

Drop-Out-Analyse: Wirkungen des Untersuchungsdesigns

F. Knapp¹, M. Heidingsfelder²

Drop-Out, der vorzeitige Ausstieg aus einer Befragung (Befragungsabbruch), ist für Web-/ Online-Befragungen ein spezielles Problem. So stehen, ähnlich wie bei schriftlichen Befragungen, keine Interviewer zur Verfügung, die das Interview steuern, Mißverständnisse ausräumen und so die Befragten bei der Stange halten bzw. im schlimmsten Fall Gründe des Drop-Outs festhalten können. Die Art der Befragtenrekrutierung und der Befragung ist oft "unpersönlicher" (z.B. Banner / Pop-Up) als bei anderen Befragungsformen. Zudem führt die zunehmende Zahl von Online-Umfragen zu einem gewissen Wear-Out bei den Befragten. Damit sinkt insgesamt die Hemmschwelle, die Beantwortung der Fragen abubrechen.

Je nach Befragungsform sind unterschiedliche Drop-Out-Formen möglich. Drop-Outs sind zunächst allgemein Ausfälle während des Befragungsprozesses. Im Offline-Bereich, also bei schriftlichen Befragen oder dem Download von Fragebögen aus dem Internet, gibt es als Ausfallmöglichkeiten nur die fehlgeschlagene Kontaktierung ausgewählter Personen und die nicht erfolgte Rücksendung des Fragebogens. Informationen über die Ausfälle liegen nur dann vor, wenn in einer vorherigen Phase solche Informationen erfaßt wurden. Im Online Bereich ist dies für fortlaufende (Html-) Fragebögen ähnlich. Im Fall des Rogator-Verfahrens „eine Frage – ein Bildschirm“ (vgl. Gräf & Heidingsfelder 1999) liegen aber auch inhaltliche Informationen (bereits beantwortete Fragen) für diese Ausfälle vor. Dann ist nämlich bekannt, an welcher Stelle der Drop-Out erfolgt ist. So können bisherige Befragungsdauer, Aufbau des Fragebogens, Art der Frage usw. als mögliche Ausfallursachen betrachtet werden.

Warum ist die Beschäftigung mit Drop-Outs / Abbrüchen überhaupt von Interesse? Im Internet gibt es ja viel eher die Möglichkeit, größere Stichproben zu ziehen, da sollte es auf einen mehr oder weniger nicht ankommen, so könnte man meinen. Das Interesse an ihrer Vermeidung hat aber folgende Gründe:

- Bei Befragungen in speziellen, schwer erreichbaren Zielgruppen sind unnötige Ausfälle zu vermeiden (vgl. S2 weiter unten).
- Die Argumentebasis gerade im qualitativen Bereich wird schmaler.

1 psyma GmbH, Fliedersteig 17, 90607 Rückersdorf, Tel. +49911-9957440, frank.knapp@psyma.com.

2 Rogator AG, mheid@rogator.de.

- Haben die Ausfälle systematische Ursachen, unterscheiden sie sich möglicherweise von den Antwortenden hinsichtlich untersuchungsrelevanter Merkmale, man erhält eine verzerrte Stichprobe.
- Schließlich scheinen Abbrüche und fehlerhafte Antworten die gleichen Ursachen zu haben, so daß hohe Drop-Outs gleichzeitig Qualitätsprobleme signalisieren können, die sich auch bei den Antwortenden niederschlagen (vgl. auch Jeavons 1999).

