

Online-Fokusgruppen als innovative Methode zur nutzerbasierten Beurteilung der Web Usability

Miriam Yom, Thorsten Wilhelm, H.Holz Müller
eResult GmbH

Zusammenfassung

Fokusgruppen und insbesondere ihre Online-Variante sind im Kontext der nutzerbasierten Usability-Beurteilung noch wenig untersucht. Es stellt sich die Frage, welche Effekte die spezielle Kommunikationssituation in Online-Fokusgruppen (im weiteren OFGs) auf die Qualität und Verwertbarkeit der Daten hat. In diesem Artikel werden die konstituierenden Merkmale der synchronen computervermittelten Kommunikation (im weiteren cvK) und somit relevanten Determinanten der Datenqualität von OFGs dargestellt. Des Weiteren stellen wir Ergebnisse eines Methodenvergleichs zwischen konventionellen Fokusgruppen im Labor und OFGs bei der Beurteilung eines Online-Reiseshops vor. Es zeigte sich, dass durch Kommunikationsstraining und technische Unterstützung die Defizite der synchronen Online-Kommunikation ausgeglichen werden können. Anhand einer deskriptiven Themenanalyse konnte gezeigt werden, dass es starke Überschneidungen der Aussagen in den beiden Kommunikationsbedingungen gibt.

1 Einleitung

Bei der nutzerbasierten Usability-Evaluation werden verschiedenste Datenerhebungsmethoden eingesetzt. Nach dem Kriterium des Erhebungszeitraums während eines Usability-Tests kann man die Verfahren in zwei Klassen einteilen. Die *interaktionsbegleitenden Verfahren* wie z.B. das Verfahren der Protokolle Lauten Denkens, Verhaltensbeobachtung und Registrierung per Videotechnik und/oder Server-/Client-Logfiles, apparative Verfahren zur Erfassung des Informationsaufnahmeverhaltens (z.B. Blickregistrierung) oder Interaktionserlebens (Mimikbasierte Mikroevaluation oder psychophysikalische Messungen) etc.

An die Nutzung des Angebots schließt sich in der Regel die *Evaluationsphase* an, in der vornehmlich eine Beurteilung des Angebots als Ganzes durchgeführt wird (*summarische Evaluation*). Dies kann quantitativ mittels standardisierter Usability-Befragungsinstrumente (vgl. z.B. KIRAKOWSKY & CIERLIK 1998, BALAZS & CHRISTOPHERSEN 2002 u.a.) erfolgen. Die summarische Evaluation kann aber auch qualitativ mittels *Online- oder Face-to-Face-Fokusgruppen* (im weiteren FtF) oder in offenen bzw. halb-strukturierten *Online- oder Offline-Einzelinterviews* erfolgen. Dabei kann über eine wiederholte Präsentation des Stimulusmaterials die Diskussion / das Gespräch auf bestimmte Bereiche des Angebots fokussiert werden.

Fokusgruppen allgemein, und OFGs im Speziellen sind im Kontext von Usability-Tests nur wenig untersucht worden. Für die innovative Methode der OFGs können jedoch Erkenntnisse aus der Erforschung der computervermittelten Kommunikation (im weiteren cvK) dazu genutzt werden, die Kommunikationssituation in den OFGs anhand konstituierender Merkmale zu beschreiben und somit die Effekte der Kommunikationsbedingung auf die Datenqualität aus theoretischer Sicht zu spezifizieren. Im Anschluss daran möchten wir ausgewählte Ergebnisse eines Methodenvergleichs (FtF vs. OFGs) vorstellen.

2 Zur Methode der Online-Fokusgruppen

Bei OFGs treffen sich die Teilnehmer und der Moderator in einem virtuellen Diskussionsraum. Über den heimischen PC loggen sich die Teilnehmer per Internetverbindung in einen Chat-Room ein, der in der Regel an die speziellen Bedürfnisse der Marktforschung technisch angepasst wurde.

Es gibt verschiedene Meinungen dazu, für welche Fragestellungen FtF-Fokusgruppen im Allgemeinen und OFGs im Speziellen geeignet sind. So betont beispielsweise GÖRTS (2001 S.153) die besondere Adäquanz von Fokusgruppen in explorativen Studienphasen, die quantitativen Studien vorgeschaltet sind. MORGAN (1997) sieht den Sinn und Zweck von Fokusgruppen jedoch nicht nur auf die explorative Phase beschränkt. Fokusgruppen können seiner Meinung auch als eine singuläre Haupterhebungsmethode angewendet werden, die *Antworten* auf die Forschungsfrage gibt (vgl. ebda S.16). Insbesondere im anwendungsorientierten Usability-Test sind Online- und Offline-Fokusgruppen ähnlich wie Einzelinterviews dazu geeignet, wahrgenommene Usability-Fehler, kognitives Verständnis und Einstellungen zu explorieren.

Die Vorteile der OFG im Vergleich zu ihrem FtF-Pendant sind insbesondere darin zu sehen, dass sie zeit- und kostenökonomischer, flexibler und, falls notwendig, über nationale Grenzen hinweg durchzuführen sind. Die Teilnehmer verbleiben in ihrer gewohnten Umgebung und fühlen sich deshalb freier, ihre Meinung kundzutun. Ein weiterer Vorteil ist, dass ansonsten sehr schwierig zu akquirierende Zielgruppen beispielsweise aus dem B-to-B-Bereich eher für Online-Diskussionen zu gewinnen sind (vgl. PALMQUIST & STUEVE 1996). Darüber hinaus sind OFGs in der Usability-Forschung von herausragendem Interesse, da eine Evaluation des Angebots *ohne Medienbruch* durchgeführt werden kann. Wie bereits aufgeführt, ist die Prototypen-Entwicklung für Websites und Shops häufig hohen zeitlichen Restriktionen unterworfen, die sich in einem für den Usability-Test zur Verfügung stehenden kurzen Zeitraum niederschlägt. Diesem Umstand tragen OFGs Rechnung, da die ansonsten bei FtF-Fokusgruppen aufwendig zu erstellenden Transkripte der Videobänder durch die automatischen Diskussionsprotokolle entfallen. Dies erleichtert die anschließende quantitative und/oder qualitative Auswertung des verbalen Materials. Diese nutzbaren Rationalisierungseffekte ermöglichen ökonomisch attraktive OFG-Studien in dichteren Zeitintervallen und/oder mit größeren Stichproben.

