

SEMINARARBEIT

Micro-Feedback

von

Jan Uwe Nolinski

eingereicht am 21. Juli 2008 beim
Institut für Angewandte Informatik
und Formale Beschreibungsverfahren (AIFB)
der Universität Karlsruhe (TH)

im Rahmen des Seminars „Social Software Engineering“ im SS 2008



Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825



Referent: Prof. Dr. Wolfried Stucky
Betreuer: Helmuth Elsner

Studienanschrift:
Görresstr.54
69126 Heidelberg
j-u-n@gmx.de

Micro-Feedback

Jan Uwe Nolinski

j-u-n@gmx.de

Abstract. Ein adäquates Mittel zum Messen der Effekte, die durch den Einsatz von Social Software entstehen, ist ein Online-Fragebogen. Im Unternehmenseinsatz wirkt sich der Zielkonflikt einer jeden Umfrage, dass sie kurz aber trotzdem aussagefähig ist, noch stärker aus. Daher wird ein Verfahren vorgeschlagen, welches die Umfrage in kleinere Teile zerlegt und somit beiden Zielen gerecht wird. Dieses Verfahren wird Micro-Feedback genannt. Zur Entwicklung des Verfahrens wird auf die Aufteilungsverfahren Multiple Matrix Sampling und Split Questionnaire eingegangen und deren Vorgehensweisen auf die hier relevante Aufgabe übertragen.

Inhaltsverzeichnis

1	Einleitung.....	3
2	Non Response.....	3
2.1	Charakteristika von Web-Befragungen.....	4
2.2	Potentielle Fehlerquellen.....	5
2.3	Formen des (Non)Response bei Web-Befragungen.....	5
2.4	Erklärungsansätze für Non-Response.....	6
3	Verwandte Arbeiten.....	7
3.1	Multiple Matrix Sampling: A Review by Jeffrey M. Gonzalez and John L. Eltinge.....	8
3.2	A Split Questionnaire Survey Design by Trivellore E. Raghunathan and James E. Grizzle.....	10
3.3	Split Questionnaire Design by Feray Adigüzel.....	12
3.3.1	Designs.....	14
3.3.2	Simulation Studies.....	17
3.3.3	Empirical Data Application.....	17
3.3.4	Field Study.....	18
4	Konzept.....	19
4.1	Vorstellung der Parameter.....	19
4.1.1	Development.....	19
4.1.2	Data Collection.....	20
4.1.3	Processing and Analysis.....	21
4.2	Aufstellung von vier Konzepten.....	22
4.2.1	Gemeinsame Parameter.....	22
4.2.2	Die vier Konzepte.....	23
4.2.2.1	Konzept 1: Random:.....	23
4.2.2.2	Konzept 2: Matrix Sampling.....	24
4.2.2.3	Konzept 3: Split Questionnaire.....	24
4.2.2.4	Konzept 4: 2-Phasen.....	24
4.3	Mein Vorschlag: Konzept 2.....	24
5	Zusammenfassung.....	25
	Literaturverzeichnis.....	26

1 Einleitung

Um die Effekte zu messen, die durch den Einsatz von Social Software in Firmen entstehen, ist der Onlinefragebogen ein akkurates Mittel. Das Problem einer jeden Umfrage ist, nur qualitativ hochwertige Aussagen getroffen werden können, wenn der Fragebogen viele Aspekte beleuchtet, was dazu führt, dass er sehr umfangreich wird. Ein langer Fragebogen bedeutet für die Teilnehmer eine hohe Bearbeitungsdauer, diese führt dazu, dass viele Teilnehmer erst gar nicht an der Befragung teilnehmen, die Befragung frühzeitig abbrechen oder einzelne Fragen unbeantwortet lassen.

Noch problematischer ist es, wenn die Teilnehmer einfach irgendetwas antworten, da dadurch die Datenqualität abnimmt. Gerade im Unternehmenskontext ist Zeit eine kritische Größe, daher ist der Ansatz folgender Arbeit, die Umfrage in Teilfragebögen aufzuteilen. Der einzelne Teilnehmer muss nicht mehr alle Fragen beantworten und dadurch weniger Zeit investieren. Diese Umfrageform nennen wir Micro-Feedback.

Die Arbeit gliedert sich in drei Teile, im ersten Teil wird das Thema der Non-Response näher untersucht. Daher wird zunächst auf die Charakteristika von Web-Befragungen, danach auf potentielle Fehler der Befragung und zum Schluss auf einige Erklärungsansätze eingegangen.

Im zweiten Kapitel werden verwandte Arbeiten vorgestellt, dabei handelt es sich nicht um Arbeiten, die sich mit der Aufteilung von Onlinebefragungen in Firmen beschäftigen, sondern um Verfahren zum Aufteilen von langen schriftlichen Fragebögen. Vorgestellt wird das Multiple Matrix Sampling und das Split Questionnaire Design. Das Multiple Matrix Sampling arbeitet mit einer mehr oder weniger zufälligen Auswahl der Fragen für die Teilfragebögen. Im Gegensatz dazu versucht das Split Questionnaire Design, die Fragen optimal auf die Teilfragebögen zu verteilen, was durch die Durchführung einer Pilotstudie und die Ermittlung der Korrelationen zwischen den einzelnen Fragen erfolgt.

Im dritten Teil wird zuerst auf die Parameter eingegangen, welche für die Durchführung einer geteilten Umfrage festgelegt werden müssen. Danach werden vier mögliche Konzepte aufgezeigt, die in Anlehnung an die verwandten Arbeiten entwickelt wurden. Zum Schluss werde ich erläutern, warum ich Konzept 2 Matrix Sampling für die Befragung einsetzen würde.

2 Non Response

Dieser Abschnitt beschäftigt sich eingehend mit dem Problem der Non-Response, welches eines der zentralen Probleme bei Befragungen darstellt. Als Grundlage dient das Buch „(Non)Response bei Web-Befragungen“ von Michael Bosnjak. Dieses handelt nicht von Befragungen in Firmen, sondern von Web-Befragungen. Trotzdem sind einige Erkenntnisse und Erfahrungen auf unseren Fall übertragbar. Im Folgenden wird zuerst auf die Charakteristika von Web-Befragungen eingegangen, der zweite Teil beschäftigt sich mit den potentiellen Fehlerquellen und im dritten Abschnitt wird auf die Formen und Erklärungsansätze für Non-Response eingegangen.

2.1 Charakteristika von Web-Befragungen

Die Sozialwissenschaften unterscheiden zwischen 4 Befragungsverfahren. Diese lassen sich einmal einteilen in Befragungen, die von einem Interviewer oder vom Befragten alleine durchgeführt werden und in solche, die computerunterstützt oder nicht computerunterstützt sind.

	Nicht computerunterstützt	Computerunterstützt
Interviewer-administriert	Persönlich-mündliche Befragung auf Papier-und-Bleistift-Basis (Paper-and-Pencil-Interview, PAPI; z.B. ALLBUS 1980 bis 1998)	<ul style="list-style-type: none"> - Telefonisches computerunterstütztes Interview (Computer Assisted Telephone Interview; CATI) - Computerunterstütztes persönlich-mündliches Interview (Computer Assisted Personal Interview; CAPI; z.B. ALLBUS ab 2000)
Selbst-administriert	(Postalisch-)Schriftliche Befragung (Mailsurveys)	Computerunterstütztes selbstadministriertes Interview (Computer Assisted Self-Administered Interview; CASI) Beispiele: <ul style="list-style-type: none"> - Disk-by-Mail - E-Mail-Befragungen - Web-Befragungen

Fig. 1. Formen Befragungsverfahren (Bosnjak, 2002)

Im Folgenden werden wir uns mit selbstadministriert computerunterstützten Verfahren beschäftigen. Diese Verfahren weisen im wesentlichen 6 Merkmale auf:

- **Asynchronität:** Der Befragte kann den Zeitpunkt der Bearbeitung selbst auswählen. Dies ist eine Eigenschaft der selbstadministrierten Befragung.
- **Alokalität:** Die Web-Befragung erlaubt es, dass sich Probanden und Untersuchungsleitung an unterschiedlichen Orten befinden, also unabhängig voneinander sind.
- **Automatisierbarkeit der Durchführung und Auswertung:** Durch die Durchführung am Rechner lassen sich Auswertungen und Durchführung automatisieren. Falls die Auswertungsverfahren schon feststehen, können immer aktuelle Reports zur Verfügung gestellt werden. Die Befragung kann an automatisch einzelne Personen anpasst werden.
- **Dokumentierbarkeit von Inhalts- und Meta-Daten:** Bei einer Web-Befragung werden nicht nur die Antwortdaten protokolliert, sondern auch zusätzlich Metadaten erhoben (z.B. Befragungsdauer, Unterbrechungen, Navigationspfade).
- **Flexibilität bei der Operationalisierung von Stimuli und Antwortoptionen:** Da die Befragung am Rechner stattfindet, lassen sich andere Stimuli wie Grafiken und Videos einbinden.
- **Objektivität der Durchführung und Auswertung:** Die Durchführung kann als objektiv bezeichnet werden, da keine direkte Interaktion mit dem Untersuchungsleiter stattfindet. Da die Daten direkt eingetragen und ausgewertet werden können, kommt es nicht zu Eingabe- oder Scanfehlern.

(vgl Busnjak 2002, 7-10)

2.2 Potentielle Fehlerquellen

Um eine aussagekräftige Umfrage zu erhalten, müssen die Fehlerquellen Coverage Error, Sampling error, Measurement Error und Non-Response Error minimiert werden.

Zum *Coverage Error* kommt es, wenn die Auswahlgesamtheit nicht mit der Grundgesamtheit übereinstimmt. Zum Beispiel spricht man von „Undercoverage“, wenn ein Gruppe in der untersuchten Menge gar nicht oder unterrepräsentiert ist.

Der *Sampling Error* bezieht sich auf die Schätzungenauigkeiten der Parameter, die entstehen, weil man nur eine Stichprobe und nicht die Grundgesamtheit untersucht.

