

## Literaturrecherche und Ergebnisdokumentation

**Autor(en):** Alyahya, S., Alrugebh, D.

**Titel:** Process Improvements for Crowdsourced Software Testing

**Jahr:** 2017

**Laufende Nr. im Literaturverzeichnis:** 2

**Schlagwörter (3-5):** Process Improvement, Crowdsourced Software Testing, Testing Process, Crowd Manager.

**Kurzbeschreibung:** Das Paper bezieht sich auf Crowdsourcing im Bereich des Software Testing. Es werden die verschiedenen Prozesse im Arbeitsablauf von Crowdttesting- Umgebungen betrachtet; vom anfänglichen Einreichen der Testbedingungen bis zur letztendlichen Veröffentlichung der Ergebnisse. Hieraus resultierend werden Einschränkungen des derzeitigen Ablaufs identifiziert und skizziert. Weiter werden Vorschläge gemacht um verschiedene Teilprozesse zu verbessern: "improving assigning the crowd manager", "improving building the test team" und "monitoring the testing progress". Hierzu wurden die genannten Verbesserungen von den Autoren eigenständig in den Ablauf des Crowdttesting implementiert und anschließend anhand einer Versuchsreihe und durch die Hilfe von Fragebögen und Workshops ausgewertet.

Das Paper ist wie folgt strukturiert: 1. Einleitung, 2. Hintergrundinformationen zu Crowdttesting, 3. Derzeitiger Stand the Crowdttesting Prozesses, 4. Diskussion derzeitiger Schwierigkeiten, 5. Vorschlag von Verbesserungen, 6. Evaluierung der Verbesserungen, 7. Implementierung der Verbesserungen im Crowdttesting Prozess. Abschließend endet das Paper mit einem Fazit in Kapitel 8.

Der Überblick über Crowdttesting, welcher im zweiten Kapitel gegeben wird, beinhaltet nicht nur eine Zusammenfassung über einzelne Akteure und Prozesse im Crowdttesting und setzt sich kritisch mit diesen auseinander, sondern auch eine Auflistung derzeitiger Anbieter im Bereich Crowdttesting. Auch in den folgenden Kapiteln wird ein guter Zusammenhang der Vor- und Nachteile innerhalb und zwischen den einzelnen Akteuren und Prozessen im Crowdttesting aufgezeigt.

**Stärken und Schwächen der Quelle:** Als Stärke sehe ich, dass Verbesserungen des Arbeitsablaufs als Diagramme dargestellt werden. Dies habe ich so in anderen Artikeln zum Thema nicht gefunden. Zwar sind Gutachten der vorgeschlagenen Verbesserungen durch Fachexperten erstellt worden, allerdings sehe ich Fragebögen und Workshops als alleinige Methoden zur Überprüfung von methodischen Konzepten als Schwäche. Wünschenswert wäre eine praktische und nicht rein theoretische Evaluation.

**Erstellt von:** Sebastian Blum (383074)

**Autor(en):** Speidel, D., Mittal, M., und Sridharan, M.

**Titel:** Testing Experience - The Magazine for Professional Testers: "Enterprise Crowd Testing"

**Jahr:** 2014

**Laufende Nr. im Literaturverzeichnis:** 28

**Schlagwörter (3-5):** Exploratory Crowd Testing Services, Enterprise Crowd Testing, Onsite Testing, Outsourced Testing.

**Kurzbeschreibung:** Der Artikel zeigt verschiedene Elemente und Möglichkeiten des Crowdtesting im Bereich von Unternehmen auf. Hierbei bezieht er sich vornehmlich auf die Vorzüge, welche Unternehmen durch die Nutzung gewinnen. Als diese werden unter anderem Kostenersparnisse und Skalierbarkeit genannt. Nichtsdestotrotz wird angeraten, Crowdtesting nicht als alleinige Testmethode zu nutzen, sondern ebenfalls auf ein internes Team, bestehend aus Experten, zu setzen.

Einleitend gibt der Artikel einen Überblick über den heutigen Aufbau von Unternehmen der Technologiebranche im Softwarebereich. Hierbei wird aufgezeigt, dass bei der Produktentwicklung nicht mehr nur die Unternehmen und deren Interessen im Vordergrund stehen. Vielmehr sollen Softwarekomponenten auf die Nutzer ausgerichtet und mithilfe dieser getestet werden. Hierbei wird CrowdTesting als eine Kernmethode genannt um direkt an End-Nutzer ein Produkt zu testen. Mögliche Testmethoden werden unterschieden zwischen "exploratory crowd testing services" und "enterprise crowd testing". Bei Ersterem wird die Software vornehmlich grundlegend und generell auf Fehler, Probleme und Defekte untersucht. Hierbei wird die Software ohne Struktur untersucht bzw. "erkundet". Bei Letzterem liegt eine konkretere Struktur der Fehlersuche zugrunde. Beide Verfahren hegen laut Artikel unterschiedliche Anforderungen an die Testnutzer. Auch wird unterschieden, ob Kriterien an die Prüfer gestellt werden oder nicht und wie selektiv die Testnutzer ausgewählt werden sollten. Weiter werden im Artikel die verschiedenen Möglichkeiten von Testing genannt. Hierbei wird zwischen „onsite“, „outsourced“ und „crowd testing“ unterschieden.

**Stärken und Schwächen der Quelle<sup>1</sup>:** Der Artikel ist in einem „Magazine for Professional Testers“ erschienen. Somit ist er auf eine konkrete Leserschaft ausgerichtet und als solcher auch zu lesen. Hierbei wird vornehmlich auf (große) Unternehmen eingegangen und aufgezeigt, wie diese das Testen ihrer Produkte auslagern können. Der Artikel liest sich demnach wie ein Werbeprospekt für Crowdsourcing. Wissenschaftliche Standards werden nicht erfüllt und Quellen sind nicht vorhanden; da der Artikel allerdings auch nicht den Anspruch an solche stellt, entziehe ich mich einer Wertung dessen, möchte es dennoch als wichtig anmerken.

**Erstellt von:** Sebastian Blum (383074)

<sup>1</sup> In der vorliegenden Quelle sieht sich dies als inhaltlichen Gesichtspunkt und nicht als Rezension der Quelle.

**Autor(en):** Leicht, N. et al.

**Titel:** When is Crowdsourcing advantageous? The Case of crowdsourced software testing.

**Jahr:** 2016

**Laufende Nr. im Literaturverzeichnis:** 18

**Schlagwörter (3-5):** Transaktionskostentheorie, In-House Testing, Crowdsourced Software Testing.

**Kurzbeschreibung:** Das Paper gibt einen Überblick darüber, ob Crowdsourcing überlegen ist gegenüber In-House Testing und wann es zu einem Einsatz von Crowdsourcing kommen sollte. Hierzu gibt das Paper einen Einblick in die aktuelle Literatur unter Berücksichtigung der Transaktionskostentheorie. Anhand der Wirksamkeit und Kosten von Crowdsourcing wird ein Überblick über verschiedene Faktoren gegeben, welche Testmanager im Einsatz von Crowdsourcing beachten sollten.

Hierbei werden anhand von zwei Fallstudien die Vor- und Nachteile von Crowdsourcing gegenüber „In-House Testing“ analysiert und bewertet. Diese sind in Unternehmen aus unterschiedlichen Bereichen angesiedelt („Swiss Bank“, „Swiss Industrial Enterprise“) um verschiedene Rahmenbedingungen zu gewährleisten. Crowdsourcing wird im Bereich mobiler Applikationen analysiert und anhand von relativer Wirksamkeit (number of bugs, quality of bug, overall time) und der relativen Kosten (direct and extra costs) evaluiert. In beiden Fallstudien zeigen Crowdsourcing und In-House Testing ähnliche Ergebnisse. Weiter wurde eine Analyse über beide verschiedenen Fallstudien durchgeführt. Insgesamt ergaben sich für die relative Wirksamkeit bessere Ergebnisse im Bereich Crowdsourcing, wohingegen nur kleinere übergreifende Unterschiede im Bereich der relativen Kosten lagen.

**Stärken und Schwächen der Quelle:** Die Einbindung von Crowdsourcing und In-House Testing anhand von konkreten Fallstudien sind eine gute Möglichkeit um Wirksamkeit und Verlässlichkeit in einer realen Umgebung zu analysieren.

**Erstellt von:** Sebastian Blum (383074)

**Autor(en):** Hoßfeld, T. et al.

**Titel:** Best Practices for QoE Crowdfunding: QoE Assessment with Crowdsourcing

**Jahr:** 2014

**Laufende Nr. im Literaturverzeichnis:** 13

**Schlagwörter (3-5):** Quality of Experience, Verlässlichkeit von Nutzerbewertungen, Crowdfunding Kampagnen.

**Kurzbeschreibung:** Der Artikel geht auf Crowdfunding im Bereich "Quality of Experience" (QoE) ein. Hierbei werden im Rahmen von Crowdfunding Kampagnen auf Vor- und Nachteile von Crowdfunding und bei der Testung von Multimedia-Apps analysiert und erörtert. Als Kriterien bei der Nutzung von Crowdfunding wird insbesondere auf die Verlässlichkeit von Nutzerbewertungen eingegangen, aber auch Fragen nach dem Einfluss von Fördergeldern, den Zahlungsschemata und eines umgebungsbedingte Kontexts der Studien erörtert. Dennoch liegt der Fokus des Artikels vornehmlich auf dem Aspekt der Verlässlichkeit von Nutzerbewertungen. Der Artikel enthält eine Sammlung der besten Methoden um sowohl konzeptuelle als auch technische Hindernisse zu behandeln. Hierbei werden Strategien aufgezeigt, um die Nachteile und Hindernisse im Umgang mit Crowdfunding zu minimieren. Diese wurden von den Autoren in einem eigenen Studiendesign implementiert und verifiziert. Die von den Autoren genannten Empfehlung werden in den Bereichen *Technical Implementation*, *Campaign and task design* und *Statistical analysis* unterschieden und als verlässlich eingestuft. Insgesamt wird eine große Anzahl an aktueller Literatur zum Thema aufgeführt und aus einer Vielzahl von Studien Resultate gezogen und Ergebnisse erzielt.

**Stärken und Schwächen der Quelle:** Als Stärke sehe ich an, dass oben genannte Kriterien wie die Verlässlichkeit von Nutzerbewertungen nicht nur unter qualitativen Gesichtspunkten, sondern auch anhand quantitativer Daten analysiert wird.