Durch Verbesserungen der Untersuchungsanlage, des Fragebogens und der Erhebungssoftware können so Drop-Outs reduziert, die Beschränkung auf „Berufsbefragte“ und „Incentive-Jäger“ gelockert oder aufgehoben werden und qualitativ bessere Antworten bewirkt werden

Drop-Outs können technisch oder benutzerinduziert sein. Im ersten Fall wird unbeabsichtigt server- oder client-seitig die Verbindung gekappt. Im zweiten Fall sinkt die Teilnahmemotivation des Befragten während der Befragung so stark ab, daß er diese abbricht. Dafür sind folgende Ursachen denkbar (vgl. Gräf 1999, Vogt 1999):

- Mangelnde Motivation des Befragten (vgl. Bosnjak & Batinic 1999),
- ein zu langer Fragebogen,
- fehlerhaften Fragen (z.B. mißverständliche Fragen),
- uninteressante optische Gestaltung der Befragung,
- sensible Fragen.

Im weiteren werden Drop-Out-Verläufe aus mehreren Befragungen zusammengetragen und analysiert. Dabei soll geprüft werden, ob sich Drop-Outs bei bestimmten Untersuchungskonstellationen besonders häufen. Insbesondere soll folgenden Fragen nachgegangen werden:

- Bei welcher Art von Studien bzw. Fragen treten Abbrüche verstärkt auf?
- Liefern die Abbrecher eher die unbrauchbaren Antworten, so daß auf sie verzichtet werden kann?
- Unterscheiden sich die Abbrecher in ihrer Struktur von den Teilnehmern, so daß eine Verzerrung der Ergebnisse erwartet werden muß?

Unter „Teilnehmer“ sind im weiteren Befragte gemeint, die den Fragebogen komplett durchgegangen sind (ohne notwendigerweise alle Fragen beantwortet zu haben) und so in der Ergebnisdatei landen. „Abbrüche“ werden bei der analysierten „eine Frage–ein Bildschirm“-Methode des verwendeten Rogator-Systems dann vermerkt, wenn ein Befragter seit 1 h keine Antwort zurückgesandt hat.

2. Untersuchte Studien

Tabelle 1 zeigt die für die Untersuchung herangezogenen Studien. Diese wurden von Mai bis Juli 1999 durchgeführt. Die erste Befragung war eine Website-Evaluation für West. Die Befragungsdauer war hier am längsten, die Abbruchquote die zweit-

höchste. Diese wurde nur noch von der zweiten Umfrage überschritten, eine Befragung zu AIDS / HIV und entsprechenden Behandlungsmethoden, für die männliche Homosexuelle auf der Website von Gayforum.de rekrutiert wurden. Befragungen 3-5 sind identisch aufgebaute Befragungen auf den Websites von 3 Talkshows. S6-9 sind wiederum Website-Evaluationen mit Fragen zu verwandten Themen. Bei S1 (Website-Evaluation West Online) wurden 50x3 Stangen Cigaretten unter den Teilnehmern ausgelobt, bei S7-S9 Karten für Autorennen. Im übrigen erfolgte die Programmierung jeweils so, daß für geschlossene Fragen eine Antwort angegeben werden mußte.

Tabelle 1: Untersuchte Studien

Studie	S1 (West)	S2 (Gay- forum)	S3	S4	S5	S6	S7	S8	S9
Teilnehmer	394	117	1805	912	3272	7825	863	126	245
Abbrüche	517	201	1129	533	1636	5957	134	33	66
ohne 1. Frage	267	95	529	273	798	2010	96	16	41
Abbruch- quote	57%	63%	38%	38%	33%	43%	13%	21%	21%
ohne 1. Frage	40%	45%	23%	23%	20%	20%	10%	11%	14%
Dauer	12m38s	4m22s	7m29s	6m12s	7m53s	9m45s	8m11s	7m39s	8m26s
#Fragen	35	16/84	12	12	12	20	15	16	15
#Matrix- fragen	3	0/-	2	2	2	3	7	7	7
#offener Fragen	7	3	5	5	5	1	0	0	0
Rekrutierung	Pop-up	Pop-up	Self	Self	Self	Self	Self	Self	Self
Incentive	ja	nein	nein	nein	nein	nein	ja	ja	ja
Ziel	WE	HIV / AIDS	Talk- show 1	Talk- show 2	Talk- show 3	WE	WE	WE	WE

