

# Toolbox

## zur kontinuierlichen Website-Evaluation und Qualitätssicherung

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

*Stand: 24. Januar 2017*

*Autor: PD Dr. Meinald T. Thielsch*

Die Toolbox wurde zur kontinuierlichen Website-Evaluation und Qualitätssicherung der Plattform inforo im Rahmen des Forschungsprojektes CAHPOT entwickelt (weitere Informationen unter: <http://www.bzga.de/forschung/vernetzung-von-online-communitys-cahpot/>). Die Toolbox beinhaltet aktuell verfügbare deutschsprachige quantitativ-empirische Instrumente 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 einem kontinuierlichen Screening und Monitoring der zu evaluierenden Plattform. Ziel dieses allgemeinen Methodensets ist eine empirisch valide Analyse der Online-Plattformgestaltung und -nutzung bei der die NutzerInnen durch die Befragung möglichst wenig belastet werden. Das Methodenset umfasst dementsprechend insgesamt 32 Fragen und ist damit in unter 10 Minuten zu beantworten. Ein erweitertes Methodenset zur vertiefenden Evaluation wird in 2017 vorgelegt.



## Inhaltsverzeichnis

Abkürzungen und Begriffsklärung .....	3
Allgemeines zu User Experience .....	4
Zentrale Konstrukte der Website User Experience .....	5
Instrumente in der Toolbox .....	7
Hintergrundinformationen zur Toolbox .....	7
Haftungsausschluss .....	8
 <i>Darstellung der Instrumente in der Toolbox</i>	
Ersteindruck.....	9
Subjektive Inhaltswahrnehmung.....	11
Subjektive Usability / Nutzerzufriedenheit .....	13
Visuelle Ästhetik .....	17
Emotionale Reaktion: Befindlichkeit, Zustimmung, Zufriedenheit .....	20
Handlungs- und Nutzungsintentionen .....	22
Gesamteindruck .....	24
FAQ zur Toolbox.....	26
Literaturverzeichnis .....	27

## Abkürzungen und Begriffsklärung

$\alpha$	Cronbachs $\alpha$ zeigt an, wie gut die Items (Fragen) innerhalb einer Skala zusammenpassen; dies wird oft als Hinweis auf die $\rightarrow$ Reliabilität interpretiert. Werte von $\alpha = .70$ gelten als noch akzeptabel für Analysen / Untersuchungen auf Gruppenebene, Werte von $\alpha = .80$ als ausreichend und $\alpha = .90$ als hervorragend.
$M$	Mittelwert einer Frage oder Skala.
$m$	Anzahl von beurteilten Websites .
$n$	Anzahl der Beurteiler.
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. 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 $\rightarrow$ Cronbachs $\alpha$ als Maß angegeben. Ferner lassen sich die Messwerte aus zwei verschiedenen Messzeitpunkten korrelieren um so die zeitliche Stabilität eines Instrumentes zu bestimmen. Die Korrelationen können zwischen -1 und 1 schwanken; 0 wäre ein reiner Zufallswert, 1 wäre eine perfekte Übereinstimmung.
$SD$	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 überhaupt wirklich das Merkmal misst das er messen 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> ).

## 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, darunter fallen 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, 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 Akzeptanz und Nutzungsverhalten determiniert.