**Erstellt von:** Sebastian Blum (383074)

**Autor(en):** Rajini, P.

**Titel:** A Technical Paper on Best Practices in Crowd Sourced Testing.

**Jahr:** 2012

**Laufende Nr. im Literaturverzeichnis:** 25

**Schlagwörter (3-5):** Crowdsourcing, Software Quality Assurance, Crowd Sourced Testing.

**Kurzbeschreibung:** Der Bericht gibt einen Überblick über grundsätzliche Fragen und Anmerkungen zu Crowdttesting und dessen Anwendung. Hierbei wird zu Anfang die Frage gestellt, wo Crowdttesting für den Bereich "Software Quality Assurance" von Relevanz ist. Es wird mehrfach betont, dass es nicht um eine Gegenüberstellung von "Crowd" und anderen Testgruppen (wie bspw. Experten) geht. Vielmehr wird auf die spezifischen Eigenheiten des Crowdttesting eingegangen und was zu beachten ist. So wird beispielsweise genannt, wann der Einsatz von Crowdttesting sinnvoll ist und auf die damit einhergehenden Herausforderungen. Außerdem werden Grundlegende Fragen geklärt; "Was" CrowdTesting ist, "Wann" es am besten eingesetzt wird und "Wie". Weiter wird ein theoretischer Überblick über Planung und Anwendung gegeben. Auch wird ein Überblick über die "Crowd" an sich gestellt und welche Voraussetzungen und Herausforderungen an diese bestehen. Hierbei wird auf deren Stärken und Schwächen eingegangen. Es wird auf unterschiedliches Nutzerverhalten bei Produkten eingegangen und in den Kategorien „In-Job Behaviors" und "Organizational Citizenship Behaviors" unterschieden. Sämtliche Ausführungen werden anhand von Beispielen dargelegt und anschaulich erläutert. Hierbei werden zu guter Letzt geläufige Mythen zum Thema genannt und auf diese eingegangen. Zusammenfassend gibt der Bericht ein ganzheitliche und einleitende Sicht zum Thema Crowdsourcing und Crowdttesting.

**Stärken und Schwächen der Quelle<sup>2</sup>:** Der Bericht dient als Grundlage für eine Präsentation einer Konferenz. Er ist als Überblick angelegt und auch als solcher zu verstehen und zu lesen. Daher finden sich keine Quellenangaben und Querverweise innerhalb des Textes. Unter dem Aspekt der Einleitung möchte ich dies nennen, aber nicht werten, obgleich es keineswegs wissenschaftlichen Standards genügt. Insbesondere wenn man neu mit der Thematik ist, gibt der Bericht eine gute Einleitung.

**Erstellt von:** Sebastian Blum (383074)

<sup>2</sup> In der vorliegenden Quelle sehe ich dies als inhaltlichen Gesichtspunkt und nicht als Rezension der Quelle.

**Autor(en):** Archambault, D., Purchase H., Hoßfeld, T.

**Titel:** Evaluation in the Crowd: An Introduction

**Jahr:** 2017

**Laufende Nr. im Literaturverzeichnis:** 1

**Schlagwörter (3-5):** Quality of Experience (QoE), Crowd, Mikroaufgaben, visuelle Repräsentation

**Kurzbeschreibung:**

Das Buch entstand als Produkt von Diskussionen und Austausch während des gleichnamigen Seminars im Schloss Dagstuhl, das als Dagstuhl-Seminar Nr. 15481 bezeichnet wird und im Jahr 2015 in Deutschland stattfand. Das Buch fokussiert sich auf ein bestimmtes Anwendungsbereich von Crowd Testing, nämlich auf visuelle Repräsentation.

Die traditionelle Methode in diesem Bereich besagt, dass eine visuelle Repräsentation im Labortest an Probanden gezeigt wird und dabei messbare Performanz-Metriken (Reaktionszeit und Fehlerrate) und/oder qualitative Rückmeldungen zum Nutzererlebnis erfasst werden. Damit sammelt man Informationen über die Vor- und Nachteile einer bestimmten visuellen Repräsentation und die Wahrnehmung der Video-Qualität. In diesem Zusammenhang ist Crowd Testing eine wichtige Methode, die während sogenannter QoE-Studien oft verwendet wird.

- In dem einleitenden Beitrag des Buches sind die relevanten Fragestellungen erläutert, die in den weiteren Kapiteln des Buches detaillierter diskutiert werden:
- Vergleich der Methoden Labortest und Crowd Testing:
  - Welche Methode ist unter welchen Bedingungen besser geeignet?
  - Testdesign bzw. Aufgabengestaltung: Wann müssen komplexere Aufgaben gestaltet werden?
- Was ist eine angemessene Incentivierung für Crowd Testing?
- Wie kann ich produktive und qualitativ gut leistende Probanden über Crowd Testing Plattformen langfristig engagieren?
- Optimierung der Methodologie:
  - Between-Subjects-Design: Was muss man besonders dabei beachten, um die Nutzergruppen zu identifizieren?
  - Within-Subjects-Design: Wie kann ich solche Experimente mit Crowd Testing Verfahren durchführen?
  - Wie kann ich bessere qualitative Daten mit Crowd Testing erheben?
  - Wie kann ich die Nutzer dazu motivieren, bessere qualitative Rückmeldungen zu geben?
- Anwendungen in HCI, Visualisierung und Quality of Experience (QoE):
  - Welche spezifischen Herausforderungen haben verschiedene Anwendungsfelder bzgl. Design, Durchführung und Analyse?
  - Gibt es Best-Practices, die in allen Feldern angewendet werden können? Wenn nicht, wie kann man solche entwickeln?
- Crowdtester besser kennenlernen:
  - Wie nehmen die Crowdtester die Welt wahr? Was sind deren Interessen und Motivationen?
- Ethische Fragestellungen:
  - Wenn Usability Testing aus dem Labor auf Crowd Testing Plattformen implementiert wird, entstehen neue ethische Anforderungen, die man bisher nicht beachtet hat?
  - Wenn die direkte Kommunikation zwischen Versuchsleiter und Proband durch Crowd Testing eliminiert wird, ignoriert man den Fakt, dass Crowdtester auch Menschen sind?
  - Was sind angemessene Incentivierungen für Mikroaufgaben?

**Schwächen:** Das Buch beschränkt sich auf den Anwendungsbereich visuelle Repräsentation. Deswegen ist das Testdesign nicht so interaktiv wie bei einem Usability Testing von Applikationen oder Webseiten.

**Stärken:** Die Einleitung erläutert die Terminologie und grundlegende Definitionen bzgl. Crowd Testing, die durch das gesamte Buch homogen verwendet werden. Besonders die ethischen und charakteristischen Fragestellungen ermöglichen, über zukünftige Strategien und Prozesse der Anwendung von Crowd Testing informierter diskutieren zu können.

**Erstellt von:** Sezen Akman (388988)

**Autor(en):** Blohm, I., Leicht, N., Leimeister, J.M., Rhyh, M.

**Titel:** Besser dank Schwarmintelligenz

**Jahr:** 2016

**Laufende Nr. im Literaturverzeichnis:** 4

**Schlagwörter (3-5):** Crowdtesting, Softwareentwicklung, Crowd-Management

**Kurzbeschreibung:**

Der Zeitschriftenartikel berichtet über Crowdtesting als eine Lösungsmöglichkeit für die Problematik in der Applikationsentwicklung, dass Softwaretests heutzutage an mehreren Endgeräten mit verschiedenen Kombinationen von Hardware und Software, ggf. in verschiedenen Orten und zu verschiedenen Zeiten durchgeführt werden müssen, um eine kompetitive Qualität gewährleisten zu können. Dabei entstehen bei traditionellen Testverfahren, unter anderem durch länger dauernde Testungen, hohe Kosten. Diese Umstände kann man laut Autoren durch Crowdtesting vermeiden. Der Artikel fasst die Crowdtesting-Verfahren wie folgt zusammen:

Phase 1:

Unternehmen definieren die Anforderungen und Rahmenbedingungen für den Crowdtesting-Anbieter, die den Testumfang, die Testart, relevante Testgeräte und die gewünschten Eigenschaften der Crowd festlegen.

Phase 2:

Crowdtesting-Anbieter bekommt die Anforderungen und Rahmenbedingungen vom Unternehmen und rekrutiert die Tester (Crowd) durch eine offene Einladung, die für einen bestimmten Zeitraum gültig ist. Crowd fängt an, an ihren eigenen Geräten zu testen.

Phase 3:

Crowd dokumentiert die Testergebnisse und Anbieter bezahlt die Tester nach dem Erhalt des Testberichtes (z.B. pro Bug).

Phase 4:

Testberichte werden seitens Anbieter an Unternehmen weitergeleitet. Unternehmen entscheidet in der Phase 1, ob Anbieter Auswertung und Aufbereitung der Ergebnisse ebenfalls übernimmt. Der Artikel beschreibt, dass Crowdtesting besonders für die folgenden Anwendungen geeignet ist: Softwaretests für Apps (Apps für Kunden, Mitarbeiter), Webseiten (Kundenportale, Online-Shops), Enterprise Software (Informationssysteme, ERP-Integrationen, Entwicklungsumgebungen), Intranet (internes Crowdtesting innerhalb eines Unternehmens), Home-Placement Systeme (selbstentwickelte Softwaremodule) oder Internet-of-Things-Anwendungen (Carsharing, SmartTV). Außerdem werden anhand eines Beispiels aus der Praxis Ergebnisse eines externen Crowdtestings und eines internen Testverfahrens in dem Unternehmen verglichen, wobei man eine neue Smartphone-Applikation getestet hat. Fragestellung der beiden Testungen war, durch zwei Testiterationen zuerst funktionale Fehler zu entdecken und dann die Bedienbarkeit der Applikation zu bewerten. Details über die Testabläufe, die einen direkten Vergleich beider Testungen ermöglichen, wurden im Artikel nicht erläutert. Jedoch berichten die Autoren, dass ein internes Testverfahren parallel zum Crowdtesting als Kontrolle lief. Die Ergebnisse zeigten, dass durch Crowdtesting 55 Fehler und durch internes Testverfahren nur 3 Fehler entdeckt wurden. Die Autoren erwähnen auch, dass durch die 55 Fehler in dem Praxisbeispiel neun schwerwiegende Softwarebugs identifiziert und korrigiert wurden.