Beim *Measurement Error* geht es um Messfehler, man unterscheidet zwischen unsystematischen (z.B. wegen Ablenkung der Probanden) und systematischen (die sich auf das Befragungsinstrument zurückführen lassen) Messfehlern.

Der *Non-Response Error* beschäftigt sich mit der Problematik, dass einige Untersuchungsobjekte keine Angaben machen, also den Fragebogen nicht beantworten. Ist dieses Fehlen rein zufällig verteilt, ist es nicht problematisch. Falls es sich jedoch um einen systematischen Fehler handelt, kann es zur Verzerrung der Umfrageergebnisse kommen. Um dem entgegenzuwirken versucht man, die Ausschöpfungsrate zu maximieren. Ein vielzitatierter Satz ist:

„Protection against nonresponse bias is the reduction of nonresponse itself...“
(Armstrong & Overton, 1977, S.396)

Wie man Non-Response reduziert, dazu gibt es umfassende Literatur, in der es entweder um Erfahrungswissen oder ‚behavioristische‘ Forschungsprogramme geht.

2.3 Formen des (Non)Response bei Web-Befragungen

Die Teilnahme an einer Web-Befragung lässt sich als mehrstufiger Selektionsprozess beschreiben. Die einzelnen Schritte sind:

- *Kenntnisnahme*: Der Proband muss darüber informiert werden, dass er bei einer Umfrage teilnehmen soll.
- *Abruf der Seite*: Der Befragte muss gewillt und in der Lage sein, die Startseite der Umfrage aufzurufen.
- *Befragungsbeginn*: Der Befragte muss gewillt und in der Lage sein, die Befragung zu beginnen.
- *Vollständige Teilnahme*: Der Befragte hat alle Fragen vollständig beantwortet, also keine übersprungen und nicht vorzeitig abgebrochen.

In jedem dieser Schritte kann und kommt es zu Ausfällen, sogenannter Non-Response. Bei dem Begriff Non-Response handelt es sich um einen Sammelbegriff, er steht für die unterschiedlichen Formen der nicht vollständigen Beantwortung der Befragung. Interessant für die Web-Befragung sind die Ausfälle ab dem Abruf der Seite bis zur vollständigen Bearbeitung.

Die frühere Grobeinteilung der Befragten in die Gruppen Teilnehmer, Unit- und Item-Nonresponders spiegelt die Teilnehmertypen nicht so gut wieder, wie die folgenden sieben Bearbeitungstypen:

1. *Complete Responder*: Die Befragten dieses Typs beantworten vollständig alle Fragen. Sie sind die Idealtypen für den Umfrageforscher.

2. *Unit-Nonresponders*: Dieser Typus ruft den Fragebogen nicht ab und nimmt somit nicht an der Befragung teil.
3. *Answering Drop-Outs*: Diese Gruppe von Personen beteiligt sich an der Befragung, bricht jedoch vor Beendigung der Umfrage ab.
4. *Lurker*: Personen dieses Typs beantworten keine Frage, klicken sich aber durch den ganzen Fragebogen.
5. *Lurking Drop-Outs*: Sind ein Typus, der eine Mischform der Answering Drop-Outs und der Lurker darstellt. Die Personen klicken sich zu Beginn durch den Fragebogen, brechen aber bevor sie das Ende erreicht haben ab.
6. *Item-Nonresponders*: Diese Personengruppe bearbeitet nicht alle Fragen, aber durchläuft den kompletten Fragebogen.
7. *Item-Nonresponding Drop-Outs*: Ist eine Zwischengruppe zwischen Item-Nonresponders und Answering Drop-Outs. Diese Personen durchlaufen nicht den ganzen Fragebogen und beantworten auch nicht alle Fragen, die sie aufgerufen haben.

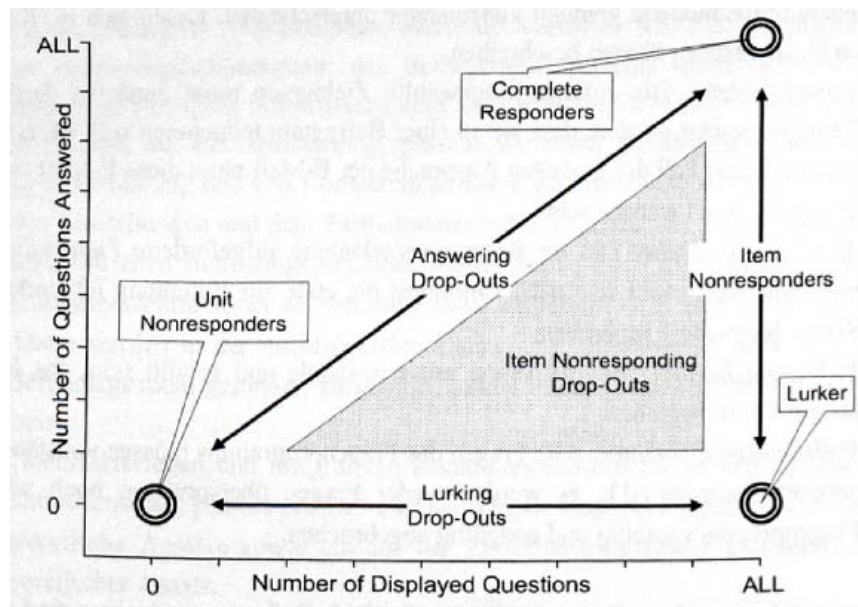


Fig. 2. Non-Response Formen bei Web-Befragungen (Bosnjak, 2002)

2.4 Erklärungsansätze für Non-Response

Arbeiten, die sich mit Erklärungsansätzen, beschäftigen lassen sich in jene einteilen, die sich mit Fragen beschäftigen, die unter direkter Kontrolle des Forschers stehen und solche, die sich auf Eigenschaften der Befragten Personen beziehen.

Ich werde hier auf zwei Merkmale der Ankündigung einer Befragung eingehen, nämlich der Wirkung von Incentives und der angekündigten Befragungsdauer.

Danach werde ich näher auf zwei Instrumente eingehen, der Verwendung einer Fortschrittsanzeige und eines Non-Response reminders.

Incentives: Dies sind Anreize, wie zum Beispiel eine Teilnahmeprämie oder die Chance auf einen Geldbetrag nach der erfolgreichen Teilnahme. Auch wenn die Bezeichnung Incentiv (Anreiz) die Vermutung nahe legt, dass es ein Anreiz ist, muss das noch nicht mit der Praxis übereinstimmen. Untersuchungen zu diesem Thema sind auf unterschiedliche Ergebnisse gekommen. Zusammenfassend kann man sagen, „dass nur im Voraus applizierte Incentives eine durchweg große positive Wirkung auf die Ausschöpfungsrate haben.“¹

Angekündigte Befragungsdauer: Hier geht es darum, welche Auswirkung die angekündigte Fragedauer auf die verschiedenen Formen der Non-Response hat. Aus einem Einzelbefund lässt sich folgende Aussage treffen: „je kürzer die angegebene Zeit, desto geringer der initiale Unit-Non-Response und je akkurater die angekündigte Zeitangabe, desto geringer der Drop-Out.“²

Fortschrittsanzeige: Dies ist eine Anzeige, die dem Teilnehmer aufzeigt, wieviele Fragen er schon beantwortet hat und wieviele er noch vor sich hat. Studien hierzu haben nicht zu einheitlichen Ergebnissen geführt. „Bei kurzen Befragungen scheinen Fortschrittsanzeigen die Drop-Out-Rate in geringem Maße zu verringern, bei langen Befragungen hingegen ist die Drop-Out-Rate bei vorhandenen Fortschrittsanzeigen höher, als wenn dieses Element fehlt.“³

Non-Response reminder: Dieser soll den Teilnehmer darauf hinweisen, dass er noch nicht alle Fragen beantwortet hat. Es gibt zwei unterschiedliche Formen, den hard reminder, dieser lässt einen Teilnehmer erst fortfahren, falls er alle Fragen beantwortet hat und den soft reminder, dieser dient nur als Erinnerung, der Teilnehmer kann auch fortfahren, wenn noch Fragen unbeantwortet sind. „Während der ‚hard reminder‘ den Drop-Out erhöhen können (Dillman & Bowker, 2001, S.173, Principle 9), konnten DeRouvray und Couper (2000) zeigen, dass mit ‚soft reminder‘ der Anteil an Item-Non-Response leicht verringert werden konnte.“⁴

3 Verwandte Arbeiten

In diesem Abschnitt werden Arbeiten vorgestellt, die sich im erweiterten Rahmen mit der Problematik, die ich bearbeite, beschäftigen. Arbeiten direkt zu meinem Thema, also aufgeteilte Onlinebefragung im Unternehmen, habe ich nicht gefunden. Die hier im Folgenden vorgestellten Arbeiten geben einen Einblick in verschiedene Modelle, die sich mit dem Aufteilen von Befragungen beschäftigen. Diese Modelle sind das Matrix Sampling Design, das Split Questionnaire Design und ein weiteres Split Questionnaire Design (between und within block design).

¹ Church (1993) S.32

² Bosnjak, Michael (2002) S.35

³ Bosnjak, Michael (2002) S.40

⁴ Bosnjak, Michael (2002) S.40

3.1 Multiple Matrix Sampling: A Review by Jeffrey M. Gonzalez and John L. Eltinge

Diese Arbeit beschäftigt sich mit der Multiple Matrix Sampling Methode, mit dieser lässt sich die Interviewzeit reduzieren. Der Fragebogen wird in kleinere Teile aufgeteilt und dann an einen Teil der Befragten verteilt. Als Anwendungsbeispiel wird der Consumer Expenditure Quarterly Interview Survey (CEQ) herangezogen, dieser sammelt Daten über das Konsumverhalten von U.S. Haushalten. Dafür werden Haushalte interviewt. Die Dauer des Interview beträgt 65 Minuten. In der Arbeit wird auf die Herkunft, die Anwendungsfelder und wie das Verfahren für eine Umfrage angewendet wird eingegangen.

