

# INUIT: The Interface Usability Instrument

Maximilian Speicher<sup>1</sup>(✉), Andreas Both<sup>2</sup>, and Martin Gaedke<sup>1</sup>

<sup>1</sup> Technische Universität Chemnitz, 09111 Chemnitz, Germany

maximilian.speicher@s2013.tu-chemnitz.de,

martin.gaedke@informatik.tu-chemnitz.de

<sup>2</sup> Research and development, Unister GmbH, 04109 Leipzig, Germany

andreas.both@unister.de

**Abstract.** Explicit user testing tends to be costly and time-consuming from a company’s point of view. Therefore, it would be desirable to infer a quantitative usability score directly from implicit feedback, i.e., the interactions of users with a web interface. As a basis for this, we require an adequate usability instrument whose items form a usability score and can be meaningfully correlated with such interactions. Thus, we present INUIT, the first instrument consisting of only seven items that have the right level of abstraction to directly reflect user behavior on the client. It has been designed in a two-step process involving usability guideline reviews and expert interviews. A confirmatory factor analysis shows that our model reasonably well reflects real-world perceptions of usability.

**Keywords:** Instrument · Metrics · Questionnaire · Usability · Interfaces

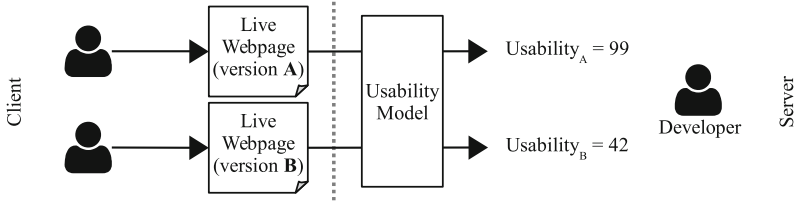
## 1 Introduction

The usability of a website is a crucial factor for ensuring customer satisfaction and loyalty [23]. However, adequate usability testing is often neglected in today’s e-commerce industry due to costliness and time consumption. Particularly user testing happens less frequently because it is “heavily constrained by available time, money and human resources” [18]. Hence, stakeholders tend to partly sacrifice usability by requesting cheaper and more efficient methods of conversion maximization (e.g., in terms of clicks on advertisements), also potentially caused by the demand for a short time-to-market. To tackle this shortcoming we require a similarly *efficient* method that is more *effective* in measuring usability. A straightforward approach would be to make use of real users’ interactions with a web interface to infer knowledge about its usability. Optimally, such knowledge would be present in terms of a *key performance indicator* (i.e., a usability score) for easier communication with stakeholders who are not usability experts.

To be able to realize such a framework (Fig. 1), it is necessary to build upon an adequate usability instrument for providing a quantitative measure that combines ratings of the contained items. A corresponding formula for such a measure

---

M. Speicher—The contents of this paper were developed while Mr. Speicher stayed at Unister GmbH as an industrial PhD student. An earlier version has been published as [24].



**Fig. 1.** A model providing a quantitative metric of usability [24].

could be  $usability = -(confusion + distraction)$ . As usability is a latent variable, we need to define factors thereof that can be meaningfully inferred from interactions, e.g., *faster and more unstructured cursor movements indicate user confusion*  $\Rightarrow$  confusion = 1. Numerous instruments for determining usability have been developed (e.g., [5, 8, 10, 22]), but none has been specifically designed for providing a key performance indicator for usability that can be directly inferred from user interactions.

Thus, we propose INUIT—a new usability instrument for web interfaces consisting of *only seven items* that have the *right level of abstraction* to directly reflect users’ client-side interactions. The items have been determined in a two-step process. First, we have reviewed more than 250 usability rules from which we created a structure of usability based on ISO 9241-11 [14]. Second, we conducted semi-structured expert interviews with nine experts working in the e-commerce industry. Based on a user study with 81 participants, results of a confirmatory factor analysis show that INUIT’s underlying model is a good approximation of real-world perceptions of usability.