**Schwächen:**

- Informationssicherheit bei Applikationen stellt ein großes Problem dar, weil bei Crowdtesting meistens ein Crowdtesting-Anbieter den ganzen Prozess ab Probandenakquise bis zum Ergebnisbericht übernimmt und dabei Zugriff auf sensible Daten hat.
- Erhöhter Aufwand durch Steuerung und Kontrolle der Crowd
  - Die Autoren forschen im Bereich Referenzmodelle und Regelprozesse für Crowdtesting, um Prozesse zu standardisieren und diese Schwäche zu eliminieren.
- Motivation und Engagement der Crowd inhomogen und meistens unbekannt.
  - Das kann Effekte auf Testergebnisse haben, die man nicht vorhersehen oder kontrollieren kann.

- Für spezifische Applikationen nicht gut geeignet, wenn für Testung bestimmte Expertise oder Erfahrung benötigt ist.
- Wenn Nutzer Fragen zum Testablauf oder Applikation haben, ist es schwierig, diese vor Abschluss des Tests zu klären.
  - Das kann Effekte auf Testergebnisse haben, die man nicht vorhersehen oder kontrollieren kann.
  - Zusätzlicher Aufwand wegen Hotlines oder Hilfe-Chat hat vermutlich geringeren Nutzen im Vergleich zu direkten Kontakt zum trainierten Versuchsleiter während des Tests.

**Stärken:**

- Besonders für große Unternehmen, die weltweit benutzte Applikationen entwickeln, hat dieses Verfahren deutliche Vorteile:
  - Diverse Nutzer werden weltweit rekrutiert und nehmen gleichzeitig an Test teil.
    - Demographische Diversität (Sprachen, Kultur, Hintergrund, Alter, etc.)
    - Testdurchführung ist unabhängig vom Ort und von Zeit
    - Testdauer wird im Vergleich zum traditionellen Verfahren um ein Vielfaches reduziert
  - Nutzer agieren in ihrer realen Umgebung und testen an ihrem eigenen Endgerät.
    - mit geringerer Incentivierung möglich
    - keine Störeinflüsse bzgl. des sozialen Kontaktes
    - realitätsnahes Erlebnis
    - Keine Fixkosten bzgl. Testgeräte, Räumlichkeiten
    - Zahlreiche Kombinationen hinsichtlich technischer Eigenschaften von Testgeräten

**Erstellt von:** Sezen Akman (388988)

**Autor(en):** Gardlo, B., Ries, M., Hoßfeld, T.

**Titel:** Impact of Screening Technique on Crowdsourcing QoE Assessments

**Jahr:** 2012

**Laufende Nr. im Literaturverzeichnis:** 10

**Schlagwörter (3-5):** Subjektive Testing, Crowdsourcing, Soziale Netzwerke, Screening Techniken, Reliabilität

**Kurzbeschreibung:**

Das Paper handelt von der Evaluation der Qualität vom Nutzererlebnis in Bezug auf Online Video Streaming, wobei man Crowd Testing als eine subjektive Testmethode in realen Bedingungen für die Bewertung des Nutzerempfindens von Video-Qualität anwenden kann. Einflussfaktoren auf Nutzerempfindens der Video-Qualität wird in drei Gruppen zusammengefasst: Inhalt der Videos (Interessen, kulturelle und soziale Hintergründe), Erwartungen (genutzte Marke, Nutzerhistorie) und technisches System (Endgerät, Mechanismus für Video-Streaming). Besonders für die Auswertung der Interessen, kulturelle und soziale Hintergründe der Probanden wird vorgeschlagen, dass die Beteiligung in sozialen Netzwerken dafür genutzt werden kann, dass die demographische Angaben der Probanden bei der Crowd Testing durch Vergleich mit entzogene Daten aus deren sozialen Netzwerkpräsenz vergleicht werden kann.

**Schwächen:**

Die ethischen Konsequenzen von Nutzen der privaten Daten sind sehr fragwürdig.

**Stärken:**

Präsentierte Ergebnisse in der Studie beweisen, dass die entwickelten Filtermechanismen sehr effektiv bei Ausfiltern der inkonsistenten Angaben von Probanden sind.

**Erstellt von:** Sezen Akman (388988)

**Autor(en):** Liu, D., Bias, R. G., Lease, M., Kuipers, R.

**Titel:** Crowdsourcing for Usability Testing

**Jahr:** 2012

**Laufende Nr. im Literaturverzeichnis:** 19

**Schlagwörter (3-5):** Crowdsourced Usability Testing, Remote Usability Testing, Mechanical Turk, CrowdFlower

**Kurzbeschreibung:**

Das Paper untersucht die Vor- und Nachteile des Crowdsourcing als eine Form von Datenerhebung bei Usability Testing im Vergleich zum Labortest, indem ein Usability Testing von derselben Webseite mit jeweils einer Datenerhebungsform durchgeführt wird.

Das Paper definiert die Testungsziele der Usability Testing als:

- Identifikation von inkonsequenten Informationen und Usabilityfehlern auf der Hochschul-Webseite,
- Bewertung der Nutzerschnittstelle durch repräsentative Nutzer unter kontrollierten Testbedingungen,
- Ausarbeitung einer Baseline bzgl. Nutzerperformanz und Nutzerschnittstelle für zukünftige Usability Testings.

Probandenakquise: Zwei Crowdsourcing-Anbieter (zweimal von Amazon Mechanical Turk selbst und einmal über CrowdFlower, aber trotzdem aus AMT-Nutzerpool) übernehmen die Probandenakquise und Testdurchführung anhand des Testdesigns und der Testdokumentation der Autoren.

Testablauf: Die Testteilnehmer benutzen ihre eigenen Webbrowser an ihren eigenen Rechnern, um zuerst eine demographische Umfrage zu beantworten und dann zu vier Aufgaben weitergeleitet zu werden. Um diese Aufgaben zu lösen, suchen die Teilnehmer nach relevanten Informationen über das Studium auf der Webseite einer Hochschule. Ihre Antworten dokumentieren sie im Testformular auf AMT. Nach Absenden des Formulars füllen die Teilnehmer eine qualitative Befragung aus und bewerten ihr Nutzererlebnis auf der Webseite. Damit wird der Test abgeschlossen.

Methode: Bei dem Testdesign (demographische Umfrage, Aufgabengestaltung und qualitative Befragung) werden Methoden von Kapelner und Chandler (2010) genutzt, um die Qualität des Nutzerfeedbacks und die Motivation der Teilnehmenden zu erhöhen. Dabei wird darauf geachtet, dass die Fragestellungen keine zufälligen Antworten ermöglichen.

Durch drei Iterationen evaluiert das Paper, welche Aufgaben sich für Crowdsourced Usability Testing besser eignen, und untersucht, ob dadurch eine bessere Kontrolle auf experimentelle Bedingungen ausgeübt werden kann.

Crowdsourced-Pilot-Studie mit elf Probanden:

Demographische Befragung zeigt, dass Teilnehmer mit höherem Abschluss schneller sind, weil sie Erfahrung mit Hochschul-Webseiten haben. Durch nicht gelöste oder falsch gelöste Aufgaben wird verständlich, dass Teilnehmenden keine Rückfragen stellen und dadurch teilweise den Test nicht richtig abarbeiten können. Das ist ein deutlicher Hinweis dafür, dass hier die klare und eindeutige Formulierung der einzelnen Aufgaben in einer Crowdsourced Usability Testing viel wichtiger ist - im Vergleich zu einem Labortest, wo die Versuchsleiter Rückfragen von Probanden direkt beantworten können. Die Ergebnisse der qualitativen Befragung zeigen, dass Teilnehmer aus AMT nicht sonderlich motiviert sind, um offene Fragen mit detaillierten Rückmeldungen zu beantworten.

Zweite Studie mit 44 Probanden:

Probanden werden danach gefragt, ob sie sich für ein Hochschulstudium beworben haben, damit die relevanten Probanden bei der Auswertung selektiert werden. Nach Abschluss jeder Aufgabe mussten Probanden die Bearbeitungsdauer angeben, so dass Spammer und nicht motivierte Probanden herausgefiltert werden können. Anzahl der Fragen zum Nutzererlebnis werden erhöht und detaillierter formuliert, um konkrete Rückmeldungen zum Nutzererlebnis zu bekommen. Durch diese Maßnahmen werden 30% der Spammer detektiert. Änderungen in der qualitativen Befragung resultieren in einer erhöhten Qualität des Nutzerfeedbacks, wobei die Erhöhung der Incentivierung im Vergleich zur Pilot-Studie auch eine Rolle spielt.

Dritte Studie mit 50 Probanden:

Durch weitere Änderungen an Testdesign werden nur 8% der Probanden als Spammer detektiert, aber die Ergebnisse hinsichtlich der Erfolgsrate, der identifizierten Usability-Probleme und qualitativer Feedback weisen keine signifikanten Unterschiede zu den ersten und zweiten Testabläufen auf.

**Schwächen:**

Die Anonymität von Probanden und fehlender direkter Kontakt zwischen Probanden und Versuchsleitern während dem Test verursachen einen Mehraufwand bei dem Testdesign. Zusätzliche Detektionsmechanismen und gezieltere Fragestellungen sind notwendig, um Betrug der Probanden zu vermeiden und Nutzerfeedback von ausreichend guter Qualität zu bekommen. Wenn die Testergebnisse von vorherigen Erfahrungen und die demographischen Merkmale von Probanden stark beeinflusst werden können, ist Labortest eine besser geeignete Methode, weil man dabei diese Störeffekte besser kontrollieren kann. Generell ist es schwieriger, Probanden aus spezifischen, demographischen Segmenten zu rekrutieren, weil die Crowdsourced-Plattform die Angaben der rekrutierten Probanden nicht auf Validität prüfen kann. Für Erfassung von time-to-task oder Nutzerperformanz ist Crowdsourcing nicht geeignet, weil die stark beschränkte experimentelle Kontrolle eine realistische Abarbeitung der einzelnen Aufgaben nicht immer gewährleisten kann. Die Diversität der Tester ist im Vergleich zum Labortest an einer Universität deutlich erhöht, obwohl dieser Fakt in dieser Studie nicht als Stärke bezeichnet werden kann, weil die Testaufgabe eine gewisse Vertrautheit mit Hochschulprozessen benötigt.