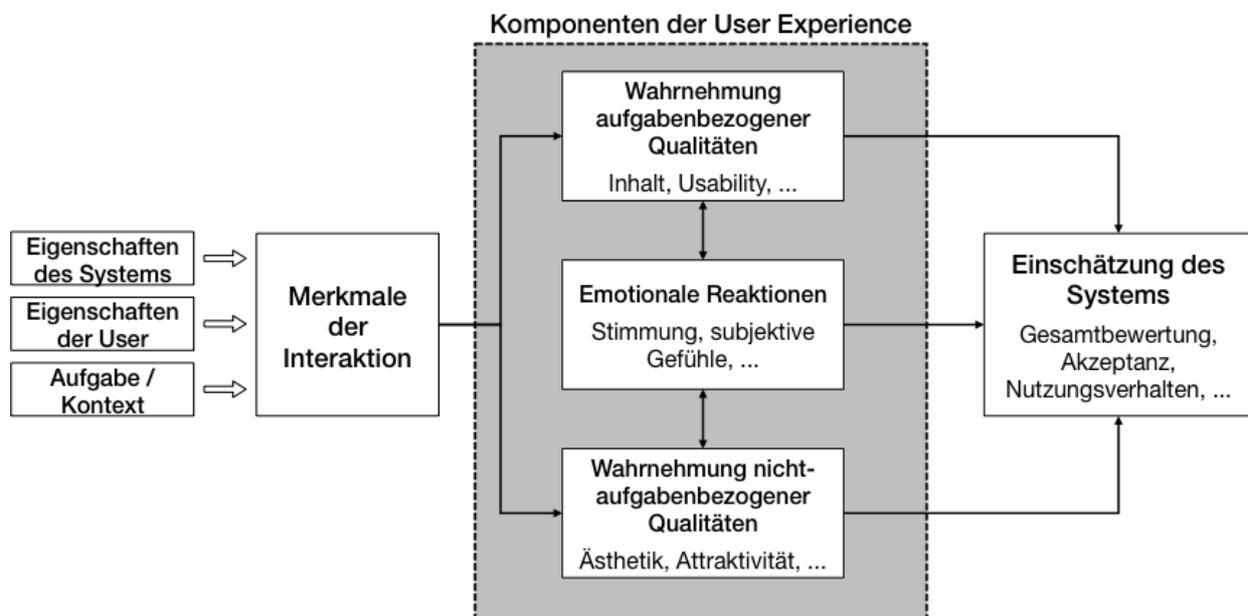


Abbildung 1: Komponenten der User Experience (Darstellung angelehnt an Thüring & Mahlke, 2007)

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. Erforscht sind hier unter

anderem Aspekte wie das biologische Alter (z.B. Sonderegger et al., 2016; Thielsch, 2008), Geschlecht (z.B. Bardzell, 2011; Tuch et al., 2010) oder Persönlichkeit (z.B. Bosnjak et al., 2007; Thielsch, 2008). Auch Erkrankungen, wie beispielsweise die Depression, können die User Experience negativ beeinflussen (Thielsch et al., 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 inforo 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 et al., 2006; Thielsch et al., 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 Benutzer oder 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 et al., 2016). Usability wird im Deutschen auch als Brauchbarkeit, Benutzerfreundlichkeit 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. Hornbaek, 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 et al., 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, hat Ästhetik eine besondere Bedeutung für den Ersteindruck (vgl. Lindgaard et al., 2006; Thielsch & Hirschfeld, 2012; Tractinsky et al., 2006), beeinflusst aber auch viele weitere Aspekte wie zum Beispiel die Zufriedenheit der User, Unternehmensattraktivität oder Vertrauen (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).

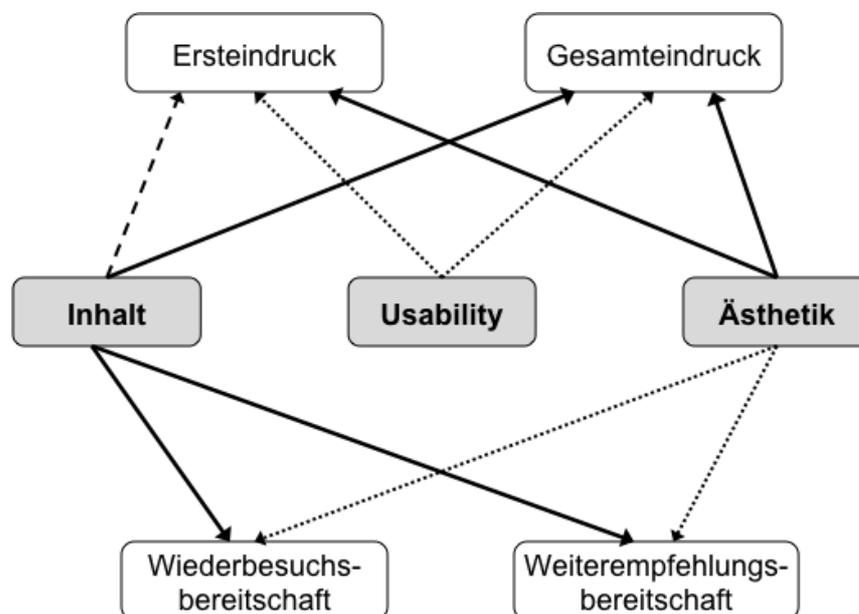


Abbildung 2: Einfluss der Wahrnehmung von Inhalt, Usability und Ästhetik auf Erst- und Gesamteindruck, bzw. Wiederbesuchs- und Weiterempfehlungsbereitschaft.

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

## Instrumente in der Toolbox

Entsprechend dem dargestellten Stand der Forschung zur Website-Wahrnehmung sind zentrale Konstrukte in der Toolbox durch verschiedene Instrumente erfasst (siehe Tabelle 1). Diese werden nachfolgend im Einzelnen dargestellt. Zudem finden sich im folgenden Abschnitt weitere Hintergrundinformationen zur Erstellung der Toolbox.