In the following, we give an overview of important background concepts and related work (Sect. 2). After that, we explain the design of our new usability instrument (Sect. 3). Section 4 presents the set-up and results of the evaluation of INUIT. In Sect. 5 we discuss results and limitations before giving concluding remarks.

## 2 Background and Related Work

**Web Interfaces.** Low-level user interactions on the *client-side* can be tracked on a per-*webpage* basis, i.e., for an HTML document delivered by a server and displayed in a web browser. Such interactions are commonly collected using Ajax technology and are valid only for the given document. Due to the stateless nature of HTTP<sup>1</sup>, they are difficult to track and put into context across multiple webpages. Contrary, user interactions in the context of a whole *website* (i.e., a set of interconnected, related webpages) are of a higher-level nature, such as navigation paths between webpages. They are usually mined from *server-side* logs.

<sup>1</sup> <http://www.w3.org/Protocols/>, retrieved June 11, 2014.

Thus, in the remainder of this article, we consider a WEB INTERFACE to be a single webpage. Particularly, this includes the HTML document's content and structure as determined within the <body> tag, and the appearance during a user's interaction with the webpage as determined by stylesheets and dynamic scripts that alter the DOM tree<sup>2</sup>.

**Usability.** In [5], Brooke states that “Usability does not exist in any absolute sense; it can only be defined with reference to particular contexts”. Thus, it is necessary that we clarify our understanding of usability in the context of our proposed approach. Orienting at ISO 25010 [13], the *internal usability* of a web application is measured in terms of static attributes (not connected to software execution); *external usability* relates to the behavior of the web application; and *usability in use* is relevant in case the web application involves real users under certain conditions. Therefore, given the fact that we intend to infer usability from real users' interactions, *usability in use* is the core concept we focus on. In accordance with this, [12] uses the notions of “do-goals” (e.g., booking a flight) and “be-goals” (e.g., being special) to distinguish between the pragmatic and hedonic dimensions of user experience, a concept that has a large intersection with usability. Particularly, he states that “*Pragmatic quality* refers to the product's perceived ability to support the achievement of ‘do-goals’ [and] calls for a focus on the product – its utility and usability” [12]. Since a user's interactions with an interface are a direct reflection of what they *do*, for our purpose the *pragmatic* dimension of usability is of particular interest.

Based on the above, in the remainder of this article USABILITY refers to the pragmatic [12] and in-use [13] dimensions of the definition given by ISO 9241-11 [14]. Internal/external usability [13] and the hedonic dimension (“the product's perceived ability to support the achievement of ‘be-goals’” [12]) of usability in use are neglected.

**Definition 1.** USABILITY: *The extent to which a web interface can be used by real users to achieve do-goals with effectiveness, efficiency and satisfaction in a specified context of use (adjusted definition by [14]).*

**Instruments for Determining Usability and Related Concepts.** Reference [22] has investigated metrics for usability, design and performance of a website. His finding is that the success of a website is a first-order construct and particularly connected to measures such as download time, navigation, interactivity and responsiveness (e.g., feedback options). The data used for analysis was collected from 1997 thru 2000, which indicates that the methods for website evaluation might be out-of-date regarding the radical changes in website appearance and thus also in the perception of usability. In particular, measures such as the download time should be less of an issue nowadays (except for slow mobile connections).

<sup>2</sup> <<http://www.w3.org/DOM/>>, retrieved June 11, 2014.

Reference [8] describe a usability instrument that is specifically aimed at websites of small businesses. They evaluated the instrument in the specific case of website navigation and found that navigation impacts ease of use and user return rates, among others. The used questionnaire (i.e., the instrument) features some factors of usability that we have identified for INUIT as well. However, it is rather elaborate and thus potentially not adequate for evaluation of online web interfaces by real users. Moreover, we do not want to focus on a specific type of website—such as small businesses—but instead provide a general instrument.

