Abteilung für Entwicklungspsychologie und Pädagogische Psychologie des Instituts für Psychologie der Uni Wien

Brigitte Wagner:

Online Research in der Psychologischen Forschung

Teil 2:
Stichprobe
Gütekriterien

Skriptum zur Lehrveranstaltung
Online Research in Theorie und Praxis

Leitung:
Ass.Prof. Dr. Marco Jirasko

Studienjahr 2001/02

2. Stichprobe und Gütekriterien

Die Erörterung internetspezifischer Charakteristika betreffend online erhobener Daten wie Aspekte der Stichprobe, Rekrutierung, Selbstselektion, Erwartungshaltungen, Gütekriterien und Generalisierbarkeit werden an dieser Stelle behandelt.

Da man nicht nur innerhalb der Online-Forschung, sondern bei vielen populationsbeschreibenden Untersuchungen auf die Ziehung reiner Zufallsstichproben verzichten muß, fordern Bortz und Döring (1995, S. 452) zumindest die Diskussion folgender Fragen:

- Für welche Population sollen die Untersuchungsergebnisse gelten?
- Nach welchem Verfahren wurden die Untersuchungsobjekte ausgewählt?
- Inwieweit ist die Generalisierung der Ergebnisse durch Besonderheiten des Auswahlverfahrens eingeschränkt?
- Gibt es strukturelle Besonderheiten der Population, die eine Verallgemeinerung der Ergebnisse auch auf andere, nicht untersuchte Populationen rechtfertigen?
- Welche Tragweite (Präzision) haben die Ergebnisse angesichts des untersuchten Stichprobenumfanges?
- Welche Überlegungen nahmen Einfluß auf die Festlegung des Stichprobenumfanges?

Zusätzlich zur Beantwortung oben genannter Fragen gilt es im Rahmen der Online-Forschung weitere wesentliche Aspekte zu beachten und Möglichkeiten zu überlegen, wie den spezifischen Anforderungen entsprochen werden kann.

2.1 Stichprobentheoretische Überlegungen

Der ersten Schritt jeder (psychologischen) Forschungsarbeit ist die Auswahl einer möglichst aussagekräftigen Stichprobe die so zu wählen ist, dass die Ergebnisse Aussagen über die Grundgesamtheit (Population) erlauben. Diese Auswahl sollte (meist) per Zufallsprinzip (Randomisierung) erfolgen.

Im Idealfall kann man mit echten Zufallsstichproben arbeiten, die Realität zwingt jedoch zu Kompromissen, wie es Birnbaum (zitiert nach Reips, 2000) anschaulich illustriert:

> *„Some say that psychological science is based on research with rats, the mentally disturbed, and college students. We study rats because they can be controlled, the disturbed because they need help, and college students because they are available“.*

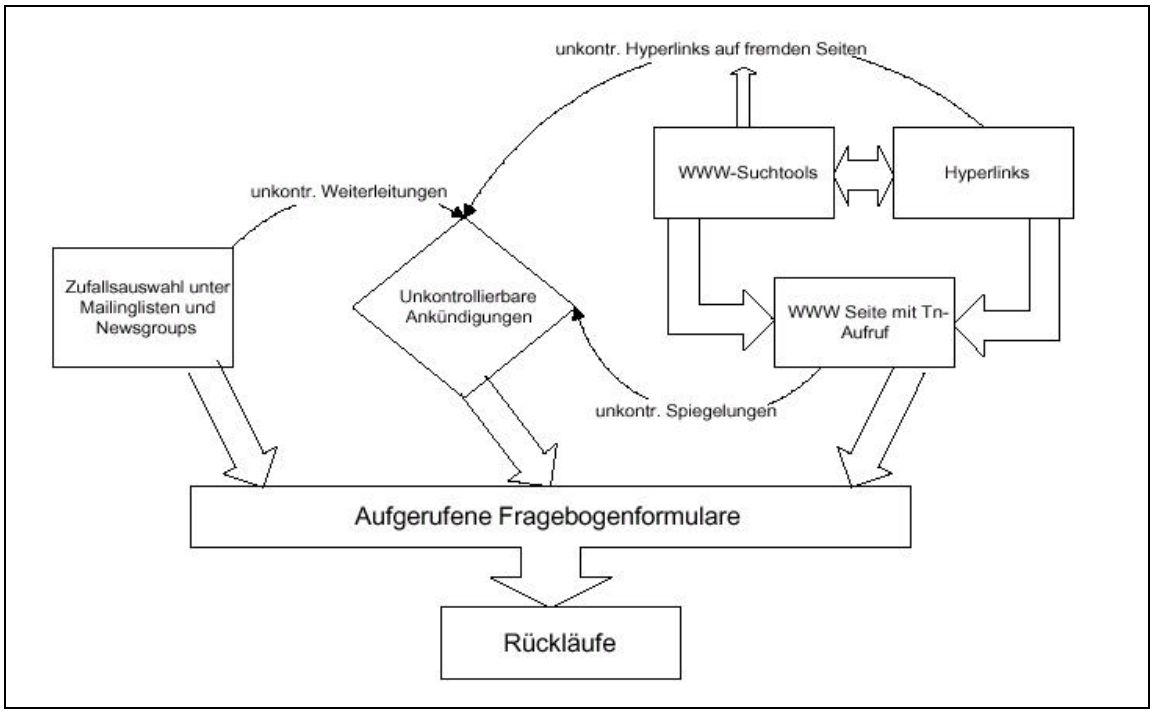

Abbildung 2.1: Illustration der eingeschränkten Verbreitungskontrolle bei Umfragen per e-Mail (Bosnjak, 1997, S. 58).

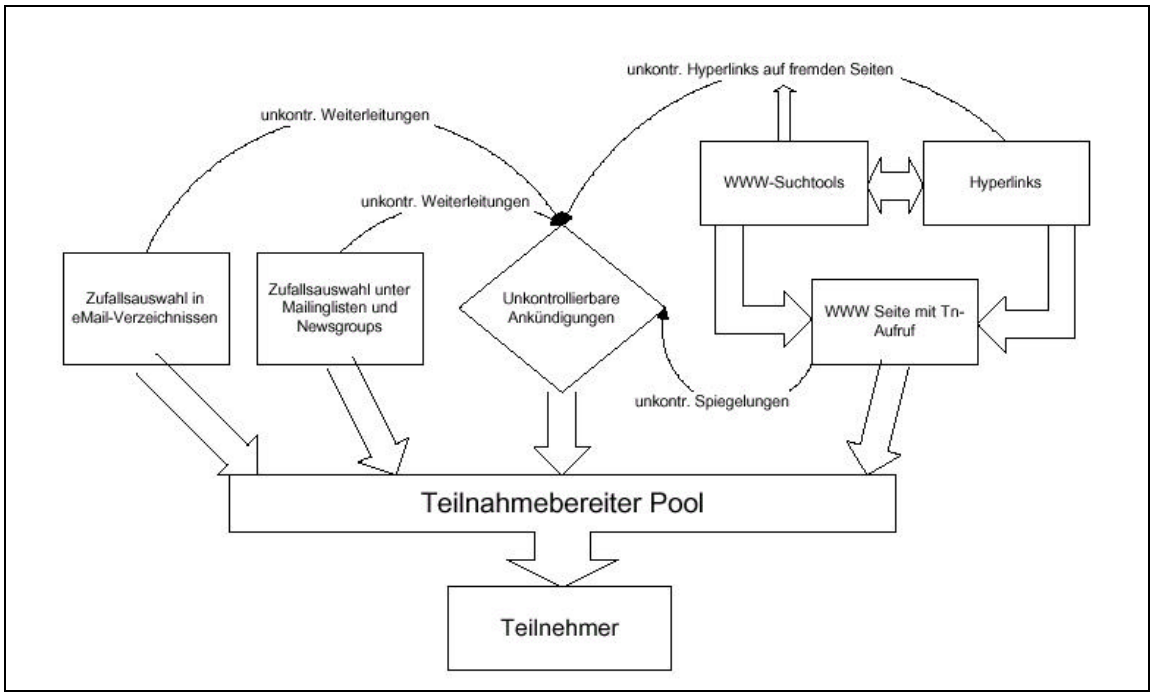

Abbildung 2.2: Probleme der Verbreitungskontrolle bei WWW – Umfragen (Bosnjak, 1997, S. 60).