Ausgangspunkt ist der CEQ, dieser wird für die Berechnung von Preisindices herangezogen. Preisindices sind sehr abhängig von der Datenqualität. Bei anderen Untersuchungen hat sich gezeigt, dass die Qualität der Daten bei einem langen Fragebogen abnimmt (Kraut et al. 1975; Johnson et al. 1974; Herzog and Bachmann 1981). Eine Umfrage unter High School Seniors hat zum Beispiel gezeigt, dass Überschreiten der optimalen Länge eines Fragebogens, die Wahrscheinlichkeit, eine korrekte Antwort zu bekommen, abnimmt. Die abnehmende Motivation wurde als Ursache ermittelt. Nach einer Motivationsabnahme steigt die Wahrscheinlichkeit, dass die Befragten einen einfacheren Weg suchen, die Fragen zu beantworten oder die Befragung abbrechen (Herzog und Bachmann, 1981). Daher ist es im Interesse des Durchführenden der Umfrage, die Zeit der Bearbeitung zu verkürzen und dadurch höhere Antwortraten und eine bessere Datenqualität zu erhalten. Außerdem lassen sich noch zusätzlich Kosten einsparen.

Die Probleme, die sich jetzt stellt ist, welche Fragen können gestrichen werden. Für diese Aufgabe kommt das Multiple Matrix Sampling Verfahren zum Einsatz. Dieses teilt den gesamte Fragebogen in Teile und verteilt diese an einen Teil der Grundgesamtheit. Somit werden alle Fragen beantwortet, nur von unterschiedlichen Befragten.

Multiple Matrix Sampling hat seinen Ursprung in den Fünfigern. Es wurde in Schulen bei Tests, ähnlich der Pisastudie, eingesetzt. Für solche Test gibt es eine Vielzahl von Fragen, die aber nicht einem einzelnen Schüler gestellt werden können. Daher kam man auf die Idee des Multiple Matrix Sampling. Man wählte jeweils einige Fragen aus dem Fragenkatalog aus und ließ diese von einigen Schülern beantworten. Ein zusätzlicher Vorteil war, dass die Dauer des Test reduziert werden konnte. Dadurch gab es mehr Schulen, die bei diesen Test teilnahmen.

In den siebzigern wurde Multiple Matrix Sampling auch von den Regierungsbehörden eingesetzt. Das U.S.Census Bureau verwendete für die Umfrage „2000 Decennial Census“ fünf verschiedene Designs. Dabei wurde darauf geachtet, dass Fragen, die eine Kreuztabulierung brauchen, zusammen in einem Fragenteil auftauchen. Ein Design bestand zum Beispiel aus 3 unterschiedlichen Fragebögen, wobei in jedem Fragebogen die wirtschaftlichen Fragen vorkamen und zusätzlich zwei von den drei anderen Fragengruppen, „Soc I“, „Soc II“ und „Housing“. Die unterschiedlichen Designs wurden ausgewertet und mit den Ergebnissen von 1990 verglichen. Dabei zeigte sich, dass die Schätzwerte meistens akzeptabel waren und dass mit Ausnahme eines Falls die Belastung der Befragten abnahm. Für das nächste Design schlugen Navarro und Griffin, die diese Untersuchungen durchführten, vor,

Fragen, die eine hohe Korrelation aufweisen, in unterschiedliche Fragebögen zu platzieren, damit man bei der Auswertung Imputationsmodelle verwenden kann.

Als letztes Anwendungsbeispiel wird die Arbeit von Raghunathan und Grizzle vorgestellt, auf die ich auch noch genauer in meiner Arbeit eingehen werde. Diese Autoren beschäftigen sich mit dem Cancer Risk Behavior Survey, für dessen Beantwortung zwischen 30 bis 50 Minuten benötigt werden. Das stellt eine hohe Belastung für die Interviewteilnehmer dar. Um dem entgegenzuwirken wird das Aufteilen des Fragebogens und das zufällige Verteilen der einzelnen Teile vorgeschlagen, das ist sehr ähnlich dem Multiple Matrix Sampling. Für die fehlenden Daten kann man von der MCAR (missing completely at random) Annahme ausgehen, diese ist Voraussetzung für eine gute statistische Analyse. Raghunathan und Grizzle erkennen, dass das Multiple Matrix Sampling einen großen Nachteil besitzt. Es kann vorkommen, dass bestimmte Fragenkombination nie zusammen auf einem Fragebogen auftauchen, und daher die statistische Auswertung nicht angewendet werden können. Aus diesem Grund entwickeln sie ein Multiple Imputations Verfahren, welches vollständige Datensätze erstellt, sodass statistische Schätzungen möglich sind. Danach untersuchen sie Ihr Konzept anhand einer Simulationsstudie und vergleichen es mit der Available Case Methode. Sie kommen zu den Ergebnissen, dass die Schätzungen der beiden Verfahren sehr ähnlich denen des vollständigen Fragebogens sind. Außerdem schneidet das Imputationsverfahren immer besser als die Available Case Methode ab.

In der Literatur wird der Ablauf einer Umfrage in drei Phasen unterteilt, die Erstellung (development) des Fragebogens, das Sammeln der Daten (data collection) und die Auswertung und Analyse (processing and analysis). Bei der Erstellung des Fragebogens geht darum, wie man den vollständigen Fragebogen in Teile aufteilt, welche Fragen man zusammen in einen Teil platziert und welche Anzahl von Teilfragebögen optimal ist. Die Beantwortung dieser Fragen ist nicht einfach, da es mehrere Ansätze gibt. Das Ziel der Umfrage sollte bei allen Entscheidungen im Vordergrund stehen. Die einfachste Aufteilungsmethode ist, die Fragen einfach zufällig in verschiedene Teilfragebögen einzuordnen, dabei kann es dazu kommen, dass logisch zusammenhängende Fragen in unterschiedlichen Teilfragebögen landen. Die nächste Methode teilt die Fragen in thematische Blöcke ein, die dann auf die einzelnen Teilfragebögen verteilt werden, dadurch kommt es nicht mehr zu dem oben erwähnten Problem. Die Einteilung der Fragen kann auch auf Grundlage statistischer Gegebenheiten erfolgen, zum Beispiel können Fragen, die eine hohe Korrelation aufweisen, in unterschiedliche Teilfragebögen eingeordnet werden. Die Idee dahinter ist, dass aus den beantworteten Fragen die fehlenden berechnet werden können. Eine weitere Variante verwendet Core Questions, die jeder Teilnehmer ausfüllen muss, also Fragen, die für den, der die Befragung initiiert von größerer Bedeutung sind. Diese Core Questions sind Bestandteil auf alle Teilfragebögen.

In der Datensammlungs-Phase geht es darum, welcher Teilnehmer welchen Fragebogen erhält. Auch diese Frage muss immer unter Berücksichtigung des Ziel der Umfrage beantwortet werden. Wenn man zum Beispiel fünf Fragebögen hat, dann kann man diese auf fünf vergleichbare Gruppen der Grundgesamtheit aufteilen. Ein anderes Verteilungskonzept wäre ein 2-Phasen Modell, jeder Teilnehmer bekommt erst einen Fragebogen und auf dessen Grundlage bekommt er einen weiteren Teilfragebogen zugeteilt.

Die Auswertung und Analyse Phase beschäftigt sich mit der Frage, was mit den Daten geschieht, nachdem sie erhoben wurden. Zu den verwendeten Verfahren zählen zum Beispiel Konsistenzchecks, Gewichtungen oder Imputationsverfahren. Auch in dieser Phase ist vor allem relevant, wofür man die Daten danach verwenden will, dies steht immer im Vordergrund. Die Entscheidungen, die in den Vorphasen Erstellung und Datensammlung getroffen wurden, beeinflussen diese Phase.

Zusammenfassend lässt sich sagen, bei allen Entscheidungen sollte das Ziel der Umfrage im Vordergrund stehen. Zur Auswertung kommen die zwei Verfahren, Available Case oder Imputation zum Einsatz. Falls man nur an den Erwartungswerten interessiert ist, kann man beide Verfahren verwenden, falls man jedoch an anderen statistischen Werten interessiert ist, muss man das Imputationsverfahren verwenden. Als nächstes Forschungsthema sehen die Autoren die noch nicht geklärten mathematischen Probleme beim Multiple Matrix Sampling.

(vgl. Jeffrey M. Gonzalez and John L. Eltinge,)

3.2 A Split Questionnaire Survey Design by Trivellore E. Raghunathan and James E. Grizzle

In dieser Arbeit wird zuerst darauf eingegangen, wie ein Fragebogen in einzelne Splits aufgeteilt und die einzelnen Splits verteilt werden sollten. Danach geht es um die Frage, wie man die fehlenden Daten ergänzt. Dafür wird ein Multiple Imputation Verfahren entwickelt. Danach werden die Ergebnisse aus zwei Simulationsstudien untersucht. In der ersten Feldstudie werden auf die Ergebnisse einer vorhandene langen Umfrage mehrere zufällige Split Questionnaire Designs angewendet, also aus einem langen Fragebogen werden nur die Antworten extrahiert, die man bei der Verwendung des Split Questionnaire Design erhalten hätte. Danach wird die Ergänzung durchgeführt und die Ergebnisse der Split Daten mit den vollständigen Datensätzen verglichen. Das Ergebnis des Vergleichs deutet darauf hin, dass in diesem Fall nur sehr wenige Informationen durch die Verwendung des Split Questionnaire Design verloren gehen. In der zweiten Studie wird die Effizienz des Split Questionnaire Design und die Robustheit der geschätzten Verteilungsannahmen, die für die Erzeugung der Imputation verwendet wird untersucht. Dabei werden vollständige und Split Datensätze auf Grundlage verschiedenster Verteilungsannahmen generiert. Die Imputationen werden unter der Annahme der Normalverteilung generiert. Der Vergleich der Daten deutet darauf hin, dass der Effizienzverlust abnimmt, wenn die Korrelation zwischen Fragen in unterschiedlichen Splits zunimmt.