Reference [10] developed a website usability instrument based on the definition given by ISO 9241-11 [14]. They have chosen five dimensions of usability: effectiveness, efficiency, level of engagement, error tolerance, and ease of learning. Along with these comes a total of 17 items to assess the dimensions. A factor analysis showed no significant difference between their usability instrument and a set of test data. However, like the above approach [8], the instrument seems to be specifically focused on e-commerce websites. In particular, they found that, e.g., error tolerance is a significant indicator for the intention to perform a transaction and that efficacy predicts the intention of further visits.

*AttrakDiff*<sup>3</sup> measures the hedonic and pragmatic user experience [12] of an e-commerce product based on a dedicated instrument. *UEQ*<sup>4</sup> follows a similar approach based on an instrument containing 26 bipolar items. In contrast to INUIT, both of these are oriented towards measuring the *user experience* of a software product as a whole. More similar to our instrument is the *System Usability Scale* (SUS) [5], which measures the usability of arbitrary interfaces by posing ten questions based on a 5-point Likert scale. The answers are then summed up and normalized to a score between 0 and 100.

There are also numerous instruments in the form of usability checklists, which can be used in terms of spreadsheets that automatically calculate usability scores (e.g., [11, 27]). However, such checklists usually contain huge amounts of items that are also very abstract in parts. They are therefore aimed at supporting inspections by experts (cf. [19]) rather than having them answered by users.

The ISO definition of usability [14] states that satisfaction is a major aspect of usability. Reference [1] present a revalidation of the well-studied End-User Computing Satisfaction Instrument (EUCS), which is an instrument for this particular aspect. While certain items of EUCS clearly intersect with those of usability instruments—e.g., in the dimension “Ease of Use”—it is clearly pointed out that EUCS specifically measures satisfaction rather than usability.

Another aspect that is closely related to usability but not mentioned in the ISO definition is the aesthetic appearance of a web interface. Reference [15] present an instrument for the concept and state that aesthetics cannot be neglected in the context of effective interaction design. The instrument is clearly focused on very subjective aspects of design and layout and shows less intersections with existing usability instruments than EUCS.

<sup>3</sup> <http://attrakdiff.de/>, retrieved July 29, 2014.

<sup>4</sup> <http://www.ueq-online.org/>, retrieved July 29, 2014.

### 3 INUIT: The *Interface Usability Instrument*

The aim of INUIT is to provide a usability instrument that is adequate for the novel concept of Usability-based Split Testing [25]. Particularly, it must be possible to meaningfully infer ratings of its contained items from client-side user interactions (e.g., unstructured cursor movements  $\Rightarrow$  confusion = 1). Also, the instrument must be consistent with Definition 1 above. All of this poses the following requirements:

- (R1). The instrument’s number of items is kept to a minimum, so that real users asked for explicit usability judgments through a corresponding questionnaire are not deterred. This helps with collecting high-quality training data.
- (R2). The contained items have the right level of abstraction, so that they can be meaningfully mapped to client-side user interactions. For example, “ease of use” is a higher-level concept that can be split into several sub-concepts while “all links should have blue color” is clearly too specific. Contrary, an item like “user confusion” can be mapped to interactions such as unstructured cursor movements.
- (R3). The contained items can be applied to a web interface as defined earlier.

Regarding these requirements, existing instruments lack meeting one or more thereof. Instruments such as those described by [5,8,10,22] feature items with a wrong level of abstraction (R2) or that cannot be applied to standalone web interfaces (R3). Similar problems arise with questionnaires like *AttrakDiff* and *UEQ* (R2, R3). Finally, usability checklists (e.g., [11,27]) usually contain huge amounts of items and therefore violate R1.

To meet the above requirements, the items contained in INUIT have been determined in a two-step process. First, we have carried out a review of popular and well-known usability guidelines that contained over 250 rules for good usability in the form of heuristics and checklists. After we eliminated all rules not consistent with the requirements above, a set of underlying factors of usability has been extracted. That is, we grouped together rules that were different expressions of the same (higher-level) factor. From these underlying factors, we have derived a structure of usability based on ISO 9241-11 [14]. Second, we asked experts for driving factors of web interface usability from their point of view and revised our usability structure accordingly.

