bei der Wahrnehmung und Bewertung einer Website. (Angepasste Abbildung nach Thielsch [24])

das schnelle Auffinden der gewünschten Informationen ermöglichen. Wichtig ist dabei die Eingrenzung auf einen zu betrachtenden Nutzungskontext, wie beispielsweise Gesundheitsinformation für Laien mittels einer Website. Wie der Inhalt ist die Usability innerhalb eines gegebenen Nutzungskontextes einerseits objektivierbar (z. B. über Ladegeschwindigkeit, Erreichbarkeit oder Linktiefe einer Website), hat aber andererseits ebenfalls eine subjektive Komponente, die sich in der Sicht der NutzerInnen widerspiegelt [13–15].

Zur Websiteästhetik liegt derzeit keine einheitliche Norm oder Definition vor. Sie ist in der Forschung zumeist als unmittelbare, angenehme und subjektive Wahrnehmung eines Webobjekts definiert, die wenig durch schlussfolgernde Prozesse beeinflusst wird [16]. Es gibt zwar Ansätze, die subjektive ästhetische Wahrnehmung mit objektiven Eigenschaften einer Website in Verbindung zu setzen [17], diese Forschung ist jedoch noch nicht so weit, dass es entsprechende Messverfahren oder Handlungsempfehlungen gäbe.

Alle drei genannten Konstrukte – Inhalt, Usability und Ästhetik – beeinflussen die Web User Experience (vgl. **Abb. 1**). Zusätzlich hat das Konstrukt „Vertrauen“ eine besondere Bedeutung in der Wahrnehmung von Websites und digitalen Informationssystemen [18–20].

In diesem Kontext definiert sich Vertrauen als die Bereitschaft, sich auf eine Website oder ein Informationssystem zu verlassen und somit „verwundbar“ zu sein [21, 22]: Da NutzerInnen die genaue Funktionsweise einer Website sowie deren Hintergründe nur sehr eingeschränkt nachvollziehen oder kontrollieren können, handeln sie unter einer gewissen Unsicherheit und gehen ein Risiko ein. Es ist offensichtlich, dass dies im Bereich von Gesundheitsinformationen eine besondere Bedeutung hat. Vertrauen entsteht in der Interaktion und beim Erleben einer Website und umfasst sowohl kognitive als auch affektive Aspekte. Es entwickelt und verändert sich in Abhängigkeit sowohl von der wahrgenommenen Vertrauenswürdigkeit einer Website als auch von der generellen Bereitschaft des Einzelnen, Onlineinformationen grundsätzlich zu vertrauen [22].

Evaluation

Die Bewertung eines Sachverhalts, Prozesses, Produkts oder Programms wird als Evaluation² bezeichnet. Anders als in der Wissenschaft, bei der der Erkenntnisgewinn im Vordergrund steht, sollen Eva-

luationen in der Praxis Entscheidungsprozesse unterstützen. Dabei sollen sie Ergebnisse liefern, die so zuverlässig und valide wie möglich sind. Evaluationen helfen, Istzustände zu beschreiben und zu bewerten. Erst wenn die vorliegende Situation genau beschrieben ist, wird klar, ob und welche weiteren Maßnahmen notwendig sind. Ist diese Beschreibung falsch, besteht die Gefahr, dass Interventionen auf falschen Annahmen beruhen oder Projekte scheitern, weil notwendige Schritte nicht identifiziert und umgesetzt werden [24, 25].

Inhaltlich zu unterscheiden ist dabei zwischen dem *Evaluationsgegenstand* (bspw. einer Gesundheitswebsite) und dem *Evaluationskriterium* [24, 25]. Zur Betrachtung eines Evaluationsgegenstandes sind oft eine Reihe von verschiedenen Evaluationskriterien denkbar (vgl. [25, S. 22]). Im Bereich E-Health könnten dies beispielsweise die Akzeptanz einer Onlineintervention, die Qualität einer webbasierten Maßnahme oder der Transfererfolg einer Website sein. Die hier dargestellte Toolbox bietet die Möglichkeit, eine Reihe von Kriterien aus dem Bereich der Sitewahrnehmung/ User Experience und damit die erlebte Websitequalität strukturiert zu erfassen.

Eine weitere Unterscheidung betrifft den Zeitpunkt einer Evaluation [25]: Eine formative Evaluation ist prozessbegleitend, diese wird beispielswei-

² Laut Duden von französisch „évaluation“ = Schätzung, zu: „évaluer“ = (ab)schätzen, zu lateinisch „valere“ = stark, wert sein [23].

se bereits während der Entwicklung oder Umgestaltung einer Gesundheitswebsite eingesetzt und erlaubt Korrekturen auf Basis von Zwischenergebnissen im laufenden Prozess. Eine summative Evaluation hingegen hat das Ziel einer abschließenden Bewertung, beispielsweise einer fertiggestellten Gesundheitswebsite. Evaluationen können durch sehr verschiedene Zielsetzungen und Rahmenbedingungen geprägt sein. Diese haben einen Einfluss auf das zu verwendende Evaluationsmodell, Evaluationszeitpunkte und Durchführungsmodi (Hilfestellungen im Kontext Onlinegesundheitsinformationen finden sich in [24]).