Diesen Vorteilen stehen jedoch auch problematische Aspekte gegenüber. So können mögliche Störeinflüsse während der Datenerhebung nur bedingt oder gar nicht kontrolliert werden. Es kann beispielsweise nicht ausgeschlossen werden, dass die Kommentare eines Teilnehmers durch Dritte beeinflusst werden. Während es bei klassischen Diskussionsrunden in der Regel keinen Drop-Out während der Datenerhebung gibt, besteht bei OFGs eine realistische Gefahr, dass Teilnehmer während der Diskussion im WWW surfen, sich vom PC entfernen etc. und als Folge davon episodisch oder gar nicht mehr an der Diskussion teilnehmen.

Ein weitreichender Kritikpunkt betrifft die Ergebnisqualität. So wird angezweifelt, dass die Güte der OFGs vergleichbar hoch sei wie die der konventionellen Fokusgruppen (vgl. z.B. GREENBAUM 1995). Korrekt ist, dass die Effekte des Mediums bzw. genauer der spezifischen Kommunikationssituation auf die Ergebnisse unklar sind. Diese Forschungslücke impliziert jedoch nicht, dass die Validität dieser Erhebungsmethode per se geringer ist als bei ihrem Offline-Pendant. Es gilt vielmehr diese Forschungslücke zu schließen, so dass auf dieser Basis die Marktforschung „...diese Effekte kontrollieren und kreativ zur Beantwortung von Fragestellungen einsetzen kann.“ (CORNELIUS 2001a S.3).

Zu diesem Zweck möchten wir zunächst kurz auf die besonderen Merkmale der Kommunikationssituation in OFGs eingehen, um dann ausgewählte Ergebnisse eines Methodenvergleichs (OFG vs. FtF-Fokusgruppen) vorzustellen.

2.1 Besonderheiten der Kommunikationssituation in Online-Fokusgruppen

Nach HERRING (1999) kann die cvK in OFGs durch folgende Charakteristika gekennzeichnet werden:

- Audio-visuelle und psychologische Anonymität,
- synchrone, textbasierte Kommunikation mit zeitlicher Verzögerung sowie
- Wegfall von unmittelbarer Rückmeldung.

Audio-visuelle und insbesondere psychologische Anonymität

Für die Kommunikation in OFGs steht ausschließlich der verbale, textbasierte Kanal zur Verfügung, da die Teilnehmer sich weder sehen noch hören können. Es wird davon ausgegangen, dass die anonyme Kommunikationssituation in OFGs im Vergleich zu FtF-Fokusgruppen zu mehr Offenheit im Antwortverhalten (vgl. PRICKARZ & URBAHN 2002), ausgeprägteren Selbstoffenbarungen und geringerer sozialer Erwünschtheit führt (vgl. JOINSON 2001). POSTMES, SPEARS & LEA (1999) weisen in diesem Zusammenhang darauf hin, dass weniger die visuelle Anonymität als vielmehr die *psychologische Anonymität* diese Effekte bewirkt. Die psychologische Anonymität bzw. soziale Nähe zu den anderen Teilnehmern kann über gezielte Informationsgabe je nach Erkenntnisziel variiert werden und damit eine für die Marktforschung nutzbare Ressource sein. So ist es z.B. möglich, durch ein gezieltes Bekanntgeben oder Zurückhalten von Informationen eine stärkere Orientierung der Teilnehmer an Gruppennormen oder an individuellen, persönlichen Normen und Einstellungen zu induzieren. Die ausführliche Erläuterung dieses Aspekts würde an dieser Stelle den Rahmen sprengen und wird an anderer Stelle im Detail dargestellt (vgl. YOM in Vorbereitung). In diesem Artikel möchten wir uns vor allem den anderen beiden Merkmalen der cvK im Detail zuwenden.

Synchrone, textbasierte Kommunikation mit zeitlicher Verzögerung

Nachrichten müssen von den Teilnehmern zunächst geschrieben, eventuell editiert und an den Server abgeschickt werden. Auf diese Weise entstehen (a) spezifische Kommunikationsbarrieren, die durch die Umsetzung von Sprache in motorische Bewegungen bedingt sind und (b) zeitliche Verzögerungen bei der Beantwortung von Nachrichten. Da alle Teilnehmer simultan Nachrichten abschicken können, ist eine Ordnung der Nachrichten in der von den Teilnehmern *inhaltlich* intendierten Reihenfolge nicht gewährleistet. Die Nachrichten erscheinen vielmehr nach ihrem Eingangszeitpunkt im zentralen Chat-Fenster. Diese technisch bedingten Eigenschaften der Kommunikationssituation führen zu einer Verletzung des sogenannten *Adjazenzprinzips*. Dieses besagt, dass Gespräche in Sequenzen segmentiert werden, z.B. in dem auf Fragen Antworten, auf Aufforderungen Reaktionen etc. folgen (vgl. CORNELIUS 2001b S.20). Kommunikationssituationen, die dem Adjazenzprinzip in praktischer und semantischer Hinsicht folgen, werden von dem Teilnehmer als "geordnet" wahrgenommen und erleichtern die Verarbeitung der Informationsinhalte.