#### 3.1 Guideline Reviews

As the first step of determining the items of INUIT, we have reviewed a set of six well-known resources concerned with usability [7,9,17,20,26,27]. They were chosen based on the commonly accepted expertise of their authors and contain guidelines by *A List Apart*<sup>5</sup> and Bruce Tognazzini (author of the first *Apple Human Interface Guidelines*), among others. The investigated heuristics and

<sup>5</sup> (<http://alistapart.com/>), retrieved June 11, 2014.

**Table 1.** Set of items derived from usability guideline reviews

Usability factor	# related rules
Aesthetic appearance	8
Amount of distraction	6
Information density	6
Informativeness	6
Reachability of desired content <sup>a</sup>	4
Readability	5
Understandability	6

<sup>a</sup> With respect to Fitt’s Law, i.e., “The time to acquire a target is a function of the distance to and size of the target” [26].

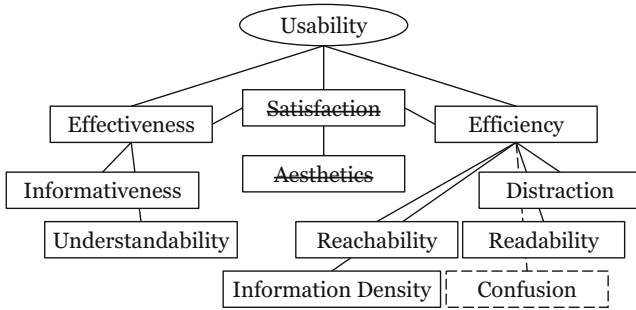
checklists contained a total of over 250 rules for good usability. In accordance with requirements R2 and R3 above, we eliminated all rules that:

- were too abstract, such as “Flexibility and efficiency of use” [20];
- were too specific, such as “Blue Is The Best Color For Links” [7];
- would not make sense when applied to a web interface in terms of a single web-page, e.g., “Because many of our browser-based products exist in a stateless environment, we have the responsibility to track state as needed” [26].

The elimination process left a total of 32 remaining rules, from which we extracted the driving factors of usability. Starting from ISO 9241-11 [14] and Definition 1, one can roughly state that the concept of *usability* features the three dimensions effectiveness, efficiency and satisfaction. Our goal was to find those factors that are one level of abstraction below these main dimensions and manifest themselves in multiple more specific usability rules. Thus, we investigated which of the remaining rules were different expressions of the same underlying principle and extracted the intended factors from these. To give just one example, “The site avoids advertisements, especially pop-ups” [27] and “Attention-attracting features [...] are used sparingly and only where relevant” [27] are expressions of the same underlying principle *distraction*, which is a driving factor of web interface usability. Moreover, distraction is to a high degree disjoint from other factors of usability at the same level of abstraction, e.g., it is different from the factor *confusion*. To complete the given example, distraction can be situated as follows regarding its relative level of abstraction (higher level of abstraction to the right): presence of advertisements → distraction → efficiency → usability.

From the remaining rules, we extracted the underlying factors of usability as shown in Table 1 (more than one related factor per rule was possible). Originally, the factor “reachability” was named “accessibility”. To prevent confusion with what is commonly understood by accessibility<sup>6</sup>, the factor was renamed later on. What we understand by “reachability” is how difficult it is for the user to

<sup>6</sup> <http://www.w3.org/TR/WCAG20/>, retrieved June 12, 2014.



**Fig. 2.** Structure of usability derived from the guideline reviews. Struck through factors were removed, factors in dashed boxes were added after the expert interviews.

find their desired content within a web interface w.r.t. the temporal and spatial distance from the initial viewport.