Website-Evaluation

In erster Linie widmet sich die Website-Evaluation der Wahrnehmung und dem Verhalten der NutzerInnen. Hier ist es sehr hilfreich, zwischen verschiedenen Phasen zu unterscheiden, in denen diese auf verschiedene Aspekte einer Website reagieren (■ **Abb. 1**). Der erste Eindruck einer Website wird zunächst vor allem durch ihre Ästhetik bestimmt: Visuelle Ästhetik wird sehr schnell wahrgenommen und bereits in einigen Hundert Millisekunden bewertet [26]. Die Beurteilung der Inhalte erfordert reflektierte kognitive Prozesse und dauert daher wahrscheinlich länger [6]. Erste Eindrücke über die Vertrauenswürdigkeit einer Website können in etwa drei bis vier Sekunden gewonnen werden [4]. Um die Usability einer Website sinnvoll zu bewerten, ist eine reale Interaktion mit ihr und damit zusätzliche Zeit notwendig [14, 16]. Alle genannten Aspekte beeinflussen den Gesamteindruck nach der Nutzung einer Website. Insbesondere der Inhalt einer Website bestimmt weitere Handlungsintentionen wie die Wiederbesuchs- oder Weiterempfehlungsbereitschaft der NutzerInnen [5].

Vor dem Hintergrund der Bedeutung des Themas Gesundheit und der möglichen teilweise weitreichenden Konsequenzen von Informationssuchen in diesem Feld ist eine Schaffung von optimalen Angeboten und damit der Bedarf an zuverlässigen und validen Instrumenten zur Bewertung von Websites offensicht-

Bundesgesundheitsbl 2020 · 63:721–728 <https://doi.org/10.1007/s00103-020-03142-7>
© Der/die Autor(en) 2020

M. T. Thielsch · M. Salaschek

Toolbox zur Website-Evaluation: Erfassung der User Experience von Onlinegesundheitsinformationen

Zusammenfassung

Das Internet ist heutzutage eine zentrale Quelle für Gesundheitsinformationen. Eine optimale Gestaltung von Webangeboten und E-Health-Maßnahmen erfordert valide Evaluationen der Wahrnehmung von Nutzenden, der sogenannten User Experience. Im Rahmen eines Forschungsprojektes der Bundeszentrale für gesundheitliche Aufklärung (BZgA) wurde zu diesem Zweck eine Toolbox zur Website-Evaluation (für Screening und Monitoring) aus frei verfügbaren, quantitativ-empirischen Fragebögen zusammengestellt. Mit dieser lassen sich zentrale Aspekte der User Experience im Internet reliabel und valide erfassen. In diesem Beitrag sollen die Komponenten der Evaluationstoolbox, die zugrunde liegenden Qualitätskriterien sowie Erfahrungen aus dem Praxiseinsatz dargestellt werden.

Das Ziel der Toolbox ist es, methodisch hochwertige, umfassende und zeitökonomische Analysen zu vereinfachen. Dokumentiert sind 9 standardisierte Fragebogeninstrumente sowie ergänzende Einzelitems aus 3 Bereichen. In der Standardvariante der Toolbox werden 35 Items verwendet, die in unter 10 min beantwortet werden können. Für eine erweiterte Untersuchung einer Website sind zusätzliche Verfahren zu Website-Inhalt, Usability (Gebrauchstauglichkeit) und Ästhetik dokumentiert. Sowohl Vorteile als auch Grenzen der Toolbox werden diskutiert und geben einen Ausblick auf mögliche zukünftige Weiterentwicklungen.

Schlüsselwörter

Digitale Gesundheitskommunikation · E-Health · Qualitätssicherung · Evaluationsmethoden · Usability

Toolbox for user-experience evaluation of online health information

Abstract

The Internet is now a central source of health information. An ideal design of web services and e-health programs requires valid evaluations of users' perceptions, the so-called user experience. As part of a research project with the German Federal Centre for Health Education (BZgA), a toolbox for website evaluation (for screening and monitoring) was compiled for this purpose from freely available quantitative empirical questionnaires. With this toolbox, key aspects of the web user experience can be reliably and validly assessed. In this article, we describe the toolbox's components, underlying quality criteria, and practical experiences. The aim of the toolbox is to facilitate methodically high-quality, comprehensive, and time-

efficient analyses. A total of nine standardized instruments and supplementary single items from three areas are documented. In the standard version of the toolbox, 35 items are used that can be answered in less than 10 min. For an extended examination of a website, the toolbox documents additional measures on website content, usability, and aesthetics. We discuss both the advantages and limitations of the toolbox and give an outlook on possible future developments.

Keywords

Digital health communication · E-health · Quality assurance · Evaluation methods · Usability

lich. Im Folgenden wird die Zusammenstellung derartiger Instrumente in der „Toolbox zur kontinuierlichen Website-Evaluation und Qualitätssicherung“ [8] dargestellt.

