

Toolbox

zur kontinuierlichen Website-Evaluation und Qualitätssicherung

*Zusammenstellung von Screening-Fragen und
Instrumenten zu User Experience und Usability*

*Autor: PD Dr. Meinald T. Thielsch
unter Mitarbeit von Dr. Martin Salaschek
Version: 2.0 | Stand: 15. August 2017*

Die Toolbox erlaubt zuverlässige und schnelle User Experience-Evaluationen von Websites mittels (Online-)Befragung der Nutzenden. Sie beinhaltet aktuell verfügbare deutschsprachige quantitativ-empirische Fragebogen-Verfahren im Bereich User Experience und Usability, mit denen sich zentrale Aspekte der Wahrnehmung von Website-NutzerInnen reliabel und valide erfassen lassen.

Der Fokus der Toolbox liegt auf einer breiten quantitativen Analyse sowie einem kontinuierlichen Screening und Monitoring der zu evaluierenden Website. Auch Vergleiche zwischen Websites sind möglich. Dabei werden die NutzerInnen durch die Befragung wenig belastet: Die Toolbox umfasst in der Standardversion insgesamt 35 Fragen und ist damit in unter 10 Minuten zu beantworten. Eine erweiterte Version mit insgesamt bis zu 68 Fragen erlaubt eine vertiefende Analyse von Website-Inhalten, Usability und Ästhetik. Die Toolbox wurde zur kontinuierlichen Website-Evaluation und Qualitätssicherung der Plattform inforo im Rahmen des BMBF-geförderten Forschungsprojektes CAHPOT entwickelt (Förderkennzeichen 01EL1426B, weitere Informationen unter: <http://www.bzga.de/forschung/vernetzung-von-online-communitys-cahpot/>), daher wird in den Interpretationshilfen speziell auf Websites im Gesundheitsbereich eingegangen. Alle Instrumente sind jedoch für Websites aller Art einsetzbar.



Inhaltsverzeichnis

Inhaltsverzeichnis	2
Abkürzungen und Begriffsklärung.....	3
Einführung und theoretischer Hintergrund	4
Allgemeines zu User Experience	4
Zentrale Konstrukte der Website User Experience	5
Struktur der Darstellung der Instrumente	7
Güte der Instrumente	7
Hintergrundinformationen zur Instrumentenauswahl	9
Darstellung der Instrumente in der Toolbox	10
1. Ersteindruck	10
2. Subjektive Inhaltswahrnehmung	12
3. Subjektive Usability / Nutzerzufriedenheit	19
4. Visuelle Ästhetik.....	26
5. Emotionale Reaktion: Befindlichkeit, Zustimmung, Zufriedenheit ...	33
6. Handlungs- und Nutzungsintentionen	35
7. Gesamteindruck	37
FAQ zur Toolbox.....	39
Gestaltungsbeispiele zu den Instrumenten der Toolbox	41
Literaturverzeichnis.....	49
Autorenprofile	53
Haftungsausschluss	53

Abkürzungen und Begriffsklärung

α	Cronbachs α zeigt an, wie gut die Items (Fragen) innerhalb einer Skala zusammenpassen („interne Konsistenz“); dies wird oft als Hinweis auf die \rightarrow Reliabilität interpretiert. Cronbachs α kann einen Wert zwischen $\alpha = 0$ und $\alpha = 1$ annehmen. Werte von $\alpha = .70$ gelten als noch akzeptabel für Analysen / Untersuchungen auf Gruppenebene, Werte von $\alpha = .80$ als ausreichend und von $\alpha = .90$ bis $.95$ als sehr gut. Werte von $\alpha > .95$ sind kritisch zu sehen (die verschiedenen Items unterscheiden sich dann meist nicht mehr ausreichend voneinander).
M	Mittelwert eines Items oder Skala. In der Darstellung von Ergebnissen wird zusätzlich die Standardabweichung ($\rightarrow SD$) berichtet.
m	Anzahl von beurteilten Websites
n	Anzahl der BeurteilerInnen / UmfrageteilnehmerInnen
Objektivität	Das Hauptgütekriterium Objektivität ist dann gegeben, wenn die Ergebnisse unabhängig von der Person sind, die eine Befragung <i>durchführt, auswertet</i> und <i>interpretiert</i> . Ein Computerfragebogen mit immer gleicher Instruktion und fester Auswertung ist folglich objektiver als ein freies Interview.
r	Korrelationskoeffizient, Maß für den Zusammenhang zwischen zwei Variablen. Korrelationen können Werte zwischen -1 und 1 annehmen; 0 wäre ein reiner Zufallswert, ± 1 eine perfekte proportionale Übereinstimmung. Zusammenhänge ab ca. $r = .30$ werden als mittelmäßig ausgeprägt angesehen, ab ca. $r = .50$ kann ein starker Zusammenhang angenommen werden.
Reliabilität	Reliabilität ist ein weiteres Hauptgütekriterium und befasst sich mit der reinen Messgenauigkeit. Meist wird hier Cronbachs $\rightarrow \alpha$ als Maß angegeben. Ferner lassen sich die Messwerte aus zwei verschiedenen Messzeitpunkten korrelieren um die zeitliche Stabilität eines Instrumentes zu bestimmen.
SD	engl. <i>Standard Deviation</i> = Standardabweichung, statistisches Maß für die Variabilität (= Streuung bzw. Varianz) der Antworten bei einer Frage oder Skala.
Validität	Das Hauptgütekriterium Validität fragt, ob ein Fragebogen oder Test wirklich das Merkmal misst, das gemessen werden soll. Validität ist schwieriger zu bestimmen als die Reliabilität. Im Idealfall kommen mehrere Verfahren zur Anwendung: Man vergleicht beispielsweise einen neuen Test mit vorhandenen Verfahren, die das gleiche messen (dann sollte es hoch korrelieren = <i>konvergente Validität</i>) oder die etwas ganz anderes messen (dann sollte es wenig bis gar nicht korrelieren = <i>divergente Validität</i>); man kann außerdem Expertenurteile oder andere vergleichbare Kriterien heranziehen (= <i>konkurrente Validität</i>); führt Experimente durch (= <i>experimentelle Validität</i>); testet die Fähigkeit des Instruments, zwischen verschiedenen Zielobjekten zu unterscheiden (= <i>diskriminante Validität</i>); oder versucht, die angenommenen Faktoren mittels einer konfirmatorischen Faktorenanalyse zu bestätigen (= <i>faktorielle Validität</i>).

Einführung und theoretischer Hintergrund

Allgemeines zu User Experience

Das Internet nutzen 83,8 % der Deutschen, über zwei Stunden täglich verbringen sie dabei durchschnittlich online. Neben E-Mail-Kommunikation ist die Informationssuche die häufigste Online-Tätigkeit (ARD/ZDF-Onlinestudie 2016; Koch & Frees 2016). Sehr schnell und spontan treffen Online-User eine Auswahl aus der Vielzahl an unterschiedlichen verfügbaren Websites. Entscheidend ist hier die persönliche Bewertung und das Erleben einer Website – die sogenannte **User Experience**. Dieser Begriff umfasst alle Erfahrungen des Users bei der Interaktion mit einer Website, was auch die Erwartungen hinsichtlich der zukünftigen Benutzung einschließt (vgl. ISO, 2009).

Ein gängiges Modell der User Experience stammt von Thüring und Mahlke (2007): Bei der Nutzung eines technischen Systems, also auch Websites, sind zunächst Eigenschaften des Systems selbst sowie seiner User und des Kontextes relevant (siehe Abbildung 1). Diese beeinflussen die Interaktionsmerkmale in der Nutzung des Systems, aus denen sich zwei zentrale Komponenten der User Experience ergeben: aufgabenbezogene und nicht-aufgabenbezogene Wahrnehmungen (vgl. Hassenzahl et al., 2000; Hassenzahl, Burmester & Koller, 2003; Thüring & Mahlke, 2007). Aufgabenbezogene Aspekte (auch bezeichnet als instrumentelle oder pragmatische Qualitäten) umfassen dabei insbesondere die Usability eines Systems oder einer Website. Nicht-aufgabenbezogene Aspekte (auch als nicht-instrumentelle oder hedonistische Qualitäten bezeichnet) umfassen die Wahrnehmung der User von Ästhetik und Attraktivität. Beide zusammen bedingen emotionale Reaktionen der User. Diese verschiedenen Komponenten von User Experience führen als Konsequenz zu einer Einschätzung des Systems oder der Website, die die Akzeptanz und das Nutzungsverhalten determiniert.

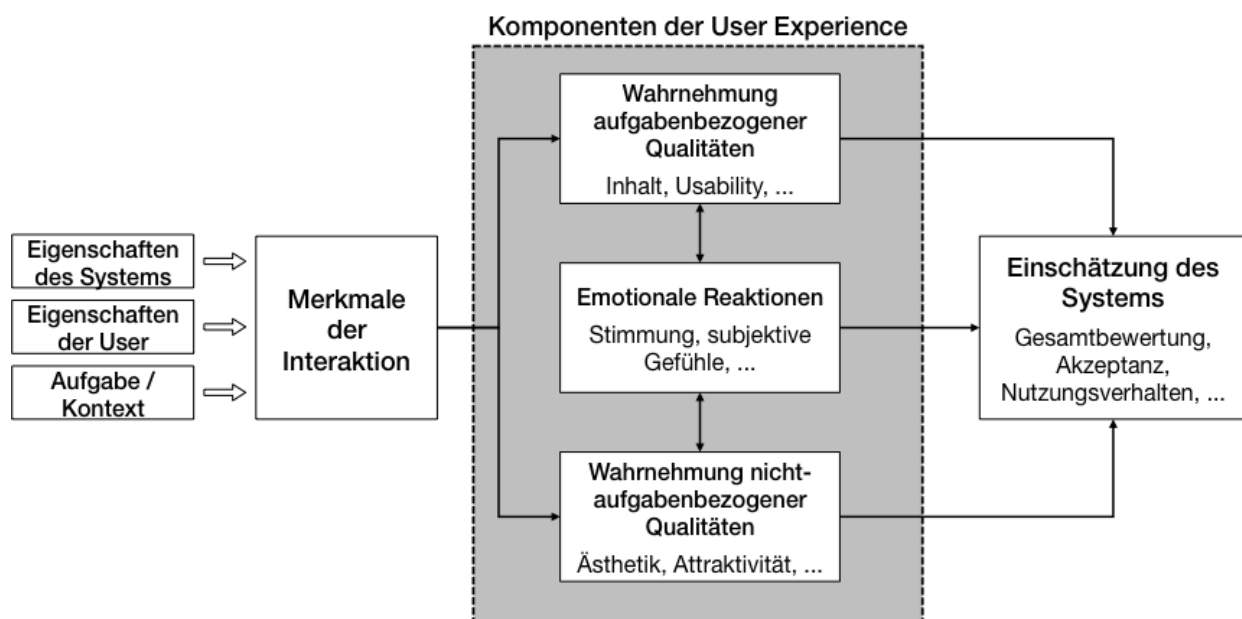


Abbildung 1: Komponenten der User Experience (Darstellung angelehnt an Thüring & Mahlke, 2007)

Abbildung 1 verdeutlicht, dass die User Experience einer Website durch Faktoren beeinflusst sein kann, die nicht durch die Website selbst determiniert sind. So gibt es beispielsweise Eigenschaften der NutzerInnen, die die User Experience verändern. Studien gibt es hier unter anderem zu Aspekten wie dem biologischen Alter (z.B. Sonderegger, Schmutz & Sauer, 2016; Thielsch, 2008), dem Geschlecht (z.B. Bardzell, 2011; Tuch et al., 2010) oder der Persönlichkeit eines Users (z.B. Bosnjak, Galesic & Tuten, 2007; Thielsch, 2008). Auch Erkrankungen, wie beispielsweise eine Depression, können die User Experience negativ beeinflussen (Thielsch & Thielsch, under review). Je spezifischer hier die Anforderungen oder die Zielgruppen einer Website sind, desto genauer sind diese Aspekte zu berücksichtigen. Aufgrund der allgemeinen, gesunden und berufstätigen Zielgruppe der Plattform infocor fokussiert die vorliegende Toolbox auf die zentralen Konstrukte der User Experience von Websites und deren Evaluation.

Zentrale Konstrukte der Website User Experience

In der gegenwärtigen Forschung erweisen sich drei Eigenschaften von Websites als zentral für die Wahrnehmung: Inhalt, Usability und Ästhetik (bspw. Cober et al., 2003; De Angeli, Sutcliffe & Hartmann, 2006; Thielsch, Blotenberg & Jaron, 2014). Alle drei Konstrukte beeinflussen die gesamte Web User Experience. Inhalt und Usability sind den aufgabenbezogenen Komponenten von User Experience zuzuordnen, Ästhetik den nicht-aufgabenbezogenen.

Inhalt

Der Inhalt ist das wichtigste Merkmal von Websites und unterscheidet diese in besonderem Maße von anderen interaktiven Produkten und klassischen Software-Anwendungen. Das World Wide Web ist ein sehr schnelllebiges Medium, in dem Unternehmen und Institutionen einem enormen Wettbewerb ausgesetzt sind. Der Inhalt einer Website wird als das wichtigste Kriterium für die Beurteilung einer Webseite genannt – noch vor Usability und Ästhetik (Thielsch et al., 2014).

ISO-Norm 9241-151 (ISO, 2006) definiert Website-Inhalt als die Zusammenstellung von Informationsobjekten, die in Form von Ton, Text oder Video präsentiert werden können. Der Inhalt zeichnet sich aber nicht nur durch objektivierbare Eigenschaften wie syntaktische Struktur, Fehlerfreiheit oder Optimierung für Suchmaschinen, sondern auch durch die subjektive Wahrnehmung der User aus (Thielsch, 2008; Thielsch & Hirschfeld, under review). Die Wahrnehmung von Webinhalten wiederum bedingt eine Vielzahl von Aspekten wie beispielsweise die Zufriedenheit der User, deren Präferenz oder Vertrauen in eine Website (vgl. bspw. Aranyi & van Schaik, 2015; Kang & Kim, 2006; Rocha, 2012).

Usability

Das zweite zentrale Konstrukt in der Website-Wahrnehmung ist Usability (allgemeine Überblicke finden sich bei Salaschek et al., 2007; Sarodnick & Brau, 2015; Shneiderman, Plaisant & Cohen, 2016). Usability wird im Deutschen auch als Brauchbarkeit, (Be)nutzerfreundlichkeit oder Gebrauchstauglichkeit bezeichnet.

In der DIN EN ISO 9241-11 (1998) ist Gebrauchstauglichkeit als Effektivität, Effizienz und Zufriedenheit beschrieben, mit der User mit einem System vorgegebene Ziele erreichen können. Websites sollten sich also leicht bedienen lassen und ermöglichen, dass User schnell die gewünschten Informationen finden können. Wie der Inhalt ist Usability einerseits

objektivierbar (z.B. Ladegeschwindigkeit oder Linktiefe einer Website), hat aber andererseits ebenfalls eine subjektive Komponente (vgl. Hornbæk, 2006; Kurosu & Kashimura, 1995). Usability beeinflusst nicht nur die Zufriedenheit der User, sondern auch viele weitere Aspekte wie zum Beispiel die subjektive Wahrnehmung einer Organisation und das Vertrauen in die Website (bspw. Cober et al., 2003; Flavián, Guinalú & Gurrea, 2006; Lee & Koubek, 2012).