Bei Split Questionnaire Design wird der Originalfragebogen in kleinere Fragebögen unterteilt, die alle ungefähr die gleiche Anzahl an Fragen haben. Das Aufsplitten kann als eine Erweiterung des Multi Matrix Sampling, das von Shoemaker (1973) und Munger und Lloyd beschrieben wurde, gesehen werden. Beim Matrix Sampling bekommt jeder Teilnehmer eine zufällige Auswahl an Fragen gestellt. Das Verfahren ist zum Schätzen der Mittelwerte verwendbar. Der Nachteil ist, dass es vorkommen kann, dass manche Fragenkombinationen nie gemeinsam beantwortet werden. Das führt dazu, dass die daraus folgenden Aussagen nicht abgeschätzt werden können. Um dem vorzubeugen gibt es beim Split Questionnaire

Einschränkungen bei der Auswahl der Fragen, zusätzlich wird ein neues Imputationverfahren entwickelt um vollständige Datensätze zu erhalten.

Bei der ersten Simulationsstudie handelt es sich um Daten aus der Cancer Risk Behavior Umfrage. Die Umfrage wurde von 1989 Befragten beantwortet, besteht aus 182 Fragen eingeteilt in 9 Gruppen. Aus diesen Fragen lassen sich 127 relevanten Variablen ableiten. Das Ziel für das Aufsplitten war, die Interviewzeit um 60% zu reduzieren, dafür wurde der lange Fragebogen in einen Kernteil, der von allen zu beantworten ist und 5 weiteren Komponenten unterteilt. Diese wurden dann überlappend verteilt, siehe Fig.3.

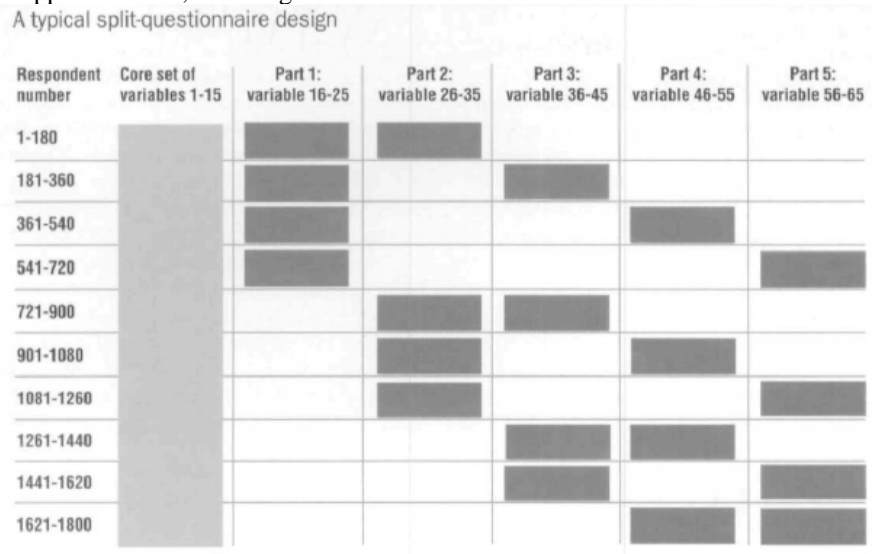


Fig. 3. A typical split-questionnaire design (Marco Vriens, Michel Wedel, and Zsolt Sándor)

Durch dieses Verfahren erhält man 10 Fragebögen. Die Umfragedaten für diese extrahiert man aus den 1989 vollständigen Datensätzen, das heißt für jeden Fragebogen werden zufällige 180 Datensätze ausgewählt.

Die Zuteilung der Variablen in die unterschiedlichen Komponenten wurde wie folgt durchgeführt. Zuerst wurden die Korrelationen zwischen allen 127 Variablen ermittelt, diese lagen im Intervall zwischen 0.08 und 0.72. Variablen mit hoher Korrelation wurden in unterschiedliche Teile eingeordnet. Nachdem alle Variablen eingeteilt wurden, wurde überprüft ob alles stimmig ist, z.B. wurden Fragen, die ein spezielles Thema abdecken als Einheit zusammengehalten.

Aus dieser Studie wird abgeleitet, dass der Schätzer den man erhält wenn man die Available Case Methode oder die Multiple Imputationmethode verwendet sehr ähnlich zu den Werten des vollständigen Fragebogens ist. Die Standardabweichung die sich aus dem Split Daten ergibt ist höher als die, die sich aus den vollständigen Daten ergibt, aber niedriger als die, die man bei Verwendung der Available Case Methode erhält. Zusammenfassen kann man sagen. Für den untersuchten Fall lässt

sich sagen das man bei der Verwendung des Split Questionnaire Design, praktisch die gleichen Rückschlüsse erhält.

Der Autor zieht folgende Schlüsse aus seiner Arbeit. Die Zuteilung der Fragebögen an den Probanden muss komplett zufällig geschehen, nur so kann man davon ausgehen das für die fehlenden Daten MCAR (missing completly at random) gilt. Die Simulationsstudien zeigen, das der Effizienzverlust am geringsten ist wenn die Korrelation zwischen den verschiedenen Komponenten hoch ist. Falls es nicht möglich ist diese hohe Korrelation zu erreichen, dann ist die Effizienz der Multi Imputationsverfahren nicht viel höher als die Available Case Methode. Das Imputationsverfahren bietet sich an wen die Daten für verschiedene Zwecke verwendet werden.

Wie der Fragebogen aufgesplittet wird hängt im Allgemeinen mit dem Zweck und dem Inhalt der Umfrage zusammen. Es gibt aber einige grundsätzlichen Dinge, jeder Split sollte schlüssig sein, das heißt zum Beispiel, falls die Beantwortung einer Frage, eine andere Frage nach sich zieht sollten sich diese im selben Split befinden. Aus statistischer Sicht sollten Variablen aus unterschiedlichen Komponenten eine hohe Korrelation aufweisen. Variablen innerhalb einer Komponente sollten dagegen nicht korrelieren, da dies dazu führt das trotz Beantwortung einer zusätzlichen Frage keine weiteres Erkenntnis gewonnen werden können. Falls keine Pilotdaten für eine Berechnung der Korrelationen vorliegen, die zur Erstellung der Splits benötigt werden, muss man das Aufsplitten auf Grundlage andere Erkenntnisse wie frühere Studie durchführen.

(vgl. Trivellore E. Raghunathan and James E. Grizzle, 1995)

3.3 Split Questionnaire Design by Feray Adigüzel

Diese Arbeit nimmt als Ausgangspunkt die Probleme, die durch den Einsatz von sehr langen Fragebögen in der Marktforschung entstehen. Dies sind hohe Unit-, Item-Non-Response und Abbruchraten. Außerdem kommt es zu Langweile und Ermüdungserscheinungen.

Als Lösung wird ein Split Questionnaire Design vorgestellt, welches den Aufwand für die Befragten reduziert ohne dabei auf wichtige Daten zu verzichten. In allen vorherigen Studien wurden Verfahren untersucht die die einzelnen Fragen rein zufällig auf verschiedene Fragebögen verteilten. Zu diesen Verfahren gehört das Time Sampling (Sikkel und Hoogendoorn, 1995) und das Multiple Matrix Sampling (Shoemaker, 1973; Thayer, 1983). Es gibt derzeit noch keine Methode die sich mit dem optimalen Zuteilen der Fragen beschäftigt, das heißt die Fragen so verteilt, das am wenigsten Information verloren geht. Eine solche Methode soll im folgenden vorgestellt und untersucht werden. Erste Ansätze sind in der Arbeit von Raghunathan und Grizzle (1995) zu erkennen, dort wird als Heuristik vorgeschlagen bei der Aufteilung auf die Zielsetzung der Umfrage zu achten, sowie darauf das die Fragen immer im richtigen Kontext stehen. Das grundsätzliche verteilen der Fragen soll auf Grundlage der Korrelationen zwischen den Fragen erfolgen. Fragen mit einer hohen Korrelation sollten in unterschiedliche Teile sortiert werden. Das Aufteilen der Fragen in Teilfragebögen ist eine sehr anspruchsvolle Aufgabe, da die Anzahl der möglichen Aufteilungen mit der Anzahl der Fragen des originalen Fragebogen exponentiell

ansteigt, deshalb wird auch bei diesem Verfahren, so wie in den andern Studien, die Struktur des ursprünglichen Fragebogen verwendet. In dieser sind die Fragen in einzelne Blöcke (z.B. thematisch) eingeteilt. Aus dieser Struktur werden jetzt die Teilfragebögen erstellt, dabei gibt es zwei unterschiedliche Designs, einmal das "between-block design", welches einen oder mehrere der Blöcke zu einem Teilfragebogen (Split) zusammenfasst, und das „within-block design“, das aus den unterschiedlichen Blöcken Fragen auswählt und diese zu einem Split zusammenfasst. In dieser Arbeit wird auf die Generierung der optimalen Split und auf die Ergänzung der fehlenden Daten eingegangen.