Using the seven factors from Table 1, we could describe all of the relevant usability rules extracted from the reviewed guidelines. Subsequently, based on the definition given by ISO 9241-11 [14] and own experience with usability evaluations, we constructed a structure of usability as shown in Fig. 2.

### 3.2 Expert Interviews

As the second step of determining the items of INUIT, we conducted semi-structured interviews with nine experts working in the e-commerce industry. The experts were particularly concerned with front-end design and/or usability testing. First, we presented them with the definition of usability given by ISO 9241-11 [14] (Fig. 3, bottom left). Based on this, we asked them to name—from their point of view—driving factors of web interface usability with the intended level of abstraction from requirement R2 in mind. That is, showing positive and negative examples on the web, they should indicate factors that potentially directly affect patterns of user interaction. All statements were recorded accordingly (Fig. 3, bottom right).

Second, we presented the experts with a pen and a sheet of paper showing the above structure of usability (Fig. 2) and asked them to modify it in such a way that it reflected their perception of usability (Fig. 3, top middle).

After the interview, the experts were asked to answer additional demographic questions (Fig. 3, top right). On average, they stated that they are *knowledgeable* ( $m=3$ ) in front-end design, interaction design and usability/UX (4-point scale, 1 = *no knowledge*, 4 = *expert*). Moreover, they indicated *passing knowledge* ( $m=2$ ) in web engineering. Two experts said they have a research background, three indicated a practitioner background and four stated that they cannot exactly tell or have both. The average age of the interviewees was 30.44 years ( $\sigma = 2.96$ ; 2 female).

Based on the interview transcripts, we mapped the usability factors identified by the experts to the seven factors shown in Table 1. The experts mentioned all of

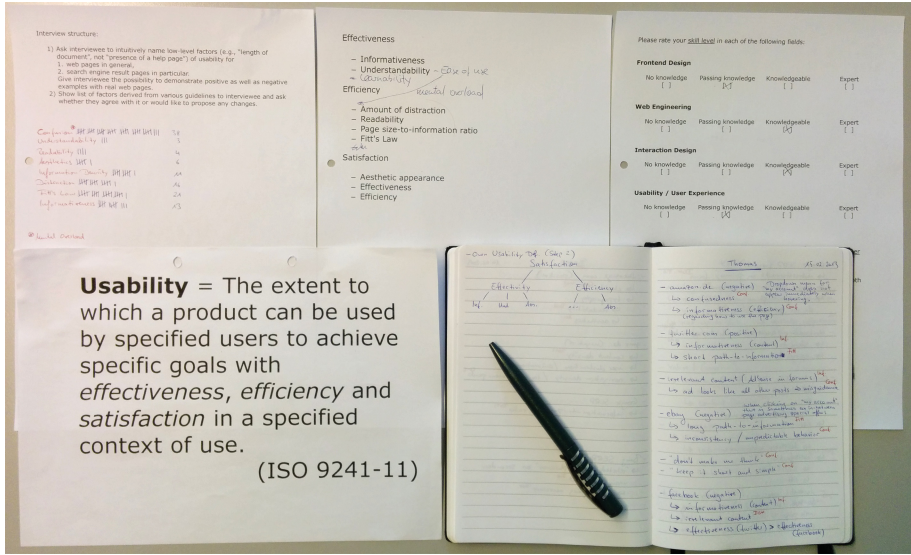


Fig. 3. Set-up of the expert interviews.

these factors multiple times, but a total of 38 statements remained that did not fit into the existing set. Rather, all of these remaining statements were expressions of an additional underlying concept *mental overload* or *user confusion*. During the second part of the interview, the experts made the following general statements:

- Aesthetic appearance goes hand in hand with both effectiveness and efficiency. Thus, it cannot be considered separate from these. Rather, the item “aesthetics” should be a sub-factor of both effectiveness and efficiency.
- An additional factor “ease of use” / “mental overload” / “user confusion” should be added as a sub-factor of efficiency since this concept is not fully reflected by the existing items.
- “Fun” should be added as a sub-factor of effectiveness or a separate higher-level factor “emotional attachment”.