Ästhetik

Das dritte zentrale Element der Website-Wahrnehmung ist die Schönheit des Designs; diese wird oft als Ästhetik bezeichnet. Ästhetik ist in der Forschung zumeist als unmittelbare, angenehme und subjektive Wahrnehmung eines Webobjekts definiert, die wenig durch schlussfolgernde Prozesse beeinflusst ist (vgl. Moshagen & Thielsch, 2010). Da sie sehr schnell wahrgenommen werden kann, etwa innerhalb der ersten halben Sekunde der Website-Nutzung (vgl. Bölte et al., 2017), hat Ästhetik eine besondere Bedeutung für den Ersteindruck (Lindgaard et al., 2006; Thielsch & Hirschfeld, 2012; Tractinsky et al., 2006). Sie beeinflusst aber auch viele weitere Aspekte wie zum Beispiel Zufriedenheit und Vertrauen der User oder zugeschriebene Attribute wie die Attraktivität einer Organisation (ein Überblick findet sich bei Moshagen & Thielsch, 2010).

Das Zusammenspiel von Inhalt, Usability und Ästhetik

Eine Studie von Thielsch und Kollegen (2014) zeigt, wie die drei Hauptmerkmale einer Website verschiedene Bewertungen der User beeinflussen: Der Ersteindruck einer Website wird vor allem durch die Ästhetik bestimmt, während Inhalt und Usability neben der Ästhetik ebenfalls signifikanten Einfluss haben (siehe Abbildung 1). Alle drei Aspekte bestimmen gemeinsam den Gesamteindruck nach der Nutzung einer Website. Vor allem der Inhalt hat ein großes Gewicht bei Wiederbesuchs- oder Weiterempfehlungsbereitschaft der User. Praktisch heißt das: Ein User wird durch hohe Ästhetik im Webdesign angezogen und dann gebunden mit guten Inhalten, die benutzerfreundlich dargeboten sind. Wenn es darum geht die Website wieder zu besuchen oder weiterzuempfehlen, ist der Inhalt zentral. In geringerem Anteil wirkt dabei die Ästhetik weiter (siehe Thielsch et al., 2014).

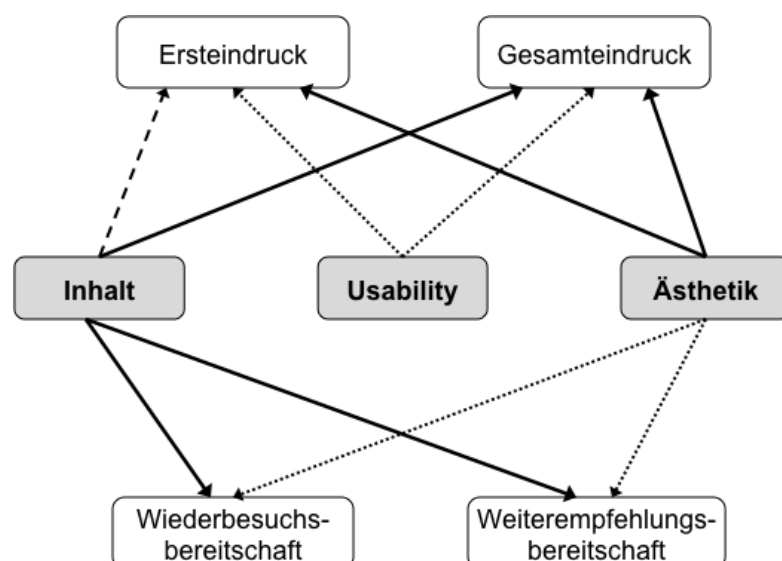


Abbildung 2: Einfluss der Wahrnehmung von Inhalt, Usability und Ästhetik auf Erst- und Gesamteindruck, bzw. Wiederbesuchs- und Weiterempfehlungsbereitschaft.

Gepunktete Linie = schwacher Einfluss, gestrichelte Linie = mittlerer Einfluss, durchgezogene Linie = großer Einfluss (Abbildung aus Thielsch, Grobien et al., 2014)

Struktur der Darstellung der Instrumente

Entsprechend dem dargestellten Stand der Forschung zur Website-Wahrnehmung sind zentrale Konstrukte in der Toolbox durch verschiedene Instrumente erfasst (siehe Tabelle 1). In der Standardversion enthält die Toolbox 35 Items aus sieben Bereichen. Für umfassendere Evaluationen werden weitere Fragebögen vorgeschlagen, sodass eine Website-Analyse mit bis zu 68 Items in der erweiterten Version vorgenommen werden kann. Die Instrumente werden nachfolgend detailliert dargestellt (ab Seite 10). Zudem finden sich im folgenden Abschnitt weitere Hintergrundinformationen zur Erstellung der Toolbox.

Die Dokumentation der eingeschlossenen Instrumente in der vorliegenden Toolbox umfasst die deutschsprachigen Instrumente; die Informationen sind wie folgt strukturiert:

- a.) Hinweise zu Autoren und Zitation
- b.) Kurzbeschreibung zum Instrument
- c.) Informationen zu Gütekriterien (Reliabilität und Validität)
- d.) Items und Itemanker
- e.) Auswertungsanweisungen und (wenn verfügbar) Benchmarks sowie Interpretationshinweise
- f.) ggf. weiterführende Informationen und Hinweise
- g.) ein Hinweis, wann das Instrument in die Toolbox aufgenommen wurde

Die digitale Version der Toolbox enthält zusätzlich a) Excel-Dateien zur halbautomatischen Auswertung der Instrumente sowie b) eine .gpx-Datei zum direkten Import der Instrumente in die Online-Befragungssoftware EFS Survey.

Güte der Instrumente

Zentrale Kriterien der psychometrischen Güte eines Fragebogenverfahrens sind Objektivität, Reliabilität und Validität (siehe bspw. Bühner, 2010; Moosbrugger & Kelava, 2012). Hinsichtlich der **Objektivität** gilt für alle in der Toolbox beschriebenen Verfahren: Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist die Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität. Die Objektivität der Interpretation eines Verfahrens nimmt sicherlich zu, wenn Entscheidungshilfen und/oder Vergleichswerte gegeben sind, daher ist dies in Tabelle 1 im Überblick dargestellt.

Hinsichtlich der Reliabilität und Validität der jeweiligen Verfahren findet sich ebenfalls ein Überblick in der nachfolgenden Tabelle 1 sowie weitere Informationen im Detail bei der jeweiligen Beschreibung des Verfahrens (ab Seite 10). Zur Reliabilität muss dabei einschränkend gesagt werden, dass Retest-Messungen im Bereich User Experience bisher leider noch unüblich sind; Angaben zur Reliabilität beziehen sich daher auf die interne Konsistenz (Cronbachs Alpha).

Konstrukt	Instrument	Nr.	# Items	Quelle	Standardversion	Erweiterte Version	Interpretationshilfen	Reliabilität	Validität
1. Ersteindruck	Einzelitems	1.1	4	Thielsch (2008)	✓	✓	-	*	*
2. Subjektive Inhaltswahrnehmung	Message Credibility Scale	2.1	3	Appelman & Sundar (2016)	✓	✓	✓	*	*
	WWI (Fragebogen zur Wahrnehmung von Website-Inhalten)	2.2	9	Thielsch (2008)	✓	✓	✓	je Skala (*) bis **	(*)
	Trusting Beliefs	2.3	11	McKnight et al. (2002)		(✓) ^a	-	**	(*)
3. Subjektive Usability / Nutzerzufriedenheit	UMUX-Lite (Usability Metric for User Experience - Lite)	3.1	2	Lewis et al. (2013)	✓		-	*	*
	PWU-G (Perceived Website Usability - German)	3.2	7	Flavián et al. (2006), Thielsch (2008)	✓	✓	✓	**	**
	System Usability Scale (SUS)	3.3	10	Brooke (1996)		✓	✓	je Studie * bis **	*
4. Visuelle Ästhetik	VisAWI-S (Visual Aesthetics of Websites Inventory - Short)	4.1	4	Moshagen & Thielsch (2013)	✓		✓	je Studie (*) bis *	**
	VisAWI (Visual Aesthetics of Websites Inventory)	4.2	18	Moshagen & Thielsch (2010)		✓	✓	Skalen: * Gesamtwert: **	**
5. Emotionale Reaktion: Befindlichkeit, Zustimmung, Zufriedenheit	Smiley-Skala	5.1	1	Jäger (2004)	✓	✓	✓	nb	**
6. Gesamteindruck	Einzelitem Gesamteindruck	6.1	1	Thielsch (2008)	✓	✓	✓	nb	(*)
7. Handlungs- und Nutzungsintentionen	Wiederbesuchs-Skala (Scale assessing the intention to revisit the website)	7.1	4	Moshagen & Thielsch (2010)	✓	✓	-	**	(*)

Tabelle 1: In die Toolbox eingeschlossene Items/Instrumente (insgesamt 35 Items in der Standardversion, bis zu 68 Items in der erweiterten Version) und deren Quellen. Hinsichtlich der Reliabilität wurden zu den Verfahren die interne Konsistenz (Cronbachs Alpha) betrachtet und wie folgt dargestellt ** = $\alpha \geq .9$; * = $\alpha \geq .8$; (*) = $\alpha \geq .7$; - = $\alpha < .7$. Bei Einzelitems kann Cronbachs Alpha nicht berechnet werden (dort „nb“), Hinweise zur Retest-Reliabilität liegen bei diesen Items derzeit nicht vor. Hinsichtlich der Validität gilt: ** = umfassende Belege für hohe Validität liegen vor; * = Validitätsbelege liegen vor; (*) = nur wenige Hinweise zur Validität liegen vor bzw. diese könnte eingeschränkt sein.
^a Die Trusting Beliefs-Skala ist in der Forschung sehr weit verbreitet; da notwendige Validitätsprüfungen noch ausstehen, ist ihr Einsatz im Einzelfall zu prüfen.

Hintergrundinformationen zur Instrumentenauswahl

Die Toolbox wurde aus frei verfügbaren, gut erprobten und validierten deutschsprachigen Evaluationsinstrumenten zusammengestellt. Lediglich im Bereich Glaubwürdigkeit bzw. Vertrauensüberzeugungen wurde auf englischsprachige Instrumente zurückgegriffen, zu denen aktuell nur erste Übersetzungen ins Deutsche aus dem Forschungskontext vorlagen. In diesem Analysebereich ließen sich leider keine in Deutsch umfassend validierten Instrumente auffinden. Ziel der Zusammenstellung der Instrumente in dieser Toolbox war eine umfassende, quantitative Evaluation der subjektiven Eindrücke von Website-NutzerInnen. Hierbei wurden nur Instrumente mit hinreichender Messgenauigkeit (Reliabilität) und ausreichend hoher Validität in Betracht gezogen (bei Erst- und Gesamteindruck ergänzt um Einzelitems, die sich in der Forschung bewährt haben). Vorgabe war zudem, dass die Fragebögen vollständig lizenzfrei durch die BZgA nutzbar sein und sich in früheren empirischen Studien bewährt haben sollen.

Die Zusammenstellung der Instrumente erfolgte unter Berücksichtigung des gegenwärtigen Forschungsstandes im Bereich User Experience und Usability von Websites (siehe oben), ergänzt um gezielte Recherchen zu verfügbaren Skalen und Fragebögen. Dabei wurden insbesondere die *Digital Library Mensch-Computer Interaktion* der Gesellschaft für Informatik e.V. gesichtet sowie eine Entwurfsversion einer Fragebogenmatrix des Arbeitskreises User Research der German UPA e.V. herangezogen. Zusätzlich erfolgte eine Sichtung des Testverzeichnisses des Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)¹, des ZIS-Portals² des GESIS – Leibniz-Institut für Sozialwissenschaften (ZIS steht für „Zusammenstellung sozialwissenschaftlicher Items und Skalen“) – und des Psychologisch-Medizinischen Testportals Psychometrikon³ sowie eine gezielte Ansprache einzelner Testautoren.

Ergänzend sei an dieser Stelle darauf hingewiesen, dass typische deutschsprachige Instrumente zur allgemeinen User Experience-Bewertung – wie beispielsweise der AttrakDiff (Hassenzahl et al., 2003), der meCUE (Minge & Riedel, 2013) oder der UEQ (Laugwitz, Schrepp & Held, 2006) – nicht in die Toolbox eingeschlossen wurden. Vorrangiger Grund ist, dass diese Instrumente vor allem zur Evaluation (interaktiver) Produkte und weniger zur Website-Analyse geeignet sind. Hinsichtlich des UEQ wird hier zukünftig (voraussichtlich im Herbst 2017) eine Kurzversion vorliegen, deren Eignung hinsichtlich der Website-Evaluation dann neu zu bewerten ist.

¹ Verfügbar unter <https://www.zpid.de/index.php?wahl=products&uwahl=printed&uuwahl=testverzeichnis>

² Verfügbar unter <http://zis.gesis.org/>

³ Verfügbar unter <http://www.psychometrikon.de>

Darstellung der Instrumente in der Toolbox

1. Ersteindruck

1.1 Einzelitems Ersteindruck

Autor

Meinald T. Thielsch

Zitation

Thielsch, M. T. (2008). *Ästhetik von Websites. Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen*. Münster: MV Wissenschaft.

Kurzbeschreibung

Der Ersteindruck einer Website ist von hoher Bedeutung und wird insbesondere durch die Ästhetik bestimmt (Thielsch et al., 2014; Tuch et al., 2012). Tractinsky et al. (2006) stellen heraus, dass es keine zweite Chance gibt, einen ersten Eindruck auf die User zu machen. Neben der Ästhetik einer Website haben aber Inhalt und Usability ebenfalls hochsignifikanten Einfluss (Thielsch et al., 2014). Ebenso bedingen alle drei Konstrukte den Gesamteindruck nach der Nutzung einer Website. Zur Erfassung des Ersteindrucks werden an dieser Stelle vier Items vorgeschlagen, die in der Forschung wiederholt zum Einsatz gekommen sind und sich bewährt haben.

Gütekriterien

Reliabilität: Es ist nicht zwingend vorgesehen aus den vier Items eine Skala zu bilden (siehe unten, Auswertungsanweisung). Zieht man jedoch die Daten von Thielsch (2008, n = 364) heran, ergibt sich für eine solche Ersteindrucksskala der vier Items ein Cronbachs α von .83.

Validität: Thielsch (2008b) berichtet hohe Korrelationen der Ersteindrucksmaße untereinander ($.51 \leq r \leq .67$), zu relevanten Website-Fragebögen ($.37 \leq r \leq .58$) sowie klare Zusammenhänge zum Gesamteindruck ($.53 \leq r \leq .59$).

Items

- Mein erster Eindruck: Ich finde den Inhalt dieser Website interessant.
- Mein erster Eindruck: Ich halte diese Website für gut benutzbar.
- Mein erster Eindruck: Ich finde, diese Website ist ansprechend gestaltet.
- Mein erster Eindruck: Ich gebe dieser Website die Gesamtnote...

Format Antwortanker

Die ersten drei Items werden mit einer 7-stufigen Likert-Skala (kodiert von 1 „stimme gar nicht zu“ bis 7 „stimme voll zu“) dargeboten. Die verbalen Itemanker sind:

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *neutral*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

Die Frage nach der Ersteindrucks-Gesamtnote ist mit folgenden Ankern versehen:

- 1 (sehr gut)
- 2 (gut)
- 3 (befriedigend)
- 4 (ausreichend)
- 5 (mangelhaft)

Auswertungsanweisung

Zu den Angaben auf den vier Items wird jeweils ein Mittelwert pro Frage gebildet. Eine weitere Verrechnung der vier Items ist nicht vorgesehen. Auf Basis der inhaltlichen Nähe, Itemkorrelationen untereinander und akzeptabler Konsistenzwerte können die Items jedoch im Bedarfsfall auch in einer Skala verrechnet werden (aufgrund der unterschiedlichen Ankerformate müssen die Items vor der Verrechnung hierfür aber z-transformiert werden⁴).