Tabelle 1: In die Toolbox eingeschlossene Items/Instrumente (insgesamt 32 Items) und deren Quellen

Konstrukt	Instrument	# Items	Quelle
<b>Ersteindruck</b>	Einzelitems	4	Thielsch (2008)
<b>Subjektive Inhaltswahrnehmung</b>	WWI (Fragebogen zur Wahrnehmung von Website-Inhalten)	9	Thielsch (2008)
<b>Subjektive Usability / Nutzerzufriedenheit</b>	UMUX-Lite (Usability Metric for User Experience - Lite)	2	Lewis et al. (2013)
	PWU-G (Perceived Website Usability - German)	7	Flavián et al. (2006), Thielsch (2008)
<b>Visuelle Ästhetik</b>	VisAWI-S (Visual Aesthetics of Websites Inventory - Short)	4	Moshagen & Thielsch (2013)
<b>Emotionale Reaktion: Befindlichkeit, Zustimmung, Zufriedenheit</b>	Smiley-Skala	1	Jäger (2004)
<b>Gesamteindruck</b>	Einzelitem Note	1	Thielsch (2008)
<b>Handlungs- und Nutzungsintentionen</b>	Wiederbesuchs-Skala (Scale assessing the intention to revisit the website)	4	Moshagen & Thielsch (2010)

## Hintergrundinformationen zur Toolbox

Die Toolbox wurde aus frei verfügbaren deutschsprachigen Evaluationsinstrumenten zusammengestellt. Ziel 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)<sup>1</sup>, des ZIS-Portals<sup>2</sup> des GESIS – Leibniz-Institut für Sozialwissenschaften (ZIS steht für „Zusammenstellung sozialwissenschaftlicher Items und Skalen“) – und des Psychologisch-Medizinischen Testportals Psychometrikon<sup>3</sup> 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 et al., 2006) – nicht in die Toolbox eingeschlossen wurden. Vorrangiger Grund ist, dass diese Instrumente vor allem zur Produktevaluation, jedoch weniger optimal zur Website-Evaluation geeignet sind. Hinsichtlich des UEQ wird hier möglicherweise zukünftig (voraussichtlich in 2017) eine Kurzversion vorliegen, deren Eignung hinsichtlich der Website-Evaluation dann neu zu bewerten ist.

Die digitale Dokumentation der eingeschlossenen Instrumente in der vorliegenden Toolbox umfasst:

1. Die deutschsprachigen Instrumente inklusive der jeweiligen Kodieranweisungen für jedes Teilinstrument, weiterführende Informationen und Hinweise zu jedem Teilinstrument und, soweit verfügbar, spezifische Manuale und Fachpublikationen;
2. Excel-Dateien zur halbautomatischen Auswertung der Instrumente;
3. eine .gpx-Datei zum direkten Import der Instrumente in EFS Survey.

## Haftungsausschluss

Die Informationen in der Toolbox wurden nach besten Wissen und dem aktuellen 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 von der Haftung für Ansprüche Dritter frei, die aufgrund einer fahrlässig oder vorsätzlich erfolgten unzulässigen Nutzung der Dokumente / Instrumente geltend gemacht werden.

---

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

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

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

# Darstellung der Instrumente in der Toolbox

## Ersteindruck

### Instrument

Einzelitems

### 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.

### Gütekriterien

**Objektivität:** Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität.

**Reliabilität:** Hierzu liegen keine Angaben vor.

**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 Itemantworten sind

- *stimme gar nicht zu*
- *stimme nicht zu*
- *stimme eher nicht zu*
- *weder noch*
- *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.

### **Benchmarks / Vergleichswerte**

Spezifische Benchmarks für Online-Communities im Gesundheitsbereich liegen derzeit nicht vor.

# Subjektive Inhaltswahrnehmung

## Instrument

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“, „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

**Objektivität:** Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität.

**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 ist zulässig und zeigt ein Cronbachs  $\alpha$  von  $.88$  (Thielsch et al., 2014).

**Validität:** Thielsch (2008) argumentiert für eine hohe inhaltliche Validität der 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 der PWU-G und dem VisAWI ( $.61 \leq r \leq .76$ ).