Self: Self Sample; WE: Website Evaluation

Die Abbruchverläufe (Anteil der Abbrüche je Frage in %) zeigen ein regelmäßig wiederkehrendes Ergebnis: Bei der 1. Frage brechen ca. 50% der Abbrecher ab, danach verteilen sich die Abbrüche in etwa. Es gibt aber bei einzelnen Fragen markanten Spitzen, gegen Ende brechen nur noch wenige ab. Je nach Untersuchungsanlage sind die Abbruchquoten unterschiedlich hoch, längere Befragungen und sensible Themen führen zu höheren Abbruchquoten. Im weiteren werden nur die Studien mit hohen Abbruchquoten betrachtet, da diese Einblicke in deren Ursachen erlauben.

3. Abbrüche nach Art der Frage

Betrachtet man für S1 die einzelnen Fragen nach der Höhe der Abbrüche so ergibt sich folgendes Bild: Die ersten beiden Fragen belegen die höchsten Abbruchquoten, danach schieben sich aber offene Textfragen und Mehrfachauswahl-Fragen in der Rangfolge

nach vorne. Fragen 3 und 4 verlangen nur eine Einfachauswahl (Tab. 2).³

Dies Ergebnisse lassen sich auch in den anderen Studien replizieren. Für die ansonsten inhaltsgleichen Fragen zu verschiedenen Talkshows (S3-5) gibt es beim Verlauf der Abbrüche einen Peak bei den Fragen 8 (bzw. 9). Diese sind Matrixfragen und erheben die Bekanntheit von jeweils 6 Talkshows auf einer 7er Skala. Die Studie S6 zeigt ein Plateau für die Fragen 9-11, die einzigen Matrixfragen der Studie, in denen jeweils 5-6 Bereiche der Website auf einer 7er Skala eingestuft werden. Die Abbrecher zeigen hier auch deutlich mehr Antwortmuster (gleiche Antwortkategorie bei allen Items einer Matrix: 19-30% vs. 13-23%). Andere Antwortmuster konnten nicht untersucht werden, weil eine Item-Randomisierung erfolgte

Tabelle 2: Abbrüche nach Fragen

Anteil an Abbrüchen gesamt in %	Anteil an Befragten je Frage in %*	Fragennummer	Bearbeitungsdauer in sec**	Art der Frage
48,4%	27,4	1	25	Einfachauswahl
7,5%	5,9	2	13	Einfachauswahl
5,4%	4,6	5	35	(offene) Textfrage
5,4%	5,5	9	33	(offene) Textfrage
5,2%	5,1	8	34	(offene) Textfrage
4,4%	4,1	7	36	(offene) Textfrage
4,3%	4,6	10	33	Mehrfachauswahl
3,9%	3,5	6	35	(offene) Textfrage
2,7%	3,1	14	19	Mehrfachauswahl
1,9%	2,4	23	37	Matrixfrage (6 items, 7er Skala)

*Teilnehmer+Abbrecher, denen die Frage vorgelegt wurde.

** ohne Befragte mit Zeiten über 150 sec pro Frage.⁴

Frage 19 schiebt sich in Spalte 2 mit 7,7, % dazwischen, enthält aber nur noch 6 Abbrecher.

Offene Textfragen und Matrixfragen sind also beschwerlich für viele Befragte, so daß hier mehr Abbrüche erfolgen.⁵ Allerdings handelt es sich bei Matrixfragen um multiple Fragen, die dadurch eine höhere Abbruchwahrscheinlichkeit aufweisen.

Tabelle 3a: Nicht-Beantwortung offener Fragen – Anteil der „Weiterklicker“

	Frage 5	Frage 6	Frage 7	Frage 8	Frage 9	Frage 22
Teilnehmer	8%	19%	24%	36%	19%	16%
Abbrecher	31%	44%	47%	55%	44%	41%
N (Abbrecher)	183	163	140	113	85	37

3 Frage 22 ist die einzige weitere offene Textfrage, die allen Befragten gestellt wurde. Bis zu diesem Zeitpunkt hat man sich offensichtlich „durchgekämpft“, so daß die Abbruchneigung abnimmt.

