Usability Testberichte gebrauchstauglicher machen

Usability Testberichte aus der CUE-9 Studie im Vergleich

Lisa Daske
Astrum IT GmbH
Am Wolfsmantel 2
91052 Erlangen
Deutschland
post@lisadaske.de

Abstract
Der Usability Test ist durchgeführt, und viele Usability Probleme wurden gefunden. Jetzt müssen diese nur noch dem Kunden übergeben werden – oft in Form eines schriftlichen Testberichts. Doch wie schreibt man eigentlich einen guten Usability Testbericht?
Im Rahmen der CUE-9 Studie bewerteten 35 erfahrene Usability Professionals aus Amerika und Deutschland unabhängig voneinander die gleichen Testvideos und erstellten schriftliche Berichte mit ihren Ergebnissen. Diese Berichte wurden in anonymisierter Form veröffentlicht und man kann viel aus ihnen lernen. Dieser Beitrag zeigt:
- Die „best practices" für den Aufbau eines Usability Berichts, in denen sich alle untersuchten Berichte ähneln
- Mittel, mit denen der Bericht lesbarer gestaltet werden kann, wie beispielsweise das Kennzeichnen verschiedener Ergebnisarten mit Icons
- Besondere Berichte, die sich von allen anderen unterscheiden

Keywords:
/// CUE-9
/// Usability Testing
/// Testberichte

1. Einleitung

Der Usability Test wird von manchen für die wichtigste aller Evaluierungsmethoden im Usability Umfeld gehalten und er wurde bereits öfter als Maßstab für andere Evaluierungsmethoden verwendet. Diese prominente Stellung des Usability Tests ist ein guter Grund, die Methode genau zu untersuchen, um ihre Stärken und Schwächen kennenzulernen und die Methode innerhalb ihrer Grenzen einzusetzen. Das ist die Motivation hinter Rolf Molichs Studienreihe CUE („Comparative Usability Evaluation", zu deutsch „Vergleichende Usability Evaluierung") (1).

Bei jeder CUE-Studie evaluieren Teams aus Usability Professionals unabhängig voneinander die selbe Webseite, Web-Anwendung oder Software. Die Studie CUE-9 ist bereits die neunte Studie in dieser Reihe. Sie wurde von Rolf Molich im Rahmen der „Mensch und Computer"-Konferenz in Chemnitz in 2011 mit 16 deutschen Usability Professionals sowie in Atlanta mit weiteren 19 Teilnehmern durchgeführt.

Der vorliegende Beitrag soll einen Einblick in die erstellten Testberichte geben und Gemeinsamkeiten sowie Besonderheiten vorstellen. Ziel des Beitrags ist es auch, Hinweise zu geben, mit welchen Mitteln man die Usability des Testberichts besser gestalten kann. Dabei stehen folgende Fragen im Mittelpunkt:
- Gibt es Übereinstimmungen zwischen den Testberichten, die auf einen „State of the Art", also eine Art Minimalanspruch bei der Erstellung von Usability Reports hinweisen?
- Welche Merkmale machen einen Testbericht besonders „usable"?
- Gibt es ungewöhnliche Berichte, die sich von den anderen unterscheiden?
- Diese Fragen werden anhand von Screenshots und Auszügen aus den Berichten beantwortet.

2. Die CUE-9 Studie

Der Fokus der CUE-9 war es, herauszufinden, ob ein „Evaluator Effect" existiert. Dieser besagt, dass auf der Grundlage derselben Daten mehrere Experten unterschiedliche Usability Probleme imselben System finden. Mehr Informationen zur CUE-Reihe sind auf der Webseite von DialogDesign zu finden (1).

Jeder der Teilnehmer erhielt im Vorfeld dieselben 5 Videos von Usability Test Sitzungen zur unabhängigen Auswertung. Grundlage der Studie war eine gemeinsame Evaluierung der Webseite der US-amerikanischen Firma U-Haul, www.u-haul.com, die LKWs an Privatpersonen zu Umzugszwecken vermietet. Den Studienteilnehmern wurden folgende Aufgaben gestellt:
- Die fünf Videos zu je 30 Minuten aus Usability Tests ansehen und auswerten
- Einen kurzen, anonymen Testbericht über das Ergebnis der Bewertung schreiben und einreichen
- Ähnliche Testberichte von anderen Studienteilnehmern lesen

Bei einem gemeinsamen Workshop trafen sich die Studienteilnehmer (getrennt in Deutschland und Amerika), um ihre Ergebnisse zu vergleichen und zu diskutieren. Diese Workshops fanden am 20. Juni 2011 in Atlanta, GA, USA statt (CUE-9a), sowie

am 11. September 2011 in Chemnitz, Deutschland (CUE-9b). In Atlanta nahmen dabei 19 Usability Experten teil, in Chemnitz waren es 16 Teilnehmer. In einem weiteren Beitrag von Rolf Molich und Lisa Daske im vorliegenden Tagungsband finden sich Details zu den Ergebnissen der Studie