Um einen optimalen Split zu erhalten wird der Federov Algorithmus angewendet, mit welchem man aus allen möglichen Designs das optimale herausfindet. Der Algorithmus funktioniert für das between-block design folgendermaßen, zuerst wird eine Matrix mit allen möglichen Splitkombinationen aufgestellt, alle unzulässigen werden aussortiert. Aus den übrigen wird dann ein Design D_0 zufällig ausgewählt. Danach werden unter Verwendung der Pilot Daten die Schätzer für die einzelnen Fragen ermittelt. Anschließend wird das Kullback-Leibler(KL) Maß, ein Maß das den Informationsverlust misst, für das Design D_0 ermittelt. Im nächsten Schritt wird der erste Split durch alle anderen möglichen Splits ausgetauscht und es wird jeweils das KL-Maß ermittelt. Der ursprüngliche Split wird durch den Split mit dem kleinsten KL-Maß ersetzt. Als nächstes wird der nächste Split, also der zweite ausgetauscht und wie beim ersten Split vorgegangen. Danach wird sichergestellt das es sich um ein Design handelt das vollständig identifizierbar ist, das bedeutet das aus den beobachteten Daten alle fehlenden Parameter ermittelt werden können. Nachdem dies sichergestellt ist wird das Vorgehen für alle weiteren Split durchgeführt. Ist man am Ende angekommen, wird wieder von vorne begonnen, solange bis es zu keinen Austausch von Splits mehr kommt. Das gesamte Vorgehen wird mit mehreren anderen Startdesigns wiederholt, um sicherzustellen das es sich beim ermittelten Design nicht nur um ein lokales Optimum handelt. Das Design mit dem kleinsten KL-Maß wird dann verwendet.

Die fehlenden Daten werden mit Hilfe des Gibbs Sampling Verfahren ergänzt. Dies ist ein Multiple Imputationsverfahren, die Grundidee von Multiple Imputationsverfahren ist die zusätzliche Unsicherheit, die durch die Erzeugung und den Gebrauch der ergänzten Werte auftritt richtig abzubilden. Dazu wird nicht ein einzelner Wert ergänzt, und somit ein Datensatz erhalten, sondern mehrere Datensätze mit unterschiedlichen imputierten Werten. Es folgt eine getrennte Auswertung mit den verschiedenen Datensätzen. Die Varianz zwischen den Auswertungsergebnissen ist ein Maß für die zusätzliche Unsicherheit, die durch das Ergänzen entsteht.

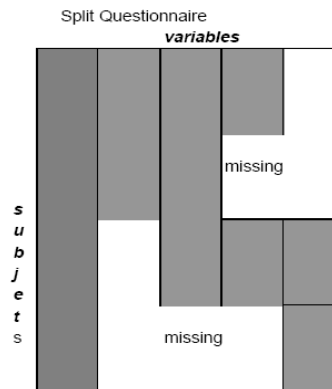


Fig. 4. Split Questionnaire Design (Adigüzel, 2006)

3.3.1 Designs

Es werden mehrere Designs vorgestellt, die bedingt durch das Design unvollständige Datensätze erzeugen, und erläutert wie damit umgegangen wird.

Two-Stage Design: Bei diesem Design besteht der Fragebogen aus zwei Teilen, einem Kernteil der von allen Befragten beantwortet werden soll und einen zweiten, der entweder rein zufällig an einen Teil der Befragten oder abhängig von den Antworten, die der Befragte im ersten Teil getroffen hat, verteilt wird. Durch die Wahl des Auswahlkriterium für den zweiten Fragebogen und die Korrelationen, die sich zwischen den beiden Teilen berechnen lassen, kann man Schlussfolgerungen für die fehlenden Fragen ziehen. (Neff, 1996)

Matrix Sampling Design: Beim Matrix Sampling wird jedem Befragten eine Auswahl an Fragen zugeteilt. Das lässt sich sehr gut in einer Matrix abbilden, die Reihen stellen die Befragten und die Spalten die Fragen dar. Wie die Fragen verteilt werden ist sehr unterschiedlich, man kann zum Beispiel ein „non-overlapping multiple matrix sampling design“ (NMS) also ein nicht überlappendes Design oder ein „balanced incomplete block design (BIB), also ein überlappendes Design wählen. Die Beispiele ist in Fig.5 in der Matrixform dargestellt. Mit diesem Design erhält man gute Erwartungswertschätzer, aber die Berechnung der Standardabweichung stellt sich kompliziert dar. Ein Unterschied zum Split Questionnaire Design(SQD) ist, das beim SQD die fehlenden Daten mit einem Imputationsverfahren ergänzt werden, das wird beim matrix sampling in der ursprünglichen Form nicht durchgeführt.

1-) Non-overlapping matrix sampling (NMS) 2-) Balanced incomplete block design (BIB)

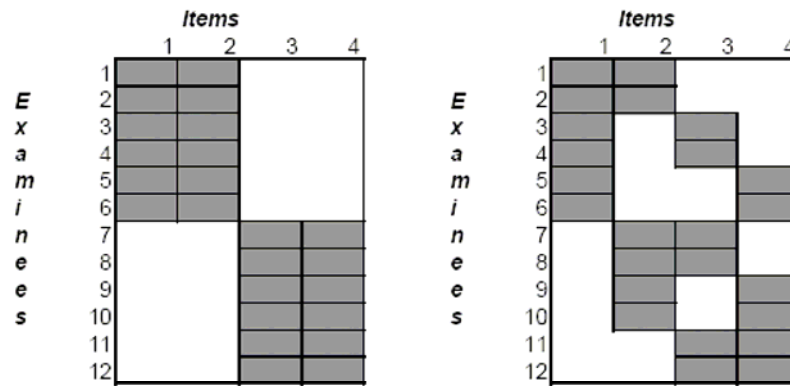


Fig. 5. Multiple Matrix Sampling (Adigüzel, 2006)

Time Sampling Design: Mit diesem Design können Informationen über Verhaltensweisen, die sich kontinuierlich über die Zeit verändern, ermittelt werden, zum Beispiel das Kaufverhalten. Anhand des Kaufverhaltens wird das Design erklärt. Es gibt zwei unterschiedliche Ansätze. Beim ersten Ansatz sind die Befragten, die über die Zeit Angaben zu ihrem Konsumverhalten machen immer die Gleichen. Es findet eine kontinuierliche Befragung statt, zum Beispiel in der Form eines Tagebuchs. Der zweite Ansatz verfolgt einen anderen Weg, bei ihm werden zu unterschiedlichen Zeitpunkten Probanden nach ihrem Konsumverhalten befragt. Durch die erhaltenen Angaben, schlussfolgert man was in den Zeiten zu denen keine Angaben vorliegen, stattgefunden hat. Die am häufigsten verwendeten Designs sind der longitudinal survey, der repeated survey und der rotating panel survey. Beim longitudinal survey werden die gleichen Probanden zu unterschiedlichen Zeitpunkten befragt. Beim repeated survey werden zu unterschiedlichen Zeitpunkten Umfragen durchgeführt, dabei wird nicht darauf geachtet, dass es die gleichen Probanden sind. Beim rotating panel survey wird zu den unterschiedlichen Zeitpunkten ein Teil Probanden aus der Befragung entfernt und durch einen anderen ersetzt. Beispiel siehe Fig.6.

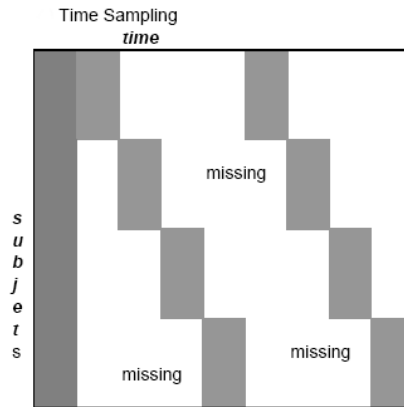


Fig. 6. Time Sampling (Adigüzel, 2006)

Data Fusion: Das Data Fusion ist ein Verfahren mit dem sich Datensätzen von unterschiedlichen Umfragen zu einem Datensatz zusammenführen lassen. Das funktioniert über Variablen, die in beiden Umfragen erhoben werden. Das Verfahren ist sehr ähnlich dem des Split Questionnaire Design, dort kann man die einzelnen Split als unterschiedliche Umfragen auffassen.



Fig. 7. Data Fusion (Adigüzel, 2006)

Split Questionnaire Design: Hier gibt es zwei unterschiedliche Formen, das „between-block design“, welches einen oder mehrere der Blöcke zu einem Teilfragebogen (Split) zusammenfasst, und das „within-block design“, das aus den unterschiedlichen Blöcken Fragen auswählt und diese zu einem Split zusammenfasst. Der Split wird so generiert das am wenigsten Information verloren geht. Dazu wird die Kullback-Leibler Distanz und der Federov Algorithmus verwendet. Die durch dieses Umfragedesign fehlenden Daten werden mit dem Gibbs Sampling Verfahren ergänzt.

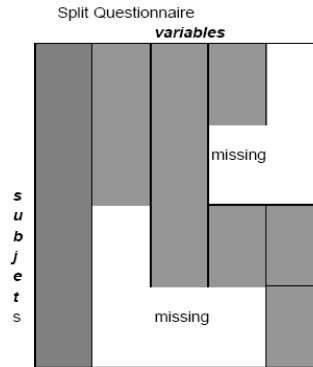


Fig. 8. Split Questionnaire Design (Adigüzel,2006)

3.3.2 Simulation Studies

Es werden zwei Simulationsstudien mit einem between-block design durchgeführt. Im ersten wird festgestellt wie gut der Fedorov Algorithmus ist das optimale Design zu identifizieren. Im zweiten wird das optimale Split Questionnaire Design (SQD) mit einem Matrix Sampling Design verglichen.

Um den Fedorov Algorithmus zu testen wird ein SQD verwendet bei dem sich alle Design berechnen lassen. Es wird ein Design mit 5 Blöcken á 4 Fragen genutzt. Die Antworten werden mit Hilfe einer multivariaten Normalverteilung erstellt. Für 5, 10 und 15 Splits wird das optimale Design ermittelt. Der Fedorov Algorithmus wird mit 10 Iterationen und 1000 unterschiedlichen Startdesigns durchgeführt. Die unterschiedlichen Startdesigns führen alle auf das gleiche optimale Ergebnis. Dies zeigt das der Fedorov Algorithmus zufriedenstellende Ergebnisse für das SQD Problem liefert.

In der zweiten Studie wird ein optimales between block split questionnaire design mit einem matrix sampling design (MSD) verglichen. Dabei wird ein Fragebogen mit sechs Blöcken und fünf Fragen pro Block verwendet. Es werden Designs mit 5 und mit 10 Splits getestet. Untersucht werden Streuungsparameter und Parameter die Aussagen über den Informationsverlust treffen. Das SQD schneidet immer besser ab als das MSD.

