

Meinald T. Thielsch

Westfälische Wilhelms-Universität Münster
Institut für Psychologie
Fliederstr. 21
48149 Münster
thielsch@uni-muenster.de

Morten Moshagen

Universität Mannheim
Lehrstuhl Psychologie III
Schloss, EO238
68165 Mannheim
moshagen@uni-mannheim.de

Abstract

Ästhetik ist ein bedeutendes Konstrukt der User Experience, sie beeinflusst eine Vielzahl von weiteren Faktoren und Konstrukte wie Usability, Zufriedenheit, Wiederbesuchs- oder Kaufbereitschaft. Dementsprechend wichtig ist, dass Ästhetik adäquat erfasst wird. Hierzu wurde der VisAWI (Visual Aesthetics of Websites Inventory) entwickelt. Dieser erfasst vier zentrale Aspekte der Webästhetik aus Sicht der Nutzer: Einfachheit, Vielfalt, Farbigekeit und Kunstfertigkeit. Die Konstruktion des VisAWI stützt sich auf vier Studien mit insgesamt 1574 Befragten. Die Faktorstruktur wurde dabei mit exploratorischen und konfirmatorischen Faktorenanalysen abgesichert. Konvergente, divergente, diskriminative und konkurrente Validität wurde in drei weiteren Studien mit insgesamt 965 Teilnehmern nachgewiesen. Hierbei wurden zwei experimentelle Validierungen durchgeführt. Es zeigen sich zudem gute bis sehr gute Reliabilitätswerte. Norm- und Benchmarkingwerte liegen für eine Stichprobe von 102 Websites und 2846 Befragte vor. Mit dem VisAWI-S wurde in drei weiteren Studien mit insgesamt 1673 Befragten eine Kurzversion geschaffen. Der VisAWI-S erfasst einen generellen Ästhetik-Faktor und zeichnet sich durch konvergente, divergente und konkurrente Validität aus. Diese 4-Item Kurzversion ist zudem ausreichend reliabel und zeigt hohe Korrelationen zur Langversion.

Keywords:

/// Ästhetik
/// Messung
/// User Experience
/// Schönheit
/// VisAWI

1. Einleitung

Das World Wide Web ist heutzutage ein zentraler Kommunikations- und Marketingkanal. Bei der empirischen Analyse von User Experience im Web hat sich visuelle Ästhetik als bedeutendes Konstrukt gezeigt: Ästhetik erfüllt dabei nicht nur wichtige Grundfunktionen und stellt ein Alleinstellungsmerkmal dar (zur Übersicht siehe Thielsch, 2008, S. 36f.). Ästhetik beeinflusst auch eine Vielzahl von weiteren Faktoren und Konstrukte wie Usability (Moshagen, Musch, & Göritz, 2009; Sonderegger & Sauer, 2010), Zufriedenheit (Cyr, Kindra & Dash, 2008; Lindgaard & Dudek, 2003), Wiederbesuchswahrscheinlichkeit (Mahlke, 2002; Moshagen & Thielsch, 2010; Yoo & Donthu, 2001) und Kaufbereitschaft (Parboteeah, Valacich & Wells, 2009). Dementsprechend wichtig ist, dass Ästhetik adäquat erfasst wird. Die wenigen existierenden Instrumente sind

jedoch meist ad hoc Skalen und leiden unter einer fraglichen Validität (Bargas-Avila et al., 2011). Eine Ausnahme stellt das von Lavie und Tractinsky (2004) entwickelte Instrument dar, welches zwei Faktoren, klassische und expressive visuelle Ästhetik, annimmt. Dieses Instrument zeigt allerdings auch verschiedene Schwächen, etwa die Vermischung von evaluativen und deskriptiven Items sowie die Abwesenheit von zentralen Merkmalen wie Farbe (Lindgaard et al., 2006; Moshagen & Thielsch, 2010). Daher wurde eine neues Instrument namens VisAWI (Visual Aesthetics of Websites Inventory) entwickelt.