Die im Rahmen der Studie erstellten Testberichte wurden veröffentlicht und können online eingesehen werden (2)(1). Für den vorliegenden Beitrag wurden diese Testberichte verglichen, um nach Gemeinsamkeiten und Auffälligkeiten zu suchen. Der Fokus ist dabei weniger eine wissenschaftliche Auswertung. Vielmehr wurde die seltene Chance genutzt, dass als „Nebenprodukt" der Studie vergleichbare Auswertungen von vielen Usability Professionals zur Verfügung stehen, in die ein Einblick sonst nicht ohne weiteres möglich wäre. Die Berichte wurden von der Autorin analysiert und verglichen und die interessantesten Aspekte werden im Weiteren vorgestellt.

3. Best practices

Dieser Abschnitt zeigt die gemeinsamen Eigenschaften eines Testberichts, die in allen oder den meisten Berichten vorhanden waren.

3.1. Executive Summary

Der Testbericht zu einem Usability-Test hat zwei Zielgruppen: einerseits dient der Bericht als Arbeitsergebnis für den Auftraggeber, andererseits als Arbeitsgrundlage für diejenigen Mitarbeiter, welche die gefundenen Probleme beseitigen sollen. Die erste Zielgruppe wird den Bericht typischerweise nicht vollständig lesen. Eine Executive Summary, also eine kurze Zusammenfassung der Testergebnisse am Anfang des Berichts, macht es dem Auftraggeber leicht, das Ergebnis des Tests zu erfassen.

Alle Testberichte in der Studie enthielten eine Executive Summary; allerdings war sie nicht in allen Fällen auf der ersten Seite des Berichts zu finden, sondern zum Teil hinter einleitenden Abschnitten zur

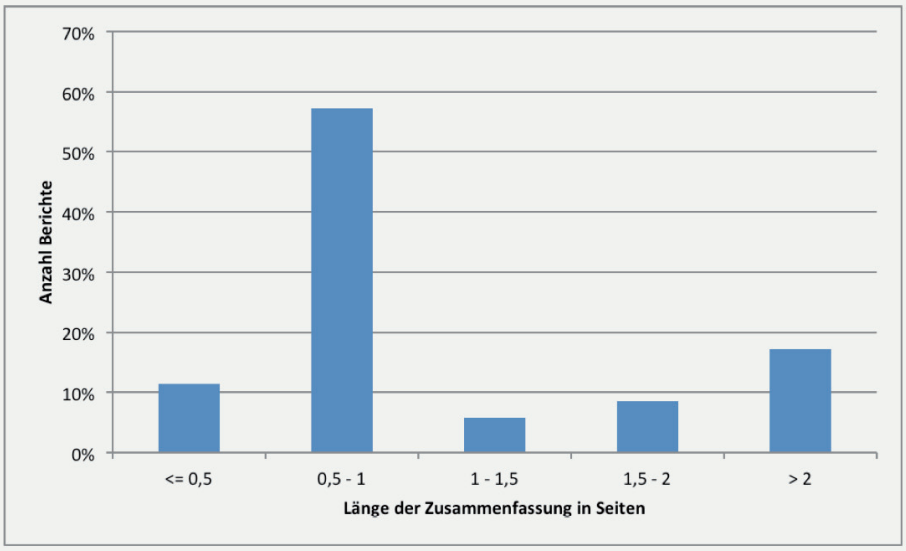
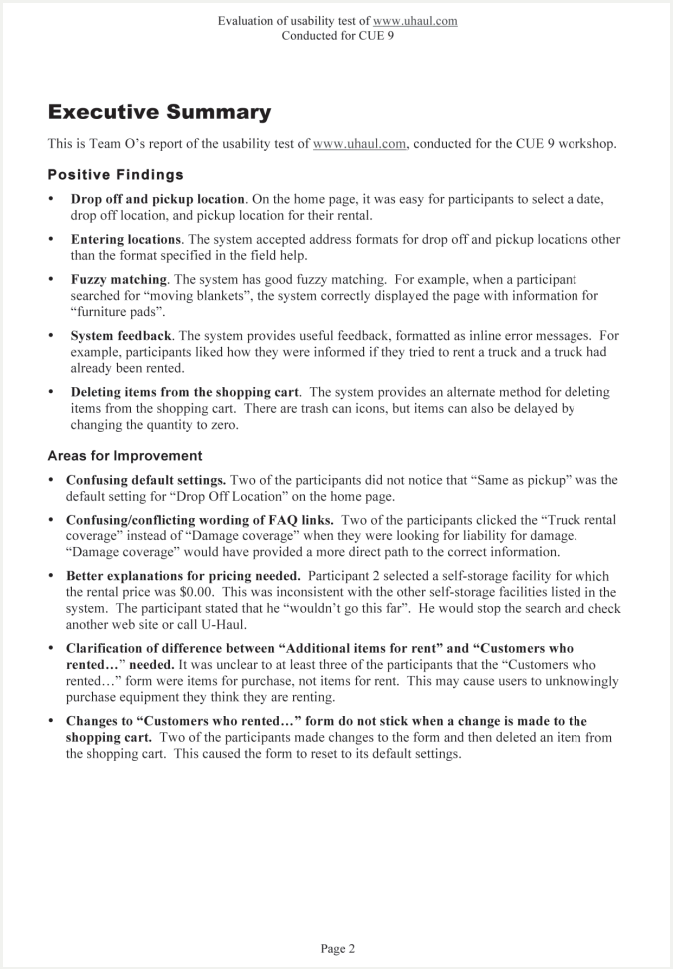