Zusammenstellung der Toolbox

Die Toolbox wurde aus frei verfügbaren und (soweit vorhanden) deutschsprachigen Fragebögen aus dem Bereich Website-Evaluation zusammengestellt. Ziel war, eine umfassende, quantitative Evaluation der subjektiven Eindrücke von WebsitenutzerInnen anhand von

Tab. 1 In die Toolbox eingeschlossene Fragebogeninstrumente und Items

Aspekte der User Experience	Fragebogeninstrumente und Items	Anzahl Items	Quelle
Ersteindruck	Einzelitems (zu Inhalt, Usability, Ästhetik und Ersteindruck insgesamt)	4	[39]
Subjektive Inhaltswahrnehmung	Message Credibility Scale	3	[40]
	Trust Measure ^a	11	[27]
	Fragebogen zur Wahrnehmung von Website-Inhalten (WWI)	9	[39]
Subjektive Usability/Nutzerzufriedenheit	Perceived Website Usability-German (PWU-G)	7	[3, 39]
	System Usability Scale (SUS) ^a	10	[28]
	Usability Metric for User Experience-Lite (UMUX-Lite)	2	[41]
Visuelle Ästhetik	Visual Aesthetics of Websites Inventory (VisAWI)	18	[16]
	Visual Aesthetics of Websites Inventory – Short (VisAWI-S)	4	[31]
Emotionale Reaktion: Befindlichkeit, Zustimmung, Zufriedenheit	Einzelitem Smiley-Skala	1	[42]
Gesamteindruck	Einzelitem Note	1	[39]
Handlungs- und Nutzungsintentionen	Wiederbesuchsskala („scale assessing the intention to revisit the website“)	4	[16]

^aInstrumente, die nicht Teil der Standardtoolbox sind, aber erweiterten Analysen dienen können

standardisierten Fragebogeninstrumenten zu ermöglichen. Hierbei wurden nur Instrumente mit hinreichender Reliabilität und hoher Validität in Betracht gezogen. Hinsichtlich der Erfassung von Erst- und Gesamteindruck wurde die Toolbox gezielt um Einzelitems ergänzt, die sich in der Forschung bewährt haben [8]. Die Toolbox soll dabei sowohl zu begleitenden, formativen Evaluationen als auch zur summativen Evaluation einsetzbar sein. Typische Anwendungsszenarien sollen unter anderem das einmalige Screening hinsichtlich einer Reihe von zentralen Konstrukten der Sitewahrnehmung oder auch ein fortlaufendes Monitoring dieser Aspekte im Zeitverlauf sein.

Verfügbare Skalen und Fragebögen wurden auf Basis der zum Erstellungszeitpunkt vorhandenen Forschung und anhand gezielter Recherchen identifiziert. Gesucht wurde dabei insbesondere in der digitalen Bibliothek des Fachbereichs „Mensch-Computer-Interaktion“ der Gesellschaft für Informatik e.V. sowie in einer Entwurfsversion der Fragebogenmatrix des Arbeitskreises „User Research“ des Berufsverbands der Deutschen Usability und User Experience Professionals (German UPA e.V.). Zusätzlich wurden das Testverzeichnis des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID), das ZIS-Portal des GESIS – Leibniz-

Instituts für Sozialwissenschaften (ZIS steht für „Zusammenstellung sozialwissenschaftlicher Items und Skalen“) – und das Psychologisch-Medizinische Testportal Psychometrikon des Universitätsklinikums der RWTH Aachen gesichtet sowie einzelne Testautorinnen und -autoren gezielt angesprochen. Die final in die Toolbox aufgenommenen Items und Instrumente zur Evaluation sind:

- Ersteindruck,
- subjektive Inhaltswahrnehmung,
- subjektive Usability/Nutzerzufriedenheit,
- visuelle Ästhetik,
- emotionale Reaktion: Befindlichkeit, Zustimmung, Zufriedenheit,
- Gesamteindruck,
- Handlungs- und Nutzungsintentionen.

Instrumente in der Toolbox

Insgesamt wurden 9 standardisierte Verfahren sowie 6 Einzelitems (zu Ersteindrücken, emotionaler Reaktion und Gesamteindruck) in die Toolbox aufgenommen (siehe [Tab. 1](#)). Hierbei gibt es eine Standardversion sowie eine erweiterte Evaluationsversion [8]. In der Standardversion wird vorgeschlagen, 6 der Fragebögen sowie die Einzelitems in einer Evaluation einzusetzen (in Summe 35 Items). Für den Anwendungsfall,

dass in den Bereichen Inhalt, Usability, Ästhetik und Vertrauenswürdigkeit weitergehende Untersuchungen vorgenommen werden sollen, können zusätzliche Instrumente zum Einsatz kommen: der Trust Measure [27], die System Usability Scale (SUS; [28]) sowie der Visual Aesthetics of Website Inventory (VisAWI; [16]). Falls SUS und VisAWI eingesetzt werden, können Usability Metric for User Experience(UMUX)-Lite und die Kurzversion VisAWI-S ausgelassen werden. Das heißt, in der erweiterten Anwendung der Toolbox werden maximal 68 Items zu einer Website abgefragt. Vor allem die Standardversion lässt Raum für weitere Abfragen (bspw. Details zur Demografie der Befragten, offene qualitative Fragen oder gezielte Evaluation spezifischer Funktionen und Aspekte einer Website) und ist auch kurz genug für den Einsatz in wiederholten Befragungen einer Zielgruppe.