2. Definition und Erfassung visueller Ästhetik

Ästhetik lässt sich als unmittelbar eintretendes, positiv bewertetes, auf ein Objekt bezogenes Erlebnis verstehen. Ein ästhetischer Eindruck ist nicht das Ergebnis eines

längeren bewussten Verarbeitungsprozesses, sondern tritt unmittelbar auf (vgl. Leder et al., 2004). Ästhetik bereitet Freude und wird daher positiv bewertet. Die klassische Ästhetikforschung spricht von einem „subjektivem Wohlgefallen“. Letztlich bezieht sich eine ästhetische Bewertung immer auf ein externes Objekt. Dem interaktionistischen Ansatz entsprechend entsteht ein ästhetischer Eindruck folglich aus dem Zusammenwirken der Eigenschaften eines Objekts und der des Beobachters (vgl. Moshagen & Thielsch, 2010).

Zur Erfassung von Usability-Bewertungen finden sich eine Reihe von etablierten Methoden (zur Übersicht siehe bspw. Gediga & Hamborg, 2002; Shneidermann & Plaisant, 2009), hierbei ist eine Trennung zwischen objektiven Merkmalen eines GUIs und subjektiven Eindrücken relativ leicht umsetzbar. Diese Trennung fällt bei der Bewertung von Ästhetik schwer: Die Erkenntnisse zu determinierenden

Variablen sind bisher nicht ausreichend um objektive Messmethoden aus diesen abzuleiten. Ästhetik bleibt somit sehr über die subjektive Wahrnehmung definiert.

Dementsprechend werden bei der Bewertung von Ästhetik zumeist die subjektiven Wahrnehmungen und Urteile der Nutzer per Fragebogen erfasst. Weitere Erhebungsverfahren und klassische Methoden wie z. B. Paarvergleich, Cognitive Walkthrough oder Checklisten-Evaluationen sind zwar denkbar, aber bisher nur an wenigen Stellen dokumentiert und wahrscheinlich seltener im Einsatz. Fragebogenmethoden sind beliebt und vergleichsweise leicht umzusetzen. Hierbei ergeben sich drei Vorgehensweisen (für eine detaillierte Übersicht siehe Thielsch, 2008, S. 53 ff.):

1. Einschätzung der Ästhetik über einzelne Fragen,
2. Einschätzung der Ästhetik über Fragebogenskalen,
3. Einschätzung der Ästhetik über standardisierte Fragebogeninstrumente.

Bei der ersten Variante ergeben sich Probleme der Messgenauigkeit (Reliabilität), sind Einzelitemmessungen doch anfällig für Messfehler (siehe Schmidt & Hunter, 1996) und Antwortverzerrungen (z. B. Akquieszenz, also die Tendenz von Probanden inhaltsunabhängig zuzustimmen), erlauben nur eine grobe Messung und erfassen nicht die Komplexität eines Konstruktes – oder werden schlicht von verschiedenen Befragten unterschiedlich verstanden. Diese Probleme können beim Einsatz von Skalen oder standardisierten Fragebogeninstrumenten deutlich verringert werden. Gerade vielen der eingesetzten ad hoc Skalen mangelt es jedoch an einer klaren Konstruktdefinition, einer Prüfung der psychometrischen Eigenschaften und einem Nachweis der inhaltlichen Gültigkeit (der Validität). Besonderes Augenmerk muss jedoch bei der metrischen Erfassung von Website-Ästhetik auf die Bestimmung der Validität gelegt werden, hier finden sich bisher nur wenige Ansätze.

3. Der VisAWI

3.1. Konstruktion und faktorielle Struktur

Der interaktionischen Definition von Ästhetik folgend, besteht der VisAWI aus subjektiven Einschätzungen auf Basis objektiver Reizmerkmale. Dem Nutzer werden Aussagen zu Gestaltungsmerkmalen einer Website präsentiert (z. B. „Die Farben haben eine angenehme Wirkung.“), welchen der Nutzer dann zustimmen oder ablehnen kann.

Der VisAWI erfasst vier zentrale Aspekte der Webästhetik aus Sicht der Nutzer: Einfachheit, Vielfalt, Farbigkeit und Kunstfertigkeit. Einfachheit umfasst dabei, wie übersichtlich und strukturiert das Layout der Website gestaltet ist. Auf der Skala Vielfalt werden Originalität und Dynamik des Designs beurteilt, während Farbigkeit nach der ästhetischen Einschätzung von Farbauswahl, -einsatz und -kombination fragt. Die Skala Kunstfertigkeit bezieht sich auf die Aktualität, Gekontheit und Durchdachtheit des Designs der Website. Zur Interpretation erzielter Werte auf den Skalen ist es wesentlich, den subjektiven Charakter der Bewertungen zu berücksichtigen.

