

Benjamin Uebel

# »Crowd-Usability-Tests – billig, aber gut?

**Franco Salvetti, Chefentwickler der Suchmaschine BING, veröffentlichte im Mai 2012 die Ergebnisse einer aufsehenerregenden Microsoft-Studie. Demnach können sog. Crowd-Usability-Tests vergleichbare oder sogar bessere Ergebnisse als klassische Usability-Tests aus dem Labor liefern – und dies zu einem Bruchteil der üblichen Kosten. Benjamin Uebel erklärt, wie solche Tests funktionieren und welche Chancen und Risiken sich bei der Nutzung ergeben können.**

Warum verlassen Besucher Ihre Webseite, ohne zu Käufern zu werden? Welche Ursachen haben hohe Absprungraten und Kaufabbrüche? Web-Analytic-Tools zeigen, WAS auf einer Webseite passiert, aber nicht, WARUM es passiert. Um das WARUM zu erklären, haben sich seit vielen Jahren Usability-Tests als Methode etabliert. Sie zeigen, mit welchen Schwierigkeiten Besucher einer Webseite zu kämpfen haben. Klassisch wurden diese Tests mit einem Testleiter in einem Labor durchgeführt, Testpersonen wurden eingeladen und bei der Nutzung der Webseite beobachtet und befragt. Da dieser Ansatz zeit- und kostenaufwendig ist, führten viele Webseitenbetreiber selten oder nie Usability-Tests durch.

## Crowd-Usability-Tests

Mit Crowd-Usability-Tests können Webseitenbetreiber selbst online Usability-Tests ihrer Seite anlegen. Sie können einfache Online-Studien erstellen, indem sie eine URL, eine Anzahl von Testpersonen sowie Aufgaben und Fragen an die Testpersonen eingeben. Zentrales Ergebnis der Tests sind 15- bis 25-minütige Videos, in denen alle Interaktionen einer Testperson mit der Webseite aufgezeichnet sind und in dem ihre laut ausgesprochenen Gedanken zu hören sind. Darin zu erkennen sind die Probleme, die den Nutzern beim Besuch der Seite entgegengetreten und zu einem Abbruch des Besuchs führen können.

## Woher kommen die Testpersonen?

Der Crowd-Service rekrutiert Teilnehmer aus einem eigenen Testerpool. Dieser ist in der Regel aus Webseitenbesuchern verschiedenster Schichten und Altersgruppen zusammengesetzt. Die Teilnehmer testen die Webseite nicht im Labor, sondern an ihren heimischen Rechnern, wobei die Aktivität des Nutzers und dessen Stimme aufgezeichnet werden. Der Webseitenbetreiber kann wählen, wie viele Probanden am Test teilnehmen sollen. Innerhalb kurzer Zeit (teilweise unter 24 Stunden) erhält er die Videos als Testergebnisse. Diese zeigen das Verhalten und die Gedanken der Nutzer beim Webseitenbesuch.

## Was kann ich damit testen?

Crowd-Usability-Tests werden am häufigsten zum Auffinden von Bedien- und Verständnisproblemen einer Webseite, eines Online-Shops oder einer Software eingesetzt. Dies kann im Zuge eines Relaunches erfolgen oder im Rahmen einer Conversion-Rate-Optimierung. Es werden Tests von Wettbewerberseiten durchgeführt, um aus den Stärken und Schwächen der Konkurrenz zu lernen. Oder es wird das Verhalten von Nutzern bei der Recherche nach bestimmten Keywords in einer Suchmaschine beobachtet. Einfache Werbewirkungstests zeigen, wie User Newsletter oder Youtube-Videos wahrnehmen und interpretieren und welche Optimierungspotenziale noch erschlossen werden können. Im begrenzten Maß sind sie auch zum Bug-Testing geeignet.

### DER AUTOR



**Benjamin Uebel** ist Geschäftsführer von

RapidUserTests, die Crowd-Usability-Tests anbieten. Er berät große und kleine Firmen in Sachen Web-Usability.