Benchmarks / Vergleichswerte

Spezifische Benchmarks oder Vergleichswerte liegen derzeit nicht vor. Grundsätzlich ist bei der gewählten Form der Items mit Likert-Skala zumindest im Durchschnitt ein Wert über dem Skalenmittel erwünscht (das heißt, dass die Befragten den Aussagen zumindest in der Tendenz zustimmen). In ihrer Formulierung ist die zudem die Note zum Ersteindruck sehr ähnlich zum Gesamteindruck, sodass hier möglicherweise ähnliche Vergleichswerte angelegt werden können (siehe dazu die Darstellung unter 7. Gesamteindruck).

Aufnahme in die Toolbox

In der Toolbox seit Version 1.0 (30. November 2016).

⁴ siehe Döring und Bortz (2016) oder bspw. https://de.wikipedia.org/wiki/Standardisierung_%28Statistik%29

2. Subjektive Inhaltswahrnehmung

2.1 Message Credibility Scale

Autoren

Alyssa Appelman & S. Shyam Sundar

Zitation

Appelman, A., & Sundar, S. S. (2016). Measuring Message Credibility: Construction and Validation of an Exclusive Scale. *Journalism & Mass Communication Quarterly*, 93(1), 59–79. <http://doi.org/10.1177/1077699015606057>

Kurzbeschreibung

Die Glaubwürdigkeit von Websites ist ein zentrales Konstrukt des Website-Inhalts und dementsprechend Teil vieler Fragebögen zur Website-Evaluation. Es finden sich verschiedene Ansätze sowohl verschiedene Facetten des Konstrukts als auch die allgemeine Glaubwürdigkeit zu erfassen (siehe bspw. Appelman & Sundar, 2016; Metzger & Flanigan, 2013). Das Ziel von Appelman und Sundar (2016) war eine kurze Skala zur Bewertung der Glaubwürdigkeit von Nachrichten zu entwickeln und zu validieren. Die Autoren grenzen dabei die Glaubwürdigkeit einer Nachricht von der Glaubwürdigkeit eines Mediums oder eines Sprechers ab und fokussieren auf den Inhalt der Kommunikation. Basierend auf einer Literaturanalyse und einer Vorstudie mit zwei Fokusgruppen stellten sie 31 Adjektive zusammen und teilten diese nach formativen und reflektiven Indikatoren auf. In einer Studie mit N=322 Personen reduzierten die Autoren diese Adjektive auf 10 bzw. 3, wobei letztere die 3-Item „Message Credibility Scale“ bilden. Eine Übersetzung dieser Skala ins Deutsche wurde im Jahr 2017 von Thielsch und Hirschfeld (under review) vorgenommen.

Gütekriterien

Reliabilität: Die Message Credibility Scale zeigt eine gute interne Konsistenz ($\alpha = .87$; siehe Appelman & Sundar, 2016).

Validität: Appelman und Sundar (2016) finden eine Übereinstimmung zwischen den Urteilen von vier professionellen Journalisten und n = 322 mit der Message Credibility Scale befragten Usern in der Bewertung von sechs Nachrichtenartikeln und interpretieren dies als Hinweis auf kriterienbezogene Validität. Konvergente Validität zeigt sich in signifikanten Korrelationen zu verwandten Konstrukten ($r = .89$ zum Nachrichtenwert, $r = .75$ zu quellenbasierten Glaubwürdigkeitseinschätzungen sowie $r = .42$ zu Vorlieben der Befragten). Weiterhin argumentieren die Autoren basierend auf dem guten Modellfit in konfirmatorischen Faktorenanalysen für Konstruktvalidität.

Anmerkungen

Die Originalskala wurde anhand der Bewertungen von Nachrichtenartikeln validiert, jedoch bereits erfolgreich in der Website-Evaluation angewendet.

Items

- akkurat
- authentisch
- glaubhaft

Die Formulierungen der Originalitems in Englisch lauten:

accurate

authentic

believable

Format Antwortanker

Die Items werden mit einer 7-stufigen Likert-Skala (die Extrempunkte sind dabei kodiert mit 1 „beschreibt sehr schlecht“ und 7 „beschreibt sehr gut“) mit der Instruktion „Wie gut beschreiben die folgenden Adjektive den Inhalt den Sie gerade gelesen haben“ dargeboten.

Die Formulierung lautet im Original:

How well do the following adjectives describe the content you just read? (from 1 = describes very poorly to 7 = describes very well): accurate, authentic, believable.

Auswertungsanweisung

Es wird ein Mittelwert gebildet: Die einzelnen Angaben auf den Items werden dazu summiert und anschließend durch drei dividiert. Dieser Mittelwert beschreibt die Glaubwürdigkeit des Inhalts einer Nachricht.

Vergleichswerte / Interpretation

Spezifische Benchmarks für Websites im Gesundheitsbereich liegen nicht vor. Thielsch und Hirschfeld (under review) testeten mit der deutschen Übersetzung Websites von drei gemeinnützigen Organisationen im Bildungsbereich. Die Websites wurden in der Studie zufällig zugeteilt und waren den TeilnehmerInnen unbekannt. Insgesamt 268 Befragte bewerteten die Websites im Mittel mit $M = 4,89$ bei einer SD von 0,95. Das heißt, mit dieser Skala erfasste Werte um 4,9 entsprechen einer generell durchschnittlichen Glaubwürdigkeitsbewertung, Werte unter 3,9 wären eindeutig unterdurchschnittlich; Werte über 5,9 hingegen eindeutig überdurchschnittlich. Die Ergebnisse zur deutschen Übersetzung stimmen damit nahezu vollständig mit den von Appelman und Sundar (2016) berichteten Skalenwerten überein. Grundsätzlich sind jedoch weitere Prüfungen der deutschen Version der Skala und zusätzliche Vergleichswerte notwendig und wünschenswert.

Aufnahme in die Toolbox

In der Toolbox seit Version 2.0 (30. Juli 2017).

2.2 Fragebogen zur Wahrnehmung von Website-Inhalten (WWI)

Autor

Meinold T. Thielsch

Zitation

Thielsch, M. T. (2008). *Ästhetik von Websites. Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen*. Münster: MV Wissenschaft.

Kurzbeschreibung

Der Inhalt einer Website ist das wichtigste Kriterium für die Beurteilung durch die NutzerInnen (Thielsch et al., 2014) und in ISO 9241-151 definiert (ISO, 2006). Der WWI ist ein Fragebogen, der insbesondere die subjektive Wahrnehmung online dargebotener Inhalte erfasst. Diese wird auf drei Skalen („Gefallen“, „Verständlichkeit“ sowie „Qualität und Nutzen“) mit jeweils drei Fragen bewertet. Erstellt wurde der WWI auf Basis umfassender Literaturrecherchen sowie zweier Studien mit Experten und Online-Usern.

Gütekriterien

Reliabilität: Der WWI ist für Auswertungen auf Gruppenebene ausreichend reliabel (je nach Studie $.71 \leq \alpha \leq .90$; siehe Thielsch, 2008). Die Berechnung eines Gesamtwertes über die drei Skalen hinweg ist zulässig und zeigt ein Cronbachs α von .88 (Thielsch et al., 2014).

Validität: Thielsch (2008) argumentiert für eine hohe inhaltliche Validität des WWI, da sich sehr deutliche Korrelation der jeweils verkürzten Skalen zur ursprünglichen vollständigen Skala zeigen ($.92 \leq r \leq .95$). Thielsch et al. (2014) berichten aber auch hohe Korrelationen des WWI mit Usability und Ästhetik-Maßen ($.61 \leq r \leq .76$), was eine eingeschränkte divergente Validität anzeigt.

Anmerkungen / weitere Informationen

Zukünftig wird eine optimierte Folgeversion des WWI zur Verfügung stehen: Ein überarbeiteter und um den Faktor Glaubwürdigkeit erweiterter Fragebogen ist derzeit in Begutachtung (Thielsch & Hirschfeld, under review). Zudem wird in 2018 eine 4-Item Kurzversion des neuen Instruments erprobt. Zukünftige Versionen dieser Toolbox sollten diese überarbeiteten Instrumente anstelle des WWI aufnehmen.

Items

Skala 1: Gefallen

- Die Website weckt mein Interesse.
- Der Inhalt der Website gefällt mir.
- Ich lese diese Website gerne.

Skala 2: Verständlichkeit

- Die einzelnen Sätze sind einfach zu lesen.
- Die Texte liefern mir kurz und bündig die wichtigsten Informationen.
- Der Sprachgebrauch in den Texten ist geläufig und allgemein verständlich.

Skala 3: Qualität und Nutzen

- Die Informationen sind qualitativ hochwertig.
- Ich finde die Informationen auf der Website nützlich.
- Die Inhalte der Website erscheinen mir so wichtig, dass ich sie mir ausdrucken oder speichern würde.

Format Antwortanker

Die Items werden mit einer 7-stufigen Likert-Skala (kodiert von 1 „stimme gar nicht zu“ bis 7 „stimme voll zu“) dargeboten. Die verbalen Itemanker sind:

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *neutral*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

Auswertungsanweisung

Je Skala wird ein Mittelwert gebildet: Die einzelnen Itemwerte werden dazu summiert und anschließend durch drei dividiert. Ein Gesamtmittelwert ergibt sich durch die Summe der drei Skalenmittelwerte geteilt durch drei; dieser Wert repräsentiert die subjektive Einschätzung eines generellen Inhalts-Faktors.

Benchmarks / Vergleichswerte / Interpretation

Spezifische Benchmarks im Gesundheitsbereich liegen nicht vor. Allgemeine Vergleichswerte zum WWI finden sich bei Thielsch et al. (2014): Insgesamt 42 Websites wurden zufällig zugeteilt von 512 Personen bewertet, dabei lag der WWI-Gesamtmittelwert bei 3,69 ($SD = 1,17$). Das heißt, WWI-Gesamtwerte um 3,7 entsprechen einer generell durchschnittlichen Inhaltsbewertung, Werte unter 2,5 wären eindeutig unterdurchschnittlich; Werte über 4,9 hingegen eindeutig überdurchschnittlich und Werte über 6,0 herausragend positiv. Zu beachten ist hier, dass die Befragten weitgehend unbekannte Websites bewertet und diese ohne ein persönliches Vorinteresse besucht haben.

Die subjektive Inhaltswahrnehmung bedingt eine Vielzahl von Aspekten, wie beispielsweise die Zufriedenheit der User oder deren Vertrauen in eine Website (vgl. bspw. Aranyi & van Schaik, 2015; Kang & Kim, 2006; Rocha, 2012). Sie hat einen starken Einfluss auf die Gesamtbewertung einer Website und zeigt den stärksten Einfluss auf die Wiederbesuchs- und Weiterempfehlungsbereitschaft der User (Thielsch et al., 2014).

Aufnahme in die Toolbox

In der Toolbox seit Version 1.0 (30. November 2016).

Erweiterte Erfassung von Inhaltswahrnehmungen: Vertrauensüberzeugungen

Ein im Rahmen der Forschung zu Vertrauen online oft eingesetztes Instrument stammt von McKnight und Kollegen (2002). Es fokussiert auf die Vertrauensüberzeugungen („Trusting Beliefs“) in Bezug auf eine konkrete Website. Soll dieser Aspekt des Website-Inhalts besonders analysiert werden, so ist eine Erfassung mithilfe dieser Skalen zu Trusting Beliefs denkbar.⁵

2.3 Trusting Beliefs

Autoren

D. Harrison McKnight, Vivek Choudhury & Charles Kacmar

Zitation

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359. <http://doi.org/10.1287/isre.13.3.334.81>

Kurzbeschreibung

Das Ziel von McKnight und Kollegen (2002) war die Schaffung eines Fragebogens zur Erfassung von Vertrauen der User in Websites. Die Vertrauensüberzeugungen in Bezug auf eine Website teilen die Autoren dabei in drei Faktoren ein: Kompetenz, Benevolenz und Integrität. Mit dieser Einteilung folgen sie einem in der Forschung sehr etablierten und oftmals verwendeten Vertrauensmodell (siehe Mayer, Davis & Schoorman, 1995). Während die Message Credibility Scale von Appelman und Sundar (2016) auf die Wahrnehmung des Inhalts fokussiert, erfassen diese Skalen zu Trusting Beliefs (Vertrauensvorstellungen) die Wahrnehmung der User hinsichtlich der Quelle des Inhalts, sprich des Akteurs, der hinter einer Website steht.

Die deutsche Version der Skala stammt aus laufenden Forschungsprojekten. Dabei wurde die Skala ins Deutsche und von einem unabhängigen Übersetzer wieder zurück ins Englische übersetzt, Abweichungen wurden nachfolgend soweit möglich behoben. Eine systematische Prüfung der deutschen Übersetzung steht derzeit allerdings noch aus. Auch im Hinblick auf die unklare faktorielle Struktur des Originals sind weitere Prüfungen der Validität erforderlich.

⁵ Ein alternatives Instrument zur Erfassung der Vertrauensüberzeugungen findet sich bei Flavián et al. (2006; siehe auch Casaló, Flavián & Guinalú, 2007): Dieses besteht aus insgesamt 15 Items auf drei Skalen („Honesty“, „Benevolence“ und „Competence“). Die Autoren berichten gute bis sehr gute Reliabilitätswerte im Sinne interner Konsistenz (je Skala $.83 \leq \alpha \leq .94$) und erste Validitätsnachweise. Jedoch liegt auch zu dieser Skala keine validierte deutschsprachige Version vor.

Gütekriterien

Reliabilität: McKnight und Kollegen (2002) berichten sehr gute Reliabilitätswerte im Sinne interner Konsistenz (je Skala $.91 \leq \alpha \leq .95$).

Validität: Es wurden verschiedene Validierungsstrategien angewendet; die Autoren finden hierbei Evidenz für konvergente und diskriminante Validität (McKnight et al., 2002). Die angenommene faktorielle Struktur der drei Skalen kann jedoch in der Originalstudie nicht einwandfrei bestätigt werden.

Anmerkungen / weitere Informationen

Eine Besonderheit des Fragebogens ist, dass gezielt nach dem Akteur sowie teilweise dem konkreten Leistungsbereich der Website gefragt wird. In der englischen Erstveröffentlichung beziehen sich daher die Fragen auf die Website LegalAdvice.com, in nachfolgenden Studien wurden hier die Fragen stets entsprechend abgewandelt. Nachfolgend sind hierzu Anmerkungen in den Items in eckigen Klammern hinterlegt.

Items

Skala 1: Benevolenz

- Ich glaube, dass [Name der Website] in meinem besten Interesse handeln würde.
- Wenn ich Hilfe bräuchte, würde [Name der Website] sein Bestes tun, um mir zu helfen.
- [Name der Website] interessiert sich für mein Wohlergehen, nicht nur für das eigene.

Skala 2: Integrität

- [Name der Website] ist aufrichtig im Umgang mit mir.
- Ich würde [Name der Website] als ehrlich beschreiben.
- [Name der Website] würde gemachte Zusagen einhalten.
- [Name der Website] ist aufrichtig und ehrlich.

Skala 3: Kompetenz

- [Name der Website] ist kompetent und effektiv in [Dienstleistung der Website].
- [Name der Website] erfüllt seine Rolle als [Dienstleistung der Website] sehr gut.
- Insgesamt ist [Name der Website] eine fähige und tüchtige [Dienstleistungsbereich der Website].
- Im Allgemeinen ist [Name der Website] sehr sachkundig was [Dienstleistung der Website] betrifft.