Apart from this, the experts generally agreed with the structure of usability that was given as a starting point (Fig. 2).

### 3.3 Items of INUIT

Based on the findings from the interviews and careful review of existing research [1, 15], we revised the structure of usability given in Fig. 2. That is, we added *user confusion* as a sub-factor of efficiency. Also, following requirement R2, we cleaned up the construct by not considering any potential factors that are higher-level latent variables themselves (i.e., satisfaction, aesthetics, emotional attachment, fun) and cannot be directly mapped to user interactions in a meaningful way.



Particularly, removing *satisfaction* as a dimension of usability is in accordance with [16], thus altering Definition 1 as originally given in Sect. 2. Taking the resulting factors, we subsequently formulated corresponding questions to form the intended usability instrument as given in Table 2.

**Table 2.** INUIT the interface usability instrument

Usability factor	Dimension	Question
Informativeness	Effectiveness	Did you find the content you were looking for?
Understandability	Effectiveness	Could you easily understand the provided content?
Confusion	Efficiency	Were you confused while using the webpage?
Distraction	Efficiency	Were you distracted by elements of the webpage?
Readability	Efficiency	Did typography and layout add to readability?
Information density	Efficiency	Was there too much information presented on too little space?
Reachability	Efficiency	Was your desired content easily and quickly reachable (concerning time and distance)?

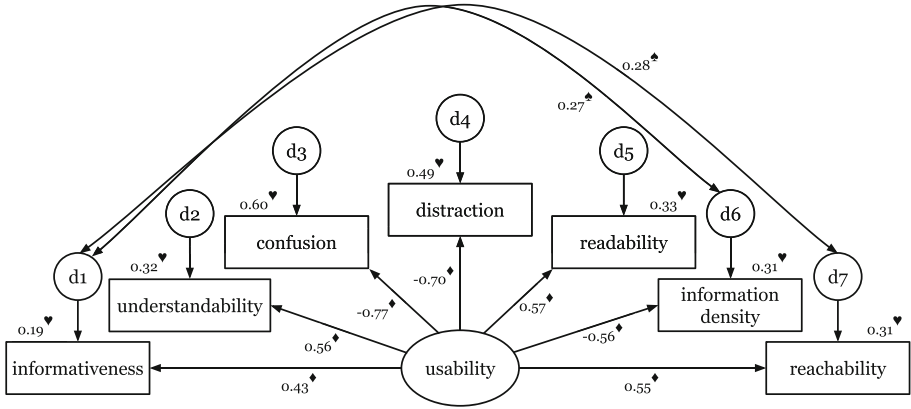
The overall usability metric of INUIT can now be formed either by directly summing up all items or by equally weighting the dimensions *effectiveness* and *efficiency*.

## 4 Evaluation

To evaluate the new usability instrument, we have conducted a confirmatory factor analysis [2, 6] with a model in which all of the seven items directly load on the latent variable *usability*.

**Method.** The data for evaluation were obtained in a user study with 81 participants recruited via Twitter, Facebook and company-internal mailing lists. Each participant was randomly presented with one of four online news articles about the Higgs boson [3] (CERN, CNN, Yahoo! News, Scientific American) and asked to find a particular piece of information within the content of the web interface<sup>7</sup>. Two of the articles did not contain the desired information (Yahoo! News, Scientific American). Having found the piece of information or being absolutely sure the article does not contain it, the participant had to indicate they finished the task. Subsequently, they were presented with a questionnaire containing the items from Table 2 and some demographic questions. As a first simple approach,

<sup>7</sup> We intended to choose a topic an average user would most probably not be familiar with to ensure equality among the participants.



**Fig. 4.** Model with standardized estimates (correlations♠, squared multiple correlations♥, regression weights♦)

the INUIT questions could only be answered with “yes” or “no” (i.e., the overall usability score has a value between 0 and 7) rather than providing a Likert scale or similar. We believe this is reasonable since it reduces the user’s perceived amount of work, which might increase the willingness to give answers in a real-world setting. It was possible to take part a maximum of four times in the study, being presented a different article each time.