4 Dies dient der Elimination künstlicher Unterbrechungen.

5 Die Abbruchrate für Textfragen ist bei S1 4mal so hoch wie für die Einfachauswahl.

Tabelle 3b: Bearbeitungsdauer bei offenen Fragen in sec

	Frage 5	Frage 6	Frage 7	Frage 8	Frage 9	Frage 22
Teilnehmer sec	35	35	36	34	33	40
N	383	382	381	380	386	380
Abbrecher sec	27	24	26	23	24	41
N	183	163	139	113	84	37

* ohne Befragte mit Zeiten über 150 sec pro Frage

4. Antwortqualität der Abbrecher

Für die Analyse der Antwortqualität haben wir zwei Maße für das „Desinteresse“ ausgewählt, nämlich die Häufigkeit, mit der bei offenen Fragen einfach weitergeklickt wurde, und die Bearbeitungsdauer für offenen Fragen.

In den Tabelle 3a/b ist zu erkennen, daß die Abbrecher deutlich weniger Neigung zur Beantwortung offener Fragen zeigen. Die Rate der Weiterklicker ist deutlich höher als die der Teilnehmer und entsprechend die Bearbeitungsdauer kürzer.⁶ Diese Hypothesen lassen sich auch stützen durch die Studien S3-S5. Die Bearbeitungsdauer der offenen Fragen ist für die Teilnehmer durchweg länger als für die Abbrecher (im Schnitt um 6 sec, vgl. Tabellen 4/5).

Tabelle 4: Anteil der „Weiterklicker“ über 5 offene Fragen in S3-S5 in %

	S3	S4	S5
Teilnehmer	6	7	5
Abbrecher	21	29	16

Tabelle 5: Differenz Bearbeitungsdauer (Teilnehmer-Abbrecher), offenen Fragen in S3-S5 in sec

	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
S3	+4	+5	+6	+4	+7
N (Teilnehmer/Abbrecher)	(1657/349)	(1744/313)	(1728/265)	(1750/232)	(1734/204)
S4	+7	+4	+8	+5 (893/94)	+10
N (Teilnehmer/Abbrecher)	(861/190)	(888/153)	(896/119)		(880/75)
S5	+6	+6	+7	+4	+4
N (Teilnehmer/Abbrecher)	(2975/573)	(3141/518)	(3133/427)	(3130/376)	(3148/330)

In Klammern die Zahl der Teilnehmer / Abbrecher; jeweils Testpersonen mit max. 150 sec Bearbeitungsdauer.

⁶ Die Unterschiede wären bis auf Frage 22 signifikant für $p = 0,05$ (T-Test auf Mittelwertsunterschiede für ungleiche Varianzen oder Varianzanalyse).

Tabelle 6: Surferfahrung auf West Online

Seit wann	Teilnehmer	Abbrecher	Wie oft	Teilnehmer	Abbrecher
Heute zum ersten Mal	13%	30%	mehrmals täglich / täglich	38%	29%
Seit 1-2 Monaten	25%	25%	mehrmals die Woche	38%	36%
Seit 3-6 Monaten	20%	18%	ca. 1 / 2mal die Woche	18%	28%
Seit 6-12 Monaten	19%	17%	ca. 1 / 2mal im Monat / seltener	6%	7%
Länger als 12 Monate	23%	11%			
N	394	226	N	342	151

5. Strukturunterschiede zwischen Teilnehmern und Abbrechern

In den ersten Fragen der Studie S1 wurden die Befragten zu ihrem Surfverhalten in Bezug auf West Online befragt. Hier zeigt sich, daß die Teilnehmer länger und häufiger dabei sind (vgl. Tabelle 6). Ein Bookmark haben 57% der Teilnehmer, aber nur 46% der Abbrecher gesetzt. Auch in S6 zeigt sich, daß bei den Teilnehmern die täglichen Besucher der Website stärker vertreten sind (31% versus 25% bei den Abbrechern). Ein deutlicher Unterschied in der Länge der wöchentlichen Online-Zeit zeigt sich bei S6 allerdings nicht.

