

In der Kürze liegt die Würze: Der Web-CLIC short zur Bewertung digitaler Inhalte

Meinald T. Thielsch
Department of Psychology
University of Münster
Münster, Germany
thielsch@uni-muenster.de

Gerrit Hirschfeld
Faculty of Business and Health
Bielefeld University of Applied Sciences
Bielefeld, Germany
gerrit.hirschfeld@fh-bielefeld.de

ABSTRACT

In digitalen Medien wie dem Internet ist der Inhalt entscheidend. Die subjektive Wahrnehmung von Inhalten durch die Nutzer*innen beeinflusst eine Vielzahl ihrer Bewertungen und Einstellungen und hat Konsequenzen für das Verhalten. Daher ist es essentiell die Inhaltswahrnehmungen der Nutzer*innen zu messen. Eine solche Erfassung sollte möglichst valide sein, muss aber gleichzeitig oft möglichst kurz sein. Hierfür haben wir eine vier Items umfassende Kurzversion des Web-CLIC-Fragebogens [18] erstellt. Dieser Web-CLIC-S wurde in einer Reihe von drei Studien getestet, in Summe $n = 1.414$ Befragte bewerteten dabei insgesamt 33 voll funktionsfähige Websites verschiedener Inhaltsbereiche. Der Web-CLIC-S spiegelt einen g-Faktor der subjektiven Wahrnehmung von Webinhalten wider: Die Unidimensionalität wurde mittels konfirmatorischer Faktoranalyse geprüft. Der Web-CLIC-S weist hohe interne Konsistenz und eine hohe kurz- bis mittelfristige Retest-Reliabilität auf. Darüber hinaus finden wir starke Evidenz für die Konstruktvalidität im Sinne von konvergenter, divergenter, diskriminativer, konkurrenzer, inkrementeller und prädiktiver Validität. Interpretationshilfen in Form von Benchmarks für 12 verschiedene Inhaltsbereiche und optimale Schwellenwerte wurden auf Basis von 12.568 Bewertungen für insgesamt 183 verschiedene Websites berechnet. Insgesamt zeigen die vorliegenden Studien, dass der Web-CLIC-S als fundiertes Screening-Instrument zur Beurteilung der subjektiven Wahrnehmung von Inhalten durch die User in Praxis und Forschung dienen kann.

CCS CONCEPTS

• Human-centered computing → Human computer interaction (HCI) → Empirical studies in HCI

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). MuC'20 Workshops, Magdeburg, Deutschland © Proceedings of the Mensch und Computer 2020 Workshop on «Quantitative Methoden zur Messung von User Experience». Copyright held by the owner/author(s). <https://doi.org/10.18420/muc2020-ws105-376>

KEYWORDS

Website-Inhalt; WWW, Evaluation, Informationsqualität, User Experience, Web-CLIC-S

1 Einleitung

Das Internet ist heutzutage ein täglicher Begleiter in Beruf und Freizeit. Der Inhalt ist dabei der wichtigste Aspekt für die Nutzer*innen (vgl. [16]). Der Spruch „Content is king“ wird durch eine Vielzahl von Studien bestätigt: Der Inhalt bestimmt unter anderem die Zufriedenheit mit der Website (z.B. [6]), das Vertrauen in den Anbieter (z.B. [12]), Loyalität von Kunden, Weiterempfehlungs- und Wiederbesuchsverhalten (z.B. [7], [16]) sowie Kaufintentionen (z.B. [13]). Darüber hinaus hat sich gezeigt, dass die Bewertungen von Inhalten durch die User auch ihr Verhalten und Entscheidungen beeinflussen [18]. Schließlich kann die Wahrnehmung von Inhalten sogar die Leistung bei der Nutzung digitaler Systeme beeinflussen: Meeßen und Kollegen [9] variierten systematisch die Glaubwürdigkeit der von einem digitalen Management-Informationssystem bereitgestellten Informationen – eine höhere Glaubwürdigkeit führte zu einem größeren Vertrauen der Nutzer*innen in das System, was zu einer besseren Leistung in Entscheidungsaufgaben und einem höheren Wohlbefinden der Nutzer*innen führte. Dies veranschaulicht, dass die subjektive Wahrnehmung von Inhalten durch die User wichtig ist und das tatsächliche Verhalten und die damit verbundenen Ergebnisse beeinflussen kann.