To make the evaluated model more realistic, we introduced covariances between the residual errors of informativeness and information density as well as between the residual errors of informativeness and reachability. This is a valid approach [2, 6] and in this case theoretically grounded since users who cannot find their desired content due to a high information density or bad reachability will probably (incorrectly) indicate a bad informativeness and vice versa.

**Results.** Of the 81 non-unique study participants, 66 were male (15 female) at an average age of 28.43 ( $\sigma = 2.37$ ). Only two of them indicated that they were familiar with the news website the presented article was taken from.

Using IBM SPSS Amos 20 [2], we performed the confirmatory factor analysis as described above. Our results (Fig. 4) suggest that the model used is a reasonably good fit to the data set, with  $\chi^2 = 15.817$  ( $df = 12$ ,  $p = 0.2$ ), a comparative fit index (CFI) of 0.971 and a root mean square error of approximation (RMSEA)<sup>8</sup> of 0.063.

**Demo.** For the complete set-up of the study and reproducing the confirmatory factor analysis, please visit <http://vsr.informatik.tu-chemnitz.de/demo/inuit>.

<sup>8</sup> According to [2], an RMSEA value of 0.08 or less is “a reasonable error of approximation”. For detailed descriptions of the measures of fit and their shortcomings, the interested reader may refer to [2].

## 5 Discussion and Conclusions

We have introduced INUIT—a novel usability instrument consisting of only seven items that has been specifically designed for meaningful correlation of its items with client-side user interactions. A corresponding CFA has been carried out based on a user study with 81 test subjects. It indicates that our instrument can reasonably well describe real-world perceptions of usability. As such, it paves the way for providing models that make it possible to infer a web interface’s usability score from user interactions alone. In fact, INUIT has already been applied in an industrial case study [25] during which we were able to directly relate interactions to usability factors, e.g., *less confusion is indicated by a lower scrolling distance from top* (Pearson’s  $r = -0.44$ ) and *better reachability is indicated by fewer changes in scrolling direction* ( $-0.31$ ).

Yet, we are aware of the fact that INUIT has several limitations. First, complex concepts like satisfaction and aesthetics have been removed from our set of items to keep the instrument simple according to the posed requirements. Particularly, INUIT can only measure the specific type of usability described in Sect. 2, which is a rather pragmatic interpretation of the concept leaving out potential hedonic qualities (cf. [12]). Second, usability itself is a difficult-to-grasp concept that cannot be forced into a structure consisting of yes/no questions in its entirety. Therefore, the mapping between our model of usability and the real world should be investigated with additional scales comprising more than two points (e.g., a Likert scale). Third, for the CFA performed we have chosen a set-up in which all factors directly load on the latent variable usability. Yet, it would be desirable to also explore set-ups in which, e.g., the factors load on the two dimensions *effectiveness* and *efficiency*, which then again load on the latent variable with equal weight. This could unveil models that even better describe real-world perceptions of usability than the one described above.

In accordance with the above, future work includes the investigation of INUIT based on different scales as well as CFAs with different set-ups. In fact, the instrument has already been applied in a separate user study [25] based on a three-point scale. The gathered data will be prepared to further investigate INUIT as intended and to confirm the good results of our CFA described in Sect. 4.

**Acknowledgements.** We thank our interviewees and all participants of the Unister Friday PhD Symposia. This work has been supported by the ESF and the Free State of Saxony.