Bezüglich der via Internet gewonnenen Stichproben sprechen Batinic und Bosnjak (1999) von nicht-probabilistischen Gelegenheitsstichproben, da im Internet keine Aussagen über den Nutzerkreis gemacht werden können. Folgende Gründe sind dafür ausschlaggebend:

- Unvollständiger symbolischer Verfügbarkeit der Untersuchungseinheiten

- Quantitativ und qualitativ dynamischer Grundgesamtheit
- Eingeschränkter Möglichkeiten zur Verbreitungskontrolle bei e-Mail Rekrutierung vs. Teilnehmeranwerbung via WWW. Abbildung 2.1 bzw. 2.2 illustrieren dieses Problem.
- Mehrstufiger Selektionsprozesse (vgl. Kapitel 2.1.3.1)

Obige Ausführungen abschließend, soll mittels Abbildung 2.3 eine anschauliche Illustration der Stichprobenauswahl im Internet geboten werden.

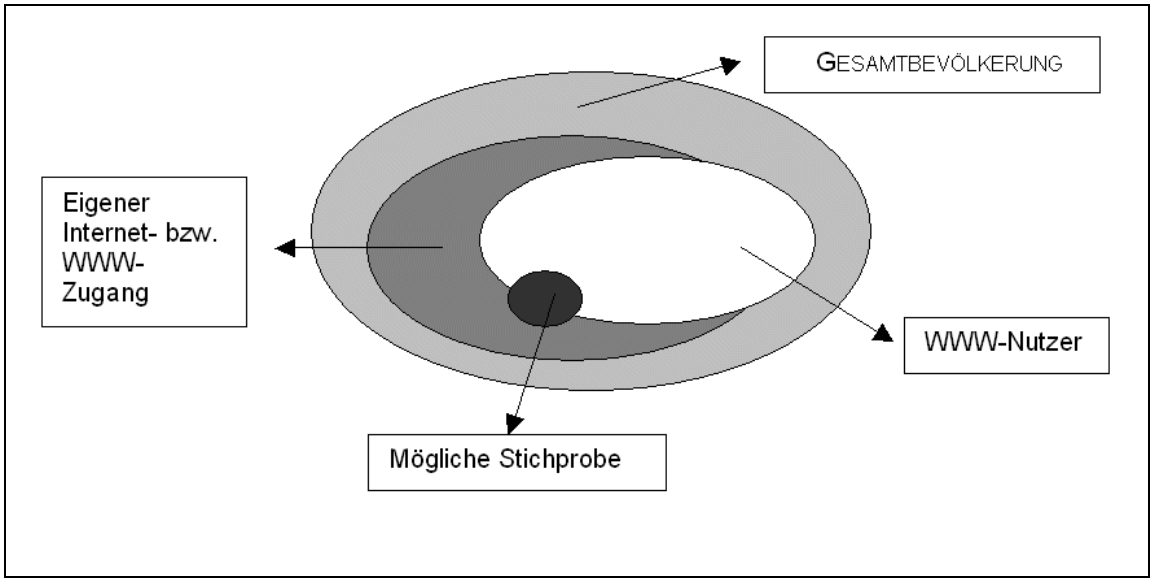

Abbildung 2.3: Modell der Stichprobenauswahl im WWW (Theobald, 2000).

2.1.1 Rekrutierungswege

Im Internet stehen dem Forscher verschiedenste Möglichkeiten zur Verfügung um Untersuchungsteilnehmer anzuwerben (Batinic & Bosnjak, 1997):

Veröffentlichung eines WWW-Teilnahmeaufrufs

Dazu müssen in erster Linie Seiten gefunden werden, die von der zu untersuchenden Zielgruppe häufig besucht werden. Dort bestehen dann grundsätzlich zwei Möglichkeiten:

- es wird ein (externer) Hyperlink auf die Untersuchung gesetzt oder
- man bringt Werbebanner dafür an (meist kostenpflichtig).

Teilnahmeaufrufe über Newsgroups und Mailinglisten

Bei Anwendung dieser Methode ist es unabdingbar, sich ausführlich mit den entsprechenden Gepflogenheiten der jeweiligen Newsgroups bzw. Mailinglisten auseinanderzusetzen

(z. B. Netiquette). Ansonsten sind mitunter schwer gutzumachende Imageverluste zu befürchten (Batinic & Bosnjak, 1997).

Rekrutierung von Teilnehmern via e-Mail bzw. e-Mail Umfragen

Zu beachten ist bei dieser Rekrutierungsmethode, dass unaufgefordert zugestellte e-Mails als Belästigung erlebt werden könnten. Daher sollte es als selbstverständlich gelten, in der e-Mail Aussendung auf die Bezugsquelle der Adresse zu verweisen. Batinic und Bosnjak (1997) stehen dem Einsatz dieses Rekrutierungsweges sehr verhalten gegenüber. So meinen sie beispielsweise, dass unaufgefordert zugestellte e-Mails, deren Adressen auf der Basis von e-Mail Verzeichnissen zusammengestellt wurden, einen groben Verstoß gegen die Netiquette darstellen und „*die Methode daher nur mit Bedacht und nach kritischer Abwägung der Wichtigkeit der Untersuchung einzusetzen ist. Beispiele hierfür sind Umfragen in innerbetrieblichen Netzen (Intranet), Expertenbefragungen oder kreuzkulturelle Umfragen [...]*“. (Batinic & Bosnjak, 1997, S. 236).

Es gibt jedoch auch andere Standpunkte als die oben genannten, die den Einsatz dieser Methode weit weniger restriktiv handhaben. Denn pragmatisch betrachtet, ist das Löschen dieser unaufgeforderten e-Mail Teilnahmeaufforderungen ein Aufwand von ein paar Sekunden und andererseits stellt sich die Frage, ob Internet-Nutzer, die ihre e-Mail Adresse in ein öffentlich zugängliches Verzeichnis eintragen lassen, nicht damit zu rechnen haben auch für solche Zwecke angeschrieben zu werden.

2.1.2 Erwartungshaltungen der Versuchspersonen

An dieser Stelle sollen Erwartungshaltungen und ihre Rollen als Störvariablen erörtert werden. Prinzipiell wird zwischen Effekten, die sich auf die Person des VL beziehen und jenen, die der VP zugeschrieben werden können, unterschieden. Eine nähere Betrachtung der Versuchsleiter-Erwartungseffekte ist nicht notwendig, da das Online-Forschen zumeist durch das Fehlen eines physisch anwesenden Versuchsleiters charakterisiert ist. Den Versuchspersonen-Effekten muß aber im Rahmen der Online-Forschung große Aufmerksamkeit geschenkt werden. Aufgrund der mangelnden Kontrollmöglichkeit im Internet müssen die Erwartungshaltungen bzw. Motive der Untersuchungsteilnehmer erörtert werden. Wie auch in Kapitel 2.1.3.1 angesprochen wird, ist das Problem der Selbstselektion eine dieser Einflußgrößen, die nicht willkürlich manipuliert werden können aber deren Vorhandensein bisweilen beträchtlichen Einfluß auf die Untersuchungsergebnisse haben kann. Generell muß man dem Online-Forschen, im Vergleich zu Laborexperimenten oder auch Face-to-Face Befragungen, zu Gute halten, dass die Freiwilligkeit der Teilnahme zu jeder Zeit gewährleistet ist. Somit sollten Erwartungen über die soziale Erwünschtheit von Verhaltensweisen bzw. Antworten weniger stark zum Ausdruck kommen als in herkömmlichen Settings, da eine anonymere Untersuchungssituation gewährleistet scheint und die Versuchspersonen nicht mit Sanktionsmaßnahmen rechnen müssen. Es bleiben jedoch

noch zwei Einflußfaktoren zu erwähnen: Die Erwartungen aufgrund der Aufforderungsvariablen bzw. Erwartungen, wie eine experimentelle Bedingung wirkt (Huber, 2000). Während man im ersten Fall die Möglichkeit nicht ausschließen kann, dass sich Untersuchungsteilnehmer aus Gründen der Gefälligkeit dem VL gegenüber ereifern, im Sinne der erörterten Instruktionshinweise zu handeln (selbst wenn dieser nicht physisch anwesend ist), so beschreibt der zweite Fall die Annahmen der VP bezüglich der zu erwartenden experimentellen Manipulation.

