

**Ein experimenteller Vergleich von Verfahren zur
Kontrolle von Antwortverzerrungen infolge sozialer
Erwünschtheit**

oder

Beichten, ohne rot zu werden

Diplomarbeit zur Erlangung
des Grades eines Diplom-Psychologen
am Institut für Experimentelle Psychologie
der Heinrich-Heine-Universität Düsseldorf

Vorgelegt von: cand. psych. Thomas Seppelfricke
Butzbacher Weg 5
40229 Düsseldorf
Matrikelnr. 169 5728
Thomas.Seppelfricke@uni-duesseldorf.de

Eingereicht im: November 2008

Prüfer: Prof. Dr. Jochen Musch
Prof. Dr. Axel Buchner

Fakultät: Mathematisch-Naturwissenschaftliche Fakultät

Danksagung

Zuerst möchte ich mich bei allen Studentinnen und Studenten der Heinrich-Heine-Universität Düsseldorf, der Fachhochschule Düsseldorf und der Bergischen Universität Wuppertal für die Teilnahme an der vorliegenden Untersuchung bedanken. Ohne sie wäre es nicht möglich gewesen, innerhalb eines überschaubaren zeitlichen Rahmens eine vierstellige Stichprobe zu befragen.

Meinem Betreuer Herrn Prof. Jochen Musch danke ich herzlich dafür, mir in allen Phasen der Diplomarbeit mit Rat und Tat zur Seite gestanden und mir das Equipment seiner Arbeitseinheit zur Lagerung der Fragebogen und zur Dateneingabe zur Verfügung gestellt zu haben. Herrn Prof. Axel Buchner danke ich für die frühzeitige und bereitwillige Zusage zur Übernahme der Zweitbegutachtung.

Meinen Kommilitoninnen und Kommilitonen Nataly Melanchuk, Marina von Cube, Till Beushausen, Johannes Weber, Andreas Michalowsky und Dirk Rampoldt danke ich für ihre tatkräftige Unterstützung beim Verteilen der Bogen in einigen der zahlreich besuchten Veranstaltungen. Der Diplomandin Christina Benesch danke ich für den begleitenden inhaltlichen Austausch.

Zu guter Letzt danke ich meiner Freundin und meiner Familie für die herzliche Unterstützung in allen Phasen des Schreibprozesses der letzten Monate.

Inhaltsverzeichnis

ZUSAMMENFASSUNG	6
1 EINLEITUNG	7
1.1 DAS PROBLEM DER ANTWORTVERZERRUNGEN BEI SELBSTAUSKÜNFTE	8
1.2 METHODEN DER INDIREKTEN BEFRAGUNG	8
1.2.1 <i>Die Randomized-Response-Technik</i>	8
1.2.1.1 Validierung der Randomized-Response-Technik	9
1.2.1.2 Verbesserung der RRT durch Verweigererdetektions-Modelle	9
1.2.2 <i>Die klassische Form der Unmatched-Count-Technik</i>	10
1.2.2.1 Validierungsstudien zur Unmatched-Count-Technik	12
1.2.2.2 Die Problematik des hohen Standardfehlers	14
1.2.2.3 Die Doppellistenteknik	15
1.2.3 <i>Das Balanced-Incomplete-Block-Design</i>	15
1.2.3.1 Grundlegende Eigenschaften eines BIBDs	16
1.2.3.2 Ein Beispiel für ein Balanced-Incomplete-Block-Design	17
1.3 PRÄVALENZRATENSCHÄTZUNGEN MITTELS MULTINOMIALER MODELLE	19
1.3.1 <i>Was sind multinomiale Modelle?</i>	19
1.3.2 <i>Ein Beispiel für ein einfaches multinomiales Modell</i>	19
1.3.3 <i>Identifizierbarkeit eines multinomialen Modells</i>	21
1.3.4 <i>Parameterschätzung für ein multinomiales Modell</i>	21
1.3.5 <i>Anpassungsgüte eines multinomialen Modells</i>	22
1.3.6 <i>Auswahl der Prüfstatistik für ein multinomiales Modell</i>	23
1.3.7 <i>Teststärkeanalyse bei multinomialen Modellen</i>	23
1.3.8 <i>Bedeutung der multinomialen Modellierung im Rahmen dieser Arbeit</i>	23
1.4 VALIDIERUNG DES BALANCED-INCOMPLETE-BLOCK-DESIGNS	24
1.5 ÜBERBLICK ÜBER DAS THEMENGEBIET „AKADEMISCHE UNEHRlichkeit“	26
1.5.1 <i>Formen des Schummelns im schulisch-akademischen Kontext</i>	27
1.5.1.1 Plagiarismus als Begleiterscheinung der ubiquitären Internetpräsenz	27
1.5.1.2 Interventionsmöglichkeiten in Bezug auf den Plagiarismus	28
1.5.2 <i>Geschlechtsbezogene Aspekte des Schummelns</i>	28
1.6 FORSCHUNGSHYPOTHESEN DER VORLIEGENDEN ARBEIT	29