Abb. 1.
Länge der Executive Summary in Seiten



Evaluation of usability test of www.uhaul.com
Conducted for CUE 9

Executive Summary

This is Team O's report of the usability test of www.uhaul.com, conducted for the CUE 9 workshop.

Positive Findings

- **Drop off and pickup location**. On the home page, it was easy for participants to select a date, drop off location, and pickup location for their rental.
- **Entering locations**. The system accepted address formats for drop off and pickup locations other than the format specified in the field help.
- **Fuzzy matching**. The system has good fuzzy matching. For example, when a participant searched for "moving blankets", the system correctly displayed the page with information for "furniture pads".
- **System feedback**. The system provides useful feedback, formatted as inline error messages. For example, participants liked how they were informed if they tried to rent a truck and a truck had already been rented.
- **Deleting items from the shopping cart**. The system provides an alternate method for deleting items from the shopping cart. There are trash can icons, but items can also be delayed by changing the quantity to zero.

Areas for Improvement

- **Confusing default settings.** Two of the participants did not notice that "Same as pickup" was the default setting for "Drop Off Location" on the home page.
- **Confusing/conflicting wording of FAQ links.** Two of the participants clicked the "Truck rental coverage" instead of "Damage coverage" when they were looking for liability for damage. "Damage coverage" would have provided a more direct path to the correct information.
- **Better explanations for pricing needed.** Participant 2 selected a self-storage facility for which the rental price was $0.00. This was inconsistent with the other self-storage facilities listed in the system. The participant stated that he "wouldn't go this far". He would stop the search and check another web site or call U-Haul.
- **Clarification of difference between "Additional items for rent" and "Customers who rented…" needed.** It was unclear to at least three of the participants that the "Customers who rented…" form were items for purchase, not items for rent. This may cause users to unknowingly purchase equipment they think they are renting.
- **Changes to "Customers who rented…" form do not stick when a change is made to the shopping cart.** Two of the participants made changes to the form and then deleted an item from the shopping cart. This caused the form to reset to its default settings.