Die zeitlichen Verzögerungen aufgrund des Schreib- und Editieraufwands in Kombination mit der Gleichzeitigkeit der Kommunikation in OFGs führen dazu, dass die Adjazenzstruktur der Diskussion unterbrochen ist. Die Verständlichkeit von thematischen Sequenzen wird dadurch erschwert, da die Zuordnung der Nachrichten aufeinander nicht mehr oder schwer nachvollziehbar ist. Es entsteht (der Eindruck) eines „Konversations-Chaos“.

Fehlen unmittelbarer Rückmeldung (Back-Channels)

Rückmeldungen in Gesprächssituationen sind Äußerungen, die den Partner nicht unterbrechen, aber Zustimmung und Ablehnung signalisieren und damit den weiteren Verlauf eines Gesprächs

beeinflussen. Sie dienen dazu, dem aktuellen Sprecher ein Feedback zu geben, dass das Gesagte aufgenommen und verstanden wurde und unterstützen die Aufrechterhaltung des Gesprächs (vgl. MCLAUGHLIN 1984). Dies kann in persönlichen Kommunikationssituationen über Gesten, Mimik, Körpersprache oder verbal erfolgen. In der cvK ist eine solche zeitgleiche Rückmeldung technisch bedingt nicht möglich. Das Fehlen bzw. die zeitliche Verzögerung von Rückmeldungen in der cvK führt zu einer zusätzlichen Reduzierung der Verständlichkeit der Nachrichten, da die Etablierung eines gemeinsamen Referenzsystems erschwert wird (vgl. CORNELIUS 2001b S.45 f.).

Bei der FtF-Fokusgruppe adressieren die Teilnehmer ihre Aussagen über die Ausrichtung des Blicks oder der Körperhaltung eindeutig an die Gruppe oder an einzelne Personen. Diese nonverbale Adressierfunktion ist in textbasierten Online-Systemen nicht mehr gegeben. Gleichzeitig sind Personen mit geringer oder keiner Chat-Erfahrung nicht gewohnt, Adressaten explizit anzusprechen, da sie dies aus konventionellen Gesprächssituationen nicht gewohnt sind. Werden Nachrichten ohne expliziten Adressat jedoch online artikuliert, sind diese zunächst unspezifisch an die Gruppe gerichtet. Als Folge dieses Kommunikationsverhalten bleibt häufig unklar, an wen eine Nachricht gerichtet ist und zu welchem Thema die Nachricht Bezug nimmt. Die dyadische Organisation des Gesprächsablauf wird hierdurch erschwert. Die Dialogsequenzen zwischen einzelnen Teilnehmern sind jedoch notwendig und wichtig, um Themen zu etablieren und weiter diskutieren zu können.

Inkohärenz in der computervermittelten Kommunikation

Sowohl der Wegfall der unmittelbaren Rückmeldung als auch die zeitliche Verzögerung und Simultanität sind Ursachen für die Entstehung von *Inkohärenz* in der cvK. So haben Analysen von unmoderierten Chats gezeigt, dass durch die Verletzung des Adjazenzprinzips Themen nur kurz aufrechterhalten oder erst gar nicht etabliert werden (vgl. HERRING 1999 S.3 ff.). Konversationale *Kohärenz* versteht CORNELIUS (2001b S.28) "... als Fähigkeit und Motivation von Gesprächsteilnehmern..., gegenseitig aufeinander einzugehen und gemeinsam Themen zu etablieren und weiterzuentwickeln".

Inwieweit ist die Kohärenz für die Fragestellung nach der Güte der Daten aus OFGs im Zusammenhang mit der Beurteilung der Benutzerfreundlichkeit von Websites von Relevanz? Zur Beantwortung dieser Frage muss die Aufgabenstellung von Fokusgruppen beim Website-Test berücksichtigt werden. Die Gruppe soll die Benutzerfreundlichkeit und das Gefallen von Shop-Prototypen *beurteilen* und *konkrete Verbesserungsvorschläge generieren*. Neben einer gemeinsamen Zieldefinition und Zielorientierung müssen die Teilnehmer dafür problemrelevantes Wissen und Meinungen austauschen, sowie vor dem Hintergrund der Problemstellung bewerten. Dabei ist der Grad der Unterschiedlichkeit der individuellen, subjektiven Bewertungen von alternativen Lösungsvorschlägen für den Usability-Praktiker von besonderem Wert. Auf diese Weise werden die zielgruppenspezifischen Bewertungsmechanismen und -kriterien sichtbar und stellen somit die Grundlage für die Herleitung von Handlungsempfehlungen dar.

Der Austausch von relevantem Wissen und die Bewertung von alternativen Verbesserungsvorschlägen setzt jedoch voraus, dass die Teilnehmer aufeinander eingehen, ein gemeinsames Thema etablieren (z.B. Beurteilung des Shops) und weiterentwickeln (z.B. Verbesserungsmöglichkeiten und -wünsche). Kohärenz in der Kommunikationssituation ist vor dem definierten Erkenntnisinteresse im Rahmen einer entsprechenden Evaluation also eine *notwendige Voraussetzung*, um zu befriedigenden Ergebnissen zu gelangen.

Als *Zwischenfazit* kann an dieser Stelle die Diskussion über die Güte von Daten aus OFGs spezifiziert werden (vgl. zusammenfassend Abb.1). Die synchrone cvK in OFGs ist aufgrund fehlender unmittelbarer Rückmeldungen und der technisch bedingten Verletzung des Adjazenzprinzips durch Inkohärenz gekennzeichnet. Wenn die Evaluation und Generierung von Verbesserungsvorschlägen Zielsetzung der OFG ist, dann ist Kohärenz eine wichtige Determinante der Datenqualität. Werden in diesem Fall keine Maßnahmen zur Reduzierung der Inkohärenz in OFGs einge-

setzt, so ist tatsächlich zu erwarten, dass die Güte der verbalen Daten in der Online-Bedingung im Vergleich zu konventionellen Fokusgruppen leidet. Umgekehrt heißt dies jedoch auch: gelingt es, die Kohärenz in der OFG über geeignete Maßnahmen dem Niveau einer FtF-Diskussion anzugleichen, dann sollten die online generierten verbalen Daten sich hinsichtlich der inhaltlichen Qualität nicht von denen der FtF-Bedingung unterscheiden.