2.1.3 Teilnahmemotive der Versuchspersonen

Orientiert an Huber (2000) sollen hier die zwei grundsätzlichen Typen von Teilnahmemotiven herausgegriffen werden, die mitunter für wissenschaftliche Untersuchungen entscheiden sein können:

- *Motive für die Teilnahme bzw. Nicht-Teilnahme* und
- *Motive, die das Verhalten während des Experimentes beeinflussen können*.

Der Fokus liegt dabei auf dem ersten Aspekt, da dieser unter dem Schlagwort der Selbstselektion immer wieder auftritt und gebührend beachtet werden sollte. Bezüglich des zweiten Aspektes bietet die Online-Forschung nur ein vergleichsweise kleines Repertoire an Möglichkeiten, um regulierend auf die Teilnehmermotivation einzugreifen.

Porst und Briel (1995) differenzieren drei Hauptaspekte der Teilnahmebereitschaft, die in weiterer Folge von Bosnjak und Batinic (1999) aufgegriffen worden und wie folgt lauten:

- *Altruistische Gründe*
  z. B. die Bereitschaft einen persönlichen Beitrag für die Forschung zu leisten oder gesellschaftliche Verantwortlichkeit
- *Befragungsbezogene Gründe*
  z. B. Neugier oder Interesse
- *Persönliche Gründe*
  Z. B. Selbsterkenntnis, Lerneffekte oder das Unvermögen „nicht nein sagen zu können“ etc.

Ein nicht unwesentlicher Teilnahmeanreiz soll schließlich auch nicht verschwiegen werden: der materielle Aspekt. Ein Experiment, das sich mit den Auswirkungen von in Aussicht gestellten Entschädigungen (Incentives)[1] beschäftigt, findet sich in einem späteren Kapitel.

[1] Der Begriff bezeichnet (vor allem) materielle Teilnahmeanreize wie Geldpreise, (Waren-) Gutscheine oder sonstige Belohnungssysteme.

2.1.3.1 Selbstselektion

Der Begriff der Selbstselektion (Batinic & Bosnjak, 1997) stellt in vielerlei Hinsicht einen wesentlichen Faktor in Zusammenhang mit internetbasierten Datenerhebungsmethoden dar. Die Selbstselektion hat nicht nur Auswirkungen auf die Teilnahmemotivation und Drop-Out Rate, sie bezieht sich auch auf die Generalisierbarkeit der Daten. Es stellt sich in diesem Zusammenhang die Frage: Wer ist interessiert an der Teilnahme an Untersuchungen im Internet und warum? Darüber hinaus scheint es wesentlich zu erörtern, welche Personen (-gruppen) kein diesbezügliches Engagement aufbringen und welche Gründe diese dafür haben. Daher erscheint die Frage wesentlich, welche Entscheidungsprozesse ein durchschnittlicher Internetuser durchläuft, bis er zu einem vollständigen Untersuchungsteilnehmer wird. Diesem Aspekt widmeten sich auch Bosnjak, Bandilla und Tuten (1998) bei ihrer Untersuchung der mehrstufigen Selbstselektion (Abbildung 2.4), bei deren Untersuchung es sich um die Gewinnung von Versuchspersonen mittels eines Werbebanners handelte. Je nach Versuchsbedingung beinhaltete ein Banner die Aufforderung

- *„internationale Studie, wichtig für die Forschung“ bzw.*
- *„internationale Studie, gewinnen Sie Preise“.*

Ermittelt wurde hierbei unter anderem die Anzahl jener Versuchsteilnehmer pro Bedingung, welche vollständig (bis zum Ende der Untersuchung) teilnahmen.

Der Begriff „Klicker“ unter Stufe drei (vgl. Abbildung 2.4), beruht auf dem Rekrutierungsweg des Banners, durch dessen „Anklicken“ auf die Untersuchungsseite verbunden wurde. Bezüglich des Teilnehmerverhaltens in den zwei verschiedenen Versuchsbedingungen zeigt sich, dass unter Bedingung „wichtig für die Forschung“ weniger vollständige Teilnehmer zu verzeichnen waren als bei „gewinnen Sie Preise“. Bei letzterer wurde jedoch auch ein größerer Prozentsatz an „Durchklickern“ (Personen, die keinerlei Angaben machten, sogenannte Lurker, vgl. dazu Kapitel 2.1.3.2) beobachtet.

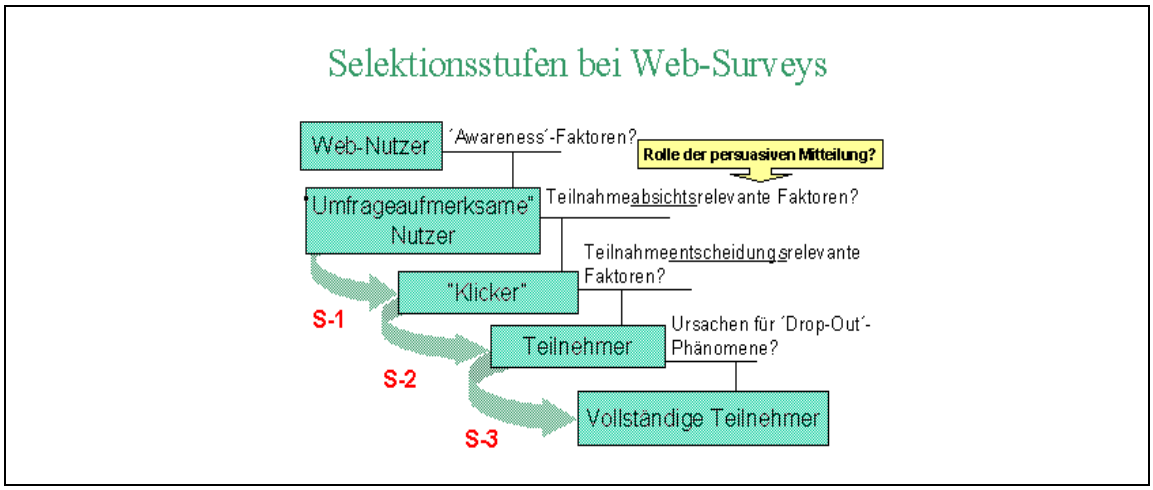

Abbildung 2.4: Selektionsstufen bei WWW–Umfragen (Bosnjak, Bandilla & Tuten, 1998).

Zwar kann man den Aspekt der Selbstselektion nicht ursächlich beheben, dennoch gibt es Möglichkeiten damit umzugehen bzw. den Einfluß dieser Größe auf die Untersuchungsergebnisse zumindest annähernd zu ermitteln: Beispielsweise durch die „multiple site entry technique“ (Reips, 2000). Dabei wird registriert, von welchen Seiten aus das Experiment angewählt wird und anschließend erörtert, ob und wie sich die Antworten bzw. Resultate der auf unterschiedlichem Wege rekrutierten Teilnehmer unterscheiden.