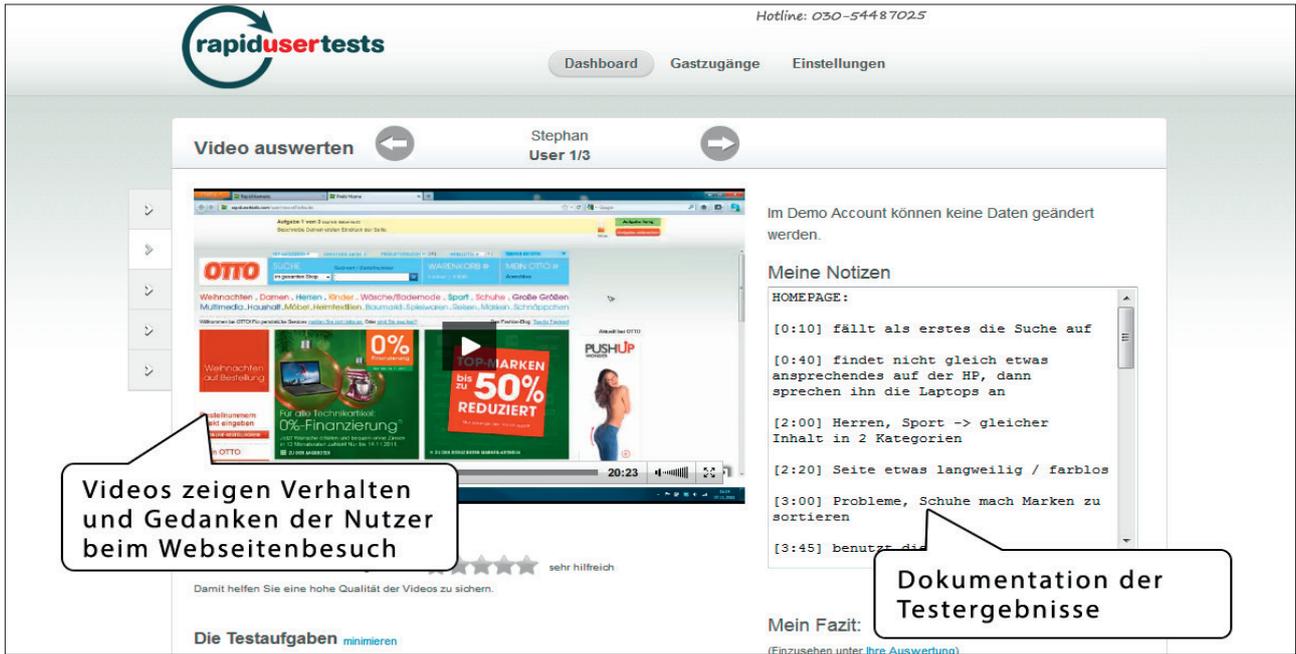


Abb. 1: 15- bis 25-minütige Videos zeichnen die Interaktionen der Testperson mit der Webseite sowie ihre laut ausgesprochenen Gedanken auf (Quelle: RapidUserTests)