Page 2

Abb. 2.
Beispiel für eine besonders übersichtliche Zusammenfassung

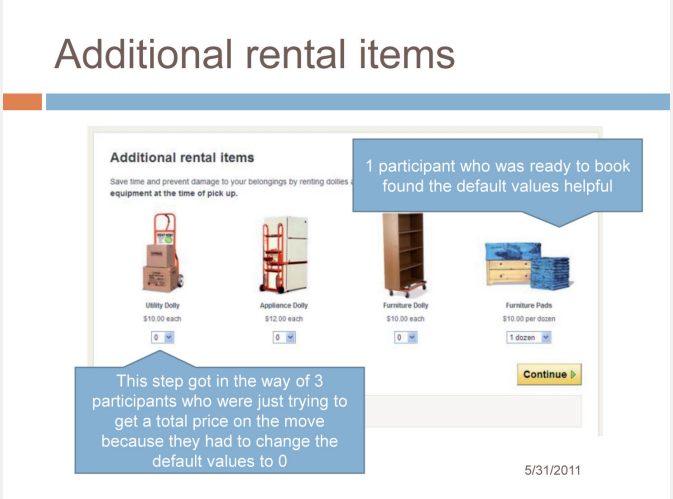


Abb. 3.
Beispiel für Problembeschreibung direkt im Screenshot

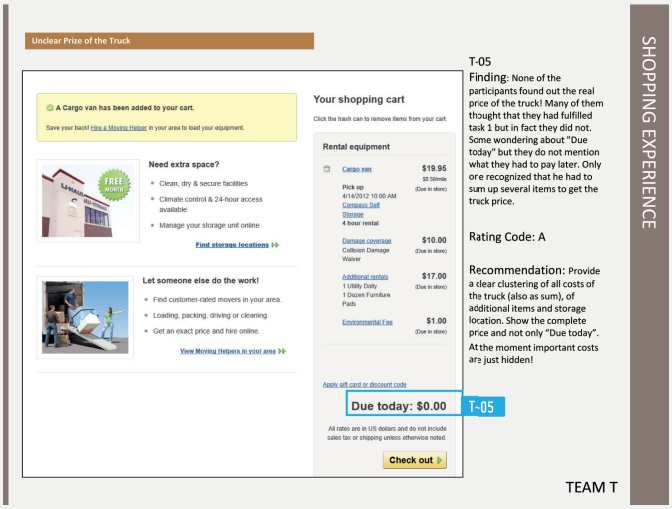


Abb. 4.
Fokus auf die visuelle Darstellung von Problemen in einem Powerpoint-Bericht

Methodik und damit schwerer aufzufinden. Eine solche Zusammenfassung gehört zu den „best practices", die in CUE-9 beobachtet werden konnten. Die Länge der Zusammenfassung variierte von wenigen Stichpunkten bis zu vier Seiten, der Durchschnitt lag bei 1,5 Seiten. **[Abb. 1]** Abbildung 2 zeigt eine besonders übersichtliche Zusammenfassung. **[Abb. 2]**

Aus der Analyse der Testberichte konnten folgende Eigenschaften abgeleitet werden, die eine gute Executive Summary kennzeichnen:

- Die Zusammenfassung sollte kurz sein; 69% der untersuchten Berichte enthielten eine Zusammenfassung von einer Seite oder weniger Länge.
- Die wichtigsten Probleme sollten genannt werden.
- Das Ergebnis zählt – Ausführliche Informationen zur Methode, den Teilnehmern, etc. gehören nicht in die Zusammenfassung und auch nicht davor

3.2. Visualisierung von Problemen mit Hilfe von Screenshots

Screenshots sind eine der wichtigsten Möglichkeiten, dem Nutzer Probleme verständlich zu machen und sollten nach Möglichkeit Teil eines Berichts sein.
13 von 35 Berichten enthielten dennoch ausschließlich Text – einige Teilnehmer berichteten allerdings, dass ihnen lediglich die Zeit zur Erstellung von Screenshots fehlte. **[Abb. 3]**

Besonders hilfreich für den Leser ist es, wenn im Screenshot der Bereich markiert ist, an dem das Problem auftrat. 25 Prozent der Teilnehmer verzichteten sogar ganz auf lange Word-Dokumente und legten mit Powerpoint-Präsentationen den Fokus klar auf die visuelle Darstellung von Problemen, wie in Abbildung 4 zu sehen. **[Abb. 4]**

3.3. Anzahl der genannten Probleme

Ein Bericht ist nicht dann gut, wenn er möglichst viele Probleme auflistet. Im Gegenteil ist es wichtig, dass der Usability Professional die berichteten Probleme priorisiert und Unwichtiges ausfiltert; denn gerade deswegen wird ein Experte engagiert.

Die höchste Zahl der berichteten Probleme war 85, der Durchschnitt lag bei 36. **[Abb. 5]** Fast drei Viertel (74 %) der Teilnehmer beschränkten sich auf weniger als 40 Probleme.

Ein Bericht enthielt sogar nur 5 Probleme auf einer Seite. **[Abb. 6]** Der Teilnehmer sagte dazu: „Ich gebe meinen Kunden nur die Probleme, die sie unbedingt beheben müssen. Wenn sie diese Probleme beseitigt habe, können sie gerne zu mir kommen um mehr Resultate zu bekommen."

3.4. Beschreibung der Usability Probleme selbst

Entscheidender als die Anzahl der Probleme ist es, dass die gefundenen Probleme verständlich formuliert sind, und dass die Auswirkungen auf den Nutzer und das Geschäft des Kunden überzeugend beschrieben werden. Das erhöht die Überzeugungskraft des Berichts und damit die Wahrscheinlichkeit, dass Probleme auch beseitigt werden.