So bedeutet ein hoher Wert beispielsweise auf der Skala Vielfalt nicht, dass das Design übermäßig originell oder dynamisch ist, sondern lediglich, dass das realisierte Maß an Vielfalt positiv bewertet wird.

Die Konstruktion des VisAWI stützt sich auf vier Studien mit insgesamt 1574 Befragten (eine detaillierte Darstellung im Detail findet sich bei Moshagen & Thielsch, 2010). Ausgehend auf einer empirischen Definition und Differenzierung des Merkmals „Website-Ästhetik“ wurde unter Einbezug von Webusern und Experten ein Itempool geschaffen. Das initiale Set von 96 Items wurde danach mehrfach überarbeitet und reduziert. Der VisAWI in der finalen Version umfasst insgesamt 18 Items (siehe Tabelle 1), die sich auf die vier Skalen zu jeweils vier bis fünf Items verteilen. Die Faktorstuktur wurde dabei mit exploratorischen und konfirmatorischen Faktorenanalysen bestimmt. Es zeigte sich ein hierarchisches Faktormodell (siehe Abbildung 1), in welchem die vier Facetten Einfachheit, Vielfalt, Farbigkeit und Kunstfertigkeit einem Generalfaktor der allgemeinen visuellen Ästhetik untergeordnet sind. Dieses Modell wurde erfolgreich an zwei unabhängigen Stichproben konfirmiert und somit kreuzvalidiert. [Abb. 1]

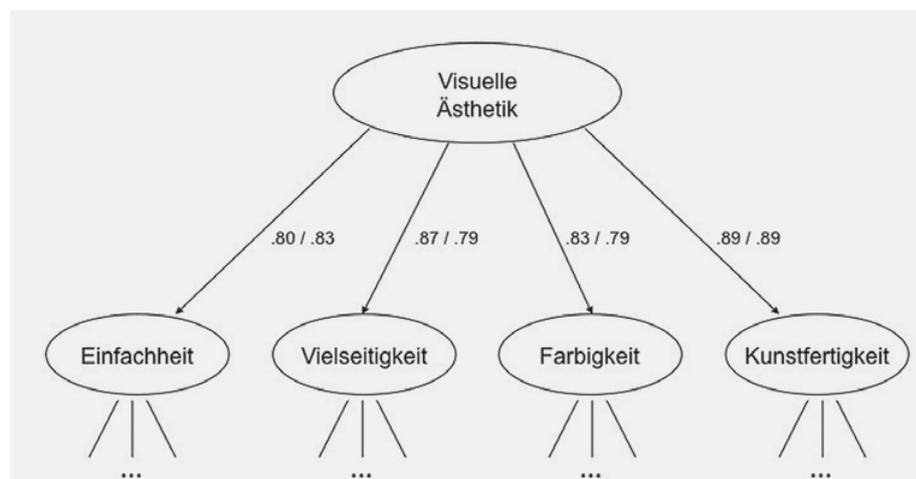


Abb. 1.
Strukturmodell des VisAWI. Ladungen der Facetten gemäß Moshagen & Thielsch (2010; Studie 3 / Studie 4).



3.2. Validierung des VisAWI

Die konvergente, divergente, diskriminative und konkurrente Validität wurde in drei Studien mit 965 Teilnehmern nachgewiesen, hierbei wurden zwei experimentelle Validierungen umgesetzt. Dabei zeigten sich moderate bis hohe Korrelationen zu verwandten Konstrukten und ähnlichen Messinstrumenten. So finden sich hohe Korrelationen zum Messinstrument von Lavie und Tractinsky (2004) im Bereich von $.52 \leq r \leq .82$ für die Skala zur klassischen Ästhetik und $.36 \leq .80$ für die zur expressiven Ästhetik. Auch die Korrelationen zur Attraktivitätsskala aus dem AttrakDiff 1 (siehe z. B. Hassenzahl, Burmester & Koller, 2003) sind hoch ($.60 \leq .80$). Dies spricht für konvergente Validität.