**Stärken:**

Das Testverfahren Crowdttesting weist einen größeren Nutzen während der iterativen Produktentwicklung von großen Konzernen hinsichtlich Testkosten und Testdauer auf. Probandenakquise ist einfacher als die für einen Labortest, weil Kommunikations- und Organisationsaufwand durch den Crowdsourcing-Anbieter stark reduziert sind. Das ermöglicht auch, dass sich viel mehr Probanden bereit erklären. Dadurch kann eine größere Menge an Daten erfasst werden. Fast keine Fixkosten und Betriebskosten werden benötigt, und dadurch ist es möglich, das Testdesign auf Testungsziele genauer anzupassen und mehrere Störeinflüsse und Verzerrungen durch schnellere und günstigere Wiederholungen zu eliminieren. Auf längere Sicht ist es einfacher möglich, dass die Prozesse von Crowdsourced Usability Testing schneller an die Testungsziele angepasst werden und dadurch bessere Ergebnisse als Labortests leisten. Weil das Testdesign viel komplexer und durchdachter formuliert werden soll, kann man auch an Auswertungsaufwand am Ende des Tests sparen.

**Erstellt von:** Sezen Akman (388988)

**Titel:** Understanding the Crowd: Ethical and Practical Matters in the Academic Use of Crowdsourcing

**Jahr:** 2017

**Laufende Nr. im Literaturverzeichnis:** 21

**Schlagwörter (3-5):** Ethics, AMT, Crowd, crowdworker, microtask

**Kurzbeschreibung:**

Das Kapitel stellt sich die Frage, ob Crowdsourcing neue ethische Konsequenzen verursacht, mit denen man sich in traditionellen Usability Testings nicht auseinandersetzen musste. Während man Kosten von Usability Testing mit dieser Methode drastisch reduzieren kann, kreierte man gleichzeitig eine neue Form eines Akkordlohns. Deswegen muss darauf geachtet werden, dass „crowdworkers“ eine angemessene Incentivierung für die „microtasks“ bekommen. Dabei wird auch hinterfragt, wie weit die Ansätze von Bürgerwissenschaft (offene Wissenschaft) ausgenutzt wird, um die eventuelle Ausnutzung der „crowdworkers“ durch geringere Incentivierungen unter Mindestlohn zu begründen. Um Motivation und Blickweise von „crowdworkers“ besser zu verstehen, wird die demographische Hintergründe von Mitglieder von einer Crowd Testing Plattform Amazon Mechanical Turk untersucht. Deren Beziehung zum Arbeitgeber („requester“ als Forscher, Wissenschaftler, Entwickler oder Unternehmer) und deren Arbeitsorganisation (Arbeitsplatz, Inhalt deren Arbeit) werden kritisch hinterfragt und deren Teilnahme und Beziehung zum sozialen Netzwerken erfasst. Durch diese Erkenntnisse wird es thematisiert, dass diese Informationen bei Richtlinien und praktische Herangehensweisen angewendet werden sollen.

**Erstellt von:** Sezen Akman (388988)

**Autor(en):** Mushtaq, M. S., Augustin, B., Mellouk, A.

**Titel:** Crowd-sourcing Framework to Assess QoE

**Jahr:** 2014

**Laufende Nr. im Literaturverzeichnis:** 24

**Schlagwörter (3-5):** QoE, QoS, Crowdsourcing, QoE Evaluation, subjektive Methode, Multimedia, Video Streaming

**Kurzbeschreibung:**

Das Paper handelt von der Evaluation der Qualität vom Nutzererlebnis in Bezug auf Multimedia Dienste, wobei man Crowd Testing als eine effiziente, subjektive Testmethode in realen Bedingungen anwenden kann. Weil Netzwerkparametern und Systemeigenschaften von den Geräten, an den die Tester diese Dienste testen, die wahrgenommene Qualität sehr stark beeinflussen können, müssen diese während des Usability Testings kontrolliert werden. Bei der Methode Crowd Testing ist es besonders schwierig, diese Bedingungen bei zahlreichen Tester orts- und zeitunabhängig zu kontrollieren. Um diese Problematik zu lösen, ein Werkzeug wird entwickelt, mit dem wichtige Netzwerkparametern, Systemeigenschaften und weitere relevante Informationen von einzelnen Nutzern während des Tests automatisch aufgezeichnet werden, die bei der Berichterstellung ggf. auch automatisch ausgewertet werden kann. Dieses Regelwerk basiert auf das Konzept „living laboratory“, das besagt, dass das Nutzererleben von freiwilligen Probanden innerhalb deren eigenen, realen Lebenssituationen unter einer experimentellen Kontrolle wie bei einem Laborexperiment erfasst werden muss. Diesen Ansatz wird operationalisiert, indem einer der größten Online Video Streaming Anbieter „Youtube“ für die Versuche benutzt wird, der die realen Nutzerbedingungen am besten repräsentiert. Das Werkzeug ist ein Fireworks Plug-In, die das Video auf der Anbieter-Webseite automatisch erkennt, und ein Button unter das Video anzeigt. Der Proband ist aufgefordert, diesen Button jedes Mal zu drücken, wenn sie mit der Video-Qualität nicht glücklich ist. Gleichzeitig werden sowohl von der Anbieter-Webseite als auch vom Endgerät des Probanden relevante technische Daten gesammelt, unter anderem, die Qualität der Internetverbindung. Wenn der Proband das Video bis Ende anschaut, wird durch einen Menu-Button das Bewertungsformular erreicht, die ausgefüllt in dem Plug-In gespeichert werden.

**Schwächen:**

Das Tool ist nur für ein Anwendungsbereich (Online Video Streaming) entwickelt. Die Richtlinien zur Bewilligung von Probanden ist unklar, die die automatische Erfassung und Speichern deren privaten Daten autorisiert.

**Stärken:**

Die Erfassung von relevanten technischen Daten ermöglicht, dass die Testergebnisse durch Ausfiltern von zahlreichen Kombinationen von Störeinflüssen valider sind, und tatsächlich das widerspiegeln, was auch bewertet werden soll. Reliabilität der technischen Daten ist gegeben.

**Erstellt von:** Sezen Akman (388988)

**Autor(en):** Eickhoff, C., & de Vries, A. P.

**Titel:** Increasing Cheat Robustness of Crowdsourcing Tasks

**Jahr:** 2012

**Laufende Nr. im Literaturverzeichnis:** 6

**Schlagwörter:** Trügerisches Verhalten in Crowdsourcing, Cheating, A priori Maßnahmen

**Kurzbeschreibung:** Die Autoren dieses Papers untersuchen ob in Crowdsourcing Studien täuschendes Verhalten bereits durch die Aufgabengestaltung reduziert werden kann. Arbeiter von verschiedenen Crowdsourcing Plattformen nahmen hierzu als Teilnehmer an der Erhebung teil. Zunächst wurden entweder Webseiten hinsichtlich ihrer Relevanz für Suchmaschineneingaben bewertet werden, oder hinsichtlich ihrer Angemessenheit für junge Kinder bewertet werden. Die abhängige Variable war stets die Anzahl an „Betrüger“. Die Ergebnisse zeigen, dass bei der simpleren Relevanz-Einordnungsaufgabe deutlich mehr trügerisches Verhalten gezeigt wird, als bei der kreativeren Angemessenheitsbewertungsaufgabe. In Folge dessen wurden die weiteren Erhebungen nur noch mit dem ersten Aufgabentyp fortgeführt. Weiterhin weisen die Ergebnisse auf, dass eine erhöhte Variabilität der Kontextfaktoren von Testaufgaben mit einer geringeren Rate an trügerischem Verhalten einhergeht. Ein Grund hierfür könnte die Schwierigkeit des Erstellens automatischer Bots sein, weshalb mehr vertrauenswürdige Teilnehmer an solchen Studien teilnehmen. Ebenfalls das vorherige Filtern kann einen Einfluss auf trügerisches Verhalten haben. So ist das Filtern der Teilnehmer nach Herkunft oder mit Hilfe von Pretests erfolgreich im Reduzieren von Betrüger Quoten, jedoch dauern die Erhebungen der Studien deutlich länger als wenn kein Filtern der Teilnehmer stattfindet. Letztlich zeigt sich, dass je größer der Stoß an sich ähnelnden Aufgaben ist, desto mehr „Betrüger“ angezogen werden hieran teilzunehmen, da ein größeres Potenzial für Bots und automatisierte Lösungen vorhanden ist.

**Stärken & Schwächen:** Das Paper untersucht geht auf Aspekte ein, die beim Erstellen von Crowdsourcing Studien von Vorteil sein können. Obwohl Crowdsourcing Studien viele Vorteile, wie bspw. schnelle und viele Ergebnisse einer breiten Masse, mit sich bringen, muss die potenziell trügerische Seite der Testteilnehmer beachtet werden. Eine Stärke der Studie ist die erhobene Stichprobe. Es handelt sich hierbei um tatsächliche Arbeiter von Crowdsourcing Plattformen, die unwissend an der Studie teilnahmen. Als Schwäche kann die Testung von lediglich zwei Arten von Testdesigns gesehen werden. Beide anfangs verwendeten Designs beinhalteten keine offenen Fragen. Es bleibt unklar inwiefern dieses Antwortschema die Rate von „Schummeln“ verändert hätte. Weiterhin bleiben weitere Fragen ungeklärt, bspw. inwiefern ein verändertes Entlohnungssystem die Betrugsrate verändern könnte. Die Studie liefert jedoch viele Ansatzpunkte für weiterführende Forschung, die mit der Zunahme an internetbasierten Testungen von hoher Wichtigkeit sind.