Europa fördert Sachsen.  
**ESF**  
 Europäischer Sozialfonds

## References

1. Abdinnour-Helm, S.F., Chaparro, B.S., Farmer, S.M.: Using the End-User Computing Satisfaction (EUCS) Instrument to Measure Satisfaction with a Web Site. *Decision Sci* 36(2) (2005)
2. Arbuckle, J.L.: IBM® SPSS® Amos™ 20 User’s Guide. IBM Corporation, Armonk, NY (2011)

3. ATLAS Collaboration: Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys Lett B* 716(1) (2012)
4. Atterer, R., Wnuk, M., Schmidt, A.: Knowing the User's Every Move - User Activity Tracking for Website Usability Evaluation and Implicit Interaction. In: *Proc. WWW.* (2006)
5. Brooke, J.: SUS: A "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) *Usability Evaluation in Industry.* Taylor and Francis (1996)
6. Byrne, B.M.: *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming.* CRC Press (2009)
7. Fadeyev, D.: 10 Useful Usability Findings and Guidelines, riptsize <http://www.smashingmagazine.com/2009/09/24/10-useful-usability-findings-and-guidelines/>
8. Fisher, J., Bentley, J., Turner, R., Craig, A.: A usability instrument for evaluating websites - navigation elements. In: *Proc. OZCHI.* (2004)
9. Goldstein, D.: Beyond Usability Testing, <http://alistapart.com/article/beyond-usability-testing>
10. Green, D., Pearson, J.M.: Development of a Website Usability Instrument based on ISO 9241-11. *JCIS* 47(1) (2006)
11. Harms, I., Schweibenz, W., Strobel, J.: Usability Evaluation von Web-Angeboten mit dem Web Usability Index [Usability evaluation of web applications using the Web Usability Index]. In: *Proc. 24. DGI-Online-Tagung.* (2002)
12. Hassenzahl, M.: User Experience (UX): Towards an experiential perspective on product quality. In: *Proc. IHM.* (2008)
13. ISO: ISO/IEC 25010:2011 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuARE) - System and software quality models. (2011)
14. ISO: ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. (1998)
15. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. *Int J Hum-Comput St* 60(3) (2004)
16. Lew, P., Olsina, L., Zhang, L.: Quality, Quality in Use, Actual Usability and User Experience as Key Drivers for Web Application Evaluation. In: Benatallah, B., Casati, F., Kappel, G., Rossi, G. (eds.) *ICWE 2010. LNCS*, vol. 6189, pp. 218-232. Springer, Heidelberg (2010)
17. Mandel, T.: *The Elements of User Interface Design.* John Wiley & Sons, Hoboken, NJ (1997)
18. Nebeling, M., Speicher, M., Norrie, M.C.: CrowdStudy: General Toolkit for Crowdsourced Evaluation of Web Interfaces. In: *Proc. EICS.* (2013)
19. Nielsen, J., Molich, R.: Heuristic Evaluation of User Interfaces. In: *Proc. CHI.* (1990)
20. Nielsen, J.: 10 Usability Heuristics for User Interface Design, <http://www.nngroup.com/articles/ten-usability-heuristics/>
21. Nielsen, J.: Putting A/B Testing in Its Place, <http://www.nngroup.com/articles/putting-ab-testing-in-its-place/>
22. Palmer, J.W.: Website Usability, Design, and Performance Metrics. *Inform Syst Res* 13(2) (2002)
23. Sauro, J.: Does Better Usability Increase Customer Loyalty? <http://www.measuringusability.com/usability-loyalty.php>
24. Speicher, M., Both, A., Gaedke, M.: Towards Metric-based Usability Evaluation of Online Web Interfaces. In: *Mensch & Computer Workshopband.* (2013)

25. Speicher, M., Both, A., Gaedke, M.: Ensuring Web Interface Quality through Usability-Based Split Testing. In: Casteleyn, S., Rossi, G., Winckler, M. (eds.) ICWE 2014. LNCS, vol. 8541, pp. 93–110. Springer, Heidelberg (2014)
26. Tognazzini, B.: First Principles of Interaction Design, <http://www.asktog.com/basics/firstPrinciples.html> (accessed Mar 22, 2013)
27. Travis, D.: 247 web usability guidelines, <http://www.userfocus.co.uk/resources/guidelines.html>