Schwächere Korrelationen zeigen sich zu divergenten Konstrukten, wie der Skala zur pragmatischen Qualität aus dem AttrakDiff ($.41 \leq .77$), einer Usability-Skala nach Flavián, Guinalú und Gurrea (2006, $.04 \leq .48$) oder dem WWI ($.48 \leq .62$), einem Instrument zur Bewertung der Inhaltsqualität (Thielsch 2008). An dieser Stelle fällt auf, dass gerade die Einfachheitsskala des VisAWI hohe Korrelationen zu pragmatischen oder Usability-Aspekten aufzeigt (siehe Moshagen & Thielsch, 2010, S. 700) und das generell teilweise mittlere und hohe Korrelationen zu divergenten Konstrukten vorliegen. Dies könnte verschiedene Gründe haben, werden doch insbesondere die Wechselwirkungen (oder auch Halo-Effekte) zwischen Ästhetik und Konstrukten wie Usability immer wieder diskutiert. Zudem werden Webdesigner versuchen Aspekte wie Inhalt, Usability und Ästhetik in gleichem Maße zu optimieren, so dass diese praktisch nicht vollkommen unabhängig sein können. An dieser Stelle ist jedoch für den VisAWI eine näher gehende, experimentelle Betrachtung notwendig, um die Validität abzusichern.

Daher wurden zwei experimentelle Validierungen durchgeführt. Hierbei zeigte sich, dass die Facetten des VisAWI einerseits responsiv für systematische Variation des Layouts einer Website sind,

aber andererseits spezifisch auf bestimmte Manipulationen reagieren. So beeinflusst bspw. eine Änderung des Farbschemas ausschließlich die Bewertungen auf der Farbigeitsfacette, nicht aber auf den anderen Facetten. Zudem zeigt sich diskriminative Validität: Der VisAWI ist in der Lage auch größere Sets von real existierenden Websites (in diesem Fall ein Pool von 42 Websites aus neun Inhaltsbereichen) signifikant hinsichtlich der Ästhetik zu unterscheiden (in einer MANOVA mit den Websites als UV und dem VisAWI als AV, siehe Moshagen & Thielsch, 2010, S. 700).

Ferner konnte die Wiederbesuchswahrscheinlichkeit (als konkurrentes Validitätsmaß) mittels des VisAWI reliabel vorhergesagt werden. Die Korrelation des Gesamtwerts mit einer Skala zur Wiederbesuchsbereitschaft liegt bei $.51$, die Subskalen des VisAWI korrelieren mit dieser im Bereich von $.40 \leq .48$. Ebenso zeigen sich gute bis sehr gute Reliabilitätswerte: Die innere Konsistenz beträgt für alle Skalen $\alpha \geq .85$, für den Gesamtwert des VisAWI ist $\alpha = .94$.

3.3. Hinweise zur Anwendung des VisAWI

Zur Darbietung: In unseren Studien haben wir oft die Einzelitems des VisAWI in einem Frame am oberen Bildrand eingeblen-det und im unteren Bereich die zu bewertende Website dargestellt (die Items waren dabei komplett randomisiert). Es sind aber auch andere Darbietungsformen möglich. Insbesondere bei einem (Live-)Nutzertest kann man auch alle Items insgesamt oder auch den Fragebogen als Papierform vorlegen. In einzelnen Projekten wurde der VisAWI zudem genutzt um andere grafische Interfaces als Websites zu beurteilen. Bei Softwarebewertungen wurden hierfür in den Items Wörter wie „Seite“ durch „Software“ ersetzt, um eine bessere Passung der Items zum Gegenstand der Evaluation herzustellen.

Entsprechend der dargestellten Struktur (siehe auch Abbildung 1) kann der VisAWI sowohl auf Skalenebene ausgewertet werden als auch als Gesamtfragebogen. Vor