Umso überraschender ist es, dass es nur wenige und zumeist unsystematische Ansätze gibt diese Wahrnehmungen von Webinhalten zu erfassen. Die meisten Studien verlassen sich auf einzelne Items oder Skalen, die oftmals ad hoc erstellt und nicht systematisch validiert wurden (siehe [18]). Eine Ausnahme ist der Web-CLIC Fragebogen [18], die Abkürzung Web-CLIC steht für „Website-Clarity, Likeability, Informativeness, Credibility“. Insgesamt 12 Fragen decken dabei die vier Bereiche Verständlichkeit, Gefallen, Informationsgehalt und Glaubwürdigkeit ab. Zudem kann ein Gesamtwert berechnet

werden, dieser spiegelt das subjektive Erleben des Website-Inhalts insgesamt wider (siehe Abbildung 1).

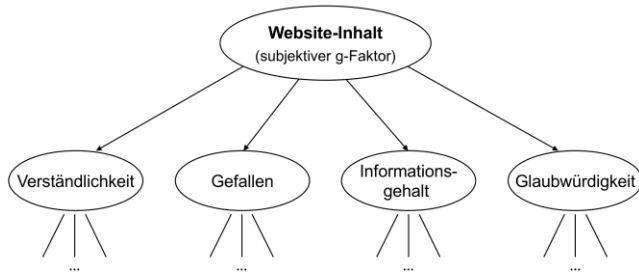


Abbildung 1: Strukturmodell des Web-CLICs (nach [18])

Konstruktion und Validierung des Web-CLICs basieren auf sechs Studien mit insgesamt $n = 3106$ Befragten und $m = 60$ getesteten Websites [18]. Der Fragebogen erweist sich als reliabel: Die interne Konsistenz (Cronbachs α) ist mindestens $> .80$ für die Skalen und $> .90$ für den Gesamtwert. Die Zeitstabilität (Retest-Reliabilität) über einen Zeitraum von 2 Wochen liegt für die Skalen im Bereich von $.69 \leq r \leq .81$, bzw. bei $r = .84$ für den Gesamtwert. Hinsichtlich der Validität findet sich eine umfassende empirische Evidenz, geprüft wurden faktorielle, konvergente, divergente, diskriminative, konkurrenente, experimentelle und prädiktive Validität. In den Studien der Autoren erfasste der Web-CLIC beispielsweise spezifisch und zielgenau experimentelle Variationen der Glaubwürdigkeit einer Gesundheitsinformationswebsite und war darüber hinaus in der Lage das Spendenverhalten der Befragten für gemeinnützige Organisationen vorherzusagen. Als weitere Auswertungshilfen liegen für den Web-CLIC optimale Schwellenwerte und Benchmarks für zehn verschiedene Website-Kategorien auf Basis von 7379 Bewertungen von $m = 120$ Websites vor [18].

Der Web-CLIC ist ein nützliches Instrument für eine detaillierte und umfassende Evaluierung, ist aber weniger geeignet für Situationen, in denen die Erhebungszeit auf ein absolutes Minimum beschränkt werden muss oder in denen nur eine Gesamtbewertung von Web-Inhalten erforderlich ist und nicht eine detaillierte Bewertung verschiedener Inhalts-Facetten notwendig ist. Dies ist besonders dann der Fall, wenn Websites im Sinne eines Monitorings kontinuierlich evaluiert werden oder wenn man eine ganze Reihe von Konstrukten gleichzeitig erhoben werden sollen. Nicht nur in der Praxis ist die Menge der Fragen in User- und Kundenbefragungen in ein höchst kritisches Thema. Auch in der Wissenschaft untersuchen Forscher*innen oftmals schwerpunktmäßig andere Aspekte von Websites, möchten aber vielleicht trotzdem beurteilen, wie die Nutzer*innen den Inhalt im Allgemeinen wahrnehmen. Beispielsweise würden Studien, die sich auf die Wirksamkeit von Online-Gesundheitsinterventionen konzentrieren, von einer kurzen aber prägnanten Messung der Inhalte profitieren, zumal die Wahrnehmung der Inhalte durch den Gesundheitszustand der Nutzer*innen beeinflusst werden kann (z.B. [1], [21]). Darüber hinaus gibt es Situationen, in denen nur ein kurzer Manipulationscheck erforderlich ist, z.B. wenn

andere Aspekte der Website manipuliert werden, aber sichergestellt sein soll, dass die Inhaltswahrnehmung nicht beeinträchtigt wurde.