## 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 mit Beginn in 2017 eine 4-Item Kurzversion erprobt.

## Items

### Skala 1: Gefallen

- Die Website weckt mein Interesse.
- Der Inhalt der Webseite 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 (kodierte 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*
- *weder noch*
- *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 für Online-Communities im Gesundheitsbereich liegen nicht vor. Allgemeine Vergleichswerte zum WWI finden sich bei Thielsch et al. (2014): Insgesamt 42 Websites wurden von 512 Personen bewertet, dabei lag der WWI-Gesamtmittelwert bei 3,69 ( $SD = 1,17$ ). Das heißt, WWI-Gesamtwerte um die 3,7 entsprechen einer generell durchschnittlichen Inhaltsbewertung, Werte unter 2,5 wären eindeutig unterdurchschnittlich; hingegen Werte über 4,9 eindeutig überdurchschnittlich und Werte über 6,0 herausragend positiv.

Die subjektive Inhaltswahrnehmung bedingt eine Vielzahl von Aspekten wie beispielsweise die Zufriedenheit der Benutzer oder deren Vertrauen in eine Website (vgl. bspw. Aranyi & van Schaik, 2015; Kang & Kim, 2006; Rocha, 2012), 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).

# Subjektive Usability / Nutzerzufriedenheit

## Instrument

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). Dieser Fragebogen wurde durch Lewis et al. (2013) im Rahmen zweier Studien auf lediglich zwei zentrale Items verkürzt.

## Gütekriterien

**Objektivität:** Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität.

**Reliabilität:** Die 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. (Original: „This system’s capabilities meet my requirements.“)
- Dieses System ist einfach zu benutzen. (Original: „This system is easy to use.“)

### 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*
- *weder noch*
- *stimme eher zu*
- *stimme zu*
- *stimme voll zu*

### Auswertungsanweisung

Die 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 für Online-Communities im Gesundheitsbereich liegen derzeit nicht vor.

## Instrument

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 et al., 2015).

## Gütekriterien

**Objektivität:** Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität.

**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 für 2017 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 Webseite 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*
- *weder noch*
- *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 für Online-Communities 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 Webblogs/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 die 4,7 entsprechen einer generell durchschnittlichen Usability-Bewertung von Informationswebsites (bzw. 3,6 für Portale), Werte unter 3,3 wären eindeutig unterdurchschnittlich (bzw. 2,2 für Portale); hingegen Werte über 6,2 eindeutig überdurchschnittlich (bzw. 5,1 für Portale).

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).

# Visuelle Ästhetik

## Instrument

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. doi: 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.<sup>4</sup>

## Gütekriterien

**Objektivität:** Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität.

**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 Korrelationen ( $r = .72$ ) zur Attraktivitätsskala aus dem AttrakDiff 1 (Hassenzahl, Burmester & Koller, 2003) und eine deutliche konkurrente Validität in Form einer hohen Korrelation des VisAWI-S mit der angegebenen Wiederbesuchsbereitschaft ( $r = .52$ ).

---

<sup>4</sup> Für eine Untersuchung spezifischer Facetten der Ästhetik ist der VisAWI-S nicht geeignet, hier empfiehlt sich der Einsatz des vollständigen VisAWI.

## Anmerkungen / weitere Informationen

Weitere Informationen sind im Internet auf [www.VisAWI.de](http://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

- 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*
- *weder noch*
- *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 für Online-Communities 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 im 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 die folgenden spezifischen 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 Webblogs/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 die 4,6 entsprechen einer generell durchschnittlichen Schönheitsbewertung von Informationswebsites (4,2 für Portale, bzw. 4,5 für Blogs), Werte unter 3,2 wären eindeutig unterdurchschnittlich (2,8 für Portale, bzw. 3,2 für Blogs); hingegen Werte über 6,0 eindeutig überdurchschnittlich (5,5 für Portale, bzw. 5,9 für Blogs).

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

# Emotionale Reaktion: Befindlichkeit, Zustimmung, Zufriedenheit

## Instrument

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) schlägt eine einfache fünfstufige Smiley-Skala vor und validiert diese. Diese 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

**Objektivität:** Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität.

**Reliabilität:** Hierzu liegen keine Angaben vor.

**Validität:** Die Smiley-Skala korreliert deutlich mit dem PANAS ( $r = .50$  bis  $r = .67$ ), zudem wurde 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.