Insgesamt ist dies ein Hinweis darauf, daß die Community wesentlich stärker bereit ist, Fragen zur Website zu beantworten, als dies Gelegenheitssurfer der Website tun.

6. Fazit

Man muß also zu dem Schluß kommen, daß in Website-Befragungen speziell die Community befragt wird, die antwortbereiter ist bzw. tiefere Antworten auf offene Fragen geben kann und will. Mithin liegen wohl strukturelle Verzerrungen durch Abbrüche vor. Dies erlaubt aber dennoch vielfältige Aussagen, etwa im Benchmarking-Bereich, nämlich:

- Gelingt überhaupt die Herausbildung einer Community? (Interessiert sich überhaupt eine größere Gruppe von Befragten für die Befragung?)
- Wie verhält sich diese Community im Vergleich zu anderen Communities?

Weiter bedarf es eines dramatischen Anreizes in der (Formulierung?) der Frage 1, damit andere Befragte motiviert werden können.

Die Ergebnisse geben Anlaß zur Hypothesenbildung über Ursachen von Abbrüchen und Ansätzen zu ihrer Reduzierung. Der nächste Schritt wäre die Konstruktion spezifischer Methodentests. Dazu zählt etwa der Vergleich der Abbrüche für Matrixfragen im Vergleich zu entsprechenden Einzelfragen. Die hohe Zahl von Antwortmustern auch bei Teilnehmern deutet allerdings auch auf ein inhaltliches Problem hin. Denkbar ist auch, eine Befragung aus offenen und geschlossenen Fragen aufzuteilen. Ein große Stichprobe wird nur mit geschlossenen Fragen konfrontiert, um eine möglichst unverzerrte Stichprobe durch weniger systematische Abbrüche zu erreichen.

Dies kann durch eine qualitative Stichprobe mit offenen und geschlossenen Fragen und entsprechenden Vergleichen ergänzt werden. Sie deckt möglicherweise schwerpunktmäßig die Community ab, ermöglicht aber immer noch die genannten Aussagen. Die vielfältigen Möglichkeiten der modernen Erhebungssoftware erlauben dabei die Überprüfung der Hypothesen, etwa durch die ständige Erhebung der technischen Ausstattung der Nutzer, des Zeitbedarfs und anderer Befragungsparameter. Dies dient der ständigen erfahrungsgestützten Optimierung von Software und Fragebogen.

Literaturangaben

- Jeavons, A. (1999). Ethology and the Web. *Marketing And Research Today*, May, 69-76.
- Gräf, L. (1999). Optimierung von WWW-Umfragen: Das Online Pretest-Studio. In B. Batinic, A. Werner, L. Gräf und W. Bandilla (Hrsg.), *Online Research*. Göttingen: Hogrefe.
- Gräf, L., & Heidingsfelder, M. Bessere Datenqualität bei WWW-Umfragen – Erfahrungen aus einem Methodenexperiment mit dem Internet-Rogator. In B. Batinic, A. Werner, L. Gräf und W. Bandilla (Hrsg.), *Online Research*. Göttingen: Hogrefe.
- Vogt, K. (1999). Verzerrungen in elektronischen Befragungen? In B. Batinic, A. Werner, L. Gräf und W. Bandilla (Hrsg.), *Online Research*. Göttingen: Hogrefe.
- Bosnjak, M., & Batinic, B. (1999). Determinanten der Teilnahmebereitschaft an internet-basierten Fragebogenuntersuchungen am Beispiel E-Mail. In B. Batinic, A. Werner, L. Gräf und W. Bandilla (Hrsg.), *Online Research*. Göttingen: Hogrefe.