Ziel der vorliegenden Studien (vgl. [20]) war es daher, eine Kurzversion des Web-CLIC zu entwickeln und zu validieren sowie entsprechende Interpretationshilfen für die Skala zur Verfügung zu stellen.

2 Konstruktion und Erprobung des Web-CLIC-S

Konstruktion und Validierung des Web-CLIC-S basieren auf drei Studien mit insgesamt $n = 1.414$ Befragten und $m = 33$ getesteten Websites. Die Studien wurden von der Ethikkommission des Fachbereichs 7 der Universität Münster genehmigt (ID 2018-13-MT). Zur Auswahl der Website-Stimuli für Studie 1 und 2 wurde eine Vorstudie durchgeführt. Ziel war ein Website-Set, das den Studienteilnehmer*innen weitgehend unbekannt war, aber dennoch die typische Bandbreite der allgemeinen Website-Inhaltsqualität widerspiegelt. Hierbei folgten wir methodisch Miniukowitsch und De Angeli [10] und nutzten einen Crowdworking-Ansatz. Dies führte zu einem endgültigen Pool von 764 Websites, aus diesem wurden die Testwebsites zufällig ausgewählt. Die Daten der Studien (einschließlich vollständiger Itemlisten in Codebüchern und der URLs aller Websites aus der Vorstudie) sind unter <https://doi.org/10.5281/zenodo.3813293> [19] frei verfügbar.

2.1 Phase 1: Schaffung der Kurzversion und konfirmatorische Faktorenanalyse

Die Schaffung der Web-CLIC Kurzversion erfolgte anhand verschiedener Kriterien: Zunächst sollte jeder im Web-CLIC enthaltene Inhaltsbereich (siehe Abb. 1) durch mindestens ein Item abgebildet werden, dabei sollten die Items so repräsentativ wie möglich für den zugehörigen Inhaltsbereich sein. Zudem sollte die resultierende Skala so kurz wie möglich sein. Dies führte zu folgenden vier Items (für Details zum Vorgehen siehe [20]):

- Die Inhalte sind anschaulich aufbereitet.
- Ich lese diese Website gerne.
- Die Website ist informativ.
- Ich kann den Informationen auf der Website vertrauen.

Mittels einer konfirmatorischen Faktorenanalyse ($n = 764$ Befragte, $m = 10$ getestete Websites) konnte gezeigt werden, dass diese den allgemeinen g-Faktor „subjektive Inhaltswahrnehmung“ eindimensional abbilden.

2.2 Phase 2: Prüfung der Reliabilität

Auf Basis der ersten Studie erfolgte eine Prüfung von Reliabilität und Retest-Reliabilität (Zeitraum: 14 Tage, $n = 551$). Dabei erweist sich der Web-CLIC-S als reliabel: Die interne Konsistenz (Cronbachs α) betrug $.85$ zum ersten Messzeitpunkt, sowie $.88$ zum zweiten Messzeitpunkt nach 14 Tagen. Die Zeitstabilität (Retest-

Reliabilität) über diesen Zeitraum von zwei Wochen liegt bei $r = 79$. Der Web-CLIC-S erweist sich damit als konsistent und stabil über kurze bis mittlere Zeiträume, die Reliabilität der Kurzversion liegt nur geringfügig unter der der Vollversion.

2.3 Phase 3: Prüfung der Validität

Hinsichtlich der Validität des Web-CLIC findet sich eine umfassende empirische Evidenz. Geprüft wurden konvergente, divergente, diskriminative, konkurrenente, und prädiktive Validität (siehe Studien 1 bis 3 in [20]). Dabei stechen besonders hohe Korrelationen zu verwandten Konstrukten und Gesamteindrücken, wie beispielsweise der Gesamtnote einer Website, der Wiederbesuchsbereitschaft oder das Vertrauen in den Websitebetreiber, heraus. Die verschiedenen Validitätskoeffizienten sind dabei mit denen der Vollversion vergleichbar und übertreffen diese sogar vereinzelt.