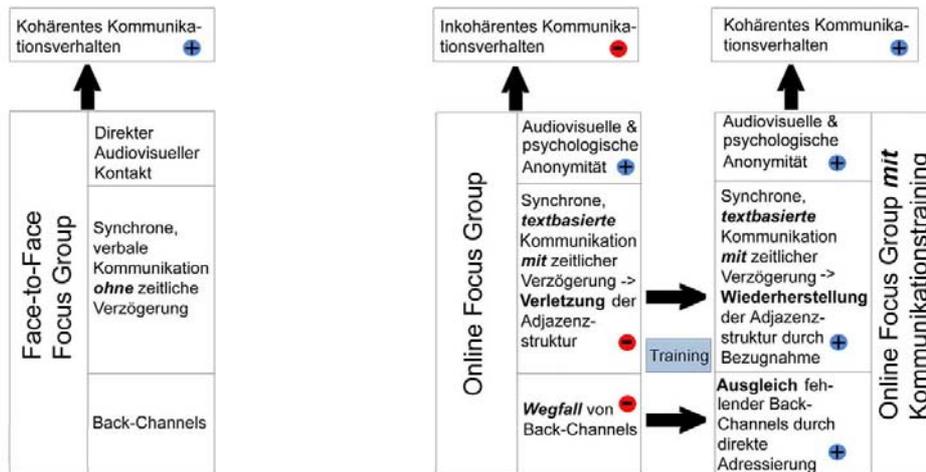


Abb. 1: Kommunikationssituation in Face-to-Face- vs. Online-Fokusgruppen mit und ohne Training

3 Ergebnisse des Methodenvergleichs: Face-to-Face vs. Online-Fokusgruppen zur Beurteilung von Online-Shops

3.1 Untersuchungsanlage

Zur Überprüfung dieser Überlegungen wurde ein Methodenvergleich durchgeführt, in dem Daten von vier konventionellen Fokusgruppen und vier OFGs verglichen wurden. Zur Erhöhung der Kohärenz in den OFGs wurde ein Kommunikationstraining eingesetzt, dessen Wirksamkeit bei Entscheidungsfindungsgruppen bereits experimentell in anderen Studien geprüft wurde (vgl. CORNELIUS 2001a). Bei dem Training handelt es sich um bild- und textbasiertes Trainingsmaterial, welches die Teilnehmer für die Probleme der computervermittelten Kommunikationssituation sensibilisiert. Es wurde des Weiteren mit den Teilnehmern trainiert, dass sie beim Versenden einer Nachricht explizit auf das referierte Thema Bezug nehmen (Wiederherstellung der Adjazenzstruktur) und ggf. den Adressaten der Nachricht benennen sollten (Kompensation des Wegfalls von unmittelbaren Rückmeldungen).

Bei der untersuchten Website handelte es sich um einen HTML-Prototypen eines Online-Reiseshops, der über redaktionelle Inhalte, ausgewählte Produktangebote und Buchungstrecken verfügte. Vor der Diskussion in der Fokusgruppe nutzten die Teilnehmer das Angebot sowohl frei als auch szenariobasiert. Interviewleitfaden, Diskussionszeit (90 Minuten) und Moderatorin waren für beide Untersuchungsbedingungen konstant.

Da die Erhebung einer möglichst großen Bandbreite an Meinungen, Einstellungen und Urteilen im Vordergrund stand, sollten die Gruppen in Bezug auf die personenbezogenen Merkmale *Alter*;

Geschlecht und Berufstätigkeit möglichst heterogen zusammengesetzt sein (vgl. LAMNEK 1998 S.96). Dies ist bei der kleinen Gruppengröße (6 Teilnehmer pro Gruppe) durch eine randomisierte Zuweisung nicht zu gewährleisten. Aus diesem Grund wurde versucht, sog. *Matched Samples* zu bilden. Diese Vorgehensweise eignet sich insbesondere für kleinere Stichproben mit einer Teilnehmerzahl von etwa 20 pro Vergleichsgruppe. Dabei werden die Teilnehmer der Stichproben einander paarweise in Bezug auf die Merkmale zugeordnet (vgl. BORTZ & DÖRING 1995 S.491 f.). Eine zeitökonomische Akquise adäquater Teilnehmer war mit Hilfe des Online-Access-Panels der eResult GmbH möglich.

3.2 Methodische Vorgehensweise bei der Aufbereitung der Daten

Zur Transformation der verbalen Diskussionsdaten in quantitative Größen wurde das Verfahren der *Inhaltsanalyse* verwendet. Zwei eingewiesene und trainierte Assistenten kodierten das Material nach vorgegebenen Kodierregeln. Die Kodierung durch die beiden Assistenten erfolgte unabhängig voneinander, so dass die Interkoder-Reliabilität berechnet werden konnte.

Zur Analyse der Interkoder-Reliabilität wurde sowohl die einfache *prozentuale Übereinstimmung* zwischen den beiden Kodierern als auch der *Kappa-Koeffizient* (κ) nach COHEN (1970 zit.n. BORTZ & DÖRING 1995 S.254) für die Variablen berechnet, die durch die Assistenten klassifiziert werden mussten. Letzterer ist ein im Vergleich zur prozentualen Übereinstimmung strengeres Kriterium. Er beinhaltet eine Zufallskorrektur des Übereinstimmungsmaßes, so dass zufällige Übereinstimmungen bei der Kodierung berücksichtigt werden. Es ergaben sich für die einfache prozentuale Übereinstimmung Werte zwischen 93,99% und 98,88% bzw. Kappa-Koeffizienten in Höhe von .71 bis .97. Dies sind gute bis sehr gute Reliabilitäten, so dass die Daten für eine weitere statistische Analyse verwendet werden konnten.