Die Formulierungen der Originalitems in Englisch lauten:

Benevolence scale:

- *I believe that LegalAdvice.com would act in my best interest.*
- *If I required help, LegalAdvice.com would do its best to help me.*
- *LegalAdvice.com is interested in my well-being, not just its own.*

Integrity scale

- *LegalAdvice.com is truthful in its dealings with me.*
- *I would characterize LegalAdvice.com as honest.*
- *LegalAdvice.com would keep its commitments.*
- *LegalAdvice.com is sincere and genuine.*

Competence scale

- *LegalAdvice.com is competent and effective in providing legal advice.*
- *LegalAdvice.com performs its role of giving legal advice very well.*
- *Overall, LegalAdvice.com is a capable and proficient Internet legal advice provider.*
- *In general, LegalAdvice.com is very knowledgeable about the law.*

Format Antwortanker

Die Items werden mit einer 7-stufigen Likert-Skala (kodiert von 1 „stimme überhaupt nicht zu“ bis 7 „stimme voll zu“) dargeboten. Die verbalen Itemanker sind:

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *neutral*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

Auswertungsanweisung

Die einzelnen Skalenwerte werden addiert, die Summe wird durch die Anzahl der Items der jeweiligen Skala dividiert (Beispiel: bei Skala „Benevolenz“ ist der Divisor 3). Die Mittelwerte zeigen mit höheren Werten eine stärkere Ausprägung der jeweiligen Skala an.

Benchmarks / Vergleichswerte / Interpretation

Spezifische Benchmarks oder Vergleichswerte für die deutsche Übersetzung der Skala liegen derzeit nicht vor. Aufgrund der starken Anpassung der Items an die Zielwebsite ist hier unsicher, inwieweit Vergleichswerte englischsprachiger Studien genutzt werden können.

Aufnahme in die Toolbox

In der Toolbox seit Version 2.0 (30. Juli 2017).

3. Subjektive Usability / Nutzerzufriedenheit

3.1 Usability Metric for User Experience – Lite (UMUX-Lite)

Autoren

James R. Lewis, Brian S. Utesch & Deborah E. Maher

Zitation

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013, April). UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2099-2102). ACM.

Kurzbeschreibung

Usability ist ein zentrales Merkmal interaktiver Produkte wie Software oder Websites. In der DIN EN ISO 9241-11 (1998) ist Usability definiert als Effektivität, Effizienz und Zufriedenheit, mit der User mit einem System vorgegebene Ziele erreichen können. Ein weit verbreitetes Instrument zur Erfassung von Software Usability, das auch zur Website-Evaluation eingesetzt wird, ist die „System Usability Scale“ (SUS; Brooke, 1996). Finstad (2010) erstellte basierend auf der SUS und zusätzlichen Items (orientiert an ISO 9241-11) einen 4-Item-Fragebogen, namens UMUX (Usability Metric for User Experience); ein (erreichtes) Ziel war unter anderem eine Korrelation zum SUS von .80 oder höher. Dieses Instrument wurde durch Lewis et al. (2013) im Rahmen zweier Studien auf lediglich zwei zentrale Items verkürzt und so der UMUX-Lite geschaffen.

Gütekriterien

Reliabilität: Der UMUX-Lite ist hinreichend reliabel (in zwei Studien der Autoren lag die innere Konsistenz im Bereich $.82 \leq \alpha \leq .83$; siehe Lewis et al., 2013).

Validität: Lewis et al. (2013) berichten konkurrente Validitäten von $r = .81$ mit dem vollständigen SUS sowie hohe Korrelationen zur Weiterempfehlungsbereitschaft ($.73 \leq r \leq .74$). Es findet sich allerdings ein leichter Mittelwertsunterschied zum vollständigen SUS, weshalb die Autoren eine Korrekturformel zur Auswertung des UMUX-Lite vorschlagen (siehe unten).

Anmerkungen / weitere Informationen

Die deutsche Version der Items wurde den Unterlagen des CPUX-UT Curriculums entnommen (siehe <http://uxqb.org/de/zertifizierung/aufbaustufen/usability-testing-and-evaluation-cpux-ut-de/>).

Items

- Die Fähigkeiten dieses Systems entsprechen meinen Anforderungen.
- Dieses System ist einfach zu benutzen.

Die Formulierungen der Originalitems in Englisch lauten:

- *This system's capabilities meet my requirements.*
- *This system is easy to use*

Hinweis: Bei Bedarf kann in den Items „System“ durch „Website“ ersetzt werden, um eine höhere Verständlichkeit des Fragebogens zu erreichen.

Format Antwortanker

Die Items werden mit einer 7-stufigen Likert-Skala (kodiert von 1 „stimme gar nicht zu“ bis 7 „stimme voll zu“) dargeboten. Die verbalen Itemanker sind:

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *neutral*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

Auswertungsanweisung

Der UMUX-Lite wird nach Angabe der Autoren auf einen Punktwert zwischen 0 und 100 skaliert. Dazu wird folgende Formel genutzt: $(\text{Item1} - 1 + \text{Item2} - 1) \times 100/12$. Weiterhin empfehlen die Autoren zur Herstellung der Vergleichbarkeit mit dem vollständigen SUS die folgende Gewichtungformel: $\text{UMUX-Lite} = .65 \times (\text{UMUX-Lite Punktwert}) + 22,9$.

Zur grundsätzlichen Vergleichbarkeit innerhalb der Toolbox könnte auch ein einfacher Mittelwert in Form von $(\text{Item1} + \text{Item 2}) / 2$ berechnet werden (Vergleichbarkeit zum SUS wäre damit allerdings nicht mehr gegeben).

Benchmarks / Vergleichswerte

Spezifische Benchmarks im Gesundheitsbereich liegen derzeit nicht vor. Auf Basis der hohen Korrelation zum SUS können möglicherweise dessen Interpretationshinweise (siehe Seite 25) als erste grobe Orientierung dienen.

Aufnahme in die Toolbox

In der Toolbox seit Version 1.0 (30. November 2016).

3.2 Perceived Website Usability – German (PWU-G)

Autoren

Carlos Flavián, Miguel Guinalú & Raquel Gurrea

Zitation

Flavián, C., Guinalú, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43(1), 1–14.

Kurzbeschreibung

Flavián et al. (2006) entwickelten eine Usability-Skala, die sich an der ISO 9241 sowie an einem etablierten kommerziellen Usability-Fragebogen (dem WAMMI, siehe Kirakowski & Cierlik, 1998) orientierte. Im Gegensatz zum UMUX-Lite erfolgte die Konstruktion mit Blick auf eine spezifische Bewertung von Websites. Zusammen mit anderen Skalen prüften Flavián et al. (2006) diese „Scale for Measuring Perceived Website Usability“ intensiv und fanden dabei eine hohe interne Konsistenz ($\alpha = .90$) sowie faktorielle Eindimensionalität. Im Rahmen der deutschen Übersetzung wurde die Skala um eine Frage verkürzt sowie weiteren empirischen Prüfungen unterzogen (siehe Thielsch, 2008; Thielsch, Engel & Hirschfeld, 2015).

Gütekriterien

Reliabilität: Die deutsche Version PWU-G zeigt eine sehr hohe interne Konsistenz ($\alpha = .95$; siehe Thielsch, 2008).

Validität: Auch die deutsche Version präsentiert sich als eindeutig eindimensional und weist hohe diskriminante, experimentelle und konvergente Validität auf (Paßlick, 2015; Thielsch, 2008). Insbesondere finden sich hohe Korrelationen zur SUS ($r = .87$) und zur pragmatischen Qualität, erfasst mit dem AttrakDiff mini ($r = .83$); weiterhin erweist sich die Skala als robust gegen Effekte von Stimmung der Befragten oder genutztem Browser (Paßlick, 2015). Es zeigen sich jedoch auch substanzielle Korrelationen zu allgemeinen Bewertungen von Inhalt, Ästhetik und hedonistischer Qualität ($.34 \leq r \leq .60$; Paßlick, 2015).

Anmerkungen / weitere Informationen

Eine umfassende Prüfung und Dokumentation der Skala ist ab 2018 geplant.

Items

- Ich finde, die Bedienung der Website ist leicht zu verstehen.
- Die Website ist einfach zu benutzen, sogar wenn sie zum ersten Mal besucht wird.
- Es fällt mir leicht, die gesuchten Informationen zu finden.
- Ich kann die Struktur der Website leicht nachvollziehen.
- Es ist einfach, sich auf der Website zurechtzufinden.
- Die Inhalte sind so organisiert, dass ich jederzeit weiß, wo ich mich befinde.
- Ich kann gesuchte Informationen schnell erreichen.

Format Antwortanker

Die Items werden mit einer 7-stufigen Likert-Skala (kodiert von 1 „stimme gar nicht zu“ bis 7 „stimme voll zu“) dargeboten. Die verbalen Itemanker sind:

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *neutral*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

Auswertungsanweisung

Es wird ein Mittelwert gebildet: Die einzelnen Angaben auf den Items werden dazu summiert und anschließend durch sieben dividiert. Dieser Mittelwert repräsentiert die subjektive Einschätzung eines allgemeinen Faktors subjektiver Usability.

Benchmarks / Vergleichswerte / Interpretation

Spezifische Benchmarks im Gesundheitsbereich liegen nicht vor. Aus früheren Studien mit der Skala lassen sich aber folgende Vergleichswerte verwandter Website-Kategorien für den PWU-G Mittelwert berechnen:

Website-Kategorie	Erläuterung Kategorie	<i>M</i>	<i>SD</i>	<i>m</i>	<i>n</i>
Information	Websites mit starkem Informationscharakter (auch Blogs/Wikis bei passiver Nutzung)	4,73	1,43	14	1080
Portale	Websites, die eine Übersicht über viele verschiedene Themen geben und hierzu jeweils Informationen und weiterführende Links/Services bieten.	3,63	1,48	9	409

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung,
m = Anzahl der beurteilten Websites in der Kategorie, *n* = Anzahl der Beurteiler.

Das heißt, PWU-G-Gesamtwerte um 4,7 entsprechen einer generell durchschnittlichen Usability-Bewertung von Informationswebsites (3,6 für Portale); Werte unter 3,3 wären eindeutig unterdurchschnittlich (2,2 für Portale); Werte über 6,2 hingegen eindeutig überdurchschnittlich (5,1 für Portale). Zu beachten ist hierbei, dass die Befragten weitgehend unbekannte Websites bewertet und diese ohne ein persönliches Vorinteresse besucht haben.

Die subjektive Usability-Bewertung beeinflusst die Zufriedenheit der User, die Wahrnehmung einer Organisation oder das Vertrauen in die Website (bspw. Cober et al., 2003; Flavián et al., 2006; Lee & Koubek, 2012) und hat grundsätzlich einen Einfluss auf den Erst- und Gesamteindruck einer Website (Thielsch et al., 2014).

Aufnahme in die Toolbox

In der Toolbox seit Version 1.0 (30. November 2016).

Erweiterte Erfassung von Usability

Im internationalen Forschungskontext wird zur Usability-Evaluation oftmals die SUS verwendet. Ist eine entsprechende Publikation angestrebt oder soll Vergleichbarkeit mit internationalen Veröffentlichungen hergestellt werden, so könnte eine Erfassung der Usability mit Hilfe der SUS (anstelle der UMUX-Lite) angeraten sein.

3.3 System Usability Scale (SUS)

Autor

John Brooke

Zitation

Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4–7. <http://doi.org/10.1002/hbm.20701>

Kurzbeschreibung

Usability ist ein zentrales Merkmal interaktiver Produkte wie Software oder Websites. In der DIN EN ISO 9241-11 (1998) ist Usability definiert als Effektivität, Effizienz und Zufriedenheit, mit der User mit einem System vorgegebene Ziele erreichen können. Ein weit verbreitetes Instrument zur Erfassung von Software Usability, das auch zur Website-Evaluation eingesetzt wird, ist die „System Usability Scale“ (SUS; Brooke, 1996). Die SUS wurde Mitte der 1980er von John Brooke geschaffen und erfreut sich seitdem einer enormen Beliebtheit (vgl. Brooke, 2013); derzeit ist das Instrument in Google Scholar rund 5000-mal zitiert (siehe <https://scholar.google.de/scholar?q=SUS-A+quick+and+dirty+usability+scale>).

Erkenntnisse über Gütekriterien für die deutsche Version liegen derzeit leider nicht vor, daher werden nachfolgend Befunde zum englischsprachigen Original berichtet.

Gütekriterien

Reliabilität: Es werden je nach Studie gute bis sehr gute Reliabilitätswerte für die englischsprachige SUS im Sinne der internen Konsistenz berichtet (je Studie $.85 \leq \alpha \leq .91$; vgl. Bangor, Kortum & Miller, 2008; Lewis, Utesch & Maher, 2015). Tullis und Stetson (2004) zeigen zudem auf, dass die SUS anderen typischen Usability-Fragebögen in einer Website-Evaluation überlegen ist und die verlässlichsten Ergebnisse in verschiedenen kleinen Teilstichproben liefert.

Validität: Die Prüfung der englischsprachigen SUS erfolgte in verschiedenen Studien, dabei finden sich vor allem hohe konvergente Validitäten zu anderen typischen Messinstrumenten im Bereich Usability (für eine Übersicht siehe Brooke, 2013; neuere Analysen bei Lewis et al., 2015). Bangor et al. (2008) zeigen weiterhin auf, dass die SUS sensitiv für Veränderungen eines Produkts ist. Allerdings ist die faktorielle Struktur der SUS unklar, bei manchen Autoren erweist sich die Skala als einfaktoriell (z.B. Bangor et al., 2008), bei anderen als zweifaktoriell (z.B. Borsci, Federici & Lauriola, 2009; Sauro & Lewis, 2011) oder gar mehrfaktoriell bzw. uneindeutig (Lewis et al., 2015). An dieser Stelle sind mehr Forschung sowie eine Prüfung der deutschen Version dringend notwendig.

Anmerkungen / weitere Informationen

Es sind verschiedene deutsche Übersetzungen verfügbar. Verschiedene davon (siehe bspw. <https://experience.sap.com/skillup/system-usability-scale-jetzt-auch-auf-deutsch/> oder <http://minds.coremedia.com/2013/09/18/sus-scale-an-improved-german-translation-questionnaire/>) basieren auf einer Crowdsourcing-basierten Übersetzung von Wolfgang Reinhardt (Universität Paderborn). Die nachfolgend dargestellten deutschen Items entstammen einer angepassten Version, dargestellt in der SAP User Experience Community (siehe [https://experience.sap.com/files/System Usability Scale A4 DE.doc](https://experience.sap.com/files/System_Usability_Scale_A4_DE.doc)).

Im englischsprachigen Raum wird zudem eine SUS-Version mit komplett positiven Aussagen diskutiert (siehe Sauro & Lewis, 2011).

Items

- Ich denke, dass ich das System gerne häufig benutzen würde.
- Ich fand das System unnötig komplex. (r)
- Ich fand das System einfach zu benutzen.
- Ich glaube, ich würde die Hilfe einer technisch versierten Person benötigen, um das System benutzen zu können. (r)
- Ich fand, die verschiedenen Funktionen in diesem System waren gut integriert.
- Ich denke, das System enthielt zu viele Inkonsistenzen. (r)
- Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit diesem System sehr schnell lernen.
- Ich fand das System sehr umständlich zu nutzen. (r)
- Ich fühlte mich bei der Benutzung des Systems sehr sicher.
- Ich musste eine Menge lernen, bevor ich anfangen konnte das System zu verwenden. (r)

Hinweis: Bei Bedarf kann in den Items „System“ durch „Website“ ersetzt werden, um eine höhere Verständlichkeit des Fragebogens zu erreichen (vgl. Brooke, 2013).