Zudem wurden hohe Korrelationen zwischen dem Web-CLIC-S und dem vollständigen Web-CLIC gefunden ($.949 \leq r \leq .972$, alle $p < .001$). Der Web-CLIC-S kann also sinnvoll als guter Proxy für den Gesamtwert des Web-CLIC dienen (eine Analyse spezifischer Inhaltsfacetten erfordert aber den Einsatz der Web-CLIC Vollversion).

Weiterhin zeigt Studie 3 [20] die Fähigkeit des Web-CLIC-S Entscheidungen von Webnutzer*innen vorherzusagen: untersucht wurde eine tatsächliche Spendenentscheidung für eine gemeinnützige Organisation. In der Vorhersage dieser Entscheidungen ist der Web-CLIC-S verschiedenen anderen Instrumenten zur Beurteilung von Usability wie UMUX-Lite [8], oder PWU ([3]; deutsche Version siehe [15], [17]), und auch dem VisAWI-S zur Erfassung der subjektiven Ästhetik [11] signifikant überlegen.

Zudem zeigt sich in weiteren Analysen der vorliegenden Daten Studien 4 in [20]), dass der Web-CLIC-S wenig beeinflusst wird durch Alter, Geschlecht oder Bildungsgrad der Befragten.

3 Erstellung von Interpretationshilfen

Abschließend wurden Daten von $n = 12.568$ Bewertungen (für $m = 183$ Websites) zusammengestellt, welche die Basis für spezifische Benchmarks für zwölf verschiedene Website-Kategorien boten. Diese Kategorien reichten von Informationswebsites und E-Health über Unternehmenswebsites bis hin zu Kategorien wie E-Commerce, Online-Recruiting oder Webportale.

Benchmarks geben jedoch keine Auskunft über die Relevanz bestimmter Werte: Selbst wenn beispielsweise der Inhalt einer bestimmten Website überdurchschnittlich gut bewertet wird, bedeutet dies nicht unbedingt, dass die User mit der präsentierten Website zufrieden sind. Diesem Problem kann durch die Verwendung optimaler Schwellenwerte begegnet werden (siehe [5], [22]). Schwellenwerte bestehen aus kritischen Werten, die z.B. angeben, ab welchem Wert ein User eine Website als allgemein gut oder schlecht einstufen wird. In dieser Analyse zeigt sich

unter anderem, dass mit dem Web-CLIC-S mindestens ein Wert von 4,66 in einer Evaluation erreicht werden sollte damit eine Website eher positiv wahrgenommen wird. Dabei unterscheiden sich die Schwellenwerte für elf geprüfte Website-Kategorien: die Kategorie Entertainment hat mit 3,70 den kleinsten Schwellenwert, E-Health mit 4,90 den höchsten Wert (siehe Studie 4 in [20]). Fragebogen, Benchmarks und weitere Interpretationshilfen (z.B. eine Excel-Auswertungsdatei) sind via www.WebCLIC.de frei verfügbar.

4 Anwendungshinweise

Die Kurzversion des Web-CLIC wurde sorgfältig geprüft und eigens validiert, ein zentraler Schritt für die Schaffung einer Kurzska (vgl. [14]). Dennoch muss einschränkend beachtet werden: Alle getesteten Teilnehmer hatten einen gemeinsamen kulturellen Hintergrund, da die Studien in deutscher Sprache durchgeführt wurden. Während in der Studie von Dames und Kollegen [2] der vollständige Web-CLIC in einer englischsprachigen Version erfolgreich angewendet wurde, fehlt für den Web-CLIC-S ein solcher Test. Kultur ist eine relevanter Einflussfaktor für die Wahrnehmung von Website-Inhalten (siehe [4]). Daher sollten zukünftige Studien kulturelle Effekte subjektiver Inhaltswahrnehmungen sowie mögliche Auswirkungen auf den Web-CLIC-S untersuchen und ebenso sind Übersetzungen und Validierungen weiterer Sprachversionen des Instruments sehr willkommen.