Wie die Diskussion im Rahmen der Workshops zeigte, waren einige der in CUE-9 beschriebenen Usability Problemefür andere Teilnehmer schwer oder gar nicht verständlich odergespickt mit Rechtschreibfehlern. Ein Beispiel hierfür ist

folgender Befund: „to less notice about miles on the top of the page in step2 after ‚get rates' miles=city to city?". Dieser Befund war für die Teilnehmer des Tests nicht nachvollziehbar, obwohl alle die selben Videos analysiert hatten. Usability Probleme sollte besonders sorgfältig formuliert werden, sonst kann leicht die Glaubwürdigkeit des gesamten Berichts in Frage gestellt werden. Insgesamt waren immerhin 55 von 1.333 Beobachtungen in CUE-9 zu unklar, um verstanden zu werden (Ergebnis der Auswertung der Studie CUE-9, siehe (3)).

Viele Berichte führten als Problembeschreibung nur Beobachtungen auf, beispielsweise „participant is searching the faq". Merkmal eines guten Berichts und Grund, einen Experten zu engagieren ist aber gerade, dass dieser das zugrundeliegende Problem herausarbeitet, eventuell sogar mit einem Hinweis auf eine mögliche Lösung. Ein Beispiel für eine solche Auswertung ist in Abbildung 7 zu sehen. **[Abb. 7]**

3.5. Priorisierung und Gruppierung der Ergebnisse

Ein Teil der Berichte führte die Probleme in der Reihenfolge ihres Auftretens nach Test-Sitzung oder durchgeführter Aufgabe auf. Diese Gruppierung ist für den Leser wenig hilfreich. Ein Usability Professional sollte die Probleme in einer nützlicheren Gruppierung darstellen. Verwendete Kategorien in den untersuchten Berichten waren beispielsweise:

- Betroffene Seiten oder Funktionalität des untersuchten Produkts
- Schweregrad des Problems
- Allgemeine Problemkategorien wie „Shopping Experience" oder „Wording and Text"

4. Weitere Gestaltungsmöglichkeiten – so begeistert ein Bericht

Es gibt viele weitere Möglichkeiten, einen Testbericht so zu gestalten, dass er

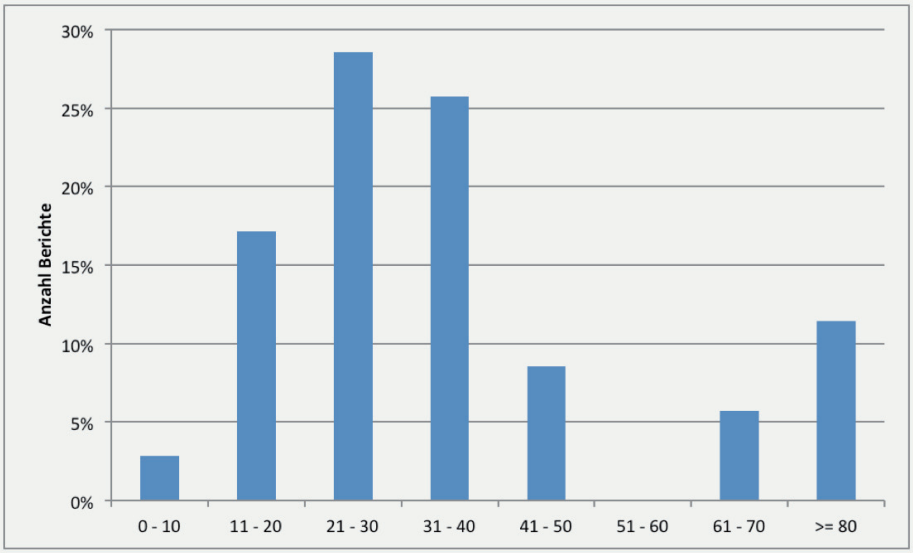


Abb. 5.
Anzahl gefundener Probleme

Usability Test Report for U-Haul (Team E)

Here are the four most serious usability problems I observed in usability test sessions U1, U2, and U3, and my recommendations for fixing them.

I observed more problems which I will be happy to report after these, the most serious ones, have been fixed.

1. **Users wanted/needed more help choosing the correct size truck.**

   Recommended immediate solution:

   - Change "Studio/Apartment" to "Apartment - Studio or 1 Bedroom"
   - Capitalize "APARTMENT" and "HOME" in headings so users are more likely to notice the difference
   - Add a small graphic for each size truck that shows a typical amount of furniture or boxes that will fit in it.

2. **Users wanted/needed more help figuring out what size storage space they needed.**

   Recommended immediate solution:

   - For each size of storage space, add graphics that show two or three typical examples of "stuff" that will fit in it. For example, one would show a stack of storage boxes that would fit, another would show a refrigerator and some boxes, another a box spring, mattress, and a few pieces of bedroom furniture.

   Possible longer term solution:

   - One participant suggested adding a calculator that would tell you the size of the storage unit you need if you enter the size and number of boxes you want to store.