2.1.3.2 Freiwilligkeit der Teilnahme – Volunteer Bias und Drop-Out Effekte

Grundsätzlich stellen die Begriffe des Drop-Outs bzw. Volunteer Bias keine ursprüngliche Neuschöpfung der Online-Forschung dar. Wie bereits in den vorangegangenen Kapiteln erörtert wurde beschäftigten Verzerrungen, die durch die Erwartungshaltungen von Versuchsleiter oder Versuchsteilnehmer auftreten können bereits in den 70ger Jahren namhafte Psychologen wie beispielsweise Rosenthal, der durch seine Beobachtungen und den dabei auftretenden Versuchsleiter-Erwartungseffekt bekannt wurde („Rosenthal-Effekt“, Zimbardo, 1992). An dieser Stelle soll jedoch noch näher eingegangen werden auf den Effekt des vorzeitigen Befragungs- bzw. Teilnahmeabbruchs, den Drop-Out. Die durch das Internet ermöglichte, ungezwungene Versuchsteilnahme wird als großer Vorteil des Online-Forschens verstanden. Wie bereits schon erörtert wurde, bringt diese Tatsache jedoch auch Nachteile mit sich, die mit dem Phänomen der Freiwilligkeit in Zusammenhang stehen. Eben durch das Fehlen eines (vor Ort) physisch vorhandenen Versuchsleiters und die möglicherweise damit verbundenen geringeren Barrieren, die Beteiligung an einer Untersuchung abzubrechen, stellt das vorzeitige Beenden einer Versuchsteilnahme ein großes Problem der Online-Forschung dar. Andererseits können Drop-Outs nicht nur Hinweise auf eine dürftige Teilnahmemotivation geben, sie sind auch in der Lage Aufschluß über mißverständlich bzw. fehlerhaft formulierte Items zu geben bzw. können sie als Indiz für einen zu langen oder optisch wenig ansprechenden Fragebogen gelten. Beispielsweise sind selektive Drop-Outs bei sensiblen Fragen häufig zu beobachten (Knapp & Heidingsfelder, 1999). Wie aber auch in der herkömmlichen (offline) Forschung des öfteren üblich, so sollten auch im Internet Bearbeitungsausfälle hinsichtlich möglicher Konfundierung mit demographischen Daten oder anderen Variablen der experimentellen Situation analysiert werden (Birnbaum & Mellers, 1989).

Welche Bedingungen letztlich vermögen Drop-Out Raten zu minimieren, dazu gibt es zum Teil widersprüchliche Resultate. Eine Frage die mit diesen Aspekten in Zusammenhang steht wäre, ob und in wie weit beispielsweise der Einsatz von Incentives (monetärer oder nicht monetärer Form) Einfluß auf die Verweildauer hat (vgl. Kapitel 5.4.3).

Bevor jedoch an dieser Stelle auf einige Vorschläge zur Reduktion von Drop-Outs eingegangen wird, noch etwas Grundsätzliches: Es ist zwischen **technischen** und **benutzerinduzierten** Teilnahmeabbrüchen zu unterscheiden, wobei im ersten Fall von unbeabsichtig-

ten server- oder clientseitigen Verbindungsabbrüchen zu sprechen ist (Knapp & Heidingsfelder, 1999).

Knapp und Heidingsfelder gingen in ihrer Arbeit nicht nur der Frage nach, ob Drop-Outs unbrauchbare Antworten liefern sondern auch ob diese sich in ihrer Struktur von den restlichen Teilnehmern unterscheiden und somit eine Verzerrung der Ergebnisse befürchtet werden muß. Dazu zogen sie insgesamt neun Studien heran, die in der Zeit von Mai bis Juli 1999 durchgeführt worden waren. Bei diesen Untersuchungen handelte es sich jedoch nicht ausschließlich um Online-Umfragen deren Teilnehmer per Selbstselektion rekrutiert wurden, es befand sich auch eine Webseiten-Evaluation[2] darunter und insgesamt zwei Versuchsbedingungen, bei denen per Pop-Up Fenster rekrutiert wurde.

Die wesentlichsten Resultate dieser Studie sind wie folgt:

- Die ersten beiden Fragen einer Untersuchung sind entscheidend:

  Ca. 50% der Abbrüche erfolgen bereits nach dem ersten Item, während sich weitere Abbrüche über den gesamten Verlauf der Untersuchung verteilen. Gegen Ende des Fragebogens brechen nur noch wenige Teilnehmer ab.

- Längere Befragungen und sensible Themen führen zu höheren Abbruchquoten.
- Offene Textfragen und Mehrfachauswahl-Fragen reduzieren vollständigen Teilnehmer.
- Potentielle Drop-Outs zeigen deutlich weniger Neigung zur Beantwortung offener Fragen.
- Die Art der Rekrutierung hat ebenfalls Einfluß auf die Antwortstruktur von Teilnehmern vs. Drop-Outs.

Je öfter eine Seite bereits vor der Untersuchung besucht wurde, desto höher war die Wahrscheinlichkeit für eine vollständige Teilnahme. Knapp und Heidingsfelder sprechen davon, dass eine solche Community wesentlich stärker bereit sei, Fragen zur entsprechenden Website zu beantworten als gelegentliche Besucher dieser Seite.

Gräf (1999) der sich vor allem mit der optimalen Gestaltung von WWW-Umfragen beschäftigt, nennt in seiner Arbeit die fünf häufigsten Fehler bei Online-Fragebögen (deren Außerachtlassung durchaus Einfluß auf die Drop-Out Rate haben kann). Darüber hinaus besteht die Möglichkeit, dass das Antwortverhalten der Versuchspersonen ebenfalls von diesen Gesichtspunkten betroffen werden könnte. In der vorliegenden Arbeit können nachfolgende Fehlerkategorien durchaus auch zu den oben genannten technisch induzierten Drop-Outs (im weitesten Sinne) gezählt werden, da von dieser technischen Warte aus, so viel

2 Dieser Begriff umfaßt vielfältige Methoden und Techniken der Beurteilung von WWW-Seiten beispielsweise hinsichtlich ihrer Benutzerfreundlichkeit, Übersichtlichkeit etc. (Usability-Kriterien).

wie möglich getan werden sollte, das Risiko des benutzerinduzierten Drop-Outs zu verringern. Anhand welcher Kategorien das möglich sein soll, faßt Gräf wie folgt zusammen:

Handwerkliche Fehler

Durch das Internets wurde es auch jenen Personen leichter möglich, sehr einfach und schnell Umfragen zu konzipieren und eigenständig durchzuführen, die nicht über das entsprechende, notwendige sozialwissenschaftliche Wissen verfügen (z. B. Itemkonstruktion).

Zu langer Fragebogen

Mit der Länge eines Fragebogens nimmt auch die Abbruchquote zu und die Antworten werden unzuverlässiger. Bosnjak und Batinic (1997) kommen aufgrund ihrer Untersuchungen zu folgendem Grundsatz: Fragebögen, die mehr als 15-25 Fragen enthalten, sind zu lang.

Wer demnach zu viele Items bzw. Inhalte vorgibt, kann sich unzuverlässige Daten einhandeln.

Zu langweiliger Fragebogen

Es scheint nicht weiter erstaunlich, dass Untersuchungsteilnehmer, die sich langweilen, ihre Teilnahme vorzeitig beenden bzw. unkonzentriert arbeiten. Somit sollte das Ziel ein intellektuell fordernder Fragebogen sein, der den Eindruck eines interessanten Gespräches vermittelt und dem Teilnehmer darüber hinaus das Gefühl gibt, Neues über sich erfahren zu können und nicht nur statistisches Material zu liefern (Gräf, 1999).

Nicht medienadäquate Formulierung bzw. Präsentation der Fragen

Hierbei ist es wesentlich einige Grundzüge im Umgang mit dem Lesen am Bildschirm bzw. Lesen von Websites zu beachten. Wie auch schon aus dem Englischen übernommen wurde, stellt das „Browsen“ im Internet die wesentliche Form der Wahrnehmung dar. Dieses „Drüberblicken“ oder „Scannen“ wie es Gräf nennt, führt dazu, dass die so wahrgenommenen Satzteile zu einem Sinnzusammenhang ergänzt werden. Somit ist das Vorwissen bzw. die Vorerfahrung der Personen ein wesentliches Kriterium. Darüber hinaus weiß man aus der Usability-Forschung, dass die Lesegeschwindigkeit von Bildschirmtexten im Vergleich zu denen auf Papier um 25 % geringer ist. Somit sollte man die kurze, knappe und möglichst übersichtliche Präsentation der Items immer vor Augen haben.