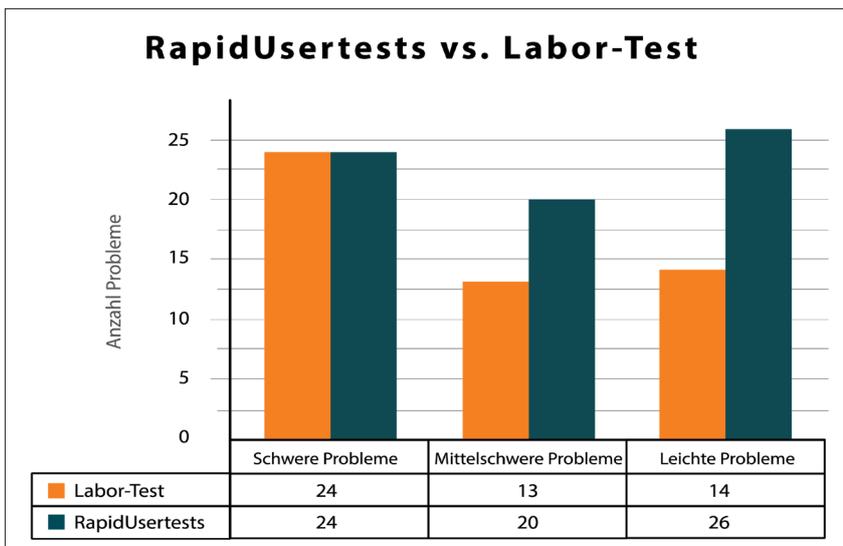


Abb. 2: Die Vergleichsstudie von eResult zeigt, dass der Crowd-Usability-Test RapidUserTests vergleichbare Ergebnisse wie klassische Labor-Usability-Tests liefert. Mit 20 Testpersonen wurden 70 Shop-Usability-Probleme identifiziert

Aufgrund Ihrer Skalierbarkeit haben Crowd-Usability-Tests das Potenzial, statistisch signifikante Ergebnisse zu produzieren. Aus qualitativen Informationen werden quantitative Daten generiert. Jedes dieser Einsatzgebiete wird im folgenden Artikel im Detail vorgestellt.

**Kann billig gut sein?**

Crowd-Usability-Tests sind im Vergleich zu klassischen Labor-Usability-Tests sehr preiswert. Sie kosten

40-50 EUR pro Testperson. Ein Test mit zehn Probanden kostet also einen dreistelligen Betrag. Dabei ist anzumerken, dass die Ergebnisse des Tests vom Webseitenbetreiber selbst ausgewertet werden müssen. Eine optionale Auswertung ist in der Regel möglich. Wenn ein Crowd-Test den Bruchteil der Kosten eines klassischen Tests ausmacht, liegt zunächst die Vermutung nahe, dass es einen Haken geben muss. Bei niedrigem Preis und hohem Bearbeitungstempo muss etwas

mit der Qualität der Ergebnisse nicht stimmen. Studien von Microsoft, STRATO und eResult zeigen jedoch, dass Crowd-Usability-Tests mit klassischen Tests mithalten können. In einigen Aspekten übertreffen sie diese sogar.

**Crowd-Tests unter der Lupe**

Unabhängig voneinander untersuchten die Agentur eResult (Quelle: <http://einfach.st/eresu1>) und der Internetdiensteanbieter STRATO (Quelle: <http://einfach.st/strat1>) den deutschen Crowd-Service RapidUserTests.com. Beide Studien zeigen, dass der Anbieter RapidUserTests dieselben Usability-Probleme aufdeckt wie ein klassischer Usability-Test im Labor. Die eResult-Studie wurde auf dem Usabilityblog veröffentlicht. Sie zeigt, dass der Crowd-Test sogar mehr kleine und mittlere Probleme identifizierte als ein klassischer Labortest. Der Crowd-Test mit 20 Testpersonen konnte dabei 70 Usability-Probleme identifizieren. Franco Salvetti, Chefentwickler bei Microsofts Suchmaschine BING, kam in einer Microsoft-Studie zu ähnlichen Ergebnissen (Quelle: <http://einfach.st/mss1>).



Abb. 3: Statt des Button „Gutscheine finden“ klickten viele Nutzer fälschlicherweise auf „erweiterte Suche“

### Beispiel eines Tests

Lassen Sie uns einen Blick darauf werfen, wie ein Test aufgebaut sein kann. Zunächst müssen die Testpersonen durch eine kurze Instruktion in die richtige Stimmung, in das richtige Mindset versetzt werden: „Stellen Sie sich vor, es ist Oktober und Sie müssen neue Winterreifen kaufen. Sie haben sich entschieden, dieses Jahr mal online nach Reifen zu suchen.“ Nach diesem Szenario zur Einstimmung werden den Testpersonen Testaufgaben gestellt, welche sie auf der Webseite bearbeiten müssen. In den Aufgaben sollte nicht der genaue Weg der User durch die Webseite beschrieben werden. Den Usern sollte stattdessen ein Ziel gegeben werden, welches sie eigenständig anstreben.