In der Darstellung der Items im Fragebogen wird (r) weggelassen, dieses zeigt lediglich zu rekodierende Items an (siehe unten, Auswertungsanweisung).

Die Formulierungen der Originalitems in Englisch lauten:

- *I think that I would like to use this system frequently.*
- *I found the system unnecessarily complex.*
- *I thought the system was easy to use.*
- *I think that I would need the support of a technical person to be able to use this system.*
- *I found the various functions in this system were well integrated.*
- *I thought there was too much inconsistency in this system.*
- *I would imagine that most people would learn to use this system very quickly.*
- *I found the system very cumbersome to use.*
- *I felt very confident using the system.*
- *I needed to learn a lot of things before I could get going with this system.*

Format Antwortanker

Die Items werden in der Regel mit einer 5-stufigen Likert-Skala dargeboten (die Extrempunkte sind dabei kodiert mit 1 „stimme überhaupt nicht zu“ und 5 „stimme voll zu“). Finstad (2010) zeigt, dass auch die Verwendung einer 7-stufigen Likert-Skala für SUS möglich und sinnvoll ist.

Auswertungsanweisung

Der SUS-Gesamtwert wird auf einen Wert zwischen 0 und 100 skaliert. Dazu wird zuerst von den Antwortwerten der Items 1, 3, 5, 7 und 9 jeweils der Wert „1“ abgezogen (die Itemantworten reichen dann von 0 bis 4). Die Antwortwerte der Items 2, 4, 6, 8 und 10 (oben markiert mit „r“) werden hingegen vom Wert „5“ subtrahiert; die Items werden so rekodiert. Danach werden die umgerechneten Werte aller 10 Items addiert und mit 2,5 multipliziert; somit ergibt sich ein Summenwert zwischen 0 und 100 (für ein Auswertungsbeispiel siehe Brooke, 1996).

Benchmarks / Vergleichswerte / Interpretation

Für die deutsche Übersetzung der SUS finden sich keine publizierten Benchmarks. Im englischsprachigen Raum liegen eine Reihe von Analysen und Vergleichswerte zur SUS vor, daher sollen im Folgenden zentrale Kennwerte aus diesen genannt werden: Bangor et al. (2008) zeigen in ihrer Studie, dass Werte unterhalb von 50 als nicht akzeptabel gelten, Werte im Bereich 80 bis 90 sind gut, Werte von 90 bis 100 als sehr gut zu bewerten. Sauro (2011) ist hier etwas weniger kritisch und stellt dar, dass ein typischer Durchschnittswert für die SUS im Bereich 70 liegt und bereits Werte ab 85 als sehr gut gelten können (vgl. auch Darstellung in Brooke, 2013 bzw. detaillierte Darstellung in Tabelle 2 bei Lewis et al., 2015).

Bangor et al. (2008) berichten eine Korrelation zwischen der SUS und dem Alter von $r = -.20$ ($p = .003$); ein steigendes Alter korrespondiert damit mit leicht schlechteren SUS-Werten. Geschlechtseffekte finden sich nicht (siehe Bangor et al., 2008).

Aufnahme in die Toolbox

In der Toolbox seit Version 2.0 (30. Juli 2017).

4. Visuelle Ästhetik

4.1 Visual Aesthetics of Websites Inventory – Short (VisAWI-S)

Autoren

Morten Moshagen & Meinald T. Thielsch

Zitation

Moshagen, M. & Thielsch, M. T. (2013). A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology*, 32(12), 1305-1311. <http://dx.doi.org/10.1080/0144929X.2012.694910>

Kurzbeschreibung

In der Forschung zu User Experience von Websites hat sich visuelle Ästhetik als bedeutender Faktor gezeigt. Website-Ästhetik kann als unmittelbare, angenehme und subjektive Wahrnehmung des Webobjekts definiert werden, die wenig durch schlussfolgernde Prozesse beeinflusst wird (Moshagen & Thielsch, 2010). Der VisAWI ist ein Fragebogen zur Erfassung der Wahrnehmung visueller Ästhetik. Dieser basiert auf dem Modell, dass Nutzer einen übergeordneten Generalfaktor Ästhetik wahrnehmen, der aus vier zugrundeliegenden Facetten besteht: Einfachheit, Vielfalt, Farbigkeit und Kunstfertigkeit. Der VisAWI-S ist die Kurzversion des VisAWI und erfasst den Generalfaktor der ästhetischen Wahrnehmung.⁶

Gütekriterien

Reliabilität: Der VisAWI-S ist für Auswertungen auf Gruppenebene ausreichend reliabel (je Studie $.76 \leq \alpha \leq .81$; siehe Thielsch & Moshagen, 2014) und zeigt eine hohe Korrelation zur Langversion ($r = .91$, siehe Moshagen & Thielsch, 2013).

Validität: Es wurden verschiedene Validierungsstrategien angewendet (vgl. Moshagen & Thielsch, 2013): Eine konfirmatorische Faktorenanalyse zeigte eine hervorragende Modellpassung, ebenso findet sich eine hohe konvergente, divergente und konkurrente Validität. Es zeigt sich insbesondere eine hohe Korrelation ($r = .72$) zur Attraktivitätsskala aus dem AttrakDiff 1 (Hassenzahl et al., 2003) und eine deutliche konkurrente Validität in Form einer hohen Korrelation des VisAWI-S mit der angegebenen Wiederbesuchsbereitschaft ($r = .52$).

Anmerkungen / weitere Informationen

Weitere Informationen sind im Internet auf www.VisAWI.de verfügbar – dort steht auch ein deutschsprachiges Manual zum Download zur Verfügung:

Thielsch, M. T. & Moshagen, M. (2014). Manual zum VisAWI (Visual Aesthetics of Websites Inventory) und der Kurzversion VisAWI-S (Short Visual Aesthetics of Websites Inventory). <http://dx.doi.org/10.13140/2.1.1926.5284>

⁶ Für eine Untersuchung spezifischer Facetten der Ästhetik ist der VisAWI-S nicht geeignet, hier empfiehlt sich der Einsatz des vollständigen VisAWI.

Items

- Auf der Seite passt alles zusammen.
- Das Layout ist angenehm vielseitig.
- Die farbliche Gesamtgestaltung wirkt attraktiv.
- Das Layout ist professionell.

Format Antwortanker

Die Items werden mit einer 7-stufigen Likert-Skala (kodiert von 1 „stimme gar nicht zu“ bis 7 „stimme voll zu“) dargeboten. Die verbalen Itemanker sind:

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *neutral*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

Auswertungsanweisung

Zu den Angaben im VisAWI-S wird ein Mittelwert gebildet: Die einzelnen Itemwerte werden dazu summiert und anschließend durch vier dividiert. Dieser Mittelwert repräsentiert die subjektive Einschätzung eines generellen Ästhetik-Faktors.

Benchmarks / Vergleichswerte / Interpretation

Spezifische Benchmarks im Gesundheitsbereich liegen nicht vor. Hirschfeld und Thielsch (2015) nennen für den VisAWI-S-Mittelwert einen allgemeinen optimalen Schwellenwert von 4,5. Das heißt, wird auf der 7-stufigen Bewertungsskala des VisAWI-S dieser Wert überschritten, bewerten die User die Website grundsätzlich und unabhängig vom spezifischen Website-Typ als attraktiv. Ziel einer Website ist damit in Hinblick auf den VisAWI-S mindestens einen Wert von 4,5 oder höher zu erreichen.

Aus dem VisAWI-Manual (Thielsch & Moshagen, 2014) lassen sich zudem für den VisAWI-S-Mittelwert spezifische Benchmarks in verwandten Website-Kategorien entnehmen:

Website-Kategorie	Erläuterung Kategorie	<i>M</i>	<i>SD</i>	<i>m</i>	<i>n</i>
Information	Websites mit starkem Informationscharakter (auch Blogs/Wikis bei passiver Nutzung)	4,60	1,36	35	2005
Portale	Websites, die eine Übersicht über viele verschiedene Themen geben und hierzu jeweils Informationen und weiterführende Links/Services bieten.	4,17	1,36	35	515
Blogs und Social Sharing	Websites, die u.a. der kollaborativen Textbearbeitung, der direkten Vernetzung und Interaktion von Usern dienen oder das Teilen von Ressourcen ermöglichen.	4,54	1,31	30	381

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung, *m* = Anzahl der beurteilten Websites in der Kategorie, *n* = Anzahl der Beurteiler.

Das heißt, VisAWI-S-Gesamtwerte um 4,6 entsprechen einer generell durchschnittlichen Schönheitsbewertung von Informationswebsites (4,2 für Portale / 4,5 für Blogs), Werte unter 3,2 wären eindeutig unterdurchschnittlich (2,8 für Portale / 3,2 für Blogs); hingegen Werte über 6,0 hingegen eindeutig überdurchschnittlich (5,5 für Portale / 5,9 für Blogs). Zu beachten ist hier, dass die Befragten weitgehend unbekannte Websites bewertet und diese ohne ein persönliches Vorinteresse besucht haben.

Die subjektive Ästhetikwahrnehmung beeinflusst viele Aspekte wie beispielsweise Zufriedenheit und Vertrauen der User oder die Wahrnehmung von Unternehmensattraktivität (ein Überblick findet sich bei Moshagen & Thielsch, 2010). Sie hat insbesondere einen sehr starken Einfluss auf den Ersteindruck sowie die Gesamtbewertung einer Website; Ästhetik zeigt zudem kleinere Effekte auf die Wiederbesuchs- und Weiterempfehlungsbereitschaft der User (vgl. Thielsch et al., 2014).

Aufnahme in die Toolbox

In der Toolbox seit Version 1.0 (30. November 2016).

Erweiterte Erfassung visueller Ästhetik

Falls eine detaillierte Analyse der visuellen Ästhetik einer Website erfolgen soll, ist die Verwendung des vollständigen VisAWI anstelle der Kurzversion VisAWI-S denkbar. In diesem Falle sind dann nicht nur Aussagen über einen generellen Ästhetikfaktor möglich, sondern zusätzlich auch auf Ebene der vier Facetten Einfachheit, Vielseitigkeit, Farbigkeit und Kunstfertigkeit.

4.2 Visual Aesthetics of Websites Inventory (VisAWI)

Autoren

Morten Moshagen & Meinald T. Thielsch

Zitation

Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68(10), 689-709. <http://dx.doi.org/10.1016/j.ijhcs.2010.05.006>

Kurzbeschreibung

In der Forschung zu User Experience von Websites hat sich visuelle Ästhetik als bedeutender Faktor gezeigt. Website-Ästhetik kann als unmittelbare, angenehme und subjektive Wahrnehmung des Webobjekts definiert werden, die wenig durch schlussfolgernde Prozesse beeinflusst wird (Moshagen & Thielsch, 2010). Der VisAWI ist ein Fragebogen zur Erfassung der Wahrnehmung visueller Ästhetik. Dieser basiert auf dem Modell, dass Nutzer einen übergeordneten Generalfaktor Ästhetik wahrnehmen, der aus vier zugrundeliegenden Facetten besteht: Einfachheit, Vielfalt, Farbigkeit und Kunstfertigkeit.

Gütekriterien

Reliabilität: Der VisAWI zeigt gute Reliabilitätswerte (je Skala $.85 \leq \alpha \leq .89$) und eine sehr gute interne Konsistenz für den Gesamtwert (Cronbachs $\alpha = .94$; siehe Moshagen & Thielsch, 2010).

Validität: Es wurden verschiedene Validierungsstrategien angewendet (vgl. Moshagen & Thielsch, 2010): Die Faktorstruktur wurde mittels explorativen und konfirmatorischen Faktorenanalysen an unabhängigen Stichproben bestätigt. Die konvergente, divergente, diskriminative und konkurrente Validität wurde in drei Studien belegt (Moshagen & Thielsch, 2010). Es finden sich insbesondere hohe Korrelationen mit dem Messinstrument von Lavie und Tractinsky (2004) im Bereich von $.52 \leq r \leq .82$ für die Skala zur klassischen Ästhetik und hohe Korrelationen ($.60 \leq r \leq .80$) zur Attraktivitätsskala aus dem AttrakDiff 1 (Hassenzahl et al., 2003). Konkurrente Validität zeigt sich in Form einer hohen Korrelation des VisAWI mit der angegebenen Wiederbesuchsbereitschaft ($r = .51$).

Zudem verdeutlichen zwei experimentelle Validierungen, dass die Facetten des VisAWI einerseits generell responsiv für systematische Variationen des Layouts einer Website sind, aber andererseits spezifisch auf bestimmte Manipulationen reagieren. So beeinflusst bspw. eine Änderung des Farbschemas ausschließlich die Bewertungen auf der Farbigkeitsfacette, nicht aber auf den anderen Facetten (vgl. Moshagen & Thielsch, 2010).

Anmerkungen / weitere Informationen

Weitere Informationen sind im Internet auf www.VisAWI.de verfügbar – dort steht auch ein deutschsprachiges Manual zum Download zur Verfügung:

Thielsch, M. T. & Moshagen, M. (2014). Manual zum VisAWI (Visual Aesthetics of Websites Inventory) und der Kurzversion VisAWI-S (Short Visual Aesthetics of Websites Inventory). <http://dx.doi.org/10.13140/2.1.1926.5284>

Items

Skala 1: Einfachheit

- Das Layout wirkt zu gedrängt. (r)
- Das Layout ist gut zu erfassen.
- Das Layout erscheint angenehm gegliedert.
- Die Seite erscheint zu uneinheitlich. (r)
- Auf der Seite passt alles zusammen.

Skala 2: Vielseitigkeit

- Die Seitengestaltung ist uninteressant. (r)
- Das Layout ist originell.
- Die Gestaltung wirkt einfallslos. (r)
- Das Layout wirkt dynamisch.
- Das Layout ist angenehm vielseitig.

Skala 3: Farbigkeit

- Die farbliche Gesamtgestaltung wirkt attraktiv.
- Die Farben passen nicht zueinander. (r)
- Der Farbeinsatz ist nicht gelungen. (r)
- Die Farben haben eine angenehme Wirkung.

Skala 4: Kunstfertigkeit

- Das Layout ist professionell.
- Das Layout ist nicht zeitgemäß. (r)
- Die Seite erscheint mit Sorgfalt gemacht.
- Das Layout wirkt konzeptlos. (r)

Hinweis: In der Darstellung der Items im Fragebogen wird „(r)“ weggelassen, dieses zeigt lediglich zu rekodierende Items an (siehe unten, Auswertungsanweisung).

Format Antwortanker

Die Items werden mit einer 7-stufigen Likert-Skala (kodiert von 1 „stimme gar nicht zu“ bis 7 „stimme voll zu“) dargeboten. Die verbalen Itemanker sind:

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *neutral*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

Auswertungsanweisung

Die Auswertung beginnt mit der Umkodierung negativ gepolter Items (siehe oben, markiert mit einem „(r)“), indem man deren Wert von 8 abzieht. Ein Beispiel: Wird bei Item 1 „Das Layout wirkt zu gedrängt“ ein Wert von 5 angegeben, wird die 5 in eine $8 - 5 = 3$ rekodiert (analog dazu: $1 \leftrightarrow 7$; $2 \leftrightarrow 6$; $3 \leftrightarrow 5$; $4 \leftrightarrow 4$).

Im Anschluss werden die einzelnen Skalenwerte von jeder Subskala addiert, und anschließend die Summe durch die Anzahl der Items der jeweiligen Subskala dividiert (Beispiel: bei Subskala „Einfachheit“ ist der Divisor 5). Die Skalenmittelwerte zeigen mit höheren Werten eine stärkere Ausprägung der jeweiligen Skala an.