Item Deutsch	Item Englisch
Einfachheit Simplicity ($\alpha = .89$)	
(r) Das Layout wirkt zu gedrängt.	The layout appears too dense.
Das Layout ist gut zu erfassen.	The layout is easy to grasp.
Das Layout erscheint angenehm gegliedert.	The layout appears well structured.
(r) Die Seite erscheint zu uneinheitlich.	The site appears patchy.
Auf der Seite passt alles zusammen.*	Everything goes together on this site.*
Vielseitigkeit Diversity ($\alpha = .87$)	
(r) Die Seitengestaltung ist uninteressant.	The design is uninteresting.
Das Layout ist originell.	The layout is inventive.
(r) Die Gestaltung wirkt einfalllos.	The design appears uninspired.
Das Layout wirkt dynamisch.	The layout appears dynamic.
Die Seite ist angenehm vielseitig.*	The layout is pleasantly varied. *
Farbigkeit Colorfulness ($\alpha = .89$)	
Die farbliche Gesamtgestaltung wirkt attraktiv.*	The color composition is attractive.*
(r) Die Farben passen nicht zueinander.	The colors do not match.
(r) Der Farbeinsatz ist nicht gelungen.	The choice of colors is botched.
Die Farben haben eine angenehme Wirkung.	The colors are appealing.
Kunstfertigkeit Craftsmanship ($\alpha = .85$)	
Das Layout ist professionell. *	The layout appears professionally designed.*
(r) Das Layout ist nicht zeitgemäß.	The layout is not up-to-date.
Die Seite erscheint mit Sorgfalt gemacht.	The site is designed with care.
(r) Das Layout wirkt konzeptlos.	The design of the site lacks a concept.

Tab. 1. Items des VisAWI und des VisAWI-S in Deutsch und Englisch.

Anmerkung: Mit * markiert sind die vier Items des VisAWI-S. (r) = vor der Bildung des Skalennittels ist dieses Item umzu-kodieren. Als Antwortskala dient eine 7-stufige Likert-Skala (von 1 „stimme überhaupt mit zu“ bis 7 „stimme voll zu“).

der Auswertung sind die negativ gepolten Items (siehe Tabelle 1) umzukodieren, danach können Skalen- und Gesamtmittelwerte gebildet werden. Um konkrete Verbesserungsvorschläge zu erhalten (besonders bei einem formativen Testen), ist das Anhängen spezifischer offener Fragen denkbar.

3.4. Normierung und Benchmarking des VisAWI

Zu Normierung und Benchmarking des VisAWI haben wir Daten aus insgesamt fünf Studien kombiniert (drei bisher unpublizierte Arbeiten, sowie je ein Validierungsdatensatz aus unseren Studien zum VisAWI und VisAWI-S). Insgesamt haben hier 2843 Personen (1747 Frauen und 1096 Männer) im Alter von 14 bis 82 Jahren ($M = 26,95$; $SD = 9,68$) 102 Websites beurteilt.

Bei einem Gegenstand wie visueller Ästhetik würden manche große Geschlechts- und Alterseffekte erwarten. Der VisAWI zeigt sich allerdings weitgehend robust gegenüber solchen Effekten. So beträgt beispielsweise der Unterschied im VisAWI-Gesamtmittelwert zwischen Männern und Frauen nur 0,11 ($M_{\text{Männer}} = 4,02$; $M_{\text{Frauen}} = 4,13$). Dieser Unterschied wird zwar aufgrund der Stichprobengröße signifikant ($F_{1, 2842} = 5,30$, $p = .02$, $\eta^2 < .01$), eine Effektgröße¹ von $d = 0,08$ zeigt aber an, dass dieser Geschlechtsunterschied praktisch unbedeutend ist. Auf Ebene der Subskala Einfachheit wird der Geschlechtseffekt gar nicht signifikant, bei den anderen drei Skalen zeigen sich ebenfalls nur sehr kleine Effekte ($d_{\text{Vielseitigkeit}} = 0,18$; $d_{\text{Farbigkeit}} = 0,10$; $d_{\text{Kunsthierarchie}} = 0,09$). Geschlechtseffekte können damit auch auf Skalenebene vernachlässigt werden.

Die Korrelationen zwischen dem VisAWI und dem Alter sind gering aber signifikant (Gesamtwert: $r = .10$; Einfachheit: $r = .12$; Vielseitigkeit: $r = .08$; Farbigkeit: $r = .07$; Kunstfertigkeit: $r = .05$). Die Alterseffekte sind zwar in Ihrer Effektgröße als klein anzusehen, insbesondere ältere Befragte kommen aber zu leicht positiveren Urteilen. Dieses mag in manchen Projekten von

Relevanz sein – eine nach sechs Altersgruppen klassierte Normierung findet sich daher im Appendix in Tabelle 2.