**1. Aufgabe:** „Suchen Sie im Internet nach Reifen für Ihr Auto.“ Diese Aufgabe ist offen formuliert und startet auf der Google-Webseite. Hier ist es interessant zu sehen, mit welchen Suchbegriffen die User nach den Reifen suchen. Man kann erkennen, welche Suchergebnisse aus welchen Gründen ansprechend auf die User wirken. Gleichsam erkennt man, wie die eigenen Suchresultate im Vergleich zu Wettbewerbsangeboten wahrgenommen und ob sie verstanden werden.

**2. Aufgabe:** „Gehen sie nun auf die Website [www.reifen.com](http://www.reifen.com). Suchen Sie dort nach einem für Ihr Auto passenden Reifen. Sobald Sie einen Reifen gefunden haben, der Ihren Wünschen entspricht, legen Sie diesen in den Warenkorb, schließen den Kauf aber nicht ab.“ An dieser Aufgabe sind mehrere Aspekte bemerkenswert. Zunächst werden die Tester auf das Angebot eines Wettbewerbers geschickt, um von dessen Stärken und Schwächen zu lernen. Die Aufforderung „suchen Sie nach einem für Ihr Auto passenden Reifen“ bzw. „ein Reifen, der Ihren Bedürfnissen entspricht“ ist auch offen formuliert. Sie ermöglicht es dem Tester, sich ungezwungen und an seinen eigenen Bedürfnissen orientiert zu verhalten. Mit dieser Art von Aufgabenstellung kommt man oft nahe an ein natürliches Nutzungsverhalten heran.

**3. Aufgabe:** „Gehen Sie nun bitte auf [www.meinereifenwebseite.de](http://www.meinereifenwebseite.de). Suchen Sie auch hier bitte nach einem Reifen für Ihr Auto, der Ihren Vorstellungen entspricht, und legen Sie diesen in den Warenkorb.“ Im eigentlichen Test der Webseite werden nun potenzielle Usability-Probleme der Webseite identifiziert.

**4. Aufgabe:** „Kaufen Sie den von Ihnen ausgewählten Reifen auf [meinereifenwebseite.de](http://meinereifenwebseite.de). Schreiben Sie als Vornamen bitte „Test“ in das Feld und wählen Sie als Zahlungsmethode Vorkasse aus.“ Prinzipiell kann man Testpersonen jede beliebige Aufgabe stellen. Man kann Testern damit individuelle Instruktionen geben. So kann ein jeder Test stark auf die Anforderungen und Fragestellungen eines Webseitenbetreibers zugeschnitten werden.

Schauen wir uns praktische Anwendungsfälle an, in denen Crowd-Usability-Tests zum Einsatz kommen.

### Anwendungsfall Conversion-Rate-Optimierung

Der Online-Shop [best-weekend.de](http://best-weekend.de) verkauft Hotelgutscheine. Um sein verfügbares Marketingbudget effizient auszunutzen, beschloss der Geschäftsführer, den Shop auf eine bessere Conversion zu optimieren. Leider ist der Traffic des Shops nicht hoch genug, um quantitative A/B-Tests durchzuführen. Ein einzelner Test würde zu lange dauern, um statistisch signifikante Unterschiede zu generieren. Um dennoch optimieren zu können, wurde die Seite mit einem Crowd-Usability-Test mit sieben Testpersonen getestet. Auf der Suche nach passenden Hotelgutscheinen traten bei den Testern Probleme auf. Der Test offenbarte, dass viele Nutzer gezielt Hotelgutscheine in bestimmten Regionen suchen wollten. Eine gezielte Suche nach Hotelgutscheinen z. B. an der Ostsee war nicht möglich. Ein weiteres kritisches Problem: Der Button der Suchfunktion war nicht erwartungsgemäß neben dem Suchschlitz positioniert. Alle Testnutzer klickten beim Bestätigen ihrer Sucheingebe unbeabsichtigt auf die erweiterte Suchfunktion und wurden von einem komplizierten Eingabeformular ent-