3. **When searching for storage after adding a truck rental to the cart, all storage rates were shown as $0.**

   Recommended immediate solution:

   - When user has added a truck rental, change the rates on storage pages and shopping cart pages to read something like:
     - $15 (First month is free with truck rental)

   Possible longer term solution:

   - Make the "Get your first month free" graphic more noticeable on pages where it appears.

4. **Quantities are already selected for items on the Additional Rental Items page and in the "Customers who rented this needed" section on the summary page.**

   Recommended immediate solution:

   - Change all of these quantities to zero.

5. **Summary page does not display the total price.**

   Recommended immediate solution:

   - Add the total price to the Summary page.

Abb. 6.
Der kürzeste Bericht in CUE-9 – mit nur 5 Ergebnissen

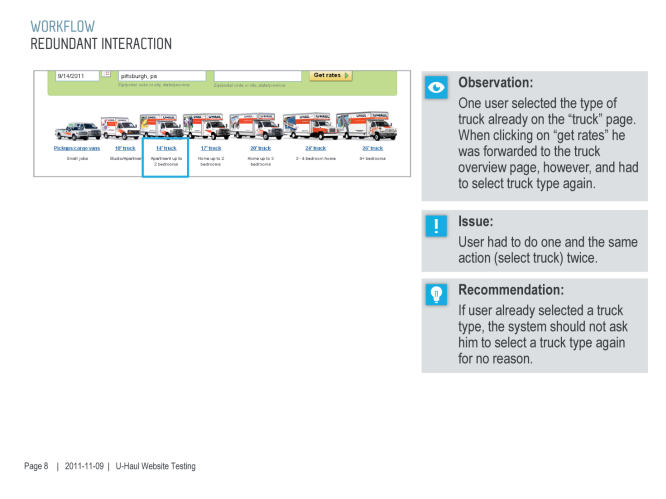


Abb. 7.
klare Beschreibung von Beobachtung, zugrundeliegendem Problem und Empfehlung

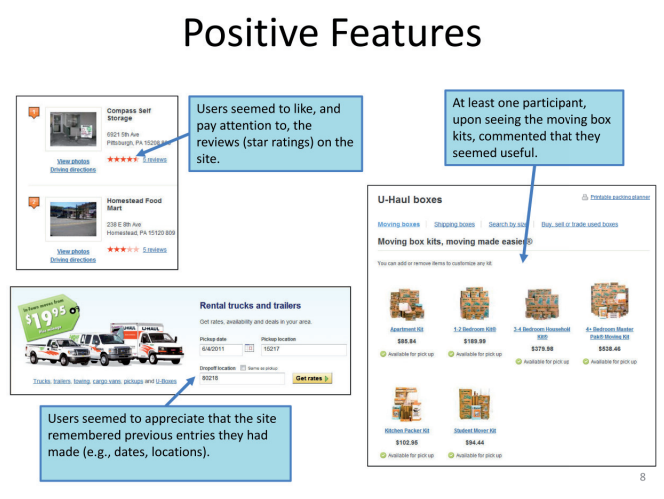


Abb. 8.
Beispiel für einen Bericht mit positiven Beobachtungen

Table 1: List of major usability problems identified during the U-Haul usability evaluation

Problem Area/Task	Problem Description	Severity
The layout and behavior of functions on the shopping cart page created problems for all participants.	Users inadvertently added extra items to their cart which they then had to delete; the accessories area was dominant visually and induced users to click on the Add to Cart button and add things that were not part of the task; To order a subset of accessories, users had to zero out most of the values in the accessories quantity field; the accessories area would refresh when users deleted something or move to another page which was very annoying.	**Critical (A)** • Reduced credibility • Extra work to correct accidental additions • Increased chance of errors

Abb. 9.
Ein gut strukturiertes Resultat mit Titel, Problembeschreibung und durch Auswirkungen auf den Nutzer begründete Klassifizierung

besonders verständlich und angenehm zu lesen ist. Die Highlights aus CUE-9 werden in diesem Abschnitt vorgestellt.