3.3.3 Empirical Data Application

Für die empirische Untersuchung werden die Daten die aus dem „Project 2000 Ninth GVU Survey Web Attitude and Perceptions Questionnaire“ verwendet. Diese Untersuchung besteht aus 65 Fragen und ist daher nicht ideal für die Verwendung des Split Questionnaire Design geeignet. Die Fragen sind nach Inhalt in neun Gruppen sortiert. Daten von zwei Erhebungen sind verfügbar, eine mit 500 und eine mit 1150 Datensätzen. Der Vorteil bei der Verwendung von Daten einer vollständigen Umfrage ist das man die Leistungsfähigkeit des SQD bewerten kann. Der Nachteil ist, dass der Vorteil der verbesserten Datenqualität, nicht zum tragen kommt. Mit den Daten der ersten Erhebung werden die Split Questionnaires erstellt. Designs mit folgenden Eigenschaften werden generiert,

- between-block und within-block
- fünf oder zehn Splits
- mit und ohne Einschränkungen (z.B. Anzahl der Blöcke pro Split)

Als Benchmark für das between-block design dient das matrix sampling design (MSD), das within-block design wird mit einem random questionnaire design (RQD) verglichen.

Die Werte, die aus dem between-block design erhalten werden, sind sehr ähnlich denen die beim vollständigen Fragebogen erhält. Beim 5 Split Design sind die Werte besser als beim 10 Split Design, was sich durch die höhere Raten an fehlenden Daten erklären lässt. Die Daten belegen das, das Design signifikant besser ist als das MSD. Beispielhaft einige Werte: bei 5 Split Design reduziert sich die Anzahl der Fragen gegenüber dem vollständigen Fragebogen um 66%, der Informationsverlust liegt dabei bei 14%, wohingegen beim MSD die Informationsverlustrate bei 18% liegt.

Die Werte beim within-block design sind besser als die Werte des RQD, aber der Unterschied zwischen den beiden sind nicht so groß wie beim between block design. Die Informationsverlustrate liegt nur noch bei 6%.

Es wird noch die Sensitivität von den Pilotdaten untersucht. Dabei schneidet das between block design sehr viel besser ab, also ist es nicht so stark von den Pilotdaten abhängig.

3.3.4 Field Study

Es wird eine Studie durchgeführt die positiven Effekte des Split Questionnaire Design untersucht. Der Fragebogen aus der empirischen Studie wird verwendet und zusätzlich werden Fragen zu den Themen Müdigkeit, Dauer, Langeweile gestellt. Verglichen wird ein vollständiges Design, ein zehn Split between block design und ein zehn Split within-block design. Die Studie wird mit 189 Teilnehmern durchgeführt, wobei jeweils ein Drittel, also 63, eins der drei Designs beantwortet. Die durchschnittliche Bearbeitungszeit für den vollständigen Fragebogen betrug 8 Minuten und ungefähr 6 Minuten für die Splitfragebögen. Das heißt die Dauer reduziert sich um 25%, wobei die Anzahl der Fragen um 50% reduziert wird.

Nach dem Auswerten der zusätzlichen Fragen lässt sich erkennen, das beim Thema Langeweile das within-block design am besten abschneidet, signifikant besser als der vollständige Design, der Unterschied zwischen between-block und vollständigem Design ist nicht signifikant. Das lässt den Schluss zu, das für die Langeweile vor allem die ähnlichen Fragen, die in einem Block zusammen stehen verantwortlich sind. Die Befragten fühlten sich nach der Beantwortung des vollständigen Fragebogens müder als nach den between block Split, die Werte waren 4,32 zu 3,57. Daraus lässt sich schließen, das Müdigkeitsgefühl von den Themenwechseln abhängig ist. Das between block design schneidet insgesamt positiver ab als das vollständige design, es wird als nicht so langweilig und wiederholend empfunden. Die Werte des within block design sind noch einmal signifikant besser als die des between block designs.

Diese positiven Werte sollten sich für die Datenqualität vorteilhaft auswirken. Bei der Befragung konnte jede Frage übersprungen werden, beim vollständigen design wurden 33, beim between-block 7 und beim withinblock 5 Fragen nicht beantwortet. Diese Zahlen deutet daraufhin das der Split Questionnaire die item Non-Response reduziert.

4 Konzept

Dieses Kapitel gliedert sich in drei Abschnitte. Der erste beschäftigt sich mit der Vorstellung der Parameter, die für eine geteilte Umfrage berücksichtigt werden müssen, darauf aufbauend werden vier Konzepte, deren Grundidee aus den verwandten Arbeiten stammt, präsentiert. Im letzten Teil erläutere ich, warum ich Konzept 2 das Matrix Sampling für die Umfrage empfehlen würde.

4.1 Vorstellung der Parameter

Zur Entwicklung eines eigenen Konzeptes werde ich zuerst auf die Parameter eingehen, die das Konzept charakterisieren. Diese Parameter übernehme ich hauptsächlich aus den verwandten Arbeiten, ergänzt um ein paar Punkte, die ich persönlich noch relevant finde. Die Einteilung der Parameter in die drei Phasen Development, Data Collection und Processing and Analysis übernehme ich aus der Arbeit von Gonzalez und Eltinge.

4.1.1 Development

In dieser Phase geht es um das Design des Fragebogens, welche Fragen kommen wie auf den Fragebogen. Dabei gibt es einige Punkte, die entschieden werden müssen.

Pilot Studie: Als erste Frage sollte man klären, ob man eine Pilot Studie durchführen will, um daraus Erkenntnisse zu gewinnen, wie zum Beispiel welche Korrelation zwischen den Fragen besteht.

Anzahl der Fragen pro Fragebogen: Man muss sich Gedanken darüber machen, wie viele Fragen jeder Teilfragebogen haben soll. Das ist die relevante Größe für die Zeit, die für die Bearbeitung eines Bogens anfällt.

Zuteilung der Fragen: Bei diesem Punkt geht es darum, wie man die Fragen auf die einzelnen Teilfragebögen verteilt. Folgende Varianten habe ich herausgesucht:

- *random:* Es werden einfach zufällig Fragen auf die Teilfragebögen verteilt.
- *between-block:* Nur komplette Fragenblöcke werden auf die Teilfragebögen verteilt, so sind inhaltlich zusammengehörige Fragen immer auf dem gleichen Fragebogen. Hier können die Blöcke zufällig oder nach einer bestimmten Methode, zum Beispiel Korrelation ausgewählt werden.
- *within-block:* Aus den einzelnen thematischen Fragenblöcken werden Fragen ausgewählt und auf einen Teilfragebogen gepackt. Dabei kann man entweder zufällig die Fragen auswählen oder zum Beispiel so, dass Fragen mit hoher Korrelation auf unterschiedlichen Teilfragebögen landen.
- *core questions:* Sehr oft ist es bei einer Umfrage so, dass es Fragen gibt, die für den Umfrageauftraggeber eine sehr hohe Relevanz haben. Diese Fragen können einen Kernteil bilden, der auf allen Fragebögen auftaucht. Die restlichen Fragen werden nach oben erläuterten Prinzip verteilt.

Fragenreihenfolge: Als letzte Frage steht die Reihenfolge der Fragen. Ordnet man sie so wie sie ursprünglich angeordnet waren oder rein zufällig.

4.1.2 Data Collection

Der Inhalt dieser Phase lässt sich grob in drei Gruppen unterteilen. Der erste beschäftigt sich mit der Frage, wie die Fragebögen zugeteilt werden. Das zweite sind äußere Rahmenbedingungen und das dritte Designelemente über deren Verwendung entschieden werden muss. Da Studien zu den einzelnen Punkten zu keinen eindeutigen Ergebnissen gekommen sind, gibt es kein richtiges Vorgehen, sondern man muss von Situation zu Situation selbst entscheiden.

1. *Zuteilung der Fragebögen:* Hier lassen sich zwei mögliche Vorgehensweisen vorstellen, entweder eine zufällige oder eine 2-Phase Zuteilung, dabei wird in der ersten Phase ein Teilfragebogen, z.B. der Kernfragebogen, von den Teilnehmern ausgefüllt und dann ausgewertet. Auf Grundlage dieser Auswertung wird dann ein weitere Teilfragebogen an die Teilnehmer verteilt.
2. *Rahmenbedingungen:* Hierunter fallen sehr viele verschiedene Aspekte über die man sich vor der Erhebung noch Gedanken machen muss.
 - *Anonymität:* Es stellt sich die Frage, ob die Umfrage anonym durchgeführt wird oder ob es nicht sogar wünschenswert ist, zu wissen, wer den Fragebogen beantwortet hat.
 - *Bearbeitung von mehreren Teilfragebögen:* Darf jeder Teilnehmer nur einen Teilfragebogen beantworten oder gibt man ihm die Möglichkeit, falls er die Zeit und das Interesse mitbringt, mehrere Teilfragebögen zu beantworten.
 - *Teilnahmebedingungen:* Im Firmen Kontext kann man sich vorstellen, dass die Mitarbeiter zur Teilnahme an einer Umfrage verpflichtet werden. Daher stellt sich die Frage ob die Umfrage verpflichtend oder freiwillig ist.
 - *Abbruchmöglichkeit:* Kann der Fragebogen zu jeden Zeitpunkt abgebrochen werden. Falls ja dann stellt sich noch zusätzlich die Frage, wie mit den Daten bis zum Abbruchzeitpunkt verfahren wird. Diese können entweder gelöscht oder in die Ergebnisse mitaufgenommen werden.
 - *Anreize(Incentives):* Der Einsatz von Anreizen z.B. monetärer Art, kann sich positiv auf die Non-response auswirken, aber auch negativ auf die Datenqualität, da nun auch Teilnehmer mitmachen, denen die Umfrage egal ist und sie daher die Fragen nicht gewissenhaft beantworten.
 - *Zeitmessung:* Ist eigentlich eine Selbstverständlichkeit, aber ich habe es noch einmal aufgenommen, da es im Kontext Micro-Feedback eine wichtige Rolle spielt. Falls die Zeitdauer nicht mit der geplanten Zeitdauer übereinstimmt, kann noch während die Daten erhoben werden der Fragebogen anpassen werden. Also falls sich rausstellt, das die Bearbeitungsdauer zu lange oder zu kurz ist kann man für folgende Teilnehmer den Fragebogen abändern.
3. *Designelemente:* Diese Bezeichnung soll ausdrücken, dass man die folgenden Elemente als Teilnehmer visuell wahrnimmt.
 - *Fortschrittsbalken:* Dieser soll dem Teilnehmer zeigen, wieviele Fragen er schon beantwortet hat und wieviele noch vor ihm liegen. Das kann dazuführen, dass, wenn nur noch wenige Fragen zu beantworten sind, der Teilnehmer diese auch noch beantwortet anstatt abzubrechen. Andererseits

wenn er schon einige Fragen ausgefüllt hat und sieht, dass er noch sehr viele vor sich hat, wird er frühzeitig abbrechen.