Gütekriterien der Instrumente in der Toolbox

Aus psychometrischer Sicht sind drei Qualitätskriterien zentral für jeden Test und Fragebogen: Objektivität, Reliabilität und Validität [29, 30]. Diese drei Kriterien sind vor den – je nach Situation ebenfalls wichtigen – Nebengütekriterien wie Kosteneffizienz, Verfügbarkeit von Vergleichswerten (Normen oder

Benchmarks) oder der Unverfälschbarkeit eines Verfahrens zu betrachten.

Das Hauptgütekriterium Objektivität ist dann gegeben, wenn die Ergebnisse unabhängig von der Person sind, die eine Befragung *durchführt, auswertet und interpretiert*. Alle Instrumente in der Toolbox sind bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, durchführungsobjektiv. Im Falle einer automatisierten Auswertung ist zudem die Auswertungsobjektivität gegeben.

Reliabilität ist das zweite Hauptgütekriterium und beinhaltet die instrumentelle Messgenauigkeit. Oft wird hier Cronbachs Alpha als Maß dafür angegeben, ob mehrere Items einer Skala dasselbe Konstrukt erfassen. Zudem lassen sich die Messwerte aus zwei verschiedenen Messzeitpunkten korrelieren, um die zeitliche Stabilität eines Instruments zu bestimmen (Retest-Reliabilität). Bei allen neun in der Toolbox dokumentierten Verfahren wurde Cronbachs Alpha mit guten ($\alpha > 0,80$) bis hervorragenden ($\alpha > 0,90$) Werten bestimmt; lediglich beim VisAWI-S zeigt eine Studie ein Cronbachs Alpha von 0,76 an, eine andere jedoch 0,81 [31]. Berechnungen zur Zeitstabilität, sprich Retest-Korrelationen, liegen jedoch bei keinem der Instrumente vor. Dies ist im Bereich der User-Experience-Evaluation leider ein typischer Umstand, hier ist zukünftige Forschung sehr wünschenswert.

Das dritte und wichtigste Hauptgütekriterium Validität fragt, ob ein Fragebogen oder Test überhaupt das Merkmal misst, das gemessen werden soll. Validität ist schwieriger zu bestimmen als Reliabilität, im Grunde genommen gibt es hierbei nur eine dichotome Entscheidung: Das Instrument ist valide oder nicht. Im Idealfall kommen zur Entscheidungsfindung mehrere Validierungsstrategien kombiniert zur Anwendung, so zum Beispiel:

- Vergleiche eines neuen Fragebogens mit vorhandenen Verfahren, die das gleiche messen (erwartet werden hohe Korrelationen = *konvergente Validität*),
- Vergleiche eines neuen Fragebogens mit vorhandenen Verfahren, die etwas ganz anderes messen (erwar-

tet werden kleine oder gar keine Korrelationen = *divergente Validität*),

- Überprüfung der angenommenen Faktoren eines Fragebogens mittels einer konfirmatorischen Faktorenanalyse (= *faktorielle Validität*),
- Vergleich mit Expertenurteilen oder anderen vergleichbaren, parallel erfassten Kriterien (= *konkurrente Validität*),
- Analysen hinsichtlich der Fähigkeit eines Instruments, zwischen verschiedenen Zielobjekten zu unterscheiden (= *diskriminative Validität*),
- gezielte Experimente zur Validitätsprüfung (= *experimentelle Validität*).

Für alle in die Toolbox aufgenommen Verfahren liegen empirische Prüfungen und Hinweise zur Validität vor, teilweise wurden verschiedene Validierungsstrategien kombiniert [8]. Lediglich die Items zu Handlungs- und Nutzungsintentionen (übernommen aus [17]) sind nicht gezielt empirisch validiert. Aufgrund mangelnder publizierter Alternativen und hoher anzunehmender inhaltlicher Validität wurden diese dennoch in die Toolbox aufgenommen.

Anwendungshilfen für die Praxis

Ein Begleitmanual zur Toolbox³ soll auch Personen aus der Praxis ohne vertieften Forschungshintergrund die Durchführung, Auswertung und sinnvolle Interpretation ermöglichen. Zu jedem Verfahren in der Toolbox finden sich neben den jeweiligen Items und Angaben zu Autoren/Zitation, theoretischem Hintergrund und Gütekriterien auch Hinweise zum inhaltlichen Hintergrund, Auswertungsanweisungen sowie Vergleichswerte und kurze Interpretationshilfen/Benchmarks. AnwenderInnen mit wenig Erfahrung im Bereich User Experience können auf das oben beschriebene Standardset von Fragen zurückgreifen, das eine schnelle Gesamtbewertung aller wesentlichen Konstrukte erlaubt. Zur weiteren Vereinfachung der Anwendung beinhaltet die Toolbox eine Importdatei für eine gängige Onlinebefragungs-

software, eine Excel-Auswertungsmatrix, in die die jeweiligen Rohdaten eingefügt werden können, sowie eine Liste der häufig gestellten Fragen (FAQ). AnwenderInnen der Toolbox benötigen somit zur Durchführung lediglich einen Zugang zu einer Onlinebefragungssoftware (mit Datenexportfunktion) und die Möglichkeit, die BesucherInnen der zu testenden Website auf die Befragung zu leiten.