**Erstellt von:** Elisabeth Neuer, Matrikelnummer: 389184

**Autor(en):** Findlater, L., Zhang, J., Froehlich, J. E. & Moffatt, K.

**Titel:** Differences in Crowdsourced vs. Lab-based Mobile Desktop Input Performance Data

**Jahr:** 2017

**Laufende Nr. im Literaturverzeichnis:** 7

**Schlagwörter:** Crowdsourcing vs. Laborstudien, Desktop vs. mobile Eingabe

**Kurzbeschreibung:** Diese Studie vergleicht Ergebnisse aus Online-Studien und Laborstudien in einem 2x2 Design. Teilnehmer nahmen entweder im Labor oder von Zuhause via Crowdsourcing teil. Dabei wurden simple interaktive Aufgaben entweder an einem Desktop Computer mit Maus, oder auf einem Tablet mit Touchscreen gelöst. In einer ersten Untersuchung wurden Fehlerrate und Eingabezeit der einzelnen Gruppen verglichen. Dabei wurde der Tablet Gruppe die explizite Anweisung gegeben das Tablet für die Bearbeitung der Aufgaben flach auf einen Tisch zu legen. Es zeigt sich, dass für Crowdsourcing Teilnehmer die Fehlerrate in der Tablet Bedingung höher ist als für jene in der Desktop Bedingung. Für Laborteilnehmer ergeben sich gegenteilige Ergebnisse: Die Fehlerrate der Desktop Bedingung ist höher als in der Tablet Bedingung. Während Teilnehmer der Labor Bedingung mehr Zeit für die Eingabe benötigen, sind Crowdsourcing Teilnehmer schneller, aber auch wenig genau. In einer zweiten Untersuchung wurden zusätzlich Neigung und Bewegungen des Tablets erhoben. Hierzu diente eine zusätzliche Gruppe an Crowdsourcing Tabletutzern, die keine explizite Anweisung zur Positionierung des Tablets erhielt, sowie die bisherigen Teilnehmer der beiden Tablet Gruppen. Die Ergebnisse zeigen, dass weniger experimentelle Kontrolle mit mehr Tablet Bewegung und Rotation einhergeht. Die Autoren folgern daraus, dass die hohe Fehlerrate der Crowdsourcing Tablet Gruppe der ersten Erhebung eine Folge von Positionsproblemen des Tablets bzw. Körpers sein können, als die Anweisung des fixierten Tablets noch gegeben war. Zwischen den beiden Crowdsourcing Tablet Gruppen wurden keine Unterschiede in Zeit oder Fehlerrate gemessen. Dies suggeriert, dass Unterschiede zwischen Labor- und Crowdsourcing Erhebungen größer sind, als kleine Veränderungen der Crowdsourcing Anweisungen.

**Stärken & Schwächen:** Die Studie weist darauf hin, dass Daten, die via Crowdsourcing erhoben wurden nicht gleichzusetzen sind mit Daten aus Laborstudien. Die Zeit, die zum Teil bei Crowdsourcing Erhebungen gewonnen wird, kann auf Grund höherer Fehlerraten nichtig werden. Ungeklärt bleiben die Gründe für die erhöhte Fehlerrate. Sollen möglichst schnell viele Daten in einem realen Setting erhoben werden, so eignen sich Crowdsourcing Studien sehr gut. Allerdings müssen stets Qualitätsabstriche der Ergebnisse auf Grund von bspw. extrinsisch motivierter Teilnehmer einkalkuliert werden. Weiterhin sollte beachtet werden, dass Teilnehmer Aufgaben nicht auf einem starren Endgerät bearbeiten, sondern dieses ständig anders positionieren. Die Ergebnisse dieser Studie zeigen zwar, dass es keinen Zusammenhang zwischen der Fehlerrate und der Bewegung des Gerätes besteht, allerdings wurde diese Erhebung mit Tablets erhoben. Unklar ist wie Ergebnisse von Crowdsourcing Studien, die auf einem deutlich kleinerem Smartphone bearbeitet wurden, ausfallen würden. Smartphones werden noch stärker bewegt und rotiert, und auch außerhalb des Hauses ständig in Gebrauch. Teilnehmer, die sich durch Crowd-Studien schnell ein paar Euro dazuverdienen wollen, könnten die Studien bspw. nebenbei in der Bahn oder beim Warten in der Einkaufsschlange bearbeiten. Neben den erhöhten Bewegungen des Smartphones, müssten in solchen Szenarien noch mehr Umweltfaktoren beachtet werden, die einen Einfluss auf die Ergebnisqualität von Crowd-Studien haben könnten. Wichtig wäre auch die Aufmerksamkeit und Konzentration, die der Studienaufgabe in solchen Situationen gewidmet wird, und wie diese variiert mit steigender Komplexität der Aufgaben.

**Erstellt von:** Elisabeth Neuer, Matrikelnummer: 389184

**Autor(en):** Guaiani, F. & Muccini, H.

**Titel:** Crowd and Laboratory Testing. Can they co-Exist? An Exploratory Study

**Jahr:** 2015

**Laufende Nr. im Literaturverzeichnis:** 11

**Schlagwörter:** Labor- und Crowdttesting

**Kurzbeschreibung:** Das Paper untersucht die Fragestellung ob Laboruntersuchungen und Crowdttesting Maßnahmen sich gegenseitig ergänzen können. Anhand einer Online-Umfrage wurde untersucht welche Art von Tests Crowdttester in der Regel bearbeiten, welche Herausforderungen und Limitationen Crowdttesting aufweist, und wie sich Crowd- und Labortestungen ergänzen können. Teilnehmer dieser Umfrage waren Testersteller, bzw. erhebende Tester, die bei verschiedenen Crowdttesting Plattformen angestellt waren. Es zeigt sich, dass Crowdttesting Erhebungen vor allem für funktionale Tests, Usability Tests sowie Leistungs- und Systemtests verwendet werden, für Stresstests, Unit und Integrationstests, sowie Sicherheitstests jedoch weniger gut geeignet sind. Als Herausforderungen von Crowdttesting erwähnten die Probanden einen Mangel an Informationen über die zu testende Applikation erwähnt, sowie Kommunikation- und Zeitdruckprobleme im Rahmen ihrer Tätigkeiten. Mehr als die Hälfte aller Teilnehmer gaben an, dass sie einen Vorteil in der Ergänzung von Labor- und Crowdttestungen sehen, aber nur 28% der Teilnehmer kannten tatsächlich Fälle in denen dieses eingetreten war.

**Stärken & Schwächen:** Eine Schwäche der Studie ist die kleine Anzahl von 18 Probanden. Auf Grund der kleinen Stichprobe haben die Ergebnisse nur geringe Power. Des Weiteren bezog sich eine der Hauptfragen auf die Schwächen und Limitationen von Crowdttesting. Anhand der erhobenen Daten lässt sich allerdings nur auf die Herausforderungen der Crowdttester schließen, die diese im Rahmen ihrer Arbeit wahrnehmen. Man erfährt nicht welche konzeptuellen Schwierigkeiten Crowdttesting mit sich bringt, und wie diese Schwierigkeiten ggf. durch Laboruntersuchungen behoben werden können. Der Titel der Studie deutet mehr an, als letztendes berichtet wird. Man erfährt nicht ob die zwei Erhebungsmethoden sich ergänzen könnten, sondern lediglich wie das Stimmungsbild einzelner Tester zu dieser Idee ist, und ob ihnen solche integrativen Studien bekannt sind. Andererseits wird die Studie auch als „explorative“ Studie betitelt, die zu weiterer Forschung anregen soll. Die Autoren selbst sehen die Schwachstellen innerhalb ihrer Studie und planen in Zukunft umfassender zu erheben.

**Erstellt von:** Elisabeth Neuer, Matrikelnummer: 389184

**Autor(en):** Gadiraju, U., Kawase, R., Dietze, S., & Demartini, G.

**Titel:** Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys

**Jahr:** 2015

**Laufende Nr. im Literaturverzeichnis:** 8

**Schlagwörter:** Crowdsourcing, Nutzerverhalten, heimtückische Absichten

**Kurzbeschreibung:** Das Paper befasst sich mit heimtückischen („malicious“) Verhalten auf Crowdsourcing Plattformen. Als heimtückisch werden Teilnehmer mit anderweitigen Motiven bezeichnet, die entweder den Test sabotieren wollen oder durch unzureichende Aufgabebearbeitung möglichst schnell ihre monetäre Entschädigung erhalten wollen. Sie sind abzugrenzen von unzuverlässigen Teilnehmern, jene Teilnehmer, die falsche Antworten bei einer oder mehreren simplen Aufmerksamkeitsüberprüfungsfragen liefern. Die Autoren befassen sich in ihrer Studie mit drei Fragestellungen. Es wird erörtert ob unzuverlässige Teilnehmer andere Methoden anwenden um Aufgaben zu bearbeiten. Des Weiteren wird untersucht ob Tester vorab von diesem Wissen profitieren können, und ob Verhaltensmuster heimtückischer Teilnehmer identifiziert und quantifiziert werden können. Die Autoren erhoben Daten von 1000 Teilnehmer einer Online Umfrage, die neben offenen Fragen Aufmerksamkeitsüberprüfungsaufgaben enthielt. Die Ergebnisse identifizieren fünf Verhaltenstypen: Unzulässige Teilnehmer, schnelle Betrüger, Regelbrecher, geschickte Betrüger und Gold-Standard „Opfer“ (aus dem engl. prey). Die Typen unterscheiden sich darin ob Anweisungen eingehalten und valide Antworten geliefert werden, oder ob auf Grund von Unaufmerksamkeiten simple Fehler gemacht werden, bzw. ob die Teilnehmer die notwendigen Voraussetzungen für den Test erfüllen. Auf Basis dieses Wissens können Tester bessere Tests erstellen. Erhebungen mit offenen Fragen bspw. sollten strenge Antwortkontrollen haben, da sie besonders empfänglich sind für Regelbrecher und schnelle Betrüger. Schnelle Betrüger geben unzulässige Antworten um mit wenig Aufwand ihren Gewinn zu maximieren. Zur Identifizierung von Heimtücke („maliciousness“) beziehen sich die Autoren auf die Angemessenheit der einzelnen Antworten eines Teilnehmers, und stellen eine Formel zur Quantifizierung dieser vor. Die Identifikation von heimtückischen Teilnehmer kann der Filterung unbrauchbarer Ergebnisse dienen. Auf Grund ihrer Einfachheit kann diese Formel auch auf andere Aufgabentypen angewendet werden. Ebenso sind die identifizierten Verhaltenstypen für andere Erhebungsarten als offene Fragen anwendbar.

**Stärken & Schwächen:** Dieses Paper betont eine der Schwierigkeiten von Crowdttesting Studien ein, deren Versuchsteilnehmer häufig durch extrinsische Faktoren motiviert werden, was zu unbrauchbaren Ergebnissen führen kann. Obwohl Maßnahmen vorgestellt werden um unterschiedliche heimtückische Teilnehmer zu identifizieren und Tests entgegen dieser Verhaltensweisen zu gestalten, gibt es keine Möglichkeiten jegliche Betrüger auszuschließen. Diese Tatsache muss bei jeder Crowd-Studie beachtet und miteingerechnet werden. Dennoch liefern die Ergebnisse dieser Studie, einschließlich dem Ansatz zur Quantifizierung von heimtückischem Verhalten, umfassendes Wissen um Crowd-Studien besser zu gestalten. Zuvor sollte jedoch bedacht werden, ob die Ergebnisse dieser Studie sich auch auf andere Aufgabentypen als Umfragen übertragen lassen. Möglicherweise lassen sich heimtückische Teilnehmer in Interpretations- oder Analyseaufgaben anders charakterisieren.