Auch der Einsatz von Filterfragen3 sollte an dieser Stelle genannt werden.

Ungeeignetes Darstellen der Items anhand von Matrixfragen in Tabellenform

Aufgrund des Wunsches, möglichst viele Fragen in einer Untersuchung realisieren zu können und dabei auch noch Platz zu sparen, kommen sogenannte Matrixfragen immer wieder zum Einsatz. Wie solche Fragekategorien aussehen, ist in Abbildung 2.5 ersichtlich.

Wie würden Sie Ihre Kenntnisse einschätzen ?	1= keine Kenntnisse				5= sehr gute Kenntnisse
WWW (Surfen....)	1				5
e-Mail	1				5
Newsgroups, Mailinglisten	1				5
Telnet	1				5
FTP	1				5
IRC ("Chat")	1				5

Abbildung 2.5: Beispiel einer Matrixfrage

Die Anwendung eines solchen Matrixformates ist nicht unumstritten, da immer wieder Befürchtungen geäußert werden, sich damit „unkalkulierbare Layout-Effekte“ einzuhandeln (Gräf, 1999). Somit würden laut Gräf die Antworten nicht mehr die Sichtweisen der Befragten repräsentieren sondern würden durch das Instrument selbst induziert werden (beispielsweise durch die Trennung des Zusammenhanges von Frage und Antwortdimension). Oftmals steht dieser Meinung auch der Ansatz des „One Screen One Item“ gegenüber, wo pro Frage bzw. Antwortkategorie eine Bildschirmseite eingeblendet wird, wodurch beispielsweise auch exaktere Aussagen über selektive Drop-Outs gemacht werden können. Andererseits kann dem durchaus entgegen gestellt werden, dass somit mögliche wünschenswerte Zusammenhänge zwischen einzelnen Fragen eher verringert werden. Ganz abgesehen davon, dass diese Art der Befragung mitunter auch sehr zeitintensiv ausfallen kann (z. B. durch den Datentransfer zwischen den einzelnen Fragen/Antworten). Selbst wenn dieses Problem gelöst ist (z. B. durch einmaliges, anfängliches Laden des gesamten Fragebogens in den Cache) nimmt man den Teilnehmern noch immer die Möglichkeit, sich einen Überblick über das bevorstehende Fragenmaterial zu verschaffen. Die genannte Methode hat jedoch auch ihre Vorzüge, welche an dieser Stelle auch nicht in Frage gestellt werden. Es ist jedoch nicht von vornherein einer von beiden Methoden der Vorzug zu geben und somit bleibt abzuwägen, welche Form der Befragung aufgrund spezieller Fragestellungen geeigneter erscheint. In Tabelle 2.1 seien an dieser Stelle nochmals die Vorteile dieser beiden obigen Darstellungsmethoden angeführt (Theobald, 2000).

[3] Filterfragen dienen der Benutzerführung. Es werden so Fragen, die den Teilnehmer nicht betreffen (wie aus den vorangegangenen Items geschlossen wird) übersprungen und nur jene Variablen erhoben, die relevant sind (adaptive Erhebung).

Tabelle 2.1: Unterschiede zwischen den Methoden „One Screen One Item“ und der Darstellung der Fragen auf einer Seite.

Fragebogen auf einer Seite	**Eine Frage pro Seite**
Fragebogen im Gesamtüberblick; höhere Transparenz; überschaubarer Fragenkontext.	Kein Scrollen notwendig; geringerer Eingabeaufwand.
Zeitbedarf für den Probanden anhand der Länge einschätzbar; kein Gefühl der „Endlosbefragung“.	Kürzere Ladezeit für die erste Frage.
Datenübertragung nur zu Beginn und am Ende der Befragung; dadurch insgesamt meist geringere Wartezeiten.	Eingabeabhängige Verzweigungen und Datenübergabe innerhalb des Fragebogens handhabbar. Isolierte Darstellung jeder einzelnen Frage.
Weniger anspruchsvolle Programmierung.	Nachträgliche Änderungen durch Rücksprünge kontrollierbar.
Ähnlichkeit mit konventionellen Papier-Fragebogen.	Abbrecherverhalten erkennbar.

Unbestritten ist, dass bei der Anwendung von Matrixfragen der Zusammenhang zwischen Frage und Antwortkategorie nicht getrennt werden darf, bzw. die Übersicht gewahrt werden sollte. Beispielsweise ist die farbliche Abgrenzung der einzelnen Fragen ein gangbarer Weg sowie auch der Einsatz von Farben zur Darstellung der Antwortdimensionen. Aus Abbildung 2.6 geht hervor, wie Matrixfragen mit Hilfe farblicher Gestaltung und Abgrenzung umgangen werden können.

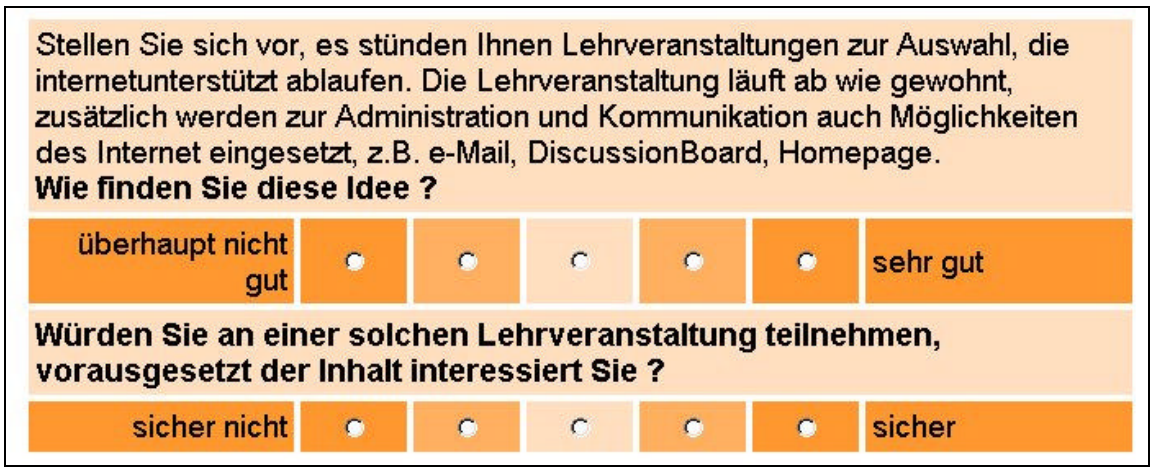

Abbildung 2.6: Illustration der Vermeidung von Matrixfragen.

Sollte man sich für eine Matrixdarstellung entscheiden, ist unbedingt zu beachten, dass keine Voreinstellung der Radio-Buttons[4] vorgenommen wurde.

Aus obigen Ausführungen resultieren schließlich folgende Leitlinien zur Gestaltung qualitativ hochwertiger Online-Fragebögen (Gräf, 1999):

- Technische Erfordernisse bedenken.
- Nach den Regeln der empirischen Sozialforschung formulieren.
- Glaubwürdige Kommunikation mit den Respondenten herstellen.
- Aufmerksamkeit erzeugen und wachhalten.
- Anspruchsvolles Design verwenden und Usability-Kriterien umsetzten.
- Zentrale Textstellen hervorheben.
- Matrixfragen vermeiden.
- Fragebögen kurz halten.
- Filterfragen einsetzen.
- Pretests durchführen.

Wie Gräf setzt sich auch Reips (1999) mit der Manipulation der Verweildauer bei Online-Untersuchungen auseinander. Sein Fokus liegt jedoch auf dem Experimentieren im Web. Seine Vorschläge bezüglich der Reduktion von Drop-Outs, im Rahmen von WWW-Experimenten, seien an dieser Stelle erörtert:

Aufwärm-Technik

Dem eigentlichen Experiment ist eine Aufwärmphase vorgeschalten, die etwaige Teilnehmer, die nur kurz einen Blick auf die Untersuchung werfen wollen, von ernsthaften Teilnehmern trennen soll. Die experimentelle Manipulation erfolgt erst nach dieser Phase, so dass sicher gestellt werden kann, dass Ausfälle zu diesem Zeitpunkt nicht in der Manipulation begründet sind.