Teilweise sehr große Effekte finden sich, wenn man die bewerteten Websites nach Kategorien sortiert und diese vergleicht. Hierfür haben wir die 102 Websites, die von der Normierungsstichprobe bewertet wurden, in zehn Kategorien eingeteilt (Informationen zum Kategorisierungsschema finden sich bei Thielsch, 2008, S. 86f.) Sowohl für den Gesamtwert ($F_{9, 2831} = 50,44$, $p < .01$, $\eta^2 = .14$) als auch für die Subskalen ($F_{36, 11324} = 22,77$, $p < .01$, $\eta^2 = .07$) ergeben sich bedeutende Unterschiede (Effekte von Alter und Geschlecht waren hierbei kontrolliert). Eine entsprechende Darstellung der Benchmarkingwerte findet sich im Appendix in Tabelle 3.

4. Kurzform VisAWI-S

Ogleich der VisAWI eine reliable und valide Einschätzung verschiedener Facetten der visuellen Ästhetik erlaubt, mag es Situationen geben, in denen man nur kurz eine generelle Ästhetik-Einschätzung erfragen möchte. Sei es, das andere Konstrukte im Vordergrund stehen oder insgesamt nur wenige Fragen gestellt werden können. In solchen Kontexten erscheint der VisAWI mit seinen 18 Items übermäßig lang.

Daher wurde in einer Serie von drei weiteren Studien mit insgesamt 1673 Befragten eine 4-Item Kurzversion (siehe Tabelle 1) entwickelt und validiert (Moshagen & Thielsch, under review).

4.1. Konstruktion und faktorielle Struktur

Der VisAWI-S erfasst lediglich den generellen Ästhetik-Faktor. Ziel war allerdings dabei jede Facette des VisAWI mit zumindest einem Item zu repräsentieren, daher wurde Items entsprechend inhaltlicher Kriterien und anhand der Faktorladung im Original-VisAWI ausgewählt. Eine konfirmatorische Faktorenanalyse mit den vier Items zeigt eine hervorragende Modellpassung und damit, dass diese Items

den latenten g-Faktor Website-Ästhetik repräsentieren.

4.2. Validierung des VisAWI-S

Die Kurzversion zeichnet sich ebenso wie die Langversion durch konvergente, divergente und konkurrente Validität aus. Wie zuvor beim VisAWI finden sich hohe Korrelationen zur Attraktivitätsskala aus dem AttrakDiff 1 ($r = .72$), sowie geringere Korrelationen zu divergenten Maßen wie Usability ($r = .54$), der Skala zur pragmatischen Qualität aus dem AttrakDiff ($r = .53$) oder verschiedenen Inhaltsbewertungen aus dem WWI wie Gefallen ($r = .49$), Verständlichkeit ($r = .34$) oder Qualität und Nutzen ($r = .41$). Divergente Validität des VisAWI-S zeigt sich weiterhin dadurch, dass die Stimmung der Befragten (erhoben mittels des MDBF von Steyer et al., 1997) keinerlei signifikanten Einfluss hat. Konkurrente Validität zeigt sich, ebenso wie beim VisAWI, in einer hohen Korrelation zur Wiederbesuchsbereitschaft ($r = .52$). Letztere ist umso beachtlicher, wenn man bedenkt, dass weitere Faktoren wie insbesondere der Inhalt die Intention eine Website wieder zu besuchen stark beeinflussen sollten.

Weiterhin ist die Kurzversion für Gruppenvergleiche ausreichend reliabel (je Studie $.76 \leq r \leq .81$) und zeigt eine hohe Korrelation zur Langversion ($r = .91$).

4.3. Hinweise zur Anwendung des VisAWI-S

Da der VisAWI-S nur aus vier Items besteht, sollte aus diesen ein Mittelwert gebildet werden. Dieser Mittelwert repräsentiert den generellen Ästhetik-Faktor, der im VisAWI-Modell (siehe Abbildung 1) gefunden wurde. Möchte man eine Auswertung auf Ebene der Subfacetten machen, so empfiehlt sich der Einsatz des vollständigen VisAWI.



5.