**Erstellt von:** Elisabeth Neuer, Matrikelnummer: 389184

**Autor(en):** Hoßfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K. & Tran-Gia, P.

**Titel:** CrowdTesting: A Novel Methodology for Subjective User Studies and QoE Evaluation

**Jahr:** 2013

**Laufende Nr. im Literaturverzeichnis:** 14

**Schlagwörter:** Crowdttesting, Crowdsourcing, Nutzerreliabilität, Vor- und Nachteile, Anwendungseinschränkungen

**Kurzbeschreibung:** Das Paper stellt Crowdttesting als Erhebungsmethode für Erlebnisqualität (nachfolgend: Quality of Experience, QoE) vor. QoE wird als empfundener Grad an Freude oder Ärger eines Nutzers über ein System oder eine Applikation definiert. Dabei sind Kontext-, Nutzer-, System- und Inhaltsfaktoren mögliche Einflussfaktoren auf QoE. Beim Crowdttesting bedient man sich der Crowdsourcing Methode: Testaufgaben werden vom Labor ins Internet verlagert. Hierbei ergeben sich Vorteile, zu denen die Autoren Kosten- und Zeitersparnisse sowie mehr und vielfältigere Versuchsteilnehmer zählen. Das Paper geht auch auf die Einschränkungen von Crowdttesting ein, und betont die Reliabilität der Erhebungsergebnisse. Als Gründe für mangelnde Reliabilität werden technische Fehler, unklare Testinstruktionen, Sprachbarrieren bzw. Verständnisprobleme, sowie „böswillige“ Teilnehmer, die auf eine schnelle Geldentschädigung hoffen, vorgestellt. Die Autoren geben in ihrem Paper gezielte Vorschläge zur Reliabilitätssicherung: Beim Design der Testaufgaben sollten, sofern möglich, „gold standard data“ (Fragen mit eindeutig richtigen Lösungen), Konsistenz- und Inhaltsüberprüfungsaufgaben eingebaut werden. Des Weiteren wird die Bedeutung von Testanreizen, Bezahlungssystemen, demografischen Daten, Hard- und Software, sowie Kontext Monitoring angesprochen. All diese Punkte können Ergebnisse von Crowdttesting Studien beeinflussen und ggf. verfälschen. In Folge einer statistischen Evaluation bekräftigen die Autoren die Notwendigkeit des frühen Filterns „nicht-reliabler“ Versuchsteilnehmer, und geben hierauf basierend Design-Empfehlungen für Crowdttesting Kampagnen. Ein 2-Phasen Design, in dessen ersten Phase reliable von nicht-reliablen Teilnehmern getrennt werden, wird vorgeschlagen. Teilnehmer, die als reliabel eingestuft wurden, werden in die zweite Phase, welche die eigentliche Erhebung beinhaltet, eingeladen.

**Stärken & Schwächen:** Crowdttesting ermöglicht höhere Teilnehmerzahlen, kürzere Erhebungszeiten und niedrigere Kosten als klassische Laborversuche. Dennoch birgt die Methode auch Herausforderungen. Aufgaben und Interfaces müssen klar verständlich und fehlerfrei konzipiert werden, da im Falle von Unklarheiten keine Testleiter für Rückfragen vorhanden sind. Treten dennoch Fehler auf, könnten diese die erhobenen Ergebnisse verfälschen. Weiterhin müssen Teilnehmerfaktoren beachtet werden. Nicht-gewissenhaft arbeitende Teilnehmer sollten möglichst früh identifiziert und aus der Erhebung herausgefiltert werden. All diese Maßnahmen erfordern neben Wissen eine gezielte Erstellung der Aufgaben, welche mit zeitlichem Aufwand verbunden ist. In Anbetracht dessen muss abgeschätzt werden ob der Mehraufwand tatsächlich noch geringer ist als der einer klassischen Laboruntersuchung. Weiterhin ist Crowdttesting nicht für alle Arten von Aufgaben geeignet. Die Autoren dieses Papers beziehen sich auf die QoE von Videos, welche sich gut ins Crowdttesting Format umwandeln lassen. Andere Test, bspw. 3D Tests oder haptische Erhebungen lassen sich schwierig außerhalb von Laboren realisieren, da Teilnehmer keine notwendige Hardware besitzen. Generell muss die Homogenität von Hard- und Software bei Crowdttesting gewährleistet oder als möglicher Störfaktor miterhoben werden. Zusammenfassend gibt es viele Punkte, die beim Anwenden von Crowdttesting Studien bedacht werden müssen, ebenso wie die Fragestellung, ob der Mehrwert des Crowdttestings zuletzt noch gegeben ist.

**Erstellt von:** Elisabeth Neuer, Matrikelnummer: 389184

**Autor(en):** Mok, R. K. P., Chang, R. K. C., & Li, W.

**Titel:** Detecting Low-Quality Workers in QoE Crowdfunding: A Worker Behavior-Based Approach

**Jahr:** 2017

**Laufende Nr. im Literaturverzeichnis:** 22

**Schlagwörter:** Low Quality Teilnehmer, Nutzerverhalten, heimtückisches Verhalten identifizieren, Crowdfunding

**Kurzbeschreibung:** Die Autoren dieser Studie stellen eine neue Methode zur Identifikation von low-quality Teilnehmern in Crowdfunding Studien, in der sie sich unter anderem auf Mauszeiger Bewegungen beziehen, vor. Die Methode wurde auf Grundlage einer Quality of Experience (QoE) Messung von Videos entwickelt. Teilnehmer bewerteten Videos anhand von Single-Item, Multiple-Choice und offenen Fragen. Zusätzlich wurden zehn Maße i.S.v. Mikro-Bewegungen des Mauszeigers sowie zeitlichen Informationen erhoben. Die Autoren begründen, dass mit Hilfe dieser Daten kognitive Prozesse der Teilnehmer erfasst werden können. Es zeigt sich, dass die folgende vier Maße am besten geeignet sind um das Verhalten der Teilnehmer widerzuspiegeln: 1. Teilbewegungen, die vom Beginn der Aufgabe bis zum ersten Klicken gemacht werden. 2. Die mittlere Zeit Mausbewegungszeit zwischen einzelnen Aufgaben 3. Zahl der zusätzlich zur Aufgabe notwendig gemachten Klicks 4. Mittlere Zeigergeschwindigkeit. Es wird jedoch auch betont, dass diese Maße nur für bestimmte Ratingskalen am besten geeignet sind, und bei anderen Skalen ggf. andere Maße verwendet werden sollten. Des Weiteren stellen die Autoren ein quantifizierendes Maß für die Qualität eines Teilnehmers vor. Dieses berechnet sich anhand der Komplexität der Texteingaben, dem Einhalten von Anweisungselementen und der Widersprüchlichkeit von Antworten. Mit Hilfe dieser beiden Maße können low-quality Teilnehmer zu 80% aufgedeckt werden, mit einer Fehlerrate von nur 30%. Dennoch betonen die Autoren, dass ihre Methode sich bisher nicht generalisieren lässt und für andere Erhebungstypen getestet werden muss.

**Stärken & Schwächen:** Die hier vorgestellte Methode bezieht bei der Detektion von low-quality Teilnehmer erstmalig Mauszeigerbewegungen und -bewegungszeiten mit ein. Diese Art von Information lässt sich ohne zusätzliche Items, wie z.B. Reliabilitäts-Überprüfungsfragen, erheben und ohne den Teilnehmern bewusst zu sein. Somit können Teilnehmer nicht versuchen die Qualitätsüberprüfungen zu betrügen. Allerdings ist die Methode spezifisch auf diese Erhebung und diesen Studientyp zugeschnitten worden. Bisher kann diese Methode noch nicht für andere Studientypen und andere Ratingskalen verwendet werden. Sie liefert jedoch einen guten Ansatz für die Identifikation der Nutzerreliabilität bei Crowdfunding Studien, und sollte ausgeweitet werden, da Crowdfunding im Zuge der Digitalisierung immer relevanter wird.

**Erstellt von:** Elisabeth Neuer, Matrikelnummer: 389184

**Autor(en):** Blohm, I., Leimeister, J. M., Krcmar, H.

**Titel:** Crowdsourcing: How to Benefit from (Too) Many Great Ideas

**Laufende Nummer im Literaturverzeichnis:** 5

**Schlagwörter:** big data, crowdsourcing, plattform, problem

**Kurzbeschreibung:** Der Artikel beschäftigt sich mit der Frage, wie man am effektivsten wertvolle Daten, aus den aus Crowdsourcing entstehenden zahlreichen Daten (Big Data), gewinnen kann. Als Beispiel werden drei verschiedene Firmen verglichen, von welchen zwei eigene Crowdsourcing-Plattformen besitzen. Besonders wird auf den Umgang und die Gestaltung dieser Plattformen eingegangen. Als Problem wird angeführt, dass die dort generierten Daten leicht zu Überforderung der bearbeitenden Mitarbeiter führen. Denn die Auswertung solch großer Datensätze ist meist sowohl zeit- als auch kostenintensiv. So sollte man das Arbeiten mit der Crowd nicht nur in den kompletten Arbeitsprozess etablieren, sondern auch beginnend bei der Personalauswahl entscheiden wer sich für eine solche Arbeit eignet. Denn nicht jeder Mitarbeiter besitzt die Motivation, oder ist in der Lage mit solch großen Datensätzen umzugehen. Entweder man macht die Arbeit mit Crowdsourcing besonders attraktiv, oder man führt eine kleine motivierte Gruppe zusammen. Ebenso wichtig wie die Personalauswahl, ist das Design der Crowdsourcing-Plattform. Erstellt man diese nach den vielen Ansätzen die in dem Artikel aufgeführt werden, so übernimmt die sich bildende Community einige Selektionsarbeit. So führen zum Beispiel mehrschrittige Einsendungssequenzen dazu, dass Spam vermieden wird. Durch das Einführen einer Hierarchie, oder eines Punktesystems, werden die Mitglieder zu besonders qualitativ hochwertigen Ideen angeregt. Außerdem kann so ein Bewertungssystem eingeführt werden, wodurch besonders schlechte Ideen schnell herausgefiltert werden können. Der Artikel gibt mithilfe der drei Fallbeispiele viele fundierte Ansätze für den Umgang mit Big Data aus dem Crowdsourcingbereich, genauso wie er die Problematik eben dieser Datenanalyse beleuchtet.