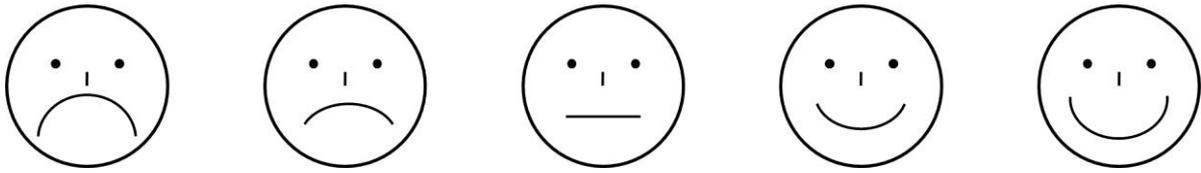
## Anmerkungen / weitere Informationen

Diese 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.<sup>5</sup>

---

<sup>5</sup> Vgl. <https://register.dpma.de/DPMAREgister/marke/register/300916140/DE> sowie <https://register.dpma.de/DPMAREgister/marke/register/300916159/DE>

## Items



## 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 nachfolgend dargestellten Aspekten überein.

						
PANAS	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	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)

# Handlungs- und Nutzungsintentionen

## Instrument

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, ist oft eine Bindung der NutzerInnen entscheidend für den langfristigen Erfolg einer Onlinepräsenz. Daher wird dieser Faktor häufig in Website-Evaluationen erfasst; ein bekanntes Beispiel mit breitem Anwendungshorizont ist der *Net promoter score* (NPS; Reichheld, 2003). Der NPS wird aber seitens der Forschung hinsichtlich Güte und Einsetzbarkeit stark kritisiert (siehe z.B. Grisaffe, 2007; Keiningham et al., 2007; Sharp, 2008). 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

**Objektivität:** Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität.

**Reliabilität:** Thielsch et al. (2014) berichten eine hohe interne Konsistenz einer verkürzten Skalenvariante mit drei Fragen ( $\alpha = .89$ ).

**Validität:** Hierzu liegen keine publizierten Angaben vor.

## 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 anzurufen.

### **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*
- *weder noch*
- *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 für Online-Communities im Gesundheitsbereich liegen derzeit nicht vor.

# Gesamteindruck

## Instrument

Einzelitem Note

## 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 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

**Objektivität:** Bei einer standardisierten Durchführung, insbesondere wenn diese computergestützt erfolgt, ist Durchführungsobjektivität als gegeben anzusehen. Im Falle einer automatisierten Auswertung gilt dies ebenso für die Auswertungsobjektivität.

**Reliabilität:** Hierzu 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 das kleine Werte positiv sind.

### **Benchmarks / Vergleichswerte / Interpretation**

Spezifische Benchmarks für Online-Communities 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 ein „befriedigend“, ein „ausreichend“ wäre unterdurchschnittlich und ein „mangelhaft“ stark unterdurchschnittlich. Ein „gut“ ist hingegen als überdurchschnittlich anzusehen, die Gesamtnote „sehr gut“ wäre herausragend positiv.

## FAQ zur Toolbox

### **Für welche Zielgruppe ist die Toolbox geeignet?**

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

### **Wie zitiere ich die jeweiligen Instrumente?**

- Siehe Zitationshinweis zu jedem Instrument.

### **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. „Software“ oder „Plattform“ statt „Website“) sind prinzipiell denkbar.

### **Wie lange braucht eine Testperson pro Frage?**

- Erfahrungsgemäß um die 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 geben?**

- Abhängig vom jeweiligen Evaluationsdesign in der Regel ja.

### **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 Prototypen als auch fertige Websites getestet werden.

### **Sollte die Zielwebsite bei der Bewertung sichtbar sein?**

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

### **Kann ich mit der BZgA-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 das die Befragten bei der Teilnahme schon erinnert werden falls ein Item vergessen wird. Aufgrund der Kürze der Instrumente in der Toolbox können ansonsten fehlende Werte sehr schnell dazu führen, dass eine Skala bei einem Befragten nicht ausgewertet werden kann.

## Literaturverzeichnis

- Aranyi, G., & van Schaik, P. (2015). Testing a model of user experience with news websites. *Journal of the Association for Information Science and Technology*.
- 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.
- 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. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1), 49-59.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- 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.
- 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.
- Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43(1), 1–14.
- 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.
- 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>
- Hornbaek, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), 79–102.
- ISO (2006). *ISO 9241: Ergonomics of Human-System Interaction – Part 151: Guidance on World Wide Web Interfaces*. Geneva: International Organization for Standardisation.
- ISO (1998). *ISO 9241: Ergonomic requirements for office work with visual display terminals, VDTS) – part 11: Guidance on usability*. Geneva: International Organization for Standardisation.
- ISO (2009). *9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems*. Geneva: International Organization for Standardisation.
- 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.

- 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, H. W., Egloff, B., Kohlmann, C. W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS). *Diagnostica*, 42, 139-156.
- 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.
- 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.
- 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.
- 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.
- 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), 46-55.
- 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.
- 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., 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., Rist, F. & 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. & 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., 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. & 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., Bargas-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., & Bargas-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.