Hürden einbauen

Dieses Vorgehen ist ähnlich der oben genannten Methode, zielt jedoch mehr auf die „inneren Hürden“ einer Versuchsperson ab. Reips führt an, dass der VP die Verbindlichkeit ihres Teilnahmeentschlusses vor Augen geführt werden soll, indem am Anfang des Experimentes relativ fordernde Ansprüche an diese gestellt werden. Beispielsweise wird verlangt e-Mail Adresse oder Telefonnummer anzugeben, was natürlich in den meisten Fällen nicht

[4] Diese dienen der Auswahl von Antwortkategorien in HTML-Fragebögen (vgl. Kapitel 4, Abbildung 4.1).

auf den Wahrheitsgehalt zu überprüfen sein wird. Doch selbst fälschlich gemachte Angaben kosten Zeit und werden somit als Hürde wahrgenommen.

Es gibt unterschiedliche Auffassungen bezüglich des Einsatzes von zwingend auszufüllenden Antwortkategorien (wie beispielsweise demographische Daten etc.). Reips (1999) verweist darauf, dass durch den Einbau von eben solchen, oben bereits dargestellten „Hürden“ die Abbruchrate verringert werden könnte, andere Autoren wiederum meinen, man würde damit potentielle Teilnehmer abschrecken bzw. auch interessierte Besucher verärgern und somit unter Umständen auch absichtlich fingiertes Datenmaterial provozieren.

Bandilla und Bosnjak (1999) untersuchten das Teilnahmeverhalten bei nicht-restringierten WWW-Umfragen, d. h. die Befragten konnten über die Reihenfolge und Vollständigkeit ihrer Angaben frei entscheiden, wobei angemerkt werden muß, dass es sich bei der genannten Untersuchung um eine Befragung mittels der Methode „Eine Frage – ein Bildschirm“ (One Screen One Item) handelte. Bandilla und Bosnjak fanden in ihrer Arbeit schließlich sechs Bearbeitungstypen vor:

- *Complete Responders*

  Diese Personengruppe beantwortete alle gestellten Fragen.

- *Unit-Nonresponders*

  Teilnehmer, die sich nach der Teilnahmeaufforderung sofort gegen eine Beteiligung entschieden.

- *Antwortende Drop-Outs*

  Personen, die zwar an der Befragung teilgenommen haben, diese jedoch nicht beendeten.

- *Lurker*

  Dies sind Personen, die zwar das gesamte Fragenmaterial sondieren aber nicht teilnehmen.

- *Lurkende Drop-Outs*

  Diese Personen, stellen eine Mischform aus Gruppe 3 und 4 dar, sehen sich einen Teil der Fragen an und steigen jedoch noch vor Beendigung aus.

- *Item-Nonresponders*

  Darunter sind Teilnehmer zu verstehen, die bis zum Ende der Befragung mitmachen, jedoch einzelne Items nicht beantworten.

Basierend auf der obigen Typisierung kommen Bandilla und Bosnjak (1999, S. 2) zu folgende Aussagen bezüglich der Interpretation des Drop-Outs bzw. Teilnahmeverhaltens:

Während bei den Befragungstypen „antwortende Drop-Outs (3) und „lurkende Drop-Outs (5) technische Artefakte, wie z. B. Time-outs aufgrund einer Überlastung des Netzverbindung nicht auszuschließen sind, kann bei den übrigen Beantwortungstypen von willentlich gesteuerten Entscheidungen ausgegangen werden. Auf der Basis dieser willentlichen Entscheidungen lassen sich im Vergleich zu „klassischen“ Befragungsformen Schlußfolgerungen über die Qualität des gesamten Befragungsinstruments ableiten.

Abschließend, soll darauf hingewiesen werden, dass die Teilnahme an Interviews oder Befragungen zwar als Regelfall und die Nichtteilnahme als Ausnahme angesehen wird, wie aus obigen Erläuterungen jedoch hervorgegangen sein sollte, ist eine differenzierte Betrachtungsweise von Drop-Out Raten und Selbstselektion ein wesentliches Kriterium der Online-Forschung.

2.2 Gütekriterien

Unter dem Begriff der Gütekriterien werden folgende Kriterien empirischer Forschungsinstrumente verstanden: Objektivität, Reliabilität und Validität.

2.2.1 Objektivität

Bortz und Döring (1995) beschreiben den Aspekt der Objektivität wie folgt:

„*Ein Test oder Fragebogen ist objektiv, wenn verschiedene Testanwender bei denselben Personen zu den gleichen Resultaten gelangen, d. h. ein objektiver Test ist vom konkreten Testanwender unabhängig*“.

Es wird dabei zwischen drei Formen dieser Anwenderunabhängigkeit differenziert: Der Durchführungs-, der Auswertungs- und der Interpretationsobjektivität. Die Daten die via Internet erhoben werden, unterscheiden sich von den herkömmlich gewonnenen primär durch die Durchführungs- und Auswertungsobjektivität. Die Auswertungsobjektivität wird insofern verbessert, als die automatische Datenübergabe in ein Statistikprogramm Tippfehler und ähnliches ausschließt. Die Durchführungsobjektivität bezieht sich auf den Zusammenhang mit den Interviewereffekten bzw. Versuchsleitereffekten (vgl. auch Kapitel 2.1.2). Durch den Wegfall der Versuchsleiter bzw. Interviewer lassen sich auch keine diesbezüglichen Effekte mehr nachweisen. Laut Lander (1998) können jedoch Auftraggebereffekte (z. B. Institution hinter der Untersuchung) auftreten. Dieses Kriterium ist bereits von den postalisch verschickten Fragebögen her bekannt (Bortz & Döring, 1995, S. 235). Lander meint weiters, dass wenn bei herkömmlichen Untersuchungen ein Antwortverhalten zur Festigung der vermuteten Hypothese angenommen wird (sozial erwünschte Antworttendenzen), so wäre im Internet auch das Gegenteil denkbar. Durch das relativ große Mißtrauen bezüglich des Datenschutzes im Internet, scheinen falsche Angaben über die

persönliche Identität naheliegend und können auch nicht überprüft werden. Somit müssen die via Internet gewonnenen Erhebungsergebnisse immer vor dem Hintergrund der genannten Institution, Zielsetzung und Verwendung der Daten gesehen werden.

2.2.2 Reliabilität

Nach der Definition von Bortz und Döring (1995) gibt die Reliabilität (Zuverlässigkeit) den Grad der Meßgenauigkeit (Präzision) eines Untersuchungsinstrumentes an. Oder kurz gesagt: Die Reliabilität eines Tests kennzeichnet den Grad der Genauigkeit, mit dem das geprüfte Merkmal gemessen wird. Lander (1998) trifft in diesem Zusammenhang folgende Aussage: „*Ist schon die Reliabilität traditioneller Erhebungsinstrumente durch die geringe Stabilität vieler sozialer Zusammenhänge problematisch, so trifft dies in noch stärkerem Maße auf das einem ständigen Wandel unterworfene Internet zu*“. Denn Reliabilität setzt (nach obiger Definition) konstante Merkmale voraus, die auch noch nach einer gewissen Zeitspanne erfaßbar sind.