**Erstellt von:** Mareen Bentrup, 396332

**Autor(en):** Hoßfeld, T., Hirth, M., Korshunov, P., Hanhart, P. Gardlo, B., Keimel, C., Timmerer, C.

**Titel:** Survey of web-based crowdsourcing frameworks for subjective quality assessment

**Laufende Nummer im Literaturverzeichnis:** 14

**Schlagwörter:** crowdsourcing, platform, framework, problems

**Kurzbeschreibung:** In diesem Artikel geht es um die technischen Hindernisse, die sich stellen, wenn Crowdsourcing betrieben wird. Zudem geben die Autoren einen guten und fundierten Überblick über verschiedene Croudsourcingplattformen und Frameworks. Die erste beschriebene Schwierigkeit besteht darin einen geeigneten Probandenpool für die Testung zu finden. Hier bieten verschiedene Plattformen unterschiedliche Möglichkeiten. Die erste Gruppe solcher Plattformen ist die Aggregator Plattform. Zu ihnen zählen z.B. Crowdfower und Crowdsource. Sie bieten vor allem eine Auswertung der Daten an, sind jedoch durch ihre hohe Hardwareabstraktion auf wenige Möglichkeiten limitiert. Ebenso kann durch die plattformeigene Filterung eine homogene Crowd zustande kommen. Die zweite Gruppe sind die sogenannten spezialisierten Plattformen, diese begrenzen sich entweder auf bestimmte Gruppen von Menschen, z.B. bestimmte Berufsgruppen, oder bestimmte Prozesse, z.B. Standortgebundene Probandenakquise, wie bei Taskrabbt. Die dritte Gruppe sind die Crowdprovider, sie sind besonders flexibel und besitzen die größten Crowds, jedoch ist durch die Direktheit dieser Plattform keine Filterung der Probanden gegeben und somit mit Einschränkungen in der Qualität zu rechnen. Zu dieser Gruppe zählen Plattformen wie Mikroworkers, TaskCN, aber auch Facebook. Die zweite Schwierigkeit geht mit der Testumgebung einher, welche im Falle des Crowdsourcings nicht, wie im Labor, kontrolliert werden kann. Faktoren wie Auflösung, Internetgeschwindigkeit und Lichtverhältnisse können die Ergebnisse verfälschen. Um die technischen Störvariablen zu kontrollieren gibt es sogenannte Frameworks. Mit ihnen kann man Testungen nicht nur erstellen, sondern auch ausführen und die dabei entstehenden Daten sammeln. Der Vorteil den diese Frameworks bieten, ist im Gegensatz zu einer eigenen Erstellung, dass verschiedene Frameworks sich unterschiedlichen technischen Problemen stellen. So fokussiert sich z. B. CrowdMOS auf Audiotestungen. Dennoch fehlt ein ganzheitliches Framework, denn viele Anbieter spezialisieren sich nur auf eine technische Problematik.

**Erstellt von:** Mareen Bentrup, 396332

**Autor(en):** Leicht, N. Blohm, I., Leimeister, J.M.

**Titel:** Leveraging the Power of the Crowd for Software Testing

**Laufende Nummer im Literaturverzeichnis:** 17

**Schlagwörter:** software, guidelines, betatesting, crowdtesting in companies

**Kurzbeschreibung:** Der Artikel beschreibt drei unterschiedliche Möglichkeiten des Crowdtestings für Software: Das Testen mit einer externen Crowd aus Internetusern, einer Crowd aus den eigenen Angestellten und einer Crowd aus Kunden. Für jedes dieser Verfahren werden besonders die Vorteile betont, die sie voneinander differenzieren und daraus abgeleitet auf ihre optimalen Einsatzgebiete geschlossen. So besteht der Vorteil einer Crowd aus Internetusern in ihrer leichten Zugänglichkeit und der Zugänglichkeit zu schnellen und kostengünstigen Ergebnissen. Nutzt man hingegen eine Crowd aus den eigenen Angestellten, ist besonders die deutlich größere Akzeptanz der neuen Software zu betonen und bei einer Crowd aus Kunden die neu entstehende Unternehmen-Kunden Beziehung. Aus den gewonnenen Erkenntnissen werden in dem Artikel nachfolgend fünf Richtlinien für das Implementieren und die ersten Schritte des Crowdtestings erstellt und definiert. Diese sind vorrangig für den Gebrauch in Firmen gedacht, aber dennoch auch auf andere Situationen gut übertragbar. Des Weiteren wird der Unterschied zwischen Crowdtesting und Betatesting einfach und prägnant erläutert. Als gravierendster Unterschied kristallisierte sich der Zeitpunkt der Testung heraus. Wo beim Betatesting schon ein fast fertiges Produkt am Ende seiner Entwicklung getestet wird, findet das Crowdtesting in jedem Entwicklungsschritt einen wirkungsvollen Nutzen.

**Erstellt von:** Mareen Bentrup, 396332

**Autor(en):** Mok, R.K.P., Li, W., Chang, R.K.C

**Titel:** A user behavior based cheat detection mechanism for crowdtesting

**Laufende Nummer im Literaturverzeichnis:** 23

**Schlagwörter:** cheating, cursor, reliability, cheat-detection

**Kurzbeschreibung:** Das Paper beschäftigt sich mit der Problematik des Schummelns (Cheating) im Bereich des Crowdtestings. Als Schummler werden Teilnehmer bezeichnet, welche zwecks Profitsteigerung mit minimalem Aufwand schnell minderwertige Ergebnisse liefern. Mithilfe einer eigens entwickelten Technik sollen diese Schummler herausgefiltert werden und so die Reliabilität des Crowdtestings erhöht werden. Da die bereits vorhandenen Anti-Cheating Methoden nur automatische Antworten herausfiltern und leicht von (semi-) menschlichen Testern erkannt und umgangen werden können ist die Notwendigkeit für dieses Programm gegeben. Ebenso versagen die vorhandenen Methoden bei Testungen kein ausschließlich absolutes Antwortschema besitzen. Der Ansatz, den die Autoren verfolgen, bezieht sich auf den Bewegungsverlauf des Mauszeigers während der Testung. Es wird angenommen und im Verlauf der Studie bestätigt, dass Schummler ein deutlich abweichendes Bewegungsverhalten zeigen. In Hinblick auf diese Erkenntnis entwickeln sie ein Programm, welches mit Hilfe einer zuvor angelegten Datenbank bestimmte Bewegungsparameter vergleicht und so Schummler herausfiltert.

**Erstellt von:** Mareen Bentrup, 396332

**Autor(en):** Redi, J., Pova, I.

**Titel:** Crowdsourcing for rating image aesthetic appeal: Better a paid or a volunteer crowd?

**Laufende Nummer im Literaturverzeichnis:** 26

**Schlagwörter:** paid, volunteer, reliability, crowdsourcing

**Kurzbeschreibung:** Testungen per Crowdsourcing bieten eine günstige, schnelle und ertragreiche Alternative zu anderen Befragungsmethoden an, dennoch treten auch bei dieser Methode Probleme auf, welche es zu lösen gilt. So zeigt, neben der Technik und den Testinstruktionen, besonders die Wahl der Probanden eine große Fehleranfälligkeit. Die Studie knüpft an letzterem Problem an. Es wird ein Vergleich zwischen einer Crowd aus freiwilligen und einer aus bezahlten Testteilnehmern im Hinblick auf die Qualität und Quantität der Ergebnisse von Crowdsourcing Testungen erstellt. Dazu wurden die bezahlten Probanden über Mikroworkers und die freiwilligen per Facebook rekrutiert. Die zu bearbeitende Aufgabe bestand darin die Attraktivität von Fotofiltern zu bewerten. Die Aufgabe enthielt weiterhin eine Verifications-Frage um Computer auszuschließen, ein Training und eine Gameification um die Aufgabe spannender zu gestalten und den Abbruch der Aufgabe zu vermeiden. Die Auswertung zeigte deutlich, dass freiwillige Teilnehmer zwar vermehrt dazu neigen die Testung abzubrechen, jedoch durch ihre intrinsische Motivation deutlich reliablere Ergebnisse als die bezahlten Teilnehmer beitrugen. Diese beendeten durchaus mehr Testungen doch durch ihre Tendenz ihre Belohnung unter der geringsten Anstrengung zu bekommen waren die Ergebnisse weniger reliabel. Die Autoren gaben daraufhin die Empfehlung bei freiwilligen Crowds besonders viele Daten zu sammeln und bei bezahlten das Augenmerk auf die Reliabilität der Ergebnisse zu legen.

**Erstellt von:** Mareen Bentrup, 396332

**Autor(en):** Manoj Narayanan

**Titel:** The Power of "Crowd" Testing

**Jahr:** 2011

**Laufende Nummer im Literaturverzeichnis:** 28

**Schlagwörter (3-5):** crowd testing, cost-effective testing, blending traditional crowd, social media

**Kurzbeschreibung:** Das Paper untersucht grundlegend inwiefern und vor allem, für welche Produkte sich die Validierungsmethode Crowdtesting einzusetzen lohnt. Verglichen wird hiermit vor allem die Größe der Untersuchungsgruppe mit der Beurteilungsqualität, da das Crowdtesting zwar weitaus mehr Probanden/Tester erreicht, diese allerdings nicht unter kontrollierten Bedingungen entstehen und somit in ihrer Ergebnisqualität durchaus schwanken können. Entsprechend wird die Idee der Kombination von konventionellen Testing-Methoden und Crowdtesting hervorgebracht und eventuelle Einsatzorte bewertet.

**Erstellt durch:** Niklas Runow 386776

**Autor(en):** Noel Wurst & Rajini Padmanaban

**Titel:** Implementing Innovative Crowdsourced Testing: An Interview with Rajini Padmanaban

**Jahr:** 2013

**Laufende Nummer im Literaturverzeichnis:** 29

**Schlagwörter (3-5):** strenght, weaknesses, operational fields, crowd testing, quality assurance

**Kurzbeschreibung:** Das Interview mit Rajini Padmanaban, Director of Engagement von QA InfoTech, beschäftigt sich mit den Stärken und Schwächen bzw. den Eigenschaften eines Projektes, welches den Einsatz von Crowdttesting in Betracht ziehen sollte und welches nicht. Dabei spielen zum einen der Inhalt, diverse Sicherheits- und Vertraulichkeitsaspekte sowie die gewünschte Spezialisierung/Qualifizierung des Crowdttesting-Teams eine Rolle. Ebenfalls thematisiert wird der Ansatz hinter der Rekrutierung/dem Aufbauen eines eigenen Crowdttesting-Pools um entsprechende Zielgruppen besser zu repräsentieren.