Grundsätzlich ist die Toolbox auf eine quantitative Website-Evaluation ausgerichtet. Dabei bleibt es in der Verantwortung der AnwenderInnen, eine Erhebungsstrategie (bspw. Screening vs. Monitoring) festzulegen [24]. Ebenso ist aus dem Entwicklungsstand und den Zielen der jeweiligen Website abzuleiten, inwieweit andere Methoden (wie bspw. qualitative Verfahren oder die Messung objektiver Nutzungsdaten) ergänzend oder anstelle der dokumentierten Fragebogenverfahren sinnvoll wären. Nicht immer werden alle erfassten Konstrukte aus der Toolbox von gleicher Relevanz sein; werden beispielsweise langjährig registrierte NutzerInnen befragt, wirken die Items zum Ersteindruck sogar fehl am Platz und sind in dieser Situation nicht mehr valide zu erfassen. In solchen Situationen (und um die Befragten möglichst wenig zu belasten) können und sollten Instrumente ausgelassen werden.

Wie bei allen quantitativen Verfahren messen auch die Instrumente aus der Toolbox umso genauer, je mehr Personen der Zielgruppe an einer Befragung teilnehmen [32, 33]. Aufgrund dieser Anforderungen können je nach Fragestellung andere methodische Ansätze und insbesondere qualitative Verfahren ressourcenschonender in der Durchführung sein [24]. Gerade aber für die summativ-evaluative und das Monitoring einer Website bietet sich eine differenzierte Analyse mithilfe von Instrumenten aus der Toolbox an. Die validierten Verfahren bieten eine solide Datenbasis beispielsweise ergänzend zu qualitativen Analysen, für Vergleiche mit anderen Anbietern und unterstützend bei Entscheidungen zu Optimierungen oder zur Neugestaltung (Relaunch) einer Website.

³ Verfügbar via <https://doi.org/10.17623/BZGA:224-2.1>.

Erfahrungen aus dem Praxiseinsatz

Ursprünglich bestand das vorrangige Ziel der Toolbox als zunächst lose Fragebogensammlung darin, im genannten Forschungsprojekt mit geringem Aufwand wiederholte valide User-Experience-Evaluationen einer BZgA-Website zur Vernetzung kommunaler Akteure im Bereich Prävention und Gesundheitsförderung zu ermöglichen. Die betreffende Website wurde zur Laufzeit des Forschungsprojekts erneuert und mit der Fragebogensammlung wurden die alte Version der Website, verschiedene Prototypen sowie die finale Website evaluiert; Problembereiche wurden identifiziert, mit den Ergebnissen weiterer Evaluationen verknüpft und in der weiteren Websiteentwicklung berücksichtigt [9, 34].

Die positiven Erfahrungen im Forschungseinsatz führten unter Praxispartnern und innerhalb der BZgA rasch zu dem Wunsch, die Fragebogensammlung für den Einsatz außerhalb des spezifischen Forschungskontextes zu öffnen. Hierfür wurde zunächst eine Basisversion mit Anwendungsmanual erstellt (Version 1 der Toolbox) und nach den ersten Rückmeldungen aus der Praxis um weitere Verfahren und Anwendungshilfen ergänzt (Version 2). Innerhalb der BZgA wurden Websiteverantwortliche zu Einsatzzweck und Nutzungsmöglichkeiten geschult und die Toolbox wurde im ersten Jahr nach Veröffentlichung der Version 2 für insgesamt sechs Websiteprojekte eingesetzt: zur ergänzenden Planung von Relaunchvorhaben, zur Erfolgsmessung (Prä- und Posttests) und zur Ergänzung anderer (meist qualitativer) Evaluationen. Die Evaluationen wurden vom Forschungsreferat gemeinsam mit den inhaltlich arbeitenden Referaten geplant, die genutzten Skalen je nach Evaluationsziel bedarfsweise angepasst bzw. weitere Einzelitems ergänzt und Konsequenzen aus den Ergebnissen besprochen. Die Durchführung der Erhebungen und die Auswertung der Daten übernahm zunächst das Forschungsteam der BZgA, um die inhaltlich arbeitenden Teams vom entsprechenden Aufwand zu befreien.

Herausforderungen im Praxiseinsatz bestanden insbesondere in drei Feldern:

1. Häufig gewünscht wurden zusammenfassende Bewertungen der Ergebnisdaten in Textform sowie ansprechendere Ergebnisgrafiken.
2. Für inferenzstatistische Vergleiche (z. B. Signifikanz, Effektstärke) zwischen zwei Testungen sind in Version 2 weiterhin Statistiksoftware und entsprechende Methodenkenntnisse erforderlich. Das Gleiche gilt für Subgruppenvergleiche (z. B. Vergleich verschiedener Alterskohorten). Insgesamt wurde eine weitere Vereinfachung von Planung, Durchführung, Auswertung und Interpretation gewünscht.
3. Teilweise wurde eine weitere Kürzung der Befragung angeregt. Bei einer Durchführungszeit von rund 10 min betrug die Abbrecherquote der Onsite-Befragungen je nach Website ca. 25–50 %.