2.2.3 Validität

Unter Validität wird die Gültigkeit eines Erhebungsinstruments verstanden. Somit wird dieses Konstrukt als das wesentlichste der drei Testgütekriterien verstanden, denn die Validität gibt den Grad der Genauigkeit an, mit dem ein Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen oder vorhersagen soll, auch tatsächlich mißt oder vorhersagt (Lienert, 1969). Validität ist somit eine relative Kategorie, sie läßt sich nur über einen Vergleich mit einem zuvor festgelegten Bezugskriterium überprüfen. Lander meint, dass traditionelle Untersuchungsmethoden häufig eine künstliche Befragungssituation darstellen, die mit eher unangenehmen Situationen wie z. B. einer Zeugenvernehmung assoziiert werden. Somit kann laut Lander eine Befragung als eine Art Gespräch gesehen werden, indem ein institutionell vorgegebener Zweck der Informationssuche verfolgt wird. Aus diesem Zweck kann sich eine Verpflichtung des Befragten auf die Wahrheit ergeben bzw. eine Neigung zur Reaktivität. Bei einer Erhebung via Internet, welches ja eine eher geringe Möglichkeit der Sanktionsfähigkeit aufweist, sinkt tendenziell diese Gefahr. Das bedeutet, die Neigung von der wahren Antwort abzuweichen sinkt. Lander führt weiter aus, dass gleichzeitig mit der Verminderung der Reaktivität, die Wahrscheinlichkeit einer anderen Art der Verzerrung, „die Inadäquanz der Untersuchungssituation gegenüber der eigentlich interessierenden Situation außerhalb des Internets“ zunehmen könne. Was hier angesprochen wird, ist die interne bzw. externe Validität.

Von manchen Autoren wird dieses Kriterium bezogen auf Daten aus dem Internet überbewertet. Oft wird vom Internet dann als „virtuelle Welt“ gesprochen, in der sich eine Versuchsperson beinahe konträr zu ihrem Verhalten im realen Leben benimmt. In vorliegender Arbeit wird diese sehr polarisierende Unterscheidung nicht getroffen. Die Differenzie-

rung zwischen externer und interner Validität hat aber auf jeden Fall Einfluß auf die Generalisierbarkeit der Daten (vgl. Kapitel 4.3.1).

2.3 Generalisierbarkeit

Manche Autoren sehen die für die Generalisierbarkeit von Daten notwendige „Stabilität von Identitäten, Einstellungen, Verhaltensdispositionen und Situationseinschätzungen“ im Internet nicht so gewährleistet, als dies bei herkömmlichen Erhebungsmethoden der Fall wäre (Lander, 1998). Folglich wären deren Ergebnisse nicht generalisierbar. Reips (1997) relativiert jedoch die Bedenken bezüglich der Generalisierbarkeit von herkömmlich gewonnenen Untersuchungsergebnissen bzw. via Internet gewonnenen, indem er darauf verweist, dass ca. 80 % aller herkömmlichen Studien auf einer Stichprobe von Studierenden basieren, deren Anteil in der Bevölkerung aber nur ca. 3 % beträgt. Dennoch übt nahe zu niemand Kritik an dieser überaus weitverbreiteten Vorgangsweise. Weiters soll an dieser Stelle nochmals an die externe Validität von Untersuchungsergebnissen erinnert werden. Denn je strikter man die Variablen einer Untersuchung kontrolliert um so artifizieller werden oftmals die Untersuchungsbedingungen. Gleichzeitig muß man damit wenig generalisierbare Ergebnisse in Kauf nehmen oder wie Martin (1991, zitiert nach Reips, 1997) es treffend beschreibt:

„As a rule of thumb, the more highly controlled the experiment, the less generelly applicable the results [...] if you want to generalize the results of your experiment, do not control all the variables.“

2.4 Repräsentativität

Für die Gewinnung repräsentativer Untersuchungsergebnissen ist die Ziehung einer entsprechenden Zufallsstichprobe aus einer vorher definierten Grundgesamtheit nötig. Das ist im Internet jedoch nur sehr selten möglich, da man über die Grundgesamtheit nichts aussagen kann. Zieht man jedoch eine Zufallsstichprobe aus einer begrenzten Grundgesamtheit, ist die Verallgemeinerung der Ergebnisse problematisch. Es sollte hier aber der eher pragmatische Standpunkt gelten, dass innerhalb der (psychologischen) Forschung nicht in allen Fällen den oben genannten Kriterien im vollem Ausmaß genüge getan werden kann, wenn eine sinnvolle Forschung gewährleistet werden soll. Somit sollte man sich auch im Rahmen der Online-Forschung an der Realität orientieren. Forschung sollte schließlich nicht ausschließlich Selbstzweck sein sondern auch anwendbare bzw. praxisrelevante Ergebnisse liefern.

Zusammenfassend läßt sich bezüglich dieser Punkte sagen: Das Internet als Erhebungsmethode verleitet dazu, Erhebungen durchzuführen und über das Medium eine große Anzahl von Versuchspersonen zu rekrutieren. Eine große Stichprobe ist jedoch noch kein

geeigneter Indikator für die Verallgemeinerbarkeit der Daten. Das schon erwähnte Problem der Selbstselektion kann zusätzlich zu systematischen Verzerrungen in der Zusammensetzung des Versuchspersonenpools führen. Somit lassen sich repräsentative Stichproben im Internet, über einen sehr engen Rahmen hinaus nicht erstellen, da eben über die Grundgesamtheit keine Aussagen getroffen werden können.

2.5 Diskussion der Vor- und Nachteile online erhobener Daten

Überblicksartig sollen an dieser Stelle all jene Vor- und Nachteile internetbasierter Datenerhebung angeführt werden, die in vorliegender Arbeit erwähnt werden bzw. generell zu erwähnen sind. Beginnend mit der Auflistung all jener Schlagwörter, die sich im Laufe der Zeit im Rahmen der Online-Forschung diesbezüglich ergeben haben werden schließlich diese Punkte diskutiert, um aufzuzeigen, dass beinahe jeder dieser Aspekte sowohl Vor- als auch Nachteile in sich birgt. Eine explizite Einteilung in Pro und Kontras wird also nicht vorgenommen, da das der Komplexität dieser Sachverhalte nicht gerecht werden würde.

So schreibt Batinic (1997) *„Internetbasierte Fragebogenuntersuchungen weisen Merkmale von „Papier-und-Bleistift“-Untersuchungen auf, erweitern diese aber auch um medienspezifische Elemente. Hieraus resultiert ein „Mehrwert“ internetbasierter Befragungssysteme, der sich aus dem informations-technischen sowie medialen Charakteristika der Internet-Dienste ableiten läßt“.*

Schlagworte internetbasierter Forschung (Batinic & Bosnjak, 1997):

Asynchronität

Darunter ist die zeitversetzte Kommunikation im Internet gemeint, wie sie beispielsweise bei e-Mail Korrespondenzen zu finden ist. Somit hat sowohl der Untersuchungsteilnehmer, als auch der Leiter der Untersuchung Zeit, um auf Reaktionen, Fragen etc. zu reagieren.

Alokalität

Der Umstand, dass sich Befragungsteilnehmer nicht mehr in ein Labor oder ähnliches begeben müssen, um an einer Untersuchung oder einem Experiment teilzunehmen, wird als großer Vorteil des Online-Forschens gesehen (nähere Angaben im Text).

Automatisierbarkeit der Durchführung und Auswertung

Dieser Aspekt kennzeichnet die technischen Möglichkeiten des Internets. So entfällt beispielsweise die Notwendigkeit, die erhobenen Daten manuell in ein Datenverarbeitungsprogramm einzuspeisen. Weiters ist die Möglichkeit der automatisierbaren Durchführung

ist in Hinblick auf eventuelle Versuchsleiter-Effekte bedeutsam und steht somit in engem Zusammenhang mit dem Aspekt der Objektivität der Durchführung und Auswertung.

Dokumentierbarkeit als medieninhärentes Merkmal

Flexibilität

Da die Versuchspersonen bzw. der Versuchsleiter nicht mehr an Untersuchungsorte oder -Uhrzeiten gebunden sind, können beispielsweise auch Berufstätige leichter teilnehmen.