- *Zeitangabe:* Zu Beginn und während der Befragung könnte die voraussichtliche Bearbeitungsdauer angezeigt werden. Es hat sich gezeigt, dass „je kürzer die angegebene Zeit, desto geringer der initiale Unit-Nonresponse und je akkurater die angekündigte Zeitangabe, desto geringer der Drop-Out.“⁵

4.1.3 Processing and Analysis

Diese Phase dient der Aufbereitung der Daten. Dabei geht es vor allem darum, wie mit Non-Response umgegangen wird. Zur Korrektur von Non-Response gibt es drei Verfahren: Fallreduktion, Gewichtung und Imputation/Ergänzung.

Fallreduktion: Hierbei werden die Datensätze, die ein fehlendes Element (Item Non-Response) haben, aus der Stichprobe gestrichen. Diese ist nur ein sinnvolles Verfahren, falls die Stichprobe genügend groß ist und es nur wenige fehlenden Werte gibt. Kein gutes Verfahren bei Mehrzweckstichproben, da es dort zu sehr vielen Datensätzen mit fehlenden Werten kommt.

Gewichtung: Dieses Verfahren findet Anwendung, wenn man mit Unit Non-Response konfrontiert ist. Man verwendet meistens Verfahren der gebundene Hochrechnung, diese sollen die Varianzreduzieren und die Verzerrungen verkleinern. Die Schwierigkeit liegt darin, die Variablen zu identifizieren, die mit den Zielvariablen und dem Auswahlprozess korrelieren. Das ist notwendig, um die Verzerrung durch die Unit Non-Response richtig zu korrigieren.

Imputation: Mit diesem Verfahren lassen sich Item Non-Response zu vollständigen Datensätzen ergänzen. Dies geschieht mit Hilfe von Schätzungen. Es kommen verschiedene Verfahren zum Einsatz.

- *Mittelwertergänzung:* Hierbei werden fehlende Daten durch das arithmetische Mittel ersetzt. Nachteilig wirkt sich das Verfahren auf die Varianz aus, diese verkleinert sich und lässt dann keine Schlüsse mehr zu.
- *Hot Deck Techniken:* Dabei wird die Stichprobe nach irgendwelchen ausgewählten Kriterien in Gruppen eingeteilt, sogenannte Imputationsgruppen. Für die Ergänzung werden diese Gruppen einzeln betrachtet. Wenn ein Wert fehlt, wird ein ähnlicher Fall gesucht und dieser Wert ergänzt. Es gibt auch Verfahren, bei denen zufällig ein Wert aus der Gruppe ausgewählt und ergänzt wird oder der nächstgelegene Wert verwendet wird.
- *Regressionsmethode:* Diese Methode ist anwendbar, falls man einen Zusammenhang zwischen fehlenden und beobachteten Werten vermutet. In diesem Fall lässt sich ein lineares oder nichtlineares Regressionmodell aufstellen, mit welchem dann die fehlenden Werte geschätzt und imputiert werden können. Durch einen Störterm lässt sich die Methode so einpassen, dass die Varianz nicht unterschätzt wird.
- *Predictive Mean Matching:* Hierfür wird ein Regressionsmodell benötigt. Diesmal werden aber nicht einfach die fehlenden Werte aus dem Regressionsmodell abgelesen, sondern es wird ein Beobachtung gesucht,

⁵ Bosnjak, Michael (2002) (Non)Response bei Web-Befragungen, S.35

die möglichst nahe an der des Schätzwertes liegt und der tatsächliche Wert der Beobachtung wird dann ergänzt.

- *Multiple Imputation:* Beim Multiple Imputationsverfahren werden für jeden fehlenden Wert mehrere Imputationswerte aus einer Verteilung erzeugt. Dadurch erhält man mehrere vollständige Datensätze. Diese werden alle getrennt analysiert und am Schluss zu einem gemeinsamen Ergebnis zusammengefasst. Die zusätzliche Unsicherheit wird durch die Höhe der Varianz der Analyseergebnisse verdeutlicht.

(vgl. Hans Kiesl)

4.2 Aufstellung von vier Konzepten

In diesem Abschnitt werde ich aufbauend auf das letzte Kapitel vier Konzepte vorstellen. Das heißt, ich werde für diese Konzepte die einzelnen Parameter festlegen. Die Grundidee der einzelnen Konzepte stammt aus den Verfahren, die ich in den verwandten Arbeiten vorgestellt habe. Daraus übernehme ich auch die Namensgebung, die ich wie folgend gewählt habe, Random, Matrix Sampling, Split Questionnaire und 2-Phasen.

Es wird zuerst auf die gemeinsame Parameter eingehen, die bei allen Konzepten gleich sein sollen und danach werde ich Unterschiede aufzeigen.

4.2.1 Gemeinsame Parameter

Hier werden die Parameter erläutert, die bei allen Konzepten die ich danach vorstelle gleich sind.

- *Anzahl der Fragen auf den Fragebogen:* Diese Frage lässt sich nicht so einfach beantworten. Festgelegt werden sollte nicht die Anzahl, sondern die Dauer für die Bearbeitung der Fragen, und daraus sollte man dann eine Anzahl von Fragen folgern. Diese Dauer sollte im Unternehmenseinsatz nicht fünf Minuten überschreiten.
- *freiwillige Teilnahme:* Die Teilnahme sollte auf jeden Fall freiwillig erfolgen, da die Teilnehmer ansonsten nicht gewissenhaft die Fragen beantworten und es somit zu einer Verfälschung der Daten kommt.
- *anonyme Teilnahme:* Die Teilnahme muss anonym stattfinden, ansonsten werden einige Teilnehmer die Fragen so ausfüllen wie es Ihr Vorgesetzter erwartet. Dadurch kommt es auch wieder zu falschen Daten.
- *Unterbrechungsmöglichkeit:* Diese sollte existieren, da Mitarbeiter sehr oft kurzfristig etwas anders erledigen müssen, danach aber weiter den Fragebogen bearbeiten und nicht wieder von Vorne anfangen wollen.
- *Abbruchmöglichkeit:* Auch diese sollte existieren. Die Daten, die bis zum Zeitpunkt des Abbruchs eingegeben wurden, sollten falls möglich mit in die Untersuchung einfließen, da man davon ausgehen kann, dass diese Wahrheitsgemäß beantwortet wurden. Damit das der Fall, ist sollte es möglich sein, sich durch die Fragen durchzuklicken ohne diese zu beantworten.
- *Möglichkeit zur Bearbeitung von mehreren Fragebogenteilen:* Dieses sollte erlaubt sein und ist wünschenswert. Mit der Einschränkung, das

sichergestellt werden muss, dass die Teilnehmer jeden Teilfragebogen nur einmal beantworten können. Kommt es zur mehrfachen Beantwortung des gleichen Teilfragebogen durch den Teilnehmer verzerrt sich das Resultat.

- *Keine Incentives:* Incentives haben sich in Studien (siehe Kapitel 2.4) nicht als durchweg positiv erwiesen. Außerdem stellt sich die Frage, in welcher Form sie überhaupt eingesetzt werden können. Beim Einsatz von Incentives kommt es zu Teilnehmern, die nur zum Erhalt dieser Anreize an der Befragung teilnehmen und die deshalb auch nicht gewissenhaft den Fragebogen ausfüllen.
- *Messung der Zeit:* Die Zeit sollte gemessen werden um zu sehen, ob die geplante Befragungsdauer nicht zu stark über- wie unterschritten wird, da man in einem solchen Fall reagieren muss.
- *Fortschrittsbalken:* Ein Fortschrittsbalken ist sinnvoll, da sich gezeigt hat das sich dieser bei kurzen Befragungen positiv auf die Non-Response ausgewirkt (siehe Kapitel 2.4). Außerdem halte ich ihn aus persönlicher Erfahrung für sinnvoll.
- *Zeitangabe:* Am Start der Befragung halte ich eine realistische Angabe der Zeitdauer für sehr forderlich. Dadurch erkennen die Teilnehmer, dass die Umfrage sehr kurz ist, was dazu beitragen sollte, dass mehr Mitarbeiter teilnehmen.

4.2.2 Die vier Konzepte

Meine Konzepte unterscheiden sich nur in wenigen der Parameter, die ich im Kapitel 4.1 vorgestellt habe. Die anderen Parameter sind für alle Konzepte gleich und wurden im Kapitel 4.2.1 beschrieben. Die Unterschiedliche sind die Einteilung der Fragen auf den Fragebogen, die Frage ob Core Questions verwendet werden, die Fragenreihenfolge, die Zuteilung der Fragebögen auf die Teilnehmer und das Auswertungsverfahren. Es folgt eine Tabelle, in der die unterschiedlichen Konzepte gegenübergestellt werden, danach werde ich kurz auf die einzelnen Konzepte eingehen und deren Unterschied erläutern. Die Grundidee für die Konzepte stammen aus den Arbeiten, die ich in Kapitel 3 vorgestellt habe.