Die BZgA entwickelt die Toolbox auch zukünftig weiter. Die in Planung befindliche nächste Version soll ein Auswertungstool beinhalten, das auf die Wünsche aus den ersten beiden genannten Punkten reagiert. Auch das Aufsetzen der Fragebögen wird weiter vereinfacht. Die Fragebogeninstrumente erfahren eine leichte Aktualisierung – für den Bereich Glaubwürdigkeit/Vertrauen fehlt es jedoch weiterhin an hochwertigen Instrumenten, die den veränderten digitalen Kommunikationsweisen Rechnung tragen. Insbesondere in diesem Bereich, aber auch allgemein bei der Evaluation abseits von klassischen Websites, bedarf es intensiver Forschung, damit die Evaluationsinstrumente mit den Entwicklungen der Digitalisierung Schritt halten [35]. In diesem Zuge wären zudem weitere Kurzversionen etablierter Instrumente wünschenswert, um gerade für erste Screenings eine noch kürzere aber dennoch qualitativ hochwertige Version der Toolbox erstellen zu können.

Diskussion

Die hier dargestellte Toolbox kann ihre AnwenderInnen bei der strukturierten Analyse der Wirkung von Onlinegesund-

heitsinformationen auf die NutzerInnen und bei der Ermittlung der erlebten Websitequalität unterstützen. Die zusammengestellten quantitativen Verfahren zeichnen sich durch eine (für den Bereich User Experience) hohe psychometrische Güte aus, sind in relativ kurzer Zeit zu beantworten und in deutscher Sprache verfügbar. Die Toolbox selbst ist ergänzt um Zusatzinformationen und vielfältige Hilfsmittel (bspw. Manual, Anwendungs- und Auswertungshilfen). In Validität, thematischer Breite und Messgenauigkeit sind die dokumentierten Instrumente gängigen Evaluationen im Webkontext weit überlegen, wenn diese zum Beispiel lediglich über Einzelfragen zur Weiterempfehlungsbereitschaft durchgeführt werden.

Der große Vorteil einer solchen fragebogenbasierten Evaluation besteht darin, dass sich Aussagen zu verschiedenen Konstruktbereichen relativ verlässlich quantifizieren lassen. Durch eine Analyse großer Zielgruppen sind zudem grobe Prognosen möglich. Die Anforderungen an die Größe der Befragungsstichprobe (vgl. [32, 33]) gehen jedoch mit entsprechendem Aufwand und Kosten einher. Nachteilig kann zudem die fehlende Reaktivität einer solchen standardisierten Befragung sein: Gezieltes Nachfragen bei Unklarheiten ist ebenso wenig möglich wie eine tiefgehende Betrachtung des individuellen Erlebens einzelner Personen. Eine Anreicherung einer Website-Evaluation um qualitative Verfahren (bspw. Interviews oder Fokusgruppen) oder auch eine Ergänzung der Toolboxinstrumente durch zusätzliche offene Fragen ist daher in manchen Kontexten angezeigt (vgl. [24]). Ebenso kann die Erweiterung einer Evaluation mit verhaltensorientierten Verfahren (wie bspw. Logfileanalysen) relevante weitere Informationen liefern. Diese verschiedenen Möglichkeiten verdeutlichen aber, dass für AnwenderInnen der Toolbox neben ihrer fachlichen Expertise ein solides Basiswissen im Bereich der Website-Evaluation sehr hilfreich ist. Auf diesem aufbauend können eine strukturierte Planung der Evaluation (vgl. [24, 25]) sowie eine sinnvolle Integration der Toolbox in das Evaluationsdesign erfolgen.

Ausblick

In der Forschung zur quantitativen Evaluation von Websites (ebenso wie bei der Technologie selbst) finden sich kontinuierliche Innovationen. Die Notwendigkeit der Aktualisierung beziehungsweise Weiterentwicklung der Toolbox ist damit absehbar und wird wie oben beschrieben bereits verfolgt. Zukünftige Updates der Toolbox werden zudem voraussichtlich eine Aktualisierung des theoretischen Evaluationsmodells sowie der Evaluationsempfehlungen auf Basis der aktuellen Forschung enthalten. Ein wichtiger Fokus wird hier unter anderem die Identifikation von Befragten sein, die bewusst oder unbewusst fehlerhafte Angaben machen (indem sie beispielsweise eine Umfrage nur durchklicken oder gegebene Aufgaben nicht korrekt bearbeiten). Ebenso bietet sich eine erneute Sichtung neuer publizierter Webevaluationsinstrumente an. Im Bereich der User Experience liegt hier beispielsweise nun eine Kurzform des Fragebogens „User Experience Questionnaire“ (UEQ short; [36]) und eine erweiterte, modulare UEQ+ Version vor [37]. In derzeitigen BZgA-internen Aktualisierungen der Toolbox wurden bereits ein neuer Fragebogen zur Bewertung von Website-Inhalten (WebCLIC; [6]) sowie aktuelle Benchmarks aus dem Gesundheitsbereich [38] integriert. Hierzu werden zukünftig in der schriftlichen Dokumentation zur Toolbox entsprechende Aktualisierungen und erweiterte Interpretationshilfen gegeben werden.