3.3 Ausgewählte Ergebnisse

Kohärenzkoeffizient

Zur quantitativen Erfassung der Kohärenz wurde ein validiertes Kodierungsschema angewendet, bei dem die kohärenten und inkohärenten Themenkoordinationsmechanismen (TKM) durch zwei Versuchsleiter kodiert wurden (zur ausführlichen Darstellung der Operationalisierung des Konstrukts der Kohärenz vgl. BOOS & CORNELIUS 2001, YOM in Vorbereitung). Als globales Maß der Kohärenz in der Fokusgruppe dient der *Kohärenzquotient*. Er wird berechnet aus der Differenz der Gesamtzahl der kohärenten TKM und der Gesamtzahl der inkohärenten TKM und dem Verhältnis dieser Differenz zur Gesamtzahl aller TKM. Das durch den Kohärenzkoeffizienten (K) abgebildete Kontinuum kann Werte zwischen 1 (extrem kohärent) und -1 (extrem inkohärent) annehmen. Für die Online-Bedingung ergab sich ein $K = .91$ und für die FtF-Bedingung ein $K = .94$. Beide Werte unterschieden sich nicht signifikant voneinander. Es ist also gelungen, die Kohärenz in der OFG dem Kohärenz-Niveau der konventionellen Fokusgruppe anzugleichen.

Anzahl der Wörter

Aufgrund des Tipp- und Editieraufwands in der Online-Bedingung zeigte sich erwartungsgemäß ein deutlicher Gruppenunterschied bei der Anzahl der Wörter. Dabei ist im Mittel ein Verhältnis von etwa 1:3 (Online:FtF) festzustellen. Die Teilnehmer der FtF-Diskussion (Mittelwert 972,75, SD 719,04) produzierten also dreimal so viele Wörter wie die Teilnehmer der OFGs (Mittelwert 316,40, SD 125,12). Aber bedeutet dies auch, dass in der Online-Bedingung nur ein Drittel an Aussagen und Meinungen artikuliert wurde?

Anzahl der Statements

Bei einem Statement bzw. einer Nachricht handelt es sich um eine in sich geschlossene verbale Äußerung oder textbasierte Mitteilung eines Teilnehmers. Ein Sprecherwechsel bzw. das Eintreffen einer neuen Nachricht markiert das Ende des Statements bzw. der Nachricht. Betrachtet man die Anzahl der Statements, so wird deutlich, dass die Teilnehmer der OFGs (Mittelwert 41,55, SD 11,60) trotz des Schreib- und Editieraufwandes im Durchschnitt sogar mehr Statements geäußert haben als die Teilnehmer der FtF-Diskussion (Mittelwert 33,79, SD 18,50). Die Anzahl der Statements in den beiden Bedingungen unterscheidet sich jedoch *nicht signifikant*.

Anzahl der angesprochenen Themen im Vergleich

Der Unterschied bei der Anzahl der Statements zwischen den beiden Kommunikationsbedingungen ist nicht signifikant. Jedoch lässt dies noch keinen Schluss darüber zu, in welcher *Variationsbreite* auf die Benutzerfreundlichkeit bezogenen Themen von den Teilnehmern in den beiden Kommunikationsbedingungen diskutiert wurden. Oder anders ausgedrückt: inwieweit haben evtl. unterschiedliche Gruppendynamiken dazu geführt, dass sich je nach Kommunikationssituation unterschiedliche Themen etabliert haben? Denn letztendlich entscheiden die Teilnehmer, welche Themen sie für diskussionswürdig halten.

Über die Gegenüberstellung der *Anzahl der angesprochenen Themen* kann die inhaltliche Dimension der Diskussionen in den Kommunikationsbedingungen stärker berücksichtigt werden. Betrachtet man die angesprochenen Themen in den Online- und Offline-Gruppen gemeinsam, dann sind insgesamt 65 Themenaspekte diskutiert worden. Dabei zeigte es sich, dass insgesamt von den Teilnehmern der OFG (n=50) etwa gleich viele Aspekte der Benutzerfreundlichkeit angesprochen und diskutiert wurden wie in der konventionellen Fokusgruppe (n=55). Die Ergebnisse zur Anzahl der angesprochenen Themen sind bemerkenswert, bedenkt man, dass aufgrund des zusätzlichen kognitiven Aufwandes durch das Schreiben und Editieren der Nachrichten für die Teilnehmer der OFGs eine geringere Produktionsgeschwindigkeit zu erwarten gewesen wäre. Wie sind diese Befunde zu erklären?

Zum einen wirkt sich die Möglichkeit zur Teilnahme an parallel laufenden Diskussionssträngen positiv auf die Zeiteffizienz aus. Des Weiteren kann in den Logfiles und Transkripten abgelesen werden, dass in den OFGs wesentlich kürzer und prägnanter von den Teilnehmern formuliert wird als in den konventionellen Gruppen (vgl. hierzu auch PRICKARZ & URBAHN 2002 S.64). Dieses Ergebnis spricht auch für die Überlegungen von WALTHER (2000), der davon ausgeht, dass die Botschaften in cvK mit mehr Bedacht formuliert werden. So beansprucht der Schreib- und Editieraufwand in der cvK zwar kognitive Ressourcen. Gleichzeitig müssen jedoch für die konventionelle Kommunikationssituation notwendige Ressourcen wie z.B. die Umweltbeobachtung, Kontrolle des eigenen nonverbalen Verhaltens etc. *nicht aktiviert* werden. Er schließt daraus, dass diese Energie vom Sender dazu genutzt werden kann, um die inhaltliche und formale Konstruktion der geschriebenen Botschaften im Sinne des intendierten Eindrucks beim Kommunikationspartner zu nutzen. Diese "überschüssigen" Ressourcen können u.E. jedoch auch der aufgabenbezogenen Dimension zugute kommen und sich in knappen, aber auch prägnanten, themenbezogenen Äußerungen niederschlagen.