Objektivität der Durchführung und Auswertung

Ökonomie

Freiwilligkeit der Teilnahme während der gesamten Untersuchung

Als Auswirkungen dieses Aspekts wird eine verringerte Anzahl von „sozial erwünschten“ Antworten vermutet, da aufgrund des fehlenden (physisch vorhandenen) Versuchsleiters kein (offensichtlicher) Druck auf die Teilnehmer ausgeübt wird und diese meist anonym bleiben. Andererseits stellt dieses Faktum jedoch auch ein kritisch betrachtetes dar, denn mit dieser Autonomie für die Befragten sinkt auch die Möglichkeit der Versuchsleiter, lenkend einzugreifen (Einstellen von weiteren Tätigkeiten während der Untersuchung, die möglicherweise Einfluß auf die Resultate haben könnten etc.).

Möglichkeiten eines sofortigen Feedbacks

Hier wird die Teilnahmemotivation angesprochen, denn für viele potentiellen Befragungsteilnehmer stellen in Aussicht gestellte Informationen, über das eigene „Abschneiden“ in einer Untersuchung, einen wesentlichen Teilnahmeanreiz dar.

Möglichkeiten der Filterführung bzw. Hinweis auf unvollständige Antworten

Diese Aspekte betreffen sowohl die Teilnehmer selbst, als auch die Untersuchung durchführenden Personen. Mittels des Einsatzes von Filterfragen kann der Fragebogen, ähnlich dem adaptiven Testen, auf den einzelnen Teilnehmer abgestimmt werden. Dem Befragten wird somit das antworten auf Items erspart, die nicht auf ihn zutreffen bzw. redundantes Fragenmaterial kann verhindert werden. Die Möglichkeit für den Versuchsleiter, auf fehlende aber nötige Angaben hinzuweisen ist jedoch ein kontrovers diskutiertes (vgl. Ausführungen im Text).

Ein Teil dieser Vorteile bezieht sich auf den **organisatorischen Nutzen** bzw. generell auf die **Erleichterungen der Durchführung für die Untersuchungsleiter**: So entfallen beispielsweise umständliche Einteilungen von Untersuchungsräumen was Reips (2000) anschaulich wie folgt beschreibt:

„No scheduling difficulties arise, and overlapping sessions don't produce organizational nightmares as well".

Unabhängig davon sind post-hoc Vergleiche in Abhängigkeit von der Tages- bzw. Nachtzeit der Teilnahme möglich sowie praktisch uneingeschränkte simultane Nutzung des Versuchsmaterials. Der als Vorteil eingestufte Aspekt der Alokalität birgt jedoch für die Untersuchungsteilnehmer auch Nachteile in sich. Da es mittels des Internets möglich wurde, das Versuchslabor zu den Teilnehmern zu bringen statt umgekehrt, wurde es auch möglich einen Teil der nicht unbeträchtlichen Kosten auf die Teilnehmer abzuwälzen. Denn beispielsweise müssen Probanden, deren Kosten für den Internetzugang pro Zeiteinheit abgerechnet werden, ihre (wenngleich auch freiwillige) Teilnahme selbst bezahlen. Ein anderer Aspekt der häufig als Pluspunkt internetbasierter Forschung genannt wird ist die Kostengünstigkeit. Wie bereits weiter oben erwähnt verringert sich der finanzielle Aufwand beispielsweise für Raummieten oder Helfer der Versuchsleitung etc. Nicht unerwähnt bleiben darf jedoch auch der Umstand, dass man mittels dieser Alokalität zumeist auf die Anwesenheit eines Versuchsleiters verzichtet und somit über weniger Kontrollmöglichkeiten verfügt. Gleichzeitig stellt dessen Absenz jedoch auch eine Möglichkeit dar, unbeeinflußtes Antwortmaterial zu bekommen (ohne Versuchsleitereffekte etc.). Ein weiterer Punkt, der die Ökonomie betrifft ist der Umstand, dass anspruchsvolles experimentelles Untersuchungsmaterial entsprechende computertechnische (Progammier-) Kenntnisse voraussetzt. Das heißt unter Umständen, dass entsprechende Leute dafür gefunden werden müssen und die zumeist auch bezahlt werden wollen.

Selbstverständlich gibt es aber auch unbestrittene Vorteile wie beispielsweise all jene, die ebenfalls dem Begriff der Ökonomie zugeschrieben werden:

Durch den Einsatz des Internets ist es möglich in kurzer Zeit bereits über Ergebnisse zu verfügen, da erstens relativ schnell Versuchspersonen rekrutiert werden können, ein schneller Rücklauf zu erwarten ist bzw. auch die manuelle Eingabe der Antworten entfällt. Zusätzlich sollte natürlich auch die Möglichkeit des adaptiven Testens genutzt werden, was hinsichtlich der Freiwilligkeit der Teilnahme und bezüglich der Motivation ausschlaggebend sein könnte.

Weiters können via Internet auch ansonsten schwer zugängliche Populationen oder Randgruppen befragt werden, wie es beispielsweise schon 1997 von Coomber, anhand von Drogendealern verschiedener Länder, anschaulich illustriert wurde. Diese explorative Studie kündigte er in insgesamt 23 drogenbezogenen Newsgroups (alt.drugs.hard, rec.drugs.misc etc). mittels dem Betreff (subject) „Have you ever sold powdered drugs? If

so, I would like your help“ an. Coomber nutzte die Ergebnisse dieser Befragung zur Ergänzung der bereits vorher durchgeführten persönlichen Gespräche. Diese seine Arbeit gilt heute als frühe Pionierarbeit für den Interneteinsatz zur Explorierung inhaltlich diffiziler Hypothesen.

Einen weiteren Vorteil internetbasierter Forschung stellt der in obiger Auflistung zuletzt genannten Punkt dar: Möglichkeiten der Filterführung bzw. des Hinweisens auf unvollständige Fragebögen bzw. Antworten. Mittels CGI-Skripts, deren Funktionsweise an späterer Stelle erklärt wird, ist die Überprüfung von Antworten möglich. Beispielsweise kann man mittels dieser kleinen Programme Fragebögen nur dann annehmen, sobald diese vollständig ausgefüllt wurden. Das heißt, der Untersuchungsteilnehmer wird auf seine fehlenden Angaben aufmerksam gemacht und gebeten diese zu ergänzen. Diese Variante, Angaben einzufordern ist jedoch nicht unumstritten und sollte differenziert betrachtet werden. Mit Sicherheit sind demographische Angaben der teilnehmenden Personen wesentlich um weitere Antworten überhaupt sinnvoll auswerten zu können. Dennoch sollte man sich bewußt sein, dass wiederholtes Hinweisen auf diesbezüglich fehlende Nennungen als durchaus störend und lähmend erlebt werden können und in nicht unerheblichen Drop-Out Raten resultieren könnten. So ist also genau abzuwägen, welche Daten als unerläßlich angesehen werden bzw. sollte in jedem Fall dann eine Option geboten werden, diese entsprechenden Angaben bewußt frei zu lassen.

Einen weiteren sehr wesentlicher Aspekt stellt die **technische Varianz** dar. Darunter sind Einflüsse der Hard- bzw. Software auf die Untersuchungsergebnisse zu verstehen. Beispielsweise kann die Verwendung unterschiedlich großer Bildschirme, Bildschirmauflösungen, Browser oder Netzverbindungen Einfluß auf die Drop-Out Rate bzw. Teilnahmemotivation haben. So kann beispielsweise eine geringe Datenübertragungsgeschwindigkeit bei einem datentransverintensivem Experiment die Motivation zur vollständigen Teilnahme beeinträchtigen. Somit erscheint es unumgänglich und sinnvoll, eine möglichst umfangreiche Abtestung des Untersuchungsmaterials innerhalb verschiedenster Systeme durchzuführen, um diese Fehlerquellen möglich gut kontrollieren zu können. Beispielsweise können auch hier CGI – Skripts unterstützend eingesetzt werden. Diese können etwa durch die Abfrage sogenannter Umgebungsvariablen die Bildschirmauflösung, Browserversion etc. des Versuchsteilnehmers registrieren und ihn (ohne Zutun oder Wissen der VP) auf die entsprechende Untersuchungsversion weiterleiten.

Diese Problematik der technischen Varianz wird auch des öfteren als „**Measurement-Error**“ (vgl. Fuchs, 1998) bezeichnet.