Die BZgA wird den Einsatz digitaler Kommunikation in Zukunft weiter verstärken [35]. Zuverlässige, einfach zu handhabende Evaluationsmethoden, die gut in die alltägliche Arbeit zu integrieren sind, sind dabei unerlässlich, um Qualität und Nutzen seriöser Gesundheitsinformation fortlaufend sicherzustellen und zu verbessern.

Korrespondenzadresse

Prof. Dr. Meinald T. Thielsch
Institut für Psychologie, Westfälische Wilhelms-Universität Münster
Fliednerstr. 21, 48149 Münster, Deutschland
thielsch@uni-muenster.de

Funding. Open Access funding provided by Projekt DEAL.

Einhaltung ethischer Richtlinien

Interessenkonflikt. M.T. Thielsch wurde durch die Bundeszentrale für gesundheitliche Aufklärung (BZgA) mit der Erstellung der dargestellten Toolbox beauftragt. M. Salaschek war bis August 2019 wissenschaftlicher Angestellter der BZgA, die Toolbox gehörte dort zu seinem Aufgabenfeld. Er war anschließend beratend für die BZgA tätig.

Für diesen Beitrag wurden von den Autoren keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

1. Beisch N, Koch W, Schäfer C (2019) ARD/ZDF-Onlinestudie 2019: Mediale Internetnutzung und Video-on-Demand gewinnen weiter an Bedeutung. *Media Perspekt* 9/2019:374–388
2. ISO (2008) DIN EN ISO 9241-151:2008-09: Ergonomie der Mensch-System-Interaktion – Teil 151: Leitlinien zur Gestaltung von Benutzungsschnittstellen für das World Wide Web (ISO 9241-151:2008); Deutsche Fassung EN ISO 9241-151:2008. Beuth, Berlin
3. Flavián C, Guinaliú M, Gurrea R (2006) The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Inf Manag* 43:1–14. <https://doi.org/10.1016/j.im.2005.01.002>

4. Robins D, Holmes J (2008) Aesthetics and credibility in web site design. *Inform Process Manag* 44:386–399. <https://doi.org/10.1016/j.ipm.2007.02.003>
5. Thielsch MT, Blotenberg I, Jaron R (2014) User evaluation of websites: from first impression to recommendation. *Interact Comput* 26:89–102. <https://doi.org/10.1093/iwc/iwt033>
6. Thielsch MT, Hirschfeld G (2019) Facets of website content. *Hum Comput Interact* 34:279–327. <https://doi.org/10.1080/07370024.2017.1421954>
7. Nölke L, Mensing M, Krämer A, Hornberg C (2015) Sociodemographic and health-(care)-related characteristics of online health information seekers: a cross-sectional German study. *BMC Public Health* 15:1–12. <https://doi.org/10.1186/s12889-015-1423-0>
8. Thielsch MT, v. Salaschek M (2018) Toolbox zur kontinuierlichen Website-Evaluation und Qualitätssicherung. Version 2. Bundeszentrale für gesundheitliche Aufklärung (BZgA), Köln <https://doi.org/10.17623/BZGA:224-2.1>
9. Salaschek M, Nöcker G (2018) Online-Community für kommunale Akteure in der Gesundheitsförderung. *Bundesgesundheitsblatt* 61:1270–1278. <https://doi.org/10.1007/s00103-018-2802-3>
10. Cober RT, Brown DA, Levy PE, Cober AB, Keeping LM (2003) Organizational web sites: Web site content and style as determinants of organizational attraction. *Int J Select Assess* 11:158–169
11. De Angeli A, Sutcliffe A, Hartmann J (2006) Interaction, usability and aesthetics: what influences users' preferences? Proceedings of the 6th conference on Designing Interactive systems. ACM, New York, S271–280
12. ISO (2018) DIN EN ISO 9241-11:2018-11: Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (ISO 9241-11:2018); Deutsche Fassung EN ISO 9241-11:2018. Beuth, Berlin
13. Hornbaek K (2006) Current practice in measuring usability: Challenges to usability studies and research. *Int J Hum Comput Stud* 64:79–102. <https://doi.org/10.1016/j.jhcs.2005.06.002>
14. Kurosu M, Kashimura K (1995) Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In: Conference companion on Human factors in computing systems. ACM, New York, S292–293
15. Thielsch MT, Engel R, Hirschfeld G (2015) Expected usability is not a valid indicator of experienced usability. *Peer J Comput Sci* 1:e19. <https://doi.org/10.7717/peerj-cs.19>
16. Moshagen M, Thielsch MT (2010) Facets of visual aesthetics. *Int J Hum Comput Stud* 68:689–709. <https://doi.org/10.1016/j.jhcs.2010.05.006>
17. Seckler M, Opwis K, Tuch AN (2015) Linking objective design factors with subjective aesthetics: an experimental study on how structure and color of websites affect the facets of users' visual aesthetic perception. *Comput Human Behav* 49:375–338. <https://doi.org/10.1016/j.chb.2015.02.056>
18. Fogg B, Marshall J, Laraki O et al (2001) What makes Web sites credible? A report on a large quantitative study. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, New York, S61–68
19. Metzger MJ, Flanagin AJ (2013) Credibility and trust of information in online environments: the use of cognitive heuristics. *J Pragmat* 59:210–220. <https://doi.org/10.1016/j.pragma.2013.07.012>
20. Thielsch MT, Meeßen SM, Hertel G (2018) Trust and distrust in information systems at the workplace.