Anzahl der Meinungen und Urteile

Diese Vermutung bestätigt sich auch in der quantitativen Analyse der Statements. So wurden Statements, die eine bewertende Komponente enthalten, nochmals nach dem Kriterium Bezugsobjekt in drei Kategorien unterschieden. Bezog sich das Statement auf einen grundsätzlichen Zusammenhang zum Thema Websites/Internet allgemein, dann handelte es sich um eine *allgemeine Meinung*. War das Statement beschreibend auf die zu untersuchende Website (Prototyp oder Liveangebot) bezogen, dann wurde diese Aussage als *objektbezogene Meinung* kodiert. Bezog sich die Aussage auf ein konkretes Gestaltungsmittel des Prototypen oder Liveangebots, dann wurde dies als *Urteil* registriert. Um Verzerrungen bei den Gruppenunterschieden aufgrund der absoluten

Anzahl an evaluativen Statements zu vermeiden, wurden die absoluten Individualwerte in relative Werte transformiert. Die Überprüfung mittels Kolmogorov-Smirnov-Test ergab, dass alle drei Variablen normalverteilt sind, so dass die Gruppenunterschiede über einen t-Test auf statistische Signifikanz überprüft werden konnten. Es zeigte sich, dass die relative Anzahl der allgemeinen Meinungen und die relative Anzahl der objektbezogenen Meinungen sich zwischen den beiden Kommunikationsbedingungen nicht signifikant unterschieden. Einen deutlichen und signifikanten Unterschied zeigte sich jedoch bei der relativen Anzahl der Urteile. Teilnehmer der OFGs äußerten signifikant mehr Urteile als Teilnehmer der FtF-Bedingung ($T=2.83$, $DF=42$, $p < .05$). Das heißt, sie machten mehr Aussagen, deren Inhalt sich auf ganz konkrete Gestaltungsmittel des Webangebots bezogen. Diese Ergebnisse sprechen also dafür, dass die Teilnehmer der OFGs prägnantere Aussagen gemacht haben als die Teilnehmer der FtF-Fokusgruppen.

Gruppenspezifische Themen

Betrachtet man die ausschließlich in einer Versuchsbedingung diskutierten Themen, also jene, die nur von den Teilnehmern der einen oder der anderen Gruppe angesprochen wurden, so zeigt sich folgende Verteilung (vgl. Abb.2): 15 Themenfelder der insgesamt 65 angesprochenen Themen (23,1%) wurden nur von Teilnehmern der FtF-Gruppe angesprochen, während sich zu 10 Themen nur Teilnehmer der OFGs äußerten (15,4%). Es ist auffällig, dass beide Methoden in ähnlichem Masse gruppenspezifische Themen generieren.

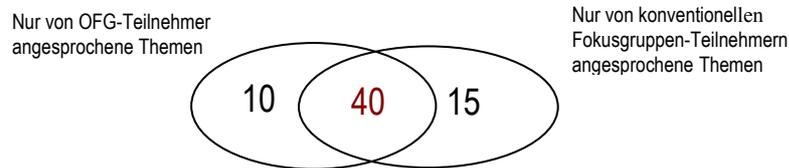


Abb.2: Anzahl der gruppenspezifischen und überlappenden Themen bei konventionellen Fokusgruppen vs. Online-Fokusgruppen

Wenden wir uns nun der inhaltlichen Dimension der gruppenspezifischen Themen zu. Gibt es möglicherweise einen charakteristischen Unterschied, welche Art von Themen eher von Online- oder FtF-Gruppen angesprochen werden? Es zeigte sich, dass sich die beiden Kommunikationsbedingungen insbesondere bei der Generierung von Themen im Bereich "Buchung" unterscheiden. Dieser Themenblock ist allerdings eine Ausnahme, da er ursprünglich keinen zentralen Aspekt der Diskussion darstellte. Aufgrund der insbesondere durch die Teilnehmer der FtF-Bedingung initiierten Etablierung dieses Themas wurde diese neue Hauptkategorie ex post notwendig. In den OFGs trat dieses Diskussionsverhalten so gut wie nicht auf, d.h. die Teilnehmer antworteten und diskutierten eng im Rahmen der vorgegebenen Themen.

Überlappende Themen

40 (61,5%) von insgesamt 65 inhaltlichen Aspekten wurden sowohl in den Online- als auch FtF-Fokusgruppen angesprochen und diskutiert (vgl. Schnittmenge in Abb.2). Auffällig ist dabei, dass bei den überlappenden Themen die *Richtung der Beurteilung bzw. Aussagen* in 31 (78%) der 40 Themen unabhängig von der Kommunikationssituation übereinstimmten. Lediglich bei neun Themen, differierten die Meinungen zwischen den beiden Kommunikationssituationen. Bei acht Themen wurden in den OFGs die Meinungen und Urteile *nicht gegenläufig* zur FtF-Bedingung, sondern lediglich *kontroverser* diskutiert. Eine kontroverse Diskussion kann die Ergiebigkeit der Ergebnisse durchaus positiv beeinflussen. Einmütige Urteile der Gruppe können durch die Beeinflussung eines charismatischen Wortführers zustande gekommen sein, die eine vertiefende Diskussion der Pro und Contras zu den einzelnen Themen im Keim erstickt oder sehr stark in eine Richtung gelenkt hat.

Nur ein Thema wurde tatsächlich gegensätzlich beurteilt. Beim Thema "Guided Tours" waren die Teilnehmer der FtF-Bedingung der Meinung, dass diese eher unnötig sei, während die Online-Teilnehmer dieses Hilfeangebot positiv hervorhoben. Für die Handlungsempfehlungen bezüglich einer benutzerfreundlichen Gestaltung wären diese gegenläufigen Meinungsbilder jedoch nicht von großer Tragweite, da „Guided Tours“ eine optional angebotene Hilfsfunktion darstellen, die nicht zu den zentralen Funktionsbereichen des Shops gehören.