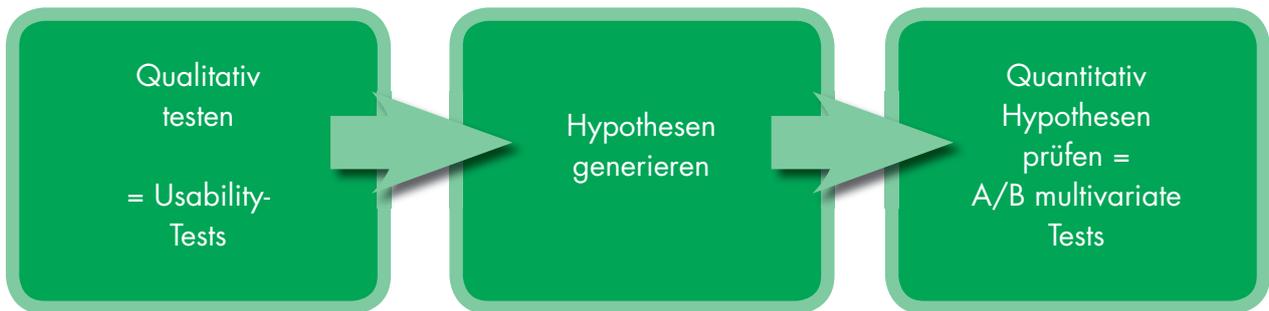


Abb. 4: Testen wie ein Wissenschaftler: Zunächst explorativ Hypothesen bilden und diese dann experimentell überprüfen

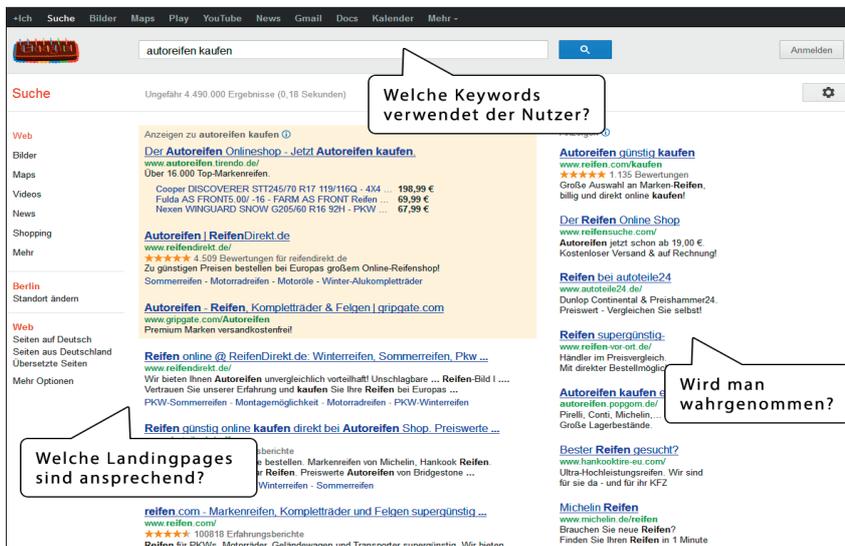


Abb. 5: Bei der Suche ab dem Google-Keyword lassen sich verschiedene Erkenntnisse gewinnen

täuscht. Nachdem diese und weitere Mängel behoben wurden, stieg die Conversion-Rate\* nach wenigen Wochen um 30 % – im selben Maß stiegen die Umsätze. Das Beispiel illustriert zwei Aspekte: Eine Webseite, die noch nie einem Usability-Test unterzogen wurde, hat häufig unentdeckte Mängel in der Nutzerführung, deren Behebung eine signifikante Verbesserung des Nutzererlebens und der Conversion-Rate bewirken kann. Als Webseitenbetreiber wird man schnell betriebsblind. Tests erweitern den Horizont des Betreibers. Sie helfen, Perspektive und Bedürfnisse der Nutzer besser zu erkennen. Andererseits zeigt das Beispiel, dass Crowd-Usability-Tests eine Alternative zu quantitativen Methoden bieten. Wenn eine Webseite nicht genügend Traffic hat, um A/B-Tests durchzuführen

ren, sind sie ein probates Mittel, um durch Verbesserung der Nutzerführung die Conversions zu optimieren. Crowd-Usability-Tests können auch in Kombination mit A/B-Tests angewendet werden. Unerfahrene Anwender von A/B-Tests optimieren häufig blind nach dem Trial-and-Error-Prinzip. Sie verändern die Farben, Positionen oder Größen von Elementen einer Landing-page\*, ohne eine Theorie dahinter zu haben. Dieses Vorgehen funktioniert, man sollte es aber nicht als systematisches Testen bezeichnen, sondern eher als „Ausprobieren“. Der Nachteil dieses Trial-and-Error-Prinzips: Es ist ineffizient und deckt selten die wirklich großen Conversion-Hebel auf. Erfahrene Conversion-Optimierer analysieren vor dem eigentlichen Testen Markt, Nutzerverhalten und