4.1. Positive Beobachtungen

Auch positive Aspekte können und sollten in einem Testbericht genannt werden: Niemand hört gern nur Kritik, besonders, wenn sie wie bei Usability Tests üblich von außen kommt. Ein Lob der positiven Aspekte eines Produkts kann erheblich dazu beitragen, dass die Testergebnisse akzeptiert und Usability Probleme behoben werden – und dass nicht bei der Überarbeitung des Produkts aus Versehen gelungene Features geändert werden. Abbildung 8 zeigt einen Bericht, der gezielt positive Beobachtungen hervorhebt. **[Abb. 8]**

4.2. Aufbereitung der Resultate

Um das „Scannen", also das Querlesen eines Berichts zu erleichtern, können die Beschreibungen der einzelnen Usability-Probleme strukturiert werden. Ein kurzer Titel hilft dabei, das Problem zu identifizieren. Bei der Beschreibung des Problems ist es wichtig, dass die Auswirkung auf den Nutzer klar wird. Abbildung 9 zeigt ein Beispiel für eine besonders gut strukturierte Problembeschreibung in tabellarischer Form. **[Abb. 9]**

In einigen Berichten wurde der Schweregrad eines Problems darüber hinaus mit Hilfe von Icons und/oder Farben visualisiert. **[Abb. 10]** Dies machte die Berichte übersichtlicher und sehr viel leichter zu überfliegen.

4.3.
Zitate

Einige Teilnehmer setzten Zitate der Teilnehmer ein. **[Abb. 11]** Dadurch lässt sich die Glaubwürdigkeit des Berichts erhöhen und der Leser gewinnt einen Einblick in die Stimmung der Nutzer. Ein Beispiel für ein solches Zitat ist der Kommentar eines Testteilnehmers, der erkannte, dass die Webseite vorselektierte zusätzliche Artikel in seinem Warenkorb platziert hatte: „Oh the fields for boxes are pre-filled? You are sneaky!".

4.4.
Ergebnislisten

Einige Berichte enthielten eine tabellarische Zusammenfassung der gefundenen Probleme als „Arbeitsliste" im Anhang oder als separates Dokument, um den Kunden bei der Bearbeitung der Probleme bestmöglich zu unterstützen. Abbildung 12 zeigt ein besonders übersichtlich gestaltetes Exemplar. **[Abb. 12]**

5.
Fazit

Gerade als Usability Experte sollte man darauf bedacht sein, besonders die eigenen Arbeitsergebnisse immer wieder auf ihre Gebrauchstauglichkeit zu untersuchen. Im Rahmen der CUE-Studie stellten sich klare Gemeinsamkeiten heraus, die eine gute Basis für die Gestaltung von Testberichten zu Usability Tests darstellen:

- Eine kurze Executive Summary am Anfang des Berichts
- Ergebnisse mit Hilfe von Screenshots visualisieren
- Die Anzahl der berichteten Probleme auf ein überschaubares Maß reduzieren
- Ergebnisse sinnvoll gruppieren und priorisieren

Darüber hinaus bleibt jedem Autor aber ein großer Gestaltungsspielraum, um einen Bericht zu verfassen, der für den Auftraggeber nützlich ist. Besonders visuelle Hilfsmittel wie Screenshots und Farbcodierungen spielen dabei eine große Rolle.Natürlich hat diese Untersuchung aufgrund der rein