Den Gesamtmittelwert des Fragebogens erhält man durch die Bildung der Summe der vier Skalenwerte mit anschließender Division durch vier.

Benchmarks / Vergleichswerte / Interpretation

Spezifische Benchmarks im Gesundheitsbereich liegen nicht vor. Hirschfeld und Thielsch (2015) nennen für den VisAWI-Gesamtwert einen allgemeinen optimalen Schwellenwert von 4,5. Das heißt, wird auf der 7-stufigen Bewertungsskala des VisAWI dieser Wert überschritten, bewerten die User die Website grundsätzlich und unabhängig vom spezifischen Website-Typ als attraktiv. Ziel einer Website ist damit in Hinblick auf den VisAWI, mindestens einen Wert von 4,5 oder höher zu erreichen.

Aus dem VisAWI-Manual (Thielsch & Moshagen, 2014) lassen sich zudem für den VisAWI-Gesamtwert sowie die Skalenmittelwerte die folgenden spezifischen Benchmarks in verwandten Website-Kategorien entnehmen:

Website-Kategorie	Gesamtwert		Einfachheit		Vielseitigkeit		Farbigkeit		Kunstfertigkeit	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Information	4,08	1,26	4,33	1,41	3,35	1,35	4,28	1,61	4,34	1,47
Portale	4,72	1,20	4,80	1,37	4,36	1,35	4,76	1,29	4,95	1,28
Blogs und Social Sharing	3,81	1,23	3,64	1,44	3,43	1,31	4,14	1,50	4,01	1,42

Das heißt beispielsweise, VisAWI-Werte auf der Skala Farbigkeit um 4,3 entsprechen einer generell durchschnittlichen Schönheitsbewertung von Informationswebsites, Werte unter 2,7 wären eindeutig unterdurchschnittlich, Werte über 5,9 hingegen eindeutig überdurchschnittlich. Hierbei ist es wesentlich, den subjektiven Charakter der Bewertungen zu berücksichtigen. So bedeutet ein hoher Wert auf der Skala Farbigkeit nicht, dass das Design besonders bunt ist, sondern, dass die realisierte Farbgestaltung von den NutzerInnen positiv bewertet wird. Die Interpretation erfolgt entsprechend analog je Skala und für den Gesamtwert.

Zu beachten ist bei diesen Vergleichswerten, dass die Befragten in den jeweiligen Studien weitgehend unbekannte Websites bewertet und diese ohne ein persönliches Vorinteresse besucht haben.

Die subjektive Ästhetikwahrnehmung beeinflusst viele Aspekte wie beispielsweise Zufriedenheit und Vertrauen der User oder die Wahrnehmung von Unternehmensattraktivität (ein Überblick findet sich bei Moshagen & Thielsch, 2010). Sie hat insbesondere einen sehr starken Einfluss auf den Ersteindruck sowie die Gesamtbewertung einer Website; Ästhetik zeigt zudem kleinere Effekte auf die Wiederbesuchs- und Weiterempfehlungsbereitschaft der User (vgl. Thielsch et al., 2014).

Aufnahme in die Toolbox

In der Toolbox seit Version 2.0 (30. Juli 2017).

5. Emotionale Reaktion: Befindlichkeit, Zustimmung, Zufriedenheit

5.1 Smiley-Skala

(im Original als „Smily-Skala“ bezeichnet)

Autorin

Ruth Jäger

Zitation

Jäger, R. (2004). Konstruktion einer Ratingskala mit Smileys als symbolische Marken. *Diagnostica*, 50(1), 31–38. <http://doi.org/10.1026/0012-1924.50.1.31>.

Kurzbeschreibung

In der Website-Evaluation werden zur Erfassung der Stimmung der User verschiedene Skalen eingesetzt. Oftmals kommen hier Instrumente wie der PANAS (Positive and Negative Affect Schedule; vgl. Krohne et al. 1996) oder SAM (Self-Assessment Manikin; Bradley & Lang, 1994) zum Einsatz. Jäger (2004) schlug eine einfache fünfstufige Smiley-Skala vor und validierte diese. Die Smileys können laut der Autorin in Selbsteinschätzungsverfahren eingesetzt werden, um Befindlichkeit zu messen. Jäger (2004, S. 37) gibt hier folgende konkrete Empfehlungen für Anwendungsfelder der Skala:

- Quantifizierung der Intensität erlebter Zufriedenheit,
- Emotionale Bewertung von Ereignissen oder Objekten,
- Zustimmung oder Ablehnung von Statements und
- Beschreibung der globalen Befindlichkeit von Personen.

Gütekriterien

Reliabilität: Bei einem Einzelitem ist keine Berechnung von internen Konsistenzmaßen möglich; zur Retest-Reliabilität liegen keine Angaben vor.

Validität: Die Smiley-Skala korreliert deutlich mit dem PANAS ($r = .50$ bis $r = .67$), zudem wurden Eindimensionalität und Äquidistanz nachgewiesen (Jäger, 2004). Jäger (2004) argumentiert weiterhin, dass fehlende Geschlechts-, Alters- und Bildungseffekte der Skala auf eine gute externe Validität der Ergebnisse hindeuten.

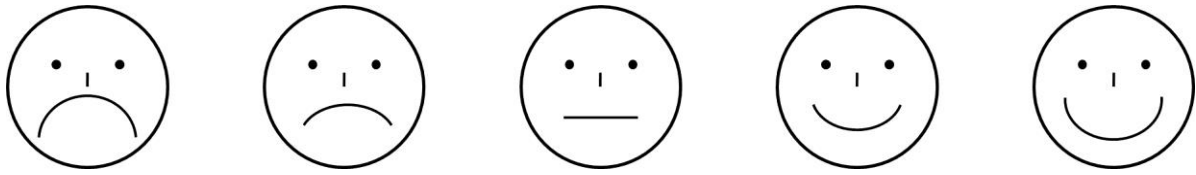
Anmerkungen / weitere Informationen

Die Skala wurde unter dem Namen „Smily-Skala“ als Wort- und Bildmarke beim Deutschen Patentamt eingetragen, dort allerdings am 01.01.2011 wieder gelöscht. Damit endete die Schutzfrist am 31.12.2010.⁷

Grundsätzlich ist die Erfassung emotionaler Reaktionen im Bereich User Experience noch nicht gut erforscht. Der Bedarf an spezifischen Instrumenten ist hoch und erste Entwicklungen sind vielversprechend, bedürfen aber noch weiterer Prüfung (siehe Sonderegger et al., 2016).

⁷ Vgl. <https://register.dpma.de/DPMAREgister/marke/register/300916140/DE> sowie <https://register.dpma.de/DPMAREgister/marke/register/300916159/DE>

Item



Instruktion




➤ Kreuzen Sie als Antwort bitte das entsprechende Gesicht an: Wie fühlen Sie sich im Moment?

Auswertungsanweisung

Die fünf Stufen der Skala können von 1 bis 5 kodiert werden, hohe Werte zeigen hierbei eine positive Ausprägung der Skala an.

Benchmarks / Vergleichswerte / Interpretation

Jäger (2004, S. 35) liefert kontextspezifische Vergleichswerte zur Ausprägung der Smileys mit Werten auf dem PANAS (1 = negative Ausprägung, 7 = positive Ausprägung). Das heißt, der traurige oder eher traurige Smiley geht einher mit entsprechend negativen PANAS-Werten, der neutrale Smiley korrespondiert mit entsprechend neutralen PANAS-Werten, und die beiden positiven Smileys stimmen mit positiven PANAS-Werten auf den in der Tabelle dargestellten Aspekten überein. So hatten beispielsweise Befragte, die den fröhlichsten Smiley ankreuzten, beim PANAS im Mittel einen Wert von 6,80 auf dem semantischen Differenzial *unzufrieden-zufrieden*.

						
PANAS	M	M	M	M	M	PANAS
unzufrieden	1,16	2,19	4,15	6,28	6,80	zufrieden
negativ	1,30	2,44	4,17	6,31	6,63	positiv
ablehnend	1,36	2,34	4,01	6,09	6,45	zustimmend
unfreundlich	1,44	2,70	4,09	6,44	6,49	freundlich
pessimistisch	1,37	2,30	4,04	6,25	6,60	optimistisch
unangenehm	1,68	2,86	4,13	6,23	6,12	angenehm
traurig	1,63	2,22	4,03	6,05	6,59	heiter

Quelle: Tabelle in Auszügen angepasst nach Jäger (2004, S. 35)

Aufnahme in die Toolbox

In der Toolbox seit Version 1.0 (30. November 2016).

6. Handlungs- und Nutzungsintentionen

6.1 Scale assessing the intention to revisit the website

Autoren

Morten Moshagen & Meinald T. Thielsch

Zitation

Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68 (10), 689-709. <http://dx.doi.org/10.1016/j.ijhcs.2010.05.006>

Kurzbeschreibung

Wiederbesuchs- und Weiterempfehlungsbereitschaft der User einer Website sind von hohem praktischen Interesse, denn oft ist eine dauerhafte Bindung der NutzerInnen entscheidend für den Erfolg einer Onlinepräsenz. Daher wird dieser Faktor häufig in Website-Evaluationen erfasst; ein bekanntes Beispiel mit breitem Anwendungshorizont ist der sogenannte *Net Promoter Score* (NPS; Reichheld, 2003). Der NPS besteht aus einem Einzelitem zur Empfehlung auf einer 11-stufigen Skala (die jedoch nur dreistufig ausgewertet wird). Der NPS ist seitens der Forschung hinsichtlich Güte und Einsetzbarkeit stark kritisiert (siehe z.B. Grisaffe, 2007; Keiningham et al., 2007; Sharp, 2008). Daher kann auf Basis der aktuellen Datenlage die allgemeine Verwendung des NPS in der Website-Evaluation nicht empfohlen werden. Andere verfügbare deutschsprachige Skalen im User Experience Bereich sind eher für Produkte geeignet (z.B. die Skala zur Nutzungsintention im meCUE; Minge & Riedel, 2013), daher soll an dieser Stelle eine 4-Item Wiederbesuchsskala aus der Website-Forschung vorgeschlagen werden.

Gütekriterien

Reliabilität: Zieht man die Daten von Moshagen und Thielsch (2010, Studie 4, n = 512) heran, ergibt sich für die Skala ein Cronbachs α von .92, was eine sehr gute interne Konsistenz darstellt. Ergebnisse zu einer verkürzten Form der Skala mit drei Items sind bei Thielsch et al. (2014) dargestellt.

Validität: Hierzu liegen keine publizierten Angaben vor. Zieht man die Daten von Moshagen und Thielsch heran (2010, Studie 4, n = 512), ergibt sich ein hoher Zusammenhang dieser Skala zum Gesamteindruck ($r = .63$).

Items

- Ich werde diese Seite wieder benutzen.
- Ich werde diese Website regelmäßig besuchen.
- Ich würde diese Website Freunden und Bekannten weiterempfehlen.
- Bei zukünftigem Interesse an solchen Themen könnte ich mir vorstellen, wieder diese Website aufzurufen.

Format Antwortanker

Die Items werden mit einer 7-stufigen Likert-Skala (kodiert von 1 „stimme gar nicht zu“ bis 7 „stimme voll zu“) dargeboten. Die verbalen Itemanker sind:

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *neutral*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

Auswertungsanweisung

Zu den Angaben auf den vier Items wird ein Mittelwert gebildet: Die einzelnen Itemwerte werden dazu summiert und anschließend durch vier dividiert.

Benchmarks / Vergleichswerte

Spezifische Benchmarks im Gesundheitsbereich liegen derzeit nicht vor. Grundsätzlich ist bei der gewählten Form der Items mit Likert-Skala zumindest im Durchschnitt ein Wert über dem Skalenmittel erwünscht (das heißt, dass die Befragten den Aussagen zumindest in der Tendenz zustimmen).

Aufnahme in die Toolbox

In der Toolbox seit Version 1.0 (30. November 2016).

7. Gesamteindruck

7.1 Einzelitem Gesamteindruck

Autor

Meinald T. Thielsch

Zitation

Thielsch, M. T. (2008). *Ästhetik von Websites. Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen*. Münster: MV Wissenschaft.

Kurzbeschreibung

Der Gesamteindruck einer Website wird in hohem Maße durch das Erleben der User von Inhalt, Usability und Ästhetik bestimmt (Thielsch et al., 2014). Ein Gesamteindruck im Evaluationsbereich umfasst zudem aber Aspekte, die in den jeweiligen Einzelmessungen nicht zwingend enthalten waren. Damit ist ein solches holistisches Maß durchaus von Wert, mögliche verzerrende Einflüsse irrelevanter Variablen sind aber zu bedenken (bspw. Kontextfaktoren der Befragungssituation, in der sich ein User gerade befindet). Zur Erfassung des Gesamteindrucks wird an dieser Stelle ein Notenitem vorgeschlagen, das in der Forschung wiederholt zum Einsatz gekommen ist. Vergleichbare Gesamtbewertungen finden sich ebenso in anderen User Experience Fragebögen (z.B. im meCUE; Minge & Riedel, 2013).

Gütekriterien

Reliabilität: Bei einem Einzelitem ist keine Berechnung von internen Konsistenzmaßen möglich; zur Retest-Reliabilität liegen keine Angaben vor.

Validität: Thielsch (2008b) berichtet hohe Korrelationen der Gesamtnote mit relevanten Fragebögen zu Website-Inhalt, Verständlichkeit, Usability und Ästhetik ($.59 \leq r \leq .76$).

Item

➤ **Alles in allem: Ich gebe dieser Website die Gesamtnote ...**

Format Antwortanker

Die Frage nach der Gesamtnote ist mit folgenden Ankern versehen:

- 1 (sehr gut)
- 2 (gut)
- 3 (befriedigend)
- 4 (ausreichend)
- 5 (mangelhaft)

Auswertungsanweisung

Es wird ein Mittelwert der Note gebildet. Dabei ist zu beachten, dass niedrige Werte positiv sind.

Benchmarks / Vergleichswerte / Interpretation

Spezifische Benchmarks im Gesundheitsbereich liegen derzeit nicht vor. Aus den kumulierten Ergebnissen früherer eigener Website-Studien lässt sich ableiten, dass die mittlere Gesamtnote für Informationswebsites im Bereich 3 liegt (bei einer SD von ca. 1). Das heißt, hier ist eine Interpretation entsprechend der verbalen Verankerung der Notenskala möglich

und zu empfehlen: Eine durchschnittliche Website erreicht im Mittel ein „befriedigend“, eine Bewertung „ausreichend“ wäre unterdurchschnittlich und ein „mangelhaft“ stark unterdurchschnittlich und keinesfalls akzeptabel. Ein „gut“ als Durchschnittswert ist hingegen als überdurchschnittlich anzusehen, die Gesamtnote „sehr gut“ wäre herausragend positiv. Zu beachten ist bei diesen Vergleichswerten, dass die Befragten weitgehend unbekannte Websites bewertet und diese ohne ein persönliches Vorinteresse besucht haben.

Aufnahme in die Toolbox

In der Toolbox seit Version 1.0 (30. November 2016).

FAQ zur Toolbox

Für welche Zielgruppen ist die Toolbox geeignet?

- Befragungsgruppen ab 14 Jahren
- begleitende und vergleichende Evaluationen mit mind. 20-50 Testpersonen pro Website
- Repräsentativstudien mit entsprechender Zielgruppenzusammensetzung.

Muss ich immer alle Instrumente aus der Toolbox verwenden?