Wettbewerb, um die vielversprechendsten Optimierungspotenziale zu identifizieren. Usability-Tests werden in dieser Vorab-Analyse eingesetzt, um Testhypothesen zu generieren. Je häufiger Testpersonen über ein störendes Problem stolpern, desto größer die Wahrscheinlichkeit, dass es einen spürbaren Einfluss auf die Conversion-Rate hat. Der Impact einer Testhypothese kann damit abgeschätzt und so effizienter optimiert werden. Der Betreiber eines Online-shops für Werkzeugartikel beauftragte z.B. eine Agentur, mit multivariaten Tests die Conversion-Rate des Shops zu steigern. Da effizient getestet werden sollte, wurden in der Analysephase Hypothesen gesammelt. 9 von 15 Besuchern verstanden beim Besuch der Landingpage nicht auf den ersten Blick das Angebot des Shops. Es fehlte ein aussagekräftiger Slogan, außerdem waren die dargestellten Produktbilder nicht repräsentativ für das Spektrum des Shops. Im multivariaten Test wurden dann gezielt Slogans und aussagekräftigere Produktbilder der Landingpage getestet.

Genau so würde ein Wissenschaftler in der Forschung vorgehen: In einem ersten Schritt durch qualitative, explorative Studien (Usability-Tests) Hypothesen bilden, welche dann durch quantitative Experimente (multivariate Tests bzw. A/B-Tests) auf Signifikanz geprüft werden. Qualitative und quantitative Methoden in Kombination als Basis für effizientes Testen.

\*siehe Online-Glossar unter [www.websiteboosting.com](http://www.websiteboosting.com)



Abb. 6: Der Test von Wettbewerbern bringt Erkenntnisse für die eigene Webseiten-Optimierung

### Anwendungsfall Google-Suchergebnisse

Welche Suchergebnisse werden wie von den Nutzern bewertet? Welchen ersten Eindruck hinterlassen die verschiedenen Landingpages der Wettbewerber? Wenn man einen Usability-Test direkt mit der Google-Suche nach einem bestimmten Keyword beginnen lässt, erhält man Erkenntnisse zu diesen Fragen.

### Anwendungsfall Wettbewerbertests

Ein brasilianisches Sprichwort sagt: „Wir müssen auch aus den Fehlern anderer lernen, denn wir leben nicht lange genug, um sie alle selbst zu machen.“ Usability-Tests von Wettbewerbern sind aufschlussreich, um aus Fehlern und Stärken der Wettbewerber Rückschlüsse für die eigene Seite zu ziehen. Der Betreiber eines Vergleichsportals von Stromanbietern entdeckte bei einem Crowd-Usability-Test seines Konkurrenten, dass die Testpersonen den Wettbewerber als vertrauenswürdiger einstufen, weil er mehr Trust-Siegel auf der Startseite positioniert hatte. Auch lernte er, dass er seinen Vergleichsrechner vereinfachen musste: Viele Testpersonen waren sich nicht über ihren jährlichen Stromverbrauch

im Klaren. Der Wettbewerber löste dieses Problem, indem er von den Nutzern alternativ die Anzahl der Personen im Haushalt abfragte – eine Zahl, die jeder Nutzer sofort wusste. Aufgrund dieses Vergleiches wurden hilfreiche Anregungen gefunden, um die eigene Seite attraktiver zu gestalten.