Anzahl der Verbesserungsvorschläge

Statements, die einen konkreten Vorschlag zur gestalterischen oder strukturellen Veränderung der Website oder deren Inhalte beinhalteten, wurden als Verbesserungsvorschlag kodiert. Auch hier wurden zur Berechnung von Gruppenunterschieden mittels t-Tests die absoluten Individualwerte in relative Werte transformiert. Es zeigte sich, dass die Teilnehmer der OFGs (MW .16, SD .13) tendenziell mehr konkrete Verbesserungsvorschläge formulierten als die Teilnehmer der FtF-Fokusgruppen (MW .13, SD .09). Der t-Test für unabhängige Stichproben mit der normalverteilten Variable "Relative Anzahl der Verbesserungsvorschläge" ergab jedoch, dass dieser Unterschied *nicht signifikant* ist. In beiden Kommunikationsbedingungen wurden also ähnlich viele konkrete Verbesserungsvorschläge generiert.

Partizipation

Betrachten wir nun die *Partizipationsrate*, die für jeden Teilnehmer berechnet wurde, in dem die absolute Anzahl der Statements in Relation zur Gesamtanzahl der Statements in der jeweiligen Gruppe berechnet wurde. In der Online-Bedingung (MW .20 SD .05) zeigten die Teilnehmer im Durchschnitt eine höhere Partizipationsrate als die Teilnehmer der FtF-Diskussion (MW .17, SD .08). Dabei sind die Gruppenunterschiede der normalverteilten Variable *tendenziell signifikant* ($T=1.81$, $df 42$, $p < .10$).

In Anlehnung an WEISBAND et al. (1995 S.1130) wurde darüber hinaus als Maß für die *Gleichverteilung der Partizipationsrate* für die gesamte Gruppe der *Gini-Koeffizient (G)* berechnet. Dieser berücksichtigt, dass beispielsweise bei einer Anzahl von sechs Teilnehmern, die optimale Gleichverteilung der Redebeiträge bei einem Sechstel aller Beiträge pro Mitglied liegt. Eine vollkommene Gleichverteilung der Partizipationsraten resultiert demnach in $G=0$, während $G=1$ bedeutet, dass die Teilnehmer in ihren Partizipationsraten maximal voneinander abweichen.

Wie die Tabelle 1 zeigt, weichen die Partizipationsraten in den OFGs sehr viel weniger voneinander ab wie die Partizipationsraten in den FtF-Fokusgruppen (höhere G-Werte).

N=20	G _{Online}	N=24	G _{Face-to-Face}
Gruppe 1 (n=5)	.05	Gruppe 5 (n=6)	.41
Gruppe 2 (n=4)	.19	Gruppe 6 (n=6)	.23
Gruppe 3 (n=6)	.13	Gruppe 7 (n=6)	.30
Gruppe 4 (n=5)	.12	Gruppe 8 (n=6)	.23

Anmerkung: $G=0.0$ (vollkommene Gleichverteilung der Partizipationsraten),
 $G=1.0$ (Partizipationsrate weichen maximal voneinander ab)

Tabelle 1: Gini-Koeffizienten Online- vs. Face-to-Face-Fokusgruppen

Da insgesamt nur acht G-Werte vorlagen, wurden die zentralen Tendenzen der Online- vs. FtF-Bedingung mittels Mann-Whitney-U-Test überprüft. Dieser wies nach, dass die Unterschiede zwischen den beiden Kommunikationsbedingungen *signifikant* ($p < .05$) sind.

3.4 Fazit

Wir haben im Rahmen unserer Studie ein ausgewähltes, aber zukunftssträchtiges Anwendungsfeld der OFGs nämlich die Beurteilung der Usability von Websites untersucht. Dabei sind erste Ergebnisse bezüglich der Güte und Verwertbarkeit der verbalen Daten aus OFGs belegt worden. Störvariablen wie Inkohärenz in der cvK können durch geeignete Maßnahmen neutralisiert werden. Es zeigte sich trotz des Tipp- und Editieraufwandes für die Teilnehmer der OFGs keine Überlegenheit der konventionellen Fokusgruppen bei der Variationsbreite der etablierten Themen. Auch bestätigte sich, dass die Teilnehmer der OFGs den Tipp- und Editieraufwand durch Reduzierung von Redundanzen und prägnanten Aussagen kompensieren können. So formulierten Online-Teilnehmer signifikant mehr Urteile, also Aussagen zu ganz konkreten Gestaltungsmitteln des Webangebots, als die Teilnehmer der FtF-Gruppen, während sich die Anzahl der allgemeinen Meinungen und der objektbezogenen Meinungen nicht unterschieden. Des weiteren konnte in der deskriptiven Themenanalyse nachgewiesen werden, dass die Diskussionsinhalte in beiden Bedingungen einen hohen Anteil gemeinsamer Themen, die zum überwiegenden Teil in gleicher Richtung bewertet bzw. von den OFG-Teilnehmern kontroverser diskutiert wurden, enthielten. Ebenso wie in den konventionellen Fokusgruppen wurde auch in den OFGs eine äquivalente Anzahl an

Verbesserungsvorschlägen generiert. Für die Online-Variante spricht, dass die Partizipationsraten online signifikant gleichverteilter war als bei den Fokusgruppen im Labor

Insgesamt zeigen die Ergebnisse, dass bei der inhaltlichen Qualität der Daten keine wesentlichen Unterschiede bestehen und OFGs im Rahmen von Web Usability-Tests eine interessante Ergänzung zum Methodenarsenal des Usability-Praktikers sein können. OFGs erlauben eine zeiteffiziente empirische Überprüfung von Website-Konzepten und werden damit der Schnellebigkeit im Webprojekt-Geschäft besonders gerecht. Unsere praktische Erfahrung hat auch gezeigt, dass die Akquise der OFG-Teilnehmer bei internetspezifischen Fragestellungen sowohl zeit- als auch kostenökonomisch mittels Online-Access-Panels erfolgen kann. Ein gepflegtes Access-Panel kann so gut wie alle Quotierungswünsche des Forschers erfüllen kann. Insbesondere bei klar definierten Zielgruppen (z.B. regelmäßige Nutzer von Reise-Websites o.ä.) bietet ein nationales Access-Panel eine gute Ausgangsbasis für eine ökonomisch durchführbare Quotierung, bei der auch das Auftreten eines „urbanen Bias“ vermieden werden kann (vgl. EPPLE & HAHN 2001 S.252).