Zusammenfassung

In einem aufwendigen Konstruktionsprozess wurden VisAWI und die Kurzform VisAWI-S geschaffen. Beide Instrumente zeichnen sich durch Reliabilität und Validität aus. Der VisAWI erfasst dabei sowohl einen g-Faktor Website-Ästhetik als auch die vier Facetten Einfachheit, Vielseitigkeit, Farbigkeit und Kunstfertigkeit; der VisAWI-S erfasst nur den g-Faktor. Beide Instrumente sind sowohl für formative als auch für summative Website-Evaluationen geeignet, insbesondere der VisAWI-S empfiehlt sich hier aufgrund seiner Kürze. So erlaubt es der VisAWI-S effizient für Webästhetik zu screenen, wenn eigentlich andere Konstrukte im Vordergrund stehen oder insgesamt nur wenig Fragen gestellt werden können.

Seit der Konstruktion der beiden Instrumente wurden diese in verschiedensten Forschungs- und Praxisprojekten zur Website-Beurteilung erfolgreich eingesetzt. Für Website-Evaluationen können wir erste Vergleichswerte zur Verfügung stellen. Wir hoffen, dass zukünftige Anwender von VisAWI und VisAWI-S uns bei der Erweiterung des Benchmarkings unterstützen. Zudem wurden die Fragebögen bereits genutzt um Prototypen, Software oder sogar andere Medienprodukte wie Zeitschriftencover zu beurteilen. Zukünftig sind hier weitere Einsatzbereiche und die Kombination mit anderen Methoden (insbesondere aus dem qualitativen Bereich) denkbar. Wir hoffen, dass dieser erfolgreiche Einsatz des VisAWI sich fortsetzt und verschiedenste Evaluationen durch ein standardisiertes Instrument zur quantitativen Ästhetikbeurteilung bereichert.

Literatur

1. Bargas-Avila, J., & Hornbæk, K. (2011). Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. CHI '11 Proceedings of the 2011 annual conference on Human factors in computing systems, 2689-2698.
2. Cohen, J. (1988). Statistical power analysis for the behavioural sciences. Hillsdale: Erlbaum.
3. Cyr, D., Kindra, G. S. & Dash, S. (2008). Web site design, trust, satisfaction and e-loyalty: the Indian experience. *Online Information Review* 32, 773-790.
4. Flavián, C., Guinaliú, M. & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43(1), 1-14. Elsevier.
5. Gediga, G., & Hamborg, K.-C. (2002): Evaluation in der Software Ergonomie: Methoden und Modelle im Software-Entwicklungs-prozess. *Zeitschrift für Psychologie*, 210(1), 40-57.
6. Hassenzahl, M., Burmester, M. & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler & G. Szwillus (Eds.), *Mensch & Computer 2003. Interaktion in Bewegung* (S. 187-196). Stuttgart: B.G. Teubner.
7. Lavie, T. & Tractinsky, N. (2004). Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. *International Journal of Human - Computer Studies*, 60, 269-298.
8. Leder, H., Belke, B., Oeberst, A. & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgements. *British Journal of Psychology*, 95, 489-508.
9. Lindgaard, G. & Dudek, C. (2003). What is this evasive beast we call user Satisfaction? *Interacting with Computers*, 15, 429-452.
10. 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.
11. Mahlke, S. (2002). Factors influencing the experience of website usage. In: *Proceedings of the CHI'02 Extended Abstracts on Human Factors in Computing Systems*. ACM, Minneapolis, pp. 846-847.
12. Moshagen, M., Musch, J. & Göritz, A. S. (2009). A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics*, 52, 1311-1320.
13. Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68 (10), 689-709. doi:10.1016/j.ijhcs.2010.05.006
14. Moshagen, M. & Thielsch, M. T. (under review). A short version of the visual aesthetics of websites inventory.
15. Parboteeah, D. V., Valacich, J. S. & Wells, J. D. (2009). The influence of website characteristics on a consumer's urge to buy impulsively. *Information Systems Research* 20, 60-78.
16. Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1(2), 199-223.
17. Shneiderman, B., & Plaisant, C. (2009). *Designing the user interface: Strategies for effective Human-Computer-Interaction* (5th ed.). Boston : Addison-Wesley.
18. Sonderegger, A. & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics* 41, 403-410.
19. Steyer, R., Schwenkmezger, O., Notz, P. & Eid, M. (1997). *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF)*. Göttingen: Hogrefe.
20. Thielsch, M. T. (2008). *Ästhetik von Websites: Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen*. Münster: MV Wissenschaft.
21. Yoo, B. & Donthu, N. (2001). Developing a scale to measure the perceived quality of an Internet shopping site (SITEQUAL). *Quarterly Journal of Electronic Commerce* 2, 31-46.
22. Zhang, P., & von Dran, G. M. (2000). Satisfiers and dissatisfiers: A two-factor model for Website design and evaluation. *Journal of the American Society for Information Science*, 51(14), 1253-1268.