Der Web-CLIC-S eignet sich für die Befragung von Erwachsenen und Jugendlichen (älter als 14 Jahre), hinsichtlich der Anwendung in Studien mit Kindern haben wir keine Erfahrungswerte. Für eine Website-Evaluation empfehlen wir eine voll funktionsfähige Version der betreffenden Website zu testen und die User relevante Aufgaben ausführen zu lassen (z.B. frei Exploration oder Suchaufgaben, siehe [2]), um eine typische Nutzung zu simulieren. Grundsätzlich sollte der Wortlaut der Items nicht geändert werden, mit Ausnahme kleiner Anpassungen, um die Verständlichkeit und eine perfekte Anpassung an die Zielwebsite zu gewährleisten. Es sollte jedoch kein Item vollständig entfernt werden, da der Web-CLIC-S bereits sehr kurz ist, und jede weitere Verkürzung die psychometrische Qualität höchstwahrscheinlich beeinträchtigt wird.

Nachdem die Evaluationsstudie durchgeführt wurde, kann der Gesamtmittelwert des Web-CLIC-S berechnet werden, indem die Einzelwerte der einzelnen Items addiert und die resultierende Summe durch vier geteilt wird. Der Web-CLIC-S sollte nur auf der Ebene dieses Summenwertes und nicht auf der Ebene der einzelnen Items interpretiert werden. Wenn bestimmte Facetten des Website-Inhalts von Interesse sind, sollte der vollständige Web-CLIC oder die jeweilige Inhaltsskala (siehe [18]) verwendet werden.

Bei der Interpretation des Web-CLIC-S ist es wichtig, den subjektiven Charakter einer solchen Evaluation zu

berücksichtigen, da es sich nicht um ein objektives algorithmisches Maß für Website-Inhalte handelt. Ein hoher Wert auf der Skala bedeutet nicht zwingend, dass der Inhalt der Website besonders hochwertig ist – sondern, dass die Website-Nutzer*innen eine positive Wahrnehmung des Inhalts hatten.

5 Fazit

In umfangreichen psychometrischen Tests zeigte die verkürzte Version des Web-CLIC [18] eine hohe Reliabilität und Konstruktvalidität. Darüber hinaus konnte der Web-CLIC-S, wie auch die Vollversion, die Absichten und das Verhalten der User vorhersagen. Folglich können wir die Verwendung der Kurzsкала in Praxis und Forschung empfehlen. Es liegt derzeit eine deutsche sowie eine englischsprachige Version vor – weitere Sprachversionen sind wünschenswert.

Zusammenfassend lässt sich sagen, dass, wenn eine kurze Skala für die Bewertung von Web-Inhalten benötigt wird, der Web-CLIC-S ein hochwertiges Maß ist, das ein valides Screening der subjektiven Inhaltswahrnehmung der User ermöglicht, dass Einzelitems weit überlegen ist.

DANKSAGUNG

Wir danken Simon Eisbach für seine Unterstützung bei der Durchführung der Studien 1 und 2 und für die Bereitstellung eines JavaScripts zur Kontrolle der Aufgabenerfüllung. Wir danken zudem David Kahre für seine Hilfe bei Studie 3 und den Benchmarking-Studien. Schließlich sind die Autoren dankbar für die Gewährung einer Zuwendung zur Unterstützung der Stichprobenrekrutierung von Studie 3 und einer zusätzlichen Studie, die in die Benchmarking- und Cut-Point-Analyse einbezogen wurde: Diese Forschung wird von der Bundeszentrale für gesundheitliche Aufklärung (BZgA) im Auftrag des Bundesministeriums für Gesundheit gefördert.