- Peer J 6:e5483. <https://doi.org/10.7717/peerj.5483>
21. Gefen D, Benbasat I, Pavlou P (2008) A research agenda for trust in online environments. *Manag Inf Syst* 24:275–286
 22. Meeßen SM, Thielsch MT, Hertel G (2020) Trust in management information systems (MIS): a theoretical model. *Z Arbeits Organisationspsychol* 64:6–16. <https://doi.org/10.1026/0932-4089/a000306>
 23. Dudenredaktion (2019) Evaluation. <https://www.duden.de/node/43225/revision/43254>. Zugriffen: 16. Dez. 2019
 24. Thielsch M, Hirschfeld G (2018) Expertise Website-Evaluation: Übersicht über bestehende Evaluationsmethoden und Entscheidungshilfe für die Evaluation bestehender sowie neu geschaffener Websites. Version 1.1. Bundeszentrale für gesundheitliche Aufklärung (BZgA), Köln <https://doi.org/10.17623/BZGA:225-EWE-1.0>
 25. Gollwitzer M, Jäger RS (2014) *Evaluation kompakt*, 2. Aufl. Beltz, Weinheim
 26. Bölte J, Hösker T, Hirschfeld G, Thielsch MT (2017) Electrophysiological correlates of aesthetic processing of webpages: a comparison of experts and laypersons. *Peer J* 5:3440. <https://doi.org/10.7717/peerj.3440>
 27. McKnight DH, Choudhury V, Kacmar C (2002) Developing and validating trust measures for e-commerce: an integrative typology. *Inf Syst Res* 13:334–359. <https://doi.org/10.1287/isre.13.3.334.81>
 28. Brooke J (1996) SUS—a quick and dirty usability scale. *Usability Eval Ind* 189:4–7. <https://doi.org/10.1002/hbm.20701>
 29. Bühner M (2010) *Einführung in die Test- und Fragebogenkonstruktion*, 3. Aufl. Pearson Studium, München
 30. Moosbrugger H, Kelava A (2012) *Testtheorie und Fragebogenkonstruktion*, 2. Aufl. Springer, Heidelberg
 31. Moshagen M, Thielsch MT (2013) A short version of the visual aesthetics of websites inventory. *Behav Inform Technol* 32:1305–1311. <https://doi.org/10.1080/0144929X.2012.694910>
 32. Hirschfeld G, Thielsch M (2019) Stichprobenumfang und Fragebogenlänge in Webevaluationen. In: *Proceedings of the Mensch und Computer 2019* <https://doi.org/10.1145/3340764.3344454>
 33. Schmettow M (2012) Sample size in usability studies. *Commun ACM* 55:64–70
 34. Salaschek M, Thielsch MT (2017) Toolbox zur kontinuierlichen Website-Evaluation & Qualitätssicherung. In: Hess S, Fischer H (Hrsg) *Mensch und Computer 2017 – Usability Professionals*. Gesellschaft für Informatik e. V. und German UPA e. V., Regensburg, S 127–137 <https://doi.org/10.18420/muc2017-up-0149>
 35. Salaschek M, Thaiss H (2019) Gesundheitskommunikation der Zukunft: Informationsvermittlung in Zeiten des digitalen Wandels. *Aktuelle Analysen* 71:19–25
 36. Schrepp M, Hinderks A, Thomaschewski J (2017) Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *Int J Interact Multimed Artif Intell* 4:103–108
 37. Schrepp M, Thomaschewski J (2019) Eine modulare Erweiterung des User Experience Questionnaire. In: Fischer H, Hess S (Hrsg) *Mensch und Computer 2019 – Usability Professionals*. Gesellschaft für Informatik e. V. und German UPA e. V., Bonn <https://doi.org/10.18420/muc2019-up-0108>
 38. Thielsch MT, Thielsch C, Hirschfeld G (2019) How Informative is Informative? Benchmarks and Optimal Cut Points for E-Health Websites. In: *Mensch und Computer 2019 – Workshopband* <https://doi.org/10.18420/muc2019-ws-642>
 39. Thielsch MT (2008) *Ästhetik von Websites. Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen*. MV Wissenschaft, Münster
 40. Appelman A, Sundar SS (2016) Measuring message credibility: construction and validation of an exclusive scale. *Journal Mass Commun* Q93:59–79. <https://doi.org/10.1177/1077699015606057>
 41. Lewis JR, Utesch BS, Maher DE (2013) UMUX-LITE: when there's no time for the SUS. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, S 2099–2102
 42. Jäger R (2004) Konstruktion einer Ratingskala mit Smilies als symbolische Marken. *Diagnostica* 50:31–38. <https://doi.org/10.1026/0012-1924.50.1.31>