**Erstellt durch:** Niklas Runow 386776

**Autor(en):** Markus Steinhauser

**Titel:** Two Approaches to Crowdsourced Software Testing

**Jahr:** 2013

**Laufende Nummer im Literaturverzeichnis:** 30

**Schlagwörter (3-5):** self-service, managed crowd testing, testing strategy, goals of testing

**Kurzbeschreibung:** Der Artikel des Head of Communications und Co-Founder von Testbirds Beschreibt den unterschiedlichen Ansatz der *managed* und *self-service* crowdsourced Testing Solutions. Es werden Stärken und Schwächen beider Verfahren betrachtet und ins Verhältnis gesetzt. Hierbei spielen vor allem der unterschiedliche Zeit- und Budgetfaktor und die bereits im eigenen Betrieb gemachte Erfahrung mit Crowdttesting eine Rolle.

**Erstellt durch:** Niklas Runow 386776

**Autor(en):** Dieter Speidel

**Titel:** How much and where does Crowd Size really matter? / Nobody owns the Crowd. And did you ever think about Crowd Activity Metrics?

**Jahr:** 2017

**Laufende Nummer im Literaturverzeichnis:** 31

**Schlagwörter (3-5):** crowd size, crowd characteristics, crowd quality,

**Kurzbeschreibung:** Der zweigeteilte Artikel des CEOs von *passbrains* beschreibt die Ansprüche an eine qualitativ ansprechende und quantitativ ausgewogene Menge an Testern für Crowdttesting Bedürfnisse. Hierbei sollte demnach nicht die Gesamtanzahl der vorhandenen Testern die entscheidende Rolle tragen, sondern Wert auf eine vollständige Kategorisierung der Profile (zur weiteren Zielgruppenspezifizierung), eine motivierte, aktive und immer zu erreichende Crowd und nicht zuletzt eine ausgewogene Auswahl an Testern/-innen bezüglich geografischer, sozialer und kultureller Hintergründe gelegt werden.

**Erstellt durch:** Niklas Runow 386776

**Autor(en):** Wibke Ziegler

**Titel:** Crowdfunding -Bugtracking via Crowd

**Jahr:** 2013

**Laufende Nummer im Literaturverzeichnis:** 32

**Schlagwörter (3-5):** crowd testing strategies, classic testing, compare, crowd testing vendor

**Kurzbeschreibung:** Der Artikel der Innovationswerkstatt der Volksbank Bühl beschreibt zunächst allgemein die Herangehensweise und den Nutzen von Crowdfunding. Unterschieden wird zwischen den beiden Anwendungsbereichen der funktionalen (explorativ und/oder Test-Cases) oder Usability-fokussierten Untersuchung. Anschließend wird auf potentielle Stärken und Schwächen eingegangen, welche Crowdfunding im Gegensatz zu konventionellen bzw. internen Testverfahren hat. Ein abschließendes Wort durch Dr. Frank Simon von der SQS-Gruppe verdeutlicht noch einmal die Wichtigkeit des internen Testens vor dem Roll-out, da bereits durch Crowdfunding gemachte Erfahrung einen erfolgreichen Start des Produkts beeinflussen kann.

**Erstellt durch:** Niklas Runow 386776

## Quellenverzeichnis

- (1) Alyahya, S., Alrugebh, D. (2017). Process Improvements for Crowdsourced Software Testing. *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 8, No. 6, 32-40.
- (2) Archambault, D. & Purchase H. & Hoßfeld, T. (2017). Evaluation in the Crowd: An Introduction. *Crowdsourcing and human-Centered Experiments* (pp. 27-69). Springer International Publishing AG
- (3) Blohm, I., Leicht, N., Leimeister, J.M., Rhyh, M. (2016). Besser Dank Schwarmintelligenz. *Swiss IT Magazine, Nr. 03/März 2016*, 40-42.
- (4) Blohm, I., Leimeister, J. M., Krcmar, H., (2013). Crowdsourcing: How to Benefit from (Too) Many Great Ideas. *MIS Q. Executive*, 4 (12), 199–211.
- (5) Eickhoff, C. & de Vries, A. (2012). Increasing cheat robustness of crowdsourcing tasks. *Information Retrieval*, 16 (2), 121-137.
- (6) Findlater, L., Zhang, J., Froehlich, J. E., & Moffatt, K. (2017). Differences in Crowdsourced vs. Lab-based Mobile Desktop Input Performance Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6813-6824). New York: ACM.
- (7) Gadiraju, U., Kawase, R., Dietze, S., & Demartini, G. (2015). Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys. In *Proceedings of the 33<sup>rd</sup> Annual ACM Conference on Human Factors in Computing Systems* (pp. 1632-1640). Seoul: ACM.
- (8) Gardlo, B. & Egger. (2014). Crowdsourcing 2.0: Enhancing execution speed and reliability of web-based QoE testing
- (9) Gardlo, B. & Ries, M. (2012). Impact of Screening Technique on Crowdsourcing QoE Assessments. *22nd International Conference Radioelektronika*
- (10) Guaiiani, F., & Muccini, H. (2015). Crowd and Laboratory Testing. Can they co-Exist? An Exploratory Study. In *Proceedings of the Second International Workshop on CrowdSourcing in Software Engineering* (pp.32-37). Florenz: IEEE.
- (11) Hoßfeld, T. & Keimel, C. (2014). Crowdsourcing Quality-of-Experience Assessments
- (12) Hoßfeld, T. et al. (2014). Best Practices for QoE Crowdttesting: QoE Assessment with Crowdsourcing. *IEEE Transactions on Multimedia*, Vol. 16, 541-558
- (13) Hoßfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K. & Tran-Gia, P. (2013). CrowdTesting: A Novel Methodology for Subjective User Studies and QoE Evaluation. *University of Würzburg, Tech. Rep, 486*.
- (14) Hoßfeld, T., Hirth, M., Korshunov, P., Hanhart, P. Gardlo, B., Keimel, C., Timmerer, C. (2014). Survey of web-based crowdsourcing frameworks for subjective quality assessment. In *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*. Indonesien: Jakarta.

- (15) Keimel, C. & Habigt, J. (2014). QualityCrowd – A Framework for Crowd-based Quality Evaluation
- (16) Leicht, N., Blohm, I., Leimeister, J.M., (2017). Leveraging the Power of the Crowd for Software Testing. *IEEE Software*, 34 (2), 62-69.
- (17) Leicht, N. et al. (2016). When is Crowdsourcing advantageous? The Case of crowdsourced software testing. *Twenty-Fourth European Conference on Information Systems (ECIS)*. Istanbul.
- (18) Liu, D. & Bias, R. (2012). Crowdsourcing for usability testing. *ASIST 2012*
- (19) Martin, D., Carpendale, S., Gupta, N., Hoßfeld, T., Naderi, B., Redi, J., Siahaan, E., Wechsung, I. (2017). Understanding the Crowd: Ethical and Practical Matters in the Academic Use of Crowdsourcing. *Evaluation in the Crowd Crowdsourcing and Human-Centered Experiments* (pp. 27-69). Springer International Publishing AG
- (20) Mok, R. K. P., Chang, R. K. C., & Li, W. (2017). Detecting Low-Quality Workers in QoE Crowdtesting: A Worker Behavior-Based Approach. *IEEE Transactions on Multimedia*, 19 (3), 530-543.
- (21) Mok, R.K.P., Li, W., Chang, R.K.C (2014). A user behavior based cheat detection mechanism for crowdtesting. In *Proceedings of the 2014 ACM conference on SIGCOMM* (pp. 213-124). USA: Chicago.
- (22) Mushtaq, M. S., Augustin, B. (2014). Crowd-sourcing Framework to Assess QoE. *IEEE ICC Communications Software, Services and Multimedia Applications Symposium, 2014*, 1705-1710.
- (23) Narayanan, M. (2011). The Power of „Crowd“ Testing. Cognizant 20-20 Insights
- (24) Rajini, P. (2012). A technical Paper on Best Practices in Crowd Sourced Testing. *Star East Testing Conference*.
- (25) Redi, J., Povoia, I., (2014). Crowdsourcing for rating image aesthetic appeal: Better a paid or a volunteer crowd? In *Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia, CrowdMM '14* (pp 25-30). USA: Orlando.
- (26) Shahid, M. & Søggaard, J. (2014). Crowdsourcing Based Subjective Quality Assessment of Adaptive Video Streaming
- (27) Speidel, Dieter. How much and where does Crowd Size really matter? / Nobody owns the Crowd. And did you ever think about Crowd Activity Metrics? (2017); <https://www.passbrains.com/top/post/how-much-and-where-does-crowd-size-really-matter.html>; <https://www.passbrains.com/top/post/nobody-owns-the-crowd-and-did-you-ever-think-about-crowd-activity-metrics.html> [17.12.2017]
- (28) Speidel, D., Mittal, M., und Sridharan, M. (2014). Testing Experience - The Magazine for Professional Testers: "Enterprise Crowd Testing". *Testing Experience* (25), 34-38.

- (29) Steinhauser, Markus (2013). Two Approaches to Crowdsourced Software Testing. <https://archive.is/20140206182859/http://www.crowdsourcing.org/editorial/two-approaches-to-crowdsourced-software-testing/28918> [16.12.2018]
- (30) Wurst, Noel; Padmanaban, Rajini (2013) Implementing Innovative Crowdsourced Testing: An Interview with Rajini Padmanaban; <https://www.stickyminds.com/interview/implementing-innovative-crowdsourced-testing-interview-rajini-padmanaban> [16.12.2017]
- (31) Ziegler, W. (2013). Crowdtesting -Bugtracking via Crowd. <http://blog.volksbank-buehl.de/blog/2013/01/17/crowdtesting-bugtracking-crowd/> [18.12.2017]
- (32) Zogaj, Shkodran, Bretschneider U. (2014). Managing crowdsourced software testing: a case study based insight on the challenges of a crowdsourcing intermediary