- Nein, es können auch einzelne Fragebögen oder Skalen je nach Fragestellung einer Studie einzeln genutzt werden.

Darf man bei einem Instrument Items weglassen?

- Nein. Die Instrumente wurden in vorliegender Form validiert.

Darf man Items an Zielobjekte anpassen?

- Geringfügige Anpassungen der Itemformulierungen an Zielobjekte (z.B. „Website“ oder „Plattform“ statt „Software“) sind prinzipiell denkbar.

Wie lange braucht eine Testperson pro Frage?

- Erfahrungsgemäß ca. 10 Sekunden. Zusätzlich sollte Zeit für Instruktionen zu einer Befragung und für eventuelle Testaufgaben mit der Zielwebsite eingeplant werden.

Sollte ich den Befragten eine Testaufgabe zur Website geben?

- Abhängig vom jeweiligen Evaluationsdesign in der Regel ja.

Sollte ich demographische Angaben der Befragten erheben?

- Ja. Um die Zusammensetzung der Befragungsstichprobe zu analysieren und mit der Zielgruppe einer Website zu vergleichen ist die Erhebung relevanter demographischer Daten angeraten.

Welche Arten der Fragebogenvorgabe sind denkbar?

- Testungen sind online und offline und sowohl stationär als auch auf mobilen Endgeräten möglich.

Was kann mit der Toolbox getestet werden?

- Es können sowohl fortgeschrittene Prototypen als auch fertige Websites getestet werden. Die hier vorgeschlagenen Instrumente eignen sich besonders für summative oder vergleichende Evaluationen und wiederholte Messungen im Sinne eines Monitorings.

Sollte die Zielwebsite bei der Bewertung sichtbar sein?

- Im Normalfall ja, am besten in einer funktionsfähigen Version.

Kann ich mit der Toolbox andere Dinge als Websites testen?

- Das kommt darauf an. Die meisten Instrumente sind lediglich für Websites, Software oder andere interaktive Systeme erprobt und validiert.

Wie ist bei der Auswertung mit fehlenden Werten umzugehen?

- Im Optimalfall ist eine Online-Befragung so angelegt, dass die Befragten bei der Teilnahme darauf aufmerksam gemacht werden, falls ein Item vergessen wird. Aufgrund der Kürze der Instrumente in der Toolbox können fehlende Werte ansonsten sehr schnell dazu führen, dass eine Skala bei einem Befragten nicht ausgewertet werden kann.

Wird aus den Instrumenten der Toolbox ein Gesamtscore errechnet?

- Nein. Dies wäre auch nicht sinnvoll, da die jeweilig evaluierten Konstrukte je Website oder Evaluationszeitpunkt verschiedene Relevanz haben können. Zum Beispiel wäre die Erfassung des Ersteindrucks bei langjährigen Usern einer Website in der Praxis nicht sinnvoll möglich.

Darf ich die jeweiligen Instrumente wirklich einfach so verwenden?

- Ja, solange Sie sich an die üblichen Zitationsregeln halten und auf die jeweiligen Quellen verweisen.

Wie zitiere ich die jeweiligen Instrumente?

- Siehe Zitationshinweis zu jedem Instrument. Wir empfehlen, alle von Ihnen eingesetzten Instrumente entsprechend einzeln zu zitieren.

Ich möchte dennoch/zusätzlich die Toolbox als Ganzes zitieren.

- Falls Sie auf die Toolbox generell oder Informationen in dieser, jenseits der dargestellten Instrumente, verweisen möchten, können Sie dies gerne tun. Die Zitation lautet:

Thielsch, M. T. (unter Mitarbeit von Salaschek, M.) (2017). *Toolbox zur kontinuierlichen Website-Evaluation und Qualitätssicherung (Version 2.0)*. Arbeitsbericht, Köln:

Bundeszentrale für gesundheitliche Aufklärung (BZgA).

<http://dx.doi.org/10.17623/BZGA:224-2.0>

Gestaltungsbeispiele zu den Instrumenten der Toolbox

Auf den folgenden Seiten finden sich beispielhafte Darstellungen, wie die jeweiligen Fragebögen und Skalen aus der Toolbox in einer Papierversion gestaltet sein könnten.

Die Darstellung auf den folgenden Seiten dient dazu, eine mögliche Darbietung der Instrumente zu visualisieren. Dies soll nicht als direkte Kopiervorlage dienen, sondern AnwenderInnen der Toolbox lediglich die Möglichkeit eines direkten Überblicks über alle möglichen Fragen und deren Formate geben. In einem Online-Fragebogen oder bei einer Papier-Stift-Version sollte eine platzsparende Darstellungsform (ohne Nennung der Instrument-Namen) gewählt werden.

Die Hinweistexte zu den Items dienen zur besseren Instruktion der Befragten, sind aber in den Originalpublikationen zumeist so nicht direkt vorgegeben worden. Änderungen zur Erhöhung der Verständlichkeit einer Befragung sind an diesen Stellen möglich.

Einzelitems Ersteindruck

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme gar nicht zu</i>	<i>stimme nicht zu</i>	<i>stimme eher nicht zu</i>	<i>neutral</i>	<i>stimme eher zu</i>	<i>stimme zu</i>	<i>stimme voll zu</i>
Mein erster Eindruck: Ich finde den Inhalt dieser Website interessant.	①	②	③	④	⑤	⑥	⑦
Mein erster Eindruck: Ich halte diese Website für gut benutzbar.	①	②	③	④	⑤	⑥	⑦
Mein erster Eindruck: Ich finde, diese Website ist ansprechend gestaltet.	①	②	③	④	⑤	⑥	⑦

Mein erster Eindruck: Ich gebe dieser Website die Gesamtnote...

- ① (sehr gut)
- ② (gut)
- ③ (befriedigend)
- ④ (ausreichend)
- ⑤ (mangelhaft)

Message Credibility Scale

Wie gut beschreiben die folgenden Adjektive den Inhalt, den Sie gerade gelesen haben?

	<i>beschreibt sehr schlecht</i>						<i>beschreibt sehr gut</i>
akkurat	①	②	③	④	⑤	⑥	⑦
authentisch	①	②	③	④	⑤	⑥	⑦
glaubhaft	①	②	③	④	⑤	⑥	⑦

Trusting Beliefs

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme überhaupt nicht zu</i>	<i>stimme nicht zu</i>	<i>stimme eher nicht zu</i>	<i>neutral</i>	<i>stimme eher zu</i>	<i>stimme zu</i>	<i>stimme voll zu</i>
Ich glaube, dass [Name der Website] in meinem besten Interesse handeln würde.	①	②	③	④	⑤	⑥	⑦
Wenn ich Hilfe bräuchte, würde [Name der Website] sein Bestes tun, um mir zu helfen.	①	②	③	④	⑤	⑥	⑦
[Name der Website] interessiert sich für mein Wohlergehen, nicht nur für das eigene.	①	②	③	④	⑤	⑥	⑦
[Name der Website] ist aufrichtig im Umgang mit mir.	①	②	③	④	⑤	⑥	⑦
Ich würde [Name der Website] als ehrlich beschreiben.	①	②	③	④	⑤	⑥	⑦
[Name der Website] würde gemachte Zusagen einhalten.	①	②	③	④	⑤	⑥	⑦
[Name der Website] ist aufrichtig und ehrlich.	①	②	③	④	⑤	⑥	⑦
[Name der Website] ist kompetent und effektiv in [Dienstleistung der Website].	①	②	③	④	⑤	⑥	⑦
[Name der Website] erfüllt seine Rolle als [Dienstleistung der Website] sehr gut.	①	②	③	④	⑤	⑥	⑦
Insgesamt ist [Name der Website] eine fähige und tüchtige [Dienstleistungsbereich der Website].	①	②	③	④	⑤	⑥	⑦
Im Allgemeinen ist [Name der Website] sehr sachkundig was [Dienstleistung der Website] betrifft.	①	②	③	④	⑤	⑥	⑦

Fragebogen zur Wahrnehmung von Website-Inhalten (WWI)

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme gar nicht zu</i>	<i>stimme nicht zu</i>	<i>stimme eher nicht zu</i>	<i>neutral</i>	<i>stimme eher zu</i>	<i>stimme zu</i>	<i>stimme voll zu</i>
Die Website weckt mein Interesse.	①	②	③	④	⑤	⑥	⑦
Der Inhalt der Website gefällt mir.	①	②	③	④	⑤	⑥	⑦
Ich lese diese Website gerne.	①	②	③	④	⑤	⑥	⑦
Die einzelnen Sätze sind einfach zu lesen.	①	②	③	④	⑤	⑥	⑦
Die Texte liefern mir kurz und bündig die wichtigsten Informationen.	①	②	③	④	⑤	⑥	⑦
Der Sprachgebrauch in den Texten ist geläufig und allgemein verständlich.	①	②	③	④	⑤	⑥	⑦
Die Informationen sind qualitativ hochwertig.	①	②	③	④	⑤	⑥	⑦
Ich finde die Informationen auf der Website nützlich.	①	②	③	④	⑤	⑥	⑦
Die Inhalte der Website erscheinen mir so wichtig, dass ich sie mir ausdrucken oder speichern würde.	①	②	③	④	⑤	⑥	⑦

Usability Metric for User Experience – Lite (UMUX-Lite)

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme gar nicht zu</i>	<i>stimme nicht zu</i>	<i>stimme eher nicht zu</i>	<i>neutral</i>	<i>stimme eher zu</i>	<i>stimme zu</i>	<i>stimme voll zu</i>
Die Fähigkeiten dieses Systems entsprechen meinen Anforderungen	①	②	③	④	⑤	⑥	⑦
Dieses System ist einfach zu benutzen.	①	②	③	④	⑤	⑥	⑦

Perceived Website Usability – German (PWU-G)

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme gar nicht zu</i>	<i>stimme nicht zu</i>	<i>stimme eher nicht zu</i>	<i>neutral</i>	<i>stimme eher zu</i>	<i>stimme zu</i>	<i>stimme voll zu</i>
Ich finde, die Bedienung der Website ist leicht zu verstehen.	①	②	③	④	⑤	⑥	⑦
Die Website ist einfach zu benutzen, sogar wenn sie zum ersten Mal besucht wird.	①	②	③	④	⑤	⑥	⑦
Es fällt mir leicht, die gesuchten Informationen zu finden.	①	②	③	④	⑤	⑥	⑦
Ich kann die Struktur der Website leicht nachvollziehen.	①	②	③	④	⑤	⑥	⑦
Es ist einfach, sich auf der Website zurechtzufinden.	①	②	③	④	⑤	⑥	⑦
Die Inhalte sind so organisiert, dass ich jederzeit weiß, wo ich mich befinde.	①	②	③	④	⑤	⑥	⑦
Ich kann gesuchte Informationen schnell erreichen.	①	②	③	④	⑤	⑥	⑦

System Usability Scale (SUS)

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme überhaupt nicht zu</i>				<i>stimme voll zu</i>
Ich denke, dass ich das System gerne häufig benutzen würde.	①	②	③	④	⑤
Ich fand das System unnötig komplex.	①	②	③	④	⑤
Ich fand das System einfach zu benutzen.	①	②	③	④	⑤
Ich glaube, ich würde die Hilfe einer technisch versierten Person benötigen, um das System benutzen zu können.	①	②	③	④	⑤
Ich fand, die verschiedenen Funktionen in diesem System waren gut integriert.	①	②	③	④	⑤
Ich denke, das System enthielt zu viele Inkonsistenzen.	①	②	③	④	⑤
Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit diesem System sehr schnell lernen.	①	②	③	④	⑤
Ich fand das System sehr umständlich zu nutzen.	①	②	③	④	⑤
Ich fühlte mich bei der Benutzung des Systems sehr sicher.	①	②	③	④	⑤
Ich musste eine Menge lernen, bevor ich anfangen konnte das System zu verwenden.	①	②	③	④	⑤

Visual Aesthetics of Websites Inventory – Short (VisAWI-S)

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme gar nicht zu</i>	<i>stimme nicht zu</i>	<i>stimme eher nicht zu</i>	<i>neutral</i>	<i>stimme eher zu</i>	<i>stimme zu</i>	<i>stimme voll zu</i>
Auf der Seite passt alles zusammen.	①	②	③	④	⑤	⑥	⑦
Das Layout ist angenehm vielseitig.	①	②	③	④	⑤	⑥	⑦
Die farbliche Gesamtgestaltung wirkt attraktiv.	①	②	③	④	⑤	⑥	⑦
Das Layout ist professionell.	①	②	③	④	⑤	⑥	⑦

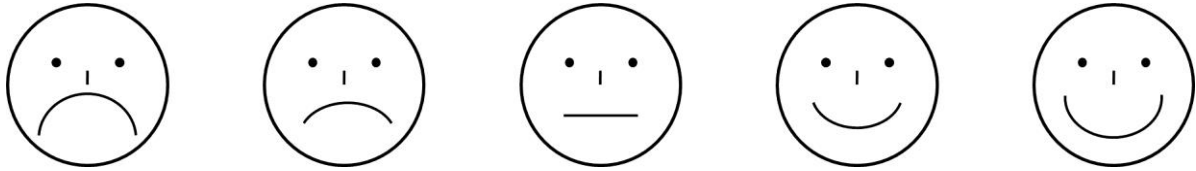
Visual Aesthetics of Websites Inventory (VisAWI)

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme gar nicht zu</i>	<i>stimme nicht zu</i>	<i>stimme eher nicht zu</i>	<i>neutral</i>	<i>stimme eher zu</i>	<i>stimme zu</i>	<i>stimme voll zu</i>
Das Layout wirkt zu gedrängt.	①	②	③	④	⑤	⑥	⑦
Das Layout ist gut zu erfassen.	①	②	③	④	⑤	⑥	⑦
Das Layout erscheint angenehm gegliedert.	①	②	③	④	⑤	⑥	⑦
Die Seite erscheint zu uneinheitlich.	①	②	③	④	⑤	⑥	⑦
Auf der Seite passt alles zusammen.	①	②	③	④	⑤	⑥	⑦
Die Seitengestaltung ist uninteressant.	①	②	③	④	⑤	⑥	⑦
Das Layout ist originell.	①	②	③	④	⑤	⑥	⑦
Die Gestaltung wirkt einfallslos.	①	②	③	④	⑤	⑥	⑦
Das Layout wirkt dynamisch.	①	②	③	④	⑤	⑥	⑦
Das Layout ist angenehm vielseitig.	①	②	③	④	⑤	⑥	⑦
Die farbliche Gesamtgestaltung wirkt attraktiv.	①	②	③	④	⑤	⑥	⑦
Die Farben passen nicht zueinander.	①	②	③	④	⑤	⑥	⑦
Der Farbeinsatz ist nicht gelungen.	①	②	③	④	⑤	⑥	⑦
Die Farben haben eine angenehme Wirkung.	①	②	③	④	⑤	⑥	⑦
Das Layout ist professionell.	①	②	③	④	⑤	⑥	⑦
Das Layout ist nicht zeitgemäß.	①	②	③	④	⑤	⑥	⑦
Die Seite erscheint mit Sorgfalt gemacht.	①	②	③	④	⑤	⑥	⑦
Das Layout wirkt konzeptlos.	①	②	③	④	⑤	⑥	⑦

Smiley-Skala

Kreuzen Sie als Antwort bitte das entsprechende Gesicht an: Wie fühlen Sie sich im Moment?



Scale assessing the intention to revisit the website

Bitte geben Sie an, wie sehr Sie den nachfolgenden Aussagen zustimmen.