### Anwendungsfall Newsletter-Tests

Ein Reiseveranstalter testete die Wirkung seines Newsletters an zehn Testpersonen, um Anregungen zur Erhöhung der Öffnungs- und Klickrate zu bekommen. Im Test wurde deutlich, dass der Newsletter zahlreiche Anglizismen aufwies. Dies missfiel nicht nur mehreren Testpersonen, sondern führte sogar zu Missverständnissen. Ältere Nutzer wünschten sich einen größeren Schriftgrad, um die Texte einfacher lesen zu können. Mehr als 75 % der Testpersonen klickten auf die Bilder der Newsletter-Artikel, um mehr Informationen zu erhalten. Da nicht die Bilder, sondern nur Überschriften verlinkt waren, führten die Klicks nicht zum gewünschten Ergebnis. Newsletter-Tests illustrieren ein weiteres Anwendungsgebiet der Crowd: die Wirkung eines Werbemittels zu verbessern. Missverständliche Werbeslogans („Come in and

find out“ – „Komm herein und finde heraus.“) oder formale Gestaltungsfehler wie zu geringe Schriftgrößen können durch diese Tests identifiziert werden. So besteht das Potenzial, mit einfachen Studien die Abverkaufswirkung von Werbemitteln zu optimieren.

### Anwendungsfall Videospot-Tests

Die deutsche Marketingabteilung eines international bekannten Getränkeherstellers erhielt den Auftrag, einen deutschen Youtube-Spot zu erstellen. Dieser sollte viral auf eine aktuelle Facebook-Kampagne aufmerksam machen. Jedoch bestand Unsicherheit, ob der amerikanische Spot für den deutschen Markt adaptiert oder ein neuer Spot erstellt werden sollte. Da der amerikanische Spot nicht gut performte, wurde ein Test beauftragt. Im Test des amerikanischen Spots an 20 Testpersonen über Youtube erkannte kein einziger User, dass für eine Facebook-Seite geworben wurde. Nur 25 % der User verstanden die Story des Spots, sodass nach kurzer Betrachtungszeit Ungeduld aufkam. Die Ergebnisse dieses Tests führten zur Optimierung der deutschen Adaptation: Der [Call-to-Action\\*](#) wurde klarer kommuniziert und eine einfachere Geschichte erzählt.

\*siehe Online-Glossar unter [www.websiteboosting.com](http://www.websiteboosting.com)

## Anwendungsfall skalierbare Tests

Durch den geringen Preis pro Testperson und den hohen Grad der Automatisierung sind Crowd-Tests skalierbar. Qualitative Testergebnisse können zu statistisch signifikanten Daten aggregiert werden. Ein großer Mittelständler testete an 80 Testpersonen zwei Designvarianten eines Webseitenprototyps. Es sollte die Entscheidung getroffen werden, welche Designvariante nutzerfreundlicher sei. Im Test wurden Aufgabenbearbeitungszeiten gemessen, die Häufigkeit der Nutzung einzelner Features gezählt sowie der Net Promoter Score erfasst. Aufgrund der großen Stichprobe von 80 Testpersonen konnten danach statistisch signifikante Aussagen zu den Unterschieden in der Nutzerfreundlichkeit der beiden Prototypvarianten getroffen werden. Auf Basis dieser quantitativen Ergebnisse wurde so eine belastbare Entscheidung für die Weiterentwicklung des Produkts gefällt.

## Anwendungsfall Bug-Testing

Die Tester arbeiten auf ihren heimischen Rechnern mit verschiedensten Systemkonfigurationen, welche in dieser Vielfalt in kaum einer Firma existieren. Daher werden beim Usability-Testen immer wieder unbekannte Bugs entdeckt. So stellte der Betreiber eines Notebook-Shops fest, dass User mit Chrome-Browser die AGBs im Checkout nicht einsehen konnten. Ein Anbieter von Ferienhäusern bemerkte im Test, dass bei einer bestimmten Filterkombination konsistent das Flash-Plug-in abstürzte.