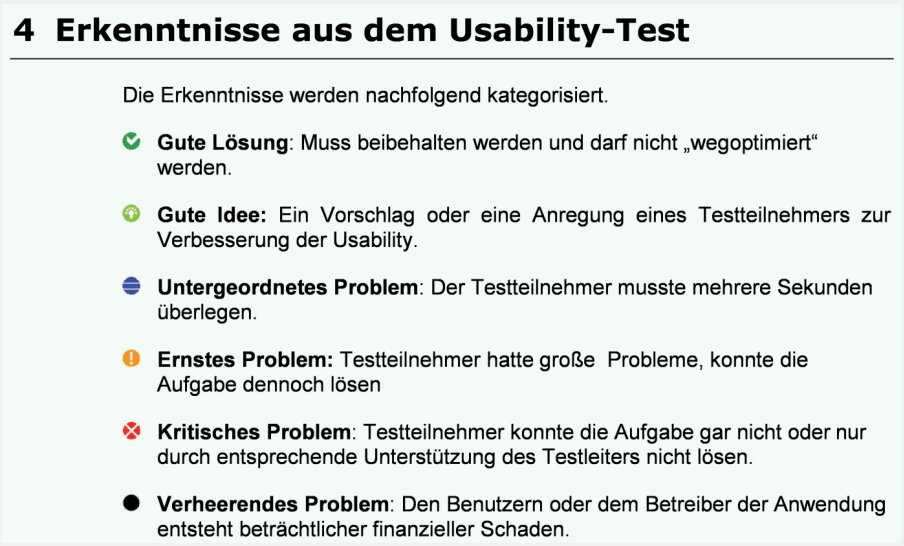


Abb. 10.
Visualisierung des Schweregrads mit Icons

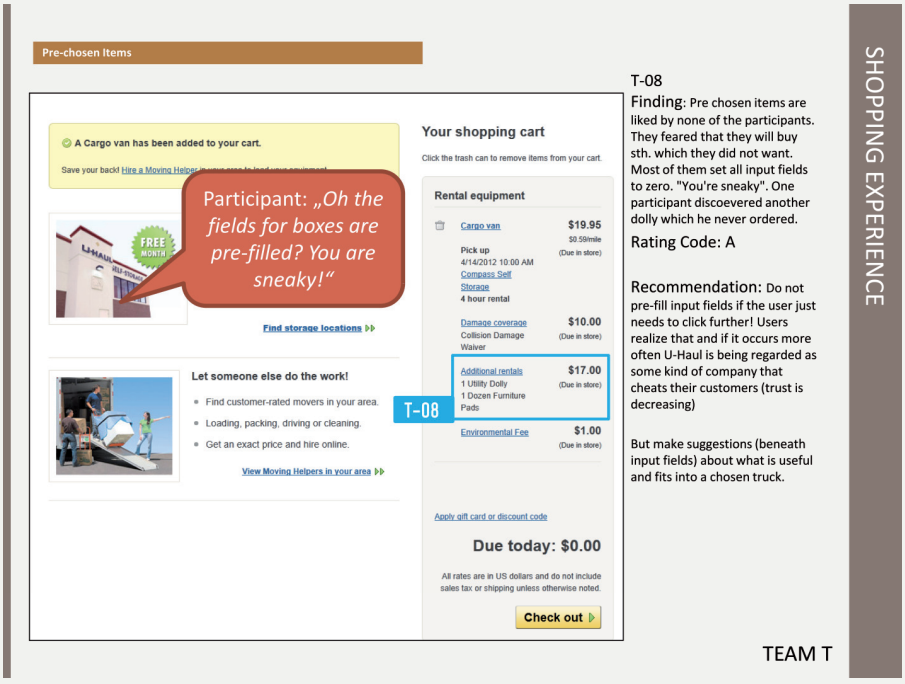


Abb. 11.
Beispiel für einen Bericht mit Zitaten der Testteilnehmer

Rate	Problem	Measures	Additional information
C	The offer to add additional stuff must be clearly recognizable as offer, not like it is already bought.	Make clear, that only stuff that the user wants, gets added to the cart.	This problem exists because of an inconsistency: some stuff is automatically added, some not.
C	Environmental fees added to the cart.	Either leave the fees out and inform on other page about the environmental initiative (price could be included in the truck price) or show better to read information.	The originally positive thought behind the environmental fee gets seen as negative – a complete contradiction to the intention.
C	Problems with icons	Provide a legend for all icons used on the page.	
C	After surfing on the site, when the user wants to add an already selected truck, a confusing question comes up.	If the same truck is already in the cart, don't ask.	This problem doesn't exist if a new page structure is used.
C	Buttons under dialog window are in the wrc	Use button orientation for „Ok", „Cancel" that most users (Windows!) know, „Ok" is left...	This is an easy change, do it.
C	Activity icon doesn't work properly.	Better to rely on the browsers activity icon, if own version acts strange.	Doing it right might be hard, leaving it out easy.
C	Company representation must override „uhaul" URL.	Increase logo size on first page, or make photo of the trucks large enough to read logo.	
C	Page changed, but cart stays the same.		
C	Calendar presentation unexpected.	Show days from other month different.	
C	Just click in drop-off location field doesn't work, check-mark must be removed		
P	Reviews are used	Keep the reviews	Most users minded the reviews and partly reacted on them.
P	In some places addresses are stored.		
P	Display of storage space is regarded nice.		
P	„Most popular" helps people – but not auto-add to cart.	If possible, show what many users bought in a similar setting, e.g. with a truck the same size.	Doing like lots of people before comforts people.
P	„Bedroom kits" are useful.		
P	Good to have all stuff on the page.	Stay with the all-in-one-location offering.	

Abb. 12.
Eine besonders übersichtliche Arbeitsliste

qualitativen Auswertung der Berichte durch die Autorin eine begrenzte Aussagekraft. In weiteren Schritten ist es vorstellbar, die Ergebnisse mit quantitativen Methoden zu untersuchen oder mit der Einbeziehung weiterer Testberichte zu Systemen aus anderen Domänen auf eine breitere Basis zu stellen.

Literatur

1. Rolf Molich, Webseite zu den CUE-Studien, http://dialogdesign.dk/CUE.html
2. Im Beitrag analysierte Testberichte, http://www.dialogdesign.dk/CUE-9.htm
3. Morten Hertzum , Niels Ebbe Jacobsen & Rolf Molich (2013): What You Get Is What You See: Revisiting the Evaluator Effect in Usability Tests, Behaviour & Information Technology, DOI:10.1080/0144929X.2013.783114