	Konzept 1 Random	Konzept 2 Matrix Sampling	Konzept 3 Split Questinnaire	Konzept 4 2-Phasen
Einteilung der Elemente	random	block	between-block	block
Core Questions	o	x	x	x(Phase1)
Fragenreihenfolge	beliebig	richtige	richtige	richtige
Zuteilung der Fragebögen	random	random	random	nach Phase 1
Auswertung	Fallreduktion	Fallreduktion	Imputation	Fallreduktion

Table 1. vier Konzepte

4.2.2.1 Konzept 1: Random:

Sowie der Name schon vermuten lässt, ist bei diesem Konzept alles zufällig. Die Anzahl zu stellender Fragen, die vorher festgelegt wurde, wird einfach zufällig aus dem Fragenpool ausgewählt. Kein Teilnehmer bekommt den gleichen Fragebogen,

daher wird auch auf die Core Questions verzichtet und auf die Fragenreihenfolge nicht geachtet. Die Zuteilung der Fragebögen erfolgt auch rein zufällig. Zur Auswertung wird auf das Fallreduktionsverfahren zurückgegriffen, da für die anderen die Daten fehlen.

4.2.2.2 *Konzept 2: Matrix Sampling*

Bei diesem Konzept wird für die Einteilung der Fragen auf die natürliche Struktur des Ursprungsfragebogens zurückgegriffen. Es werden nur Blöcke von thematisch zusammenhängenden Fragen auf die einzelnen Teilfragebögen verteilt, nicht einzelne Fragen. Da die Fragen nur Blockweise verteilt werden, sollte die Reihenfolge eingehalten werden, das erleichtert die Bearbeitung für den Teilnehmer. Dieses Konzept arbeitet mit Core Questions und die Zuteilung der Fragebögen erfolgt zufällig. Zur Auswertung wird, wie schon beim Konzept 1, mit der Fallreduktion gearbeitet, für eine gute Imputation fehlen die notwendigen Daten. Außerdem liefert die Fallreduktion gute Ergebnisse.

4.2.2.3 *Konzept 3: Split Questionnaire*

Hier wird mit einer Pilotstudie gearbeitet. Dadurch lassen sich optimale Teilfragebögen des Typs Between-Block zusammenstellen. Between-Block bedeutet, nur ganze Fragenblöcke werden auf unterschiedliche Teilfragebögen verteilt und das erfolgt, verkürzt gesagt, so, dass Blöcke mit hoher Korrelation auf unterschiedlichen Teilfragebögen erscheinen. Die fehlenden Daten werden durch ein passendes Imputationsverfahren ergänzt, das durch die optimale Gestaltung der Fragebögen auch gute Ergebnisse liefert. Auch in diesem Konzept kommt es zum Einsatz von Core Questions. Die Fragenreihenfolge wird nicht geändert und die Zuteilung erfolgt wie bei den beiden vorherigen Konzepten rein zufällig.

4.2.2.4 *Konzept 4: 2-Phasen*

Dieses Konzept unterscheidet sich von den anderen in wesentlichen Punkten. Das Vorgehen erfolgt in zwei Phasen. In der ersten Phase wird jedem Teilnehmer die Core-Questions übermittelt und von diesem bearbeitet. Danach wird in Abhängigkeit, der dort beantworteten Fragen, ein Teilfragebogen zugeteilt. Die Auswahl dieses Teilfragebogens kann nach verschiedensten Kriterien erfolgen, zum Beispiel auf Grundlage der Position des Mitarbeiters oder des Geschlechts oder so, dass alle Teilfragebögen gleich oft ausgefüllt wurden. Für die Auswertung wird wieder auf das bewährte Fallreduktionsverfahren zurückgegriffen.

4.3 Mein Vorschlag: Konzept 2

Für den konkreten Anwendungsfall, einer relativ kurzen Befragung im Firmenkontext, empfehle ich das Konzept Matrix Sampling. Dieses Verfahren hat sich im Praxiseinsatz bereits bewährt. Es ist bei normalen Umfragen schon seit über 50 Jahren im Einsatz und es gibt mehrere wissenschaftlich Untersuchungen darüber. Die Erstellung der Teilfragebögen erfolgt schnell und unkompliziert, dadurch kann es nicht zu Fehlern bei der Erstellung kommen, die Einfluss auf die Umfrage haben. Eine Pilotstudie, die ich im Unternehmensalltag für unrealistisch oder verfälschend halte, da nur ein bestimmter Typus Mitarbeiter daran teilnehmen würde, wenn

überhaupt, wird nicht benötigt. Core-Questions halten ich für ein sehr wichtiges Element, da es bei jeder Umfrage Fragen gibt, die bedeutender sind als der Rest. Die natürliche Blockstruktur kommt dem Teilnehmer entgegen, er muss gedanklich nicht zwischen verschiedenen Themen hin und herwechseln, sodass es zu einer fließenden Beantwortung des Teilfragebogens kommen sollte. Da nur der Erwartungswert interessant ist und dieser sich sehr gut aus dem Fallreduktionsverfahren ermitteln lässt, eignet sich dieses Verfahren gut für die Analyse und Auswertung. Diese erfolgt dadurch schnell, einfach und unkompliziert, und es ist möglich zu jedem beliebigen Zeitpunkt eine Auswertung zu erhalten. Da bei diesem Verfahren keine Parameter, wie zum Teil bei Imputationsverfahren, festgelegt werden müssen, kommt es nicht zur Verfälschung der Daten durch falsche Parameterwahl. Zusammenfassend lässt sich sagen, dass dieses Konzept die besten Ergebnisse liefern sollte, da es auf bewährte Methoden setzt und eine geringe Fehleranfälligkeit hat. Für denjenigen, der die Untersuchung durchführt hat es den Vorteil, dass es einfach einzusetzen ist, kein Vorwissen verlangt und sehr gut an die konkrete Aufgabe angepasst werden kann.

5 Zusammenfassung

Die Problematik, die sich durch die widersprüchliche Forderung nach einer aussagekräftigen und einer kurzen Umfragen ergibt, kann durch die Aufteilung des Fragebogens entschärft werden. Dadurch lassen sich qualitativ hochwertige Umfrageergebnisse mit einer reduzierten Bearbeitungsdauer für den Teilnehmer erzielen.

Ein Konzept zum konkreten Vorgehen bei einer Firmenonlinebefragung gibt es nicht, aber es gibt wissenschaftliche Arbeiten, die sich mit dem Aufteilen von Fragebögen beschäftigen. Aber auch hier fehlen weitangelegte Studien, die belegen, dass die Konzepte den an sie gestellten Anforderungen gerecht werden. In der Arbeit wird zuerst auf die Problematik der Non-Response eingegangen, danach werden die oben erwähnten wissenschaftlichen Arbeiten vorgestellt. Daraus wird ein Parameterkonzept entwickelt. Unter Verwendung der Parameter werden vier mögliche Konzepte aufgezeigt und dann eines für den Anwendungsfall ausgewählt.

Auf dem Gebiet der geteilten Fragebögen sowie der Onlineumfrage in Unternehmen müssen weitere Forschungsbemühungen unternommen werden, um ein fundiertes Micro-Feedback Verfahren zu entwickeln.

Literaturverzeichnis

Adigüzel, Feray (2006). "Essays on customisation applications in marketing". Thesis.

Adigüzel Feray & Wedel, Michel (2005). "Split Questionnaire Design". Working Paper No.966. University of Michigan.

Armstrong, J.S. & Overton, T.S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 16, 396-402.

Bosnjak, Michael (2002). "(Non)Response bei Web-Befragungen". Aachen. Shaker Verlag.

Church, A.H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, 57, 62-79.

DeRouvray, C. & Couper, M.P. (2000). Designing a strategy for capturing 'respondent uncertainty' in Web-based surveys. Manuscript submitted for publication

Dillmann, D.A. & Bowker, D.K. (2001). The Web questionnaire challenge to survey methodologists. In U.D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (S.159-178). Lengerich: Pabst Science Publishers.

Gonzalez, Jeffrey M. & Eltinge, John L. (2007). "Multiple Matrix Sampling: A Review", American Statistical Association, Session 36: Sampling Design—Methods, Adaptation, Redesign

Herzog, A.R. & Bachman, J.G. (1981). Effects of Questionnaire Length on Response Quality. *The Public Opinion Quarterly*, 45, 549-59.

Johnson, W.R., Sieveking, N.A. and Clanton, E.S. (1974). Effects of Alternative Positioning of Open-Ended Questions in Multiple-Choice Questionnaires. *Journal of Applied Psychology*, 59, 776-8.

Kiesl, Hans (2007). "Wirtschafts- und Sozialstatistik". Katholische Universität Eichstätt Ingolstadt. Vorlesung

Kraut, A.I., Wolfson, A.D. and Rothenberg, A. (1975). Some Effects of Position on Opinion Survey Items. *Journal of Applied Psychology*, 60, 774-6.

Rässler, Susanne & Koller, Florian & Mäenpää, Christine (2002). "A Split Questionnaire Survey Design applied to German Media and Consumer Surveys". The

International Conference on Improving Surveys, 25-28 August 2002. University of Copenhagen – Denmark.

Raghunathan, Trivellore E. & Grizzle, James (1995). "A Split Questionnaire Survey Design". *Journal of the American Statistical Association*, 90, 54-63.

Rodrigues, Nadia M.C. & Iwakami Kaizô (2005). A split questionnaire design applied to the census of Brazil. *International Union for the Scientific Study of Population XXV International Population Conference Tours, France, July 18-23, 2005*.

Vriens, Marco, Michel Wedel, and Zsolt Sandor (2001). "Split-Questionnaire Design: A New Tool in Survey Design and Panel Management". *Marketing Research*, 13 (Summer), 14-19.