## Risiken bei der Nutzung

Beim Einsatz von Crowd-Usability-Tests gibt es Risiken. So besteht die Gefahr, die Ergebnisse eines Tests falsch zu interpretieren. Usability-Tests

sind KEINE Meinungsumfragen. Im Test sollen nicht subjektive Meinungen erfasst, sondern auftretende Bedien- und Verständnisprobleme identifiziert werden. Usability-Papst Jakob Nielsen sagt dazu: „Pay attention to what users do, not what they say.“ (Übersetzung: „Achte darauf, was deine User tun, nicht was sie sagen.“) Die Erfahrung zeigt, dass die Behebung konkreter Usability-Probleme einen größeren Einfluss auf die Conversion-Rate hat als die Umsetzung subjektiver Meinungen zu Farbe, Design und Wording. Wenn man dies berücksichtigt, können in der Auswertung unerfahrene Personen eigenständig gute Rückschlüsse aus Testergebnissen ziehen. Ein weiteres Risiko sind zu geringe Stichprobengrößen. Auch wenn die Verführung eines preiswerten Low-Budget-Tests groß ist, sollten Websitebetreiber auf eine ausreichende Anzahl an Testern achten. Analysen mit zwei Testpersonen mögen interessante Resultate bringen. Es ist jedoch nicht einschätzbar, ob die gefundenen Probleme einzelne Ereignisse darstellen oder verallgemeinerbar sind. So besteht die Gefahr, bei zu kleinen Testerzahlen Arbeit in die Optimierung irrelevanter Probleme zu stecken. Die Erfahrung zeigt, dass sich Usability-Probleme in Tests ab sieben Testpersonen systematisch wiederholen. Dann kann die Relevanz der gefundenen Probleme zuverlässig abgeschätzt werden. Die Abschätzung wird umso genauer, je mehr Testpersonen verwendet werden. 10 bis 20 Testpersonen liefern in der Regel belastbare Resultate.

## Fazit

Crowd-Usability-Tests sind eine günstige Alternative zu klassischen Labor-Tests und liefern vergleichbar aussagekräftige Ergebnisse. Sie können zur einfachen Usability-Analyse, zur Conversion-Rate-Optimierung und als Werbewirkungstests einge-

## TIPP:

Zur Überprüfung der Usability gibt es prinzipiell zwei Methoden. Zum einen sind dies die analytischen Verfahren, wie z. B. der Cognitive Walkthrough. Hier versucht ein Experte einen typischen oder auch definierten Nutzer(-Typ) zu simulieren und durchläuft dabei den angenommenen Weg durch eine Website. Aus der Erfahrung heraus weiß er, wo Nutzer Schwierigkeiten bei der Erkennung von (Auswahl-)Möglichkeiten haben oder wo Verständnisprobleme bei Begriffen und Prozessen auftreten könnten. Solche Problemstellen werden anschließend mit dem Designer besprochen und ggf. entschärft.

Auf der anderen Seite gibt es die sog. empirischen Ansätze. Hier werden z. B. Tests im Labor mittels Eye-Tracking-Tools, Analysen aufgrund von Trackingdaten oder auch Methoden der aufgabengestützten Online-Befragung durchgeführt. Eine oft latent vorhandene Gefahr ist das Unterschätzen des tatsächlich notwendigen Planungsaufwands. Die Qualität und Verlässlichkeit der Ergebnisse hängt sehr stark von einer überlegten Vorgehensweise und den Frage- und Aufgabenstellungen ab. Der oft beobachtbare falsche Einsatz solcher Online-Fragebögen auf Webseiten hängt sicher zu einem nicht geringen Teil davon ab, dass man solche Tools recht schnell parametrisieren und ohne große Methodenkenntnisse starten kann. Fraglich bleibt zusätzlich, ob sich denn da auch immer die Richtigen veranlasst oder motiviert sehen, an Umfragen teilzunehmen. Anders gesagt: Misst man tatsächlich im richtigen Zielgruppenanteil? Daher sollte man sich durch einen einfachen und schnellen Einsatz von Befragungssystemen – gleich welcher Art – nicht davon entbunden sehen, sich auch fachlich damit zu beschäftigen oder entsprechende Hilfe heranzuziehen.

setzt werden. Achtet man auf eine ausreichende Stichprobengröße und missinterpretiert die Ergebnisse nicht als Meinungsumfrage, sind sie ein probates Mittel, die Nutzerfreundlichkeit von Webseiten zuverlässig zu analysieren.¶