Für die Zukunft wünschenswert sind weitere Methodentests, um die Anwendbarkeit bei verschiedenen Fragestellungen, Zielgruppen und Anwendungen zu prüfen.

Literaturverzeichnis

- BALAZS, B. & CHRISTOPHERSEN, T. (2002). Entwicklung und Überprüfung eines benutzerzentrierten Fragebogens zur Erfassung der Usability im E-Commerce, Präsentation auf der German Online Research 2002, Stuttgart
- BOOS, M. & CORNELIUS, C. (2001). Zur Erfassung konversationaler Kohärenz in direkter und computervermittelter Kommunikation, in: F.W. HESSE & H.F. FRIEDRICH (Hrsg.) Partizipation und Interaktion im virtuellen Seminar, Münster: Waxmann, S.55-80
- BORTZ, J. & DÖRING, N. (1995). Forschungsmethoden und Evaluation für Sozialwissenschaftler, 2.Aufl., Berlin: Springer
- CORNELIUS, C. (2001a). Best Match zwischen Forschungsfrage und Design von Online-Fokus-Gruppen, in: K.J. JONAS, P. BREUER, B. SCHAUENBURG & M. BOOS (Hrsg.). Perspectives on Internet Research: Concepts and Methods (Online-Document), URL: <http://server3.unipsych.gwdg.de/gor/contrib/cornelius-caroline/cornelius-caroline.html>, Stand 31.01.02, 9 Seiten
- CORNELIUS, C. (2001b). Gegenseitiges Verständnis in Computerkonferenzen. Voraussetzungen und Folgen konversationaler Kohärenz in Entscheidungsfindungsgruppen im Medienvergleich, Münster: Waxmann
- EPPLE, M. & HAHN, G. (2001). Dialog im virtuellen Raum: Die Online-Focus Group in der Praxis der Marktforschung, in: A. Theobald , M. Dreyer & T. Starsetzki (Hrsg.). Online-Marktforschung - Theoretische Grundlagen und praktische Erfahrungen, Wiesbaden: Gabler, S.249-260
- GÖRTS, T. (2001). Gruppendiskussionen - Ein Vergleich von Online- und Offline-Focus-Groups, in: A. THEOBALD , M. DREYER & T. STARSETZKI (Hrsg.). Online-Marktforschung - Theoretische Grundlagen und praktische Erfahrungen, Wiesbaden: Gabler, S.149-164

- GREENBAUM, T. (1995). Focus Groups on the Internet: An Interesting Idea but not a Good One, URL: <http://www.quirks.com/CGI-BIN/SM40i.exe?docid=3000:58911&%70assArticleID=136>, Stand: 30.11.99, 3 Seiten
- HERRING, S.C. (1999). Interactional Coherence in CMC, in: Journal of Computer Mediated Communication, Vol.4, No.4, URL: <http://www.ascus.org/jcmc/vol4/issue4/herring.html>, Stand: 17.06.01, 25 Seiten
- JOINSON, A. (2001). Self-Disclosure in Computer-Mediated Communication: The Role of Self-Awareness and Visual Anonymity, in: European Journal of Social Psychology, Vol.31, No.2, pp.177-192.
- KIRAKOWSKI, J. & CIERLIK, B. (1998). Measuring the Usability of Web Sites, Paper presented at the Human Factors and Ergonomics Society Annual Conference, Chicago, URL: <http://www.ucc.ie/hfrg/questionnaires/wammi/research.html>, Stand: 07.11.02, 6 Seiten
- MCLAUGHLIN, M.L.(1984). Conversation: How Talk is Organized, Beverly Hills
- MORGAN, D.L. (1997). Focus Groups as Qualitative Research, 2. Aufl., Newsbury Park: Sage Publications
- PALMQUIST, J. & STUEVE, A. (1996). Stay Plugged into New Opportunities, in: Marketing Research, Vol.8, No.1, pp.13-15
- POSTMES, T., SPEARS, R. & LEA, M. (1999). Social Identity, Normative Content and Deindividuation in Computer-Mediated Groups, in: N. ELLEMERS, R. SPEARS & B. DOOSJE (Eds.). Social Identity: Context, Commitment and Content, Oxford: Blackwell, pp.164-183
- PRICKARZ, H. & URBAHN, J. (2002). Qualitative Datenerhebung mit Online-Fokusgruppen, in: Planung & Analyse, 28.Jg., Nr.1, S.63-70
- WALTHER, J.B. (2000). Die Beziehungsdynamik in virtuellen Teams, in: M. BOOS, K. JONAS & K. SASSENBERG (HRSG.). Computervermittelte Kommunikation in Organisationen, Göttingen, Hogrefe, S.11-25
- WEISBAND, S.P., SCHEIDER, S.K. & CONNOLLY, T. (1995). Computer-Mediated Communication and Social Information: Status Salience and Status Differences, in: Academy of Management Journal, Vol.38, No.4, pp.1124-1151
- YOM, M. (in Vorbereitung). Web Usability von Online-Shops - Erkenntnisse & Testmethoden, Dissertationsprojekt am Institut für Marketing, Universität Dortmund