¹ Nach Cohen (1998) zeigt einen Effektgröße von $d = 0,2$ einen kleinen Effekt an, $d = 0,5$ einen mittleren und ab $d = 0,8$ wird von einem großen Effekt gesprochen.

Appendix

Altersgruppe		Gesamtwert	Einfachheit	Vielseitigkeit	Farbigkeit	Kunstfertigkeit
14-19 Jahre	M	4,10	4,31	3,44	4,24	4,53
n = 194	SD	1,05	1,30	1,20	1,41	1,19
20-29 Jahre	M	4,04	4,13	3,47	4,23	4,44
n = 2034	SD	1,22	1,47	1,30	1,57	1,43
30-39 Jahre	M	4,15	4,29	3,58	4,38	4,45
n = 315	SD	1,24	1,57	1,36	1,51	1,45
40-49 Jahre	M	4,24	4,52	3,60	4,44	4,49
n = 156	SD	1,28	1,48	1,45	1,60	1,45
50-59 Jahre	M	4,53	4,82	3,94	4,51	4,95
n = 100	SD	1,38	1,56	1,63	1,59	1,41
60 Jahre und älter	M	4,55	4,85	3,89	4,91	4,63
n = 44	SD	1,44	1,59	1,63	1,48	1,73
Gesamtstichprobe	M	4,09	4,21	3,51	4,28	4,47
N = 2843	SD	1,23	1,49	1,33	1,56	1,43

Tab. 2.

Normierung des VisAWI: Gesamt- und Skalenmittelwerte in Abhängigkeit von der Altersgruppe.

Anmerkung: Da Geschlechtseffekte aufgrund ihrer geringen Größe praktisch unbedeutend sind, finden sich hier Werte für Männer und Frauen zusammengefasst.

Kategorie		Gesamtwert	Einfachheit	Vielseitigkeit	Farbigkeit	Kunstfertigkeit
Download & Software	M	3,48	3,38	2,85	3,96	3,91
(m = 10; n = 97)	SD	0,99	1,30	1,18	1,41	1,15
E-Commerce	M	3,72	3,92	3,09	3,90	4,08
(m = 11; n = 237)	SD	1,11	1,43	1,21	1,44	1,32
Entertainment	M	3,84	3,72	3,63	3,69	4,40
(m = 11; n = 232)	SD	1,13	1,42	1,22	1,48	1,38
E-Learning	M	4,24	4,36	3,52	4,52	4,70
(m = 10; n = 111)	SD	1,19	1,57	1,35	1,53	1,44
E-Recruiting & E-Assessment	M	3,73	4,05	3,18	3,73	4,02
(m = 12; n = 355)	SD	1,31	1,41	1,39	1,71	1,55
Information	M	4,44	4,64	3,64	4,84	4,79
(m = 12; n = 339)	SD	1,11	1,37	1,27	1,33	1,25
Portale	M	3,39	3,13	3,21	3,63	3,72
(m = 12; n = 396)	SD	1,18	1,39	1,20	1,48	1,46
Präsentation & Selbstdarstellung	M	4,61	4,83	3,88	4,80	5,05
(m = 14; n = 873)	SD	1,07	1,25	1,34	1,41	1,21
Weblogs und Social Sharing	M	3,99	3,94	3,62	4,29	4,21
(m = 5; n = 81)	SD	1,28	1,41	1,32	1,57	1,49
Suchmaschinen	M	4,30	4,86	3,42	4,60	4,42
(m = 5; n = 122)	SD	1,11	1,30	1,37	1,50	1,33
Gesamtstichprobe	M	4,09	4,21	3,51	4,28	4,47
(m = 102; N = 2843)	SD	1,23	1,49	1,33	1,56	1,43

Tab. 3.

Benchmarking des VisAWI: Gesamt- und Skalenmittelwerte in Abhängigkeit von der Website-kategorie.

Anmerkung: m = Anzahl der beurteilten Websites in der Kategorie, n = Anzahl der Beurteiler.

Hinweise zum Kategorisierungsschema finden sich bei Thielsch (2008, S. 86f.).