LITERATUR

- [1] G. Bansal, F. M. Zahedi, & D. Gefen (2010). The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems*, 49(2), 138–150. doi:10.1016/j.dss.2010.01.010.
- [2] H. Dames, G. Hirschfeld, T. Sackmann, & M. T. Thielsch (2019). Searching vs. Browsing - The influence of consumers' goal directedness on website evaluations. *Interacting with Computers*, 31(1), 95-112. doi:10.1093/iwc/iwz006.
- [3] C. Flavián, M. Guinaliú, & R. Gurrea (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43(1), 1–14. doi:10.1016/j.im.2005.01.002.
- [4] R. Fletcher (2006). The impact of culture on web site content, design, and structure: An international and a multicultural perspective. *Journal of Communication Management*, 10(3), 259–273. doi:10.1108/13632540610681158.
- [5] G. Hirschfeld, & M. T. Thielsch (2015). Establishing meaningful cut points for online user ratings. *Ergonomics*, 58(2), 310-320. <http://dx.doi.org/10.1080/00140139.2014.965228>.
- [6] Y. Kang, & Y. Kim (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.
- [7] H. Kim, & L. S. Niehm (2009). The Impact of Website Quality on Information Quality, Value, and Loyalty Intentions in Apparel Retailing. *Journal of Interactive Marketing*, 23(3), 221–233. <http://doi.org/10.1016/j.intmar.2009.04.009>.
- [8] J. R. Lewis, B. S. Utesch, & D. E. Maher (2013, April). UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (S. 2099-2102). ACM.
- [9] S. Meeßen, M. T. Thielsch, D. Riehle, & G. Hertel (2020). Trust is Essential: Positive Effects of Information Systems on Users' Memory require Trust in the System. *Ergonomics*, 63(7), 909-926. <https://doi.org/10.1080/00140139.2020.1758797>.
- [10] A. Miniukovich, & A. De Angeli (2015, April). Computation of interface aesthetics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (S. 1163-1172).
- [11] M. Moshagen, & M. T. Thielsch (2013). A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology*, 32(12), 1305-1311. doi:10.1080/0144929X.2012.694910.
- [12] F. Rahimnia, & J. F. Hassanzadeh (2013). The impact of website content dimension and e-trust on e-marketing effectiveness: The case of Iranian commercial saffron corporations. *Information and Management*, 50(5), 240–247. <http://doi.org/10.1016/j.im.2013.04.003>.
- [13] A. Shukla, N. K. Sharma, & S. Swami (2010). Website characteristics, user characteristics and purchase intention: mediating role of website satisfaction. *International Journal of Internet Marketing and Advertising*, 6(2), 142. <http://doi.org/10.1504/IJIMA.2010.032479>.
- [14] G. T. Smith, D. M. McCarthy, & K. G. Anderson (2000). On the sins of short-form development. *Psychological Assessment*, 12(1), 102–111. doi:10.1037/1040-3590.12.1.102.
- [15] M. T. Thielsch (2008). *Ästhetik von Websites. Wahrnehmung von Ästhetik und deren Beziehung zu Inhalt, Usability und Persönlichkeitsmerkmalen*. Münster: MV Wissenschaft.
- [16] M. T. Thielsch, I. Blotenberg, & R. Jaron (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>.
- [17] M. T. Thielsch, R. Engel, & G. Hirschfeld (2015). Expected usability is not a valid indicator of experienced usability. *PeerJ Computer Science*, 1-19. doi:10.7717/peerj-cs.19.
- [18] M. T. Thielsch, & G. Hirschfeld (2019). Facets of website content. *Human-Computer Interaction*, 34(4), 279-327. <https://doi.org/10.1080/07370024.2017.1421954>.
- [19] M. T. Thielsch, & G. Hirschfeld (2020). Quick assessment of web content perceptions [Data set]. *Zenodo*. doi:10.5281/zenodo.3813293.
- [20] M. T. Thielsch, & G. Hirschfeld (in press). Quick assessment of web content perceptions. *International Journal of Human-Computer Interaction*.
- [21] M. T. Thielsch, & C. Thielsch (2018). Depressive symptoms and web user experience. *PeerJ*, 1-26. doi:10.7717/peerj.4439.
- [22] M. T. Thielsch, C. Thielsch, & G. Hirschfeld (2019). How informative is informative? Benchmarks and optimal cut points for E-Health Websites. *Mensch und Computer 2019 – Workshopband* (S. 448-452). Bonn: Gesellschaft für Informatik e.V. <https://doi.org/10.18420/muc2019-ws-642>.