	<i>stimme gar nicht zu</i>	<i>stimme nicht zu</i>	<i>stimme eher nicht zu</i>	<i>neutral</i>	<i>stimme eher zu</i>	<i>stimme zu</i>	<i>stimme voll zu</i>
Ich werde diese Seite wieder benutzen.	①	②	③	④	⑤	⑥	⑦
Ich werde diese Website regelmäßig besuchen.	①	②	③	④	⑤	⑥	⑦
Ich würde diese Website Freunden und Bekannten weiterempfehlen.	①	②	③	④	⑤	⑥	⑦
Bei zukünftigem Interesse an solchen Themen könnte ich mir vorstellen, wieder diese Website aufzurufen.	①	②	③	④	⑤	⑥	⑦

Gesamteindruck

Alles in allem: Ich gebe dieser Website die Gesamtnote...

- ① (sehr gut)
- ② (gut)
- ③ (befriedigend)
- ④ (ausreichend)
- ⑤ (mangelhaft)

Literaturverzeichnis

- Appelman, A., & Sundar, S. S. (2016). Measuring Message Credibility: Construction and Validation of an Exclusive Scale. *Journalism & Mass Communication Quarterly*, 93(1), 59–79. <http://doi.org/10.1177/1077699015606057>
- Aranyi, G., & van Schaik, P. (2015). Testing a model of user experience with news websites. *Journal of the Association for Information Science and Technology* 67(7), 1555-1575. <http://doi.org/10.1002/asi.23462>
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(March 2015), 574–594. <http://doi.org/10.1080/10447310802205776>
- Bardzell, S., & Churchill, E. F. (2011). IwC special issue “Feminism and HCI: new perspectives” Special Issue Editors’ introduction. *Interacting with Computers*, 23(5), iii-xi. [http://doi.org/10.1016/S0953-5438\(11\)00089-0](http://doi.org/10.1016/S0953-5438(11)00089-0)
- Bölte, J., Hösker, T., Hirschfeld, G. & Thielsch, M. T. (2017). Electrophysiological correlates of aesthetic processing of webpages: A comparison of experts and laypersons. *PeerJ*, 5:e3440. <http://dx.doi.org/10.7717/peerj.3440>
- Borsci, S., Federici, S., & Lauriola, M. (2009). On the dimensionality of the System Usability Scale: A test of alternative measurement models. *Cognitive Processing*, 10(3), 193–197. <http://doi.org/10.1007/s10339-009-0268-9>
- Bosnjak, M., Galesic, M., & Tuten, T. (2007). Personality determinants of online shopping: Explaining online purchase intentions using a hierarchical approach. *Journal of Business Research*, 60(6), 597–605. <http://doi.org/10.1016/j.jbusres.2006.06.008>
- Bradley, M., & Lang, P. J. (1994). Measuring Emotion: The Self-Assessment Semantic Differential Manikin and the. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [http://doi.org/10.1016/0005-7916\(94\)90063-9](http://doi.org/10.1016/0005-7916(94)90063-9)
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7. <http://doi.org/10.1002/hbm.20701>
- Brooke, J. (2013). SUS: A Retrospective. *Journal of Usability Studies*, 8(2), 29–40. <http://doi.org/10.1074/jbc.R115.675280>
- Bühner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.
- Casaló, L. V., Flavián, C., & Guinalú, M. (2007). The role of security, privacy, usability and reputation in the development of online banking. *Online Information Review*, 31(5), 583–603. <http://doi.org/10.1108/14684520710832315>
- Cober, R.T., Brown, D.A., Levy, P.E., Cober, A.B. & Keeping, L.M. (2003). Organizational web sites: Web site content and style as determinants of organizational attraction. *International Journal of Selection and Assessment*, 11(2/3), 158-169. <http://doi.org/10.1111/1468-2389.00239>
- 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*, p. 271–280. New York: ACM.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation* (5. Aufl.). Heidelberg: Springer.
- Finstad, K. (2010). Response interpolation and scale sensitivity: Evidence against 5-point scales. *Journal of Usability Studies*, 5(3), 104–110.

- Finstad, K. (2010). The Usability Metric for User Experience. *Interacting with Computers*, 22(5), 323–327. <http://doi.org/10.1016/j.intcom.2010.04.004>
- Flavián, C., Guinalú, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43(1), 1–14. <http://doi.org/10.1016/j.im.2005.01.002>
- Grisaffe, D. B. (2007). Questions about the ultimate question: conceptual considerations in evaluating Reichheld's net promoter score (NPS). *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 20, 36. <http://doi.org/10.1016/j.jcps.2014.06.001>
- Hassenzahl, M., Platz, A., Burmester, M., & Lehner, K. (2000). Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. *Proceedings of the CHI2000 Conference on Human Factors in Computing Systems*, 201–208.
- Hassenzahl, M., Burmester, M. & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler & G. Szwillus (Hrsg.), *Mensch & Computer 2003. Interaktion in Bewegung* (S. 187–196). Stuttgart: B.G. Teubner.
- Hirschfeld, G. & Thielsch, M. T. (2015). Establishing meaningful cut points for online user ratings. *Ergonomics*, 58(2), 310-320. <http://dx.doi.org/10.1080/00140139.2014.965228>
- Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), 79–102. <http://doi.org/10.1016/j.ijhcs.2005.06.002>
- ISO (1998). *ISO 9241: Ergonomic requirements for office work with visual display terminals, VDTs – part 11: Guidance on usability*. Geneva: International Organization for Standardization.
- ISO (2006). *ISO 9241: Ergonomics of Human-System Interaction – Part 151: Guidance on World Wide Web Interfaces*. Geneva: International Organization for Standardization.
- ISO (2009). *9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems*. Geneva: International Organization for Standardization.
- Jäger, R. (2004). Konstruktion einer Ratingskala mit Smilies als symbolische Marken. *Diagnostica*, 50(1), 31–38. <http://doi.org/10.1026/0012-1924.50.1.31>
- Kang, Y., & Kim, Y. (2006). Do visitors' interest level and perceived quantity of web page content matter in shaping the attitude toward a web site? *Decision Support Systems*, 42(2), 1187–1202. <http://dx.doi.org/10.1016/j.dss.2005.10.004>
- Keiningham, T. L., Cooil, B., Andreassen, T. W., & Aksoy, L. (2007). A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing*, 71(3), 39-51. <http://doi.org/10.2307/30163980>
- Kirakowski, J., & Cierlik, B. (1998, October). Measuring the usability of web sites. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 42, No. 4, pp. 424-428). SAGE Publications.
- 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* (pp. 292–293). ACM.
- Koch, W., & Frees, B. (2016). Dynamische Entwicklung bei mobiler Internetnutzung sowie Audios und Videos. *Media Perspektiven*, 9/2016, 418–437.
- Krohne, W. H., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der „Positive and Negative Affect Schedule“ (PANAS). *Diagnostica*, 42(2), 139–156. <http://doi.org/1996-05083-003>
- Laugwitz, B., Schrepp, M., Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In A. M. Heinecke & H. Paul (Hrsg.), *Mensch und Computer 2006: Mensch und Computer im StrukturWandel* (S. 125–134). München: Oldenbourg.

- Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human Computer Studies*, 60(3), 269–298. <http://doi.org/10.1016/j.ijhcs.2003.09.002>
- Lee, S., & Koubek, R. J. (2012). Users' perceptions of usability and aesthetics as criteria of pre- and post-use preferences. *European Journal of Industrial Engineering*, 6(1), 87–117. <http://doi.org/10.1504/EJIE.2012.044812>
- Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013, April). UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2099–2102). ACM.
- Lewis, J. R., Utesch, B. S., & Maher, D. E. (2015). Measuring Perceived Usability: The SUS, UMUX-LITE, and AltUsability. *International Journal of Human-Computer Interaction*, 31(8), 496–505. <http://doi.org/10.1080/10447318.2015.1064654>
- Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, 25(2), 115–126. <http://doi.org/10.1080/01449290500330448>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <http://doi.org/10.5465/AMR.1995.9508080335>
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359. <http://doi.org/10.1287/isre.13.3.334.81>
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210–220. <http://doi.org/10.1016/j.pragma.2013.07.012>
- Minge, M., & Riedel, L. (2013). meCUE-Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In S. Boll, S. Maaß & R. Malaka (Hrsg.), *Mensch & Computer 2013* (S. 89–98). München: Oldenbourg.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2. Aufl.). Heidelberg: Springer.
- Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68(10), 689–709. <http://dx.doi.org/10.1016/j.ijhcs.2010.05.006>
- Moshagen, M. & Thielsch, M. T. (2013). A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology*, 32(12), 1305–1311. <http://dx.doi.org/10.1080/0144929X.2012.694910>
- Paßlick, N. (2015). Konstruktvalidierung der Skala zur Wahrgenommenen Website-Usability. Unveröffentlichte Masterarbeit, Westfälische Wilhelms-Universität Münster.
- Reichheld, F. F. (2003). The One Number You Need to Grow. *Harvard Business Review*, 81(12), 46–54+124. <http://doi.org/10.1111/j.1467-8616.2008.00516.x>
- Rocha, Á. (2012). Framework for a global quality evaluation of a website. *Online Information Review*, 36(3), 374–382. <http://dx.doi.org/10.1108/14684521211241404>
- Salaschek, M., Holling, H., Freund, P. A., & Kuhn, J.-T. (2007). Benutzbarkeit von Software: Vor- und Nachteile verschiedener Methoden und Verfahren. *Zeitschrift Für Evaluation*, 6(2), 247–276.
- Sarodnick, F., & Brau, H. (2015). *Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung*. Göttingen: Hogrefe.
- Sauro, J. (2011). *A practical guide to the System Usability Scale: Background, benchmarks, & best practices*. Denver, CO: Measuring Usability LLC.

- Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? *Proceedings of the 29th SIGCHI Conference on Human Factors in Computing Systems*, (January 2011), 2215–2223. <http://doi.org/10.1145/1978942.1979266>
- Sharp, B. (2008). Net promoter score fails the test. *Marketing research*, 20(4), 28-30.
- Shneiderman, B., Plaisant, C., & Cohen, M. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (Vol. 6th). London: Pearson.
- Sonderegger, A., Heyden, K., Chavaillaz, A., & Sauer, J. (2016, May). AniSAM & AniAvatar: Animated Visualizations of Affective States. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4828-4837). ACM.
- Sonderegger, A., Schmutz, S., & Sauer, J. (2016). The influence of age in usability testing. *Applied Ergonomics*, 52, 291–300. <http://doi.org/10.1016/j.apergo.2015.06.012>
- Thielsch, C. & Thielsch, M. T. (under review). Depressive web user: A different user experience?
- Thielsch, M. T. (2008). *Ästhetik von Websites. Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen*. Münster: MV Wissenschaft.
- Thielsch, M. T. (2008b). Inhalt, Usability und Ästhetik in der Bewertung durch Webnutzer. In M. Herczeg & M. C. Kindsmüller (Hrsg.), *Mensch & Computer 2008: Viel mehr Interaktion* (S. 441-444). München: Oldenbourg
- Thielsch, M. T., Blotenberg, I. & Jaron, R. (2014). User evaluation of websites: From first impression to recommendation. *Interacting with Computers*, 26 (1), 89-102. <http://dx.doi.org/10.1093/iwc/iwt033>
- Thielsch, M. T., Engel, R. & Hirschfeld, G. (2015). Expected usability is not a valid indicator of experienced usability. *PeerJ Computer Science*, 1:e19. <http://dx.doi.org/10.7717/peerj-cs.19>
- Thielsch, M. T., Grobien, F. R., Jahn, M. & Schmidt-Bussmann, S. (2014). Gestaltung und Evaluation von Karriere-Webseiten – Inhalt, Usability und Ästhetik. In P. Mehlich, T. Brandenburg & M. T. Thielsch (Hrsg.), *Praxis der Wirtschaftspsychologie III: Themen und Fallbeispiele für Studium und Anwendung* (S. 353-368). Münster: MV Wissenschaft.
- Thielsch, M. T. & Hirschfeld, G. (under review). Facets of website content.
- Thielsch, M. T. & Hirschfeld, G. (2012). Spatial frequencies in aesthetic website evaluations – explaining how ultra-rapid evaluations are formed. *Ergonomics*, 55 (7), 731-742. <http://dx.doi.org/10.1080/00140139.2012.665496>
- Thielsch, M. T. & Moshagen, M. (2014). *VisAWI Manual (Visual Aesthetics of Websites Inventory) and the short form VisAWI-S (Short Visual Aesthetics of Websites Inventory)*. <http://dx.doi.org/10.13140/RG.2.1.3985.6169>
- Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology*, 42(4), 253–264. <http://doi.org/10.1080/00207590701396674>
- Tractinsky, N., Cokhavi, A., Kirschenbaum, M., & Sharfi, T. (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. *International Journal of Human - Computer Studies*, 64(11), 1071–1083.
- Tuch, A. N.,argas-Avila, J. A., & Opwis, K. (2010). Symmetry and aesthetics in website design: It's a man's business. *Computers in Human Behavior*, 26(6), 1831–1837. <http://doi.org/10.1016/j.chb.2010.07.016>
- Tuch, A. N., Presslauer, E. E., Stöcklin, M., Opwis, K., &argas-Avila, J. A. (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies*, 70(11), 794-811. <http://doi.org/10.1016/j.ijhcs.2012.06.003>
- Tullis, T. S. & Stetson, J. N. (2004, June). *A Comparison of Questionnaires for Assessing Website Usability*. Usability Professionals Association (UPA) 2004 Conference, Minneapolis, USA.

Autorenprofile

PD Dr. **Meinald T. Thielsch**, Dipl.-Psych.; seit 2004 Mitarbeiter am Institut für Psychologie der Westfälischen Wilhelms-Universität Münster; dort 2008 Promotion zur „Ästhetik von Websites“ und 2013 Habilitation im Themenfeld „Mensch-Computer Interaktion“. Seit 2014 als Akademischer Rat in der Organisations- und Wirtschaftspsychologie an der Universität Münster tätig. Seine Forschungsschwerpunkte liegen in den Bereichen Human-Computer Interaction und User Experience, Wirtschaftspsychologie, Evaluation und Online-Forschung. Weitere Informationen finden sich unter www.meinald.de.

Dr. **Martin Salaschek**, Dipl.-Psych.; wissenschaftlicher Referent bei der Bundeszentrale für gesundheitliche Aufklärung (BZgA); koordiniert das BMBF-geförderte Forschungsprojekt CAHPOT (Community Access to Health Promotion via Online Tools). 2009 Abschluss als Diplom-Psychologe an der Westfälischen Wilhelms-Universität Münster, 2014 dort Promotion. Seine Arbeitsschwerpunkte liegen in den Bereichen Evaluation / Qualitätssicherung und Online-Forschung im Bereich Gesundheitsförderung.

Haftungsausschluss

Die Informationen in der Toolbox wurden nach besten Wissen und dem aktuell verfügbaren Stand der Forschung zusammengestellt. Soweit zum Zeitpunkt der Erstellung bekannt, wurden Hinweise auf weitere relevante Instrumentenentwicklungen gegeben. Zukünftige Veränderungen und methodische Innovationen sind im Bereich der Website-Evaluation / User Experience zu erwarten. Eine Gewährleistung für die inhaltliche Richtigkeit der Dokumente, die dem Auftraggeber übergeben wurden, sowie für die Eignung der Instrumente für den von NutzerInnen intendierten Zweck wird nicht übernommen. NutzerInnen dieser Toolbox stellen den Auftragnehmer / die Autoren von der Haftung für Ansprüche Dritter frei, die aufgrund einer fahrlässig oder vorsätzlich erfolgten unzulässigen Nutzung der Dokumente / Instrumente geltend gemacht werden.