2	METHODISCHES VORGEHEN	31
2.1	VERSUCHSPLAN	31
2.1.1	<i>Auswahl eines geeigneten Teilverschachtelungsdesigns</i>	31
2.1.2	<i>Auswahl der nicht-sensitiven Füll- bzw. Kontrollfragen</i>	34
2.1.3	<i>Auswahl der sensitiven Fragen</i>	35
2.1.4	<i>Eine zusammenfassende Übersicht über das Untersuchungsdesign</i>	37
2.1.4.1	Positionierung der sensitiven Fragen in den indirekten Bedingungen	37
2.1.4.2	Begründung der Formalisierung sieben direkter Bedingungen	37
2.2	WESENTLICHE ASPEKTE DER FRAGEBOGENGESTALTUNG	39
2.3	STICHPROBE	40
2.3.1	<i>Beschreibung der Stichprobe</i>	40
2.3.2	<i>Planung der Stichprobengröße</i>	42
2.4	VERSUCHSDURCHFÜHRUNG	42
2.4.1	<i>Zeitpunkt der Datenerhebung</i>	43
2.4.2	<i>Ablauf der Datenerhebung</i>	43
2.4.3	<i>Dateneingabe</i>	44
2.4.3.1	Aufbau der Datenmaske	44
2.4.3.2	Kriterien für den Ausschluss von Fragebogen	45
2.4.3.3	Nacherhebungen	46
2.5	DIE DATENAUSWERTUNGS-SOFTWARE HMMTREE	46
3	ERGEBNISSE	47
3.1	KONZEPTION DER AUSWERTUNGS-RELEVANTEN FILES	47
3.1.1	<i>Generierung des Datenfile</i>	47
3.1.2	<i>Generierung des Gleichungsfile</i>	49
3.1.3	<i>Auftretende Modellparameter in der eigenen Studie</i>	52
3.2	VORGEHEN BEI DER AUSWERTUNG MIT HMMTREE	53
3.2.1	<i>Kennwerte des zugrunde liegenden multinomialen Modells</i>	53
3.2.2	<i>Parameterrestriktionen des Grundmodells</i>	54
3.2.3	<i>Betrachtung der Standardfehler für die Parameter des Grundmodells</i>	56
3.2.4	<i>Hypothesenkonforme Parameterrestriktionen</i>	58
3.2.4.1	Grundsätzliches Vorgehen bei der Parameterrestriktion	58
3.2.4.2	Die Ergebnisse der sechs Parameterrestriktionen im Einzelnen	59
3.3	KLASSISCHE AUSWERTUNG DER UNMATCHED COUNT-BEDINGUNGEN	61

3.4	GESCHLECHTSSPEZIFISCHE AUSWERTUNG	63
3.4.1	<i>Deskriptive Betrachtung der geschlechtsspezifischen Parameterschätzwerte</i>	63
3.4.2	<i>Interferenzstatistische Absicherung der auffälligen Geschlechterdifferenzen</i>	65
4	DISKUSSION	67
4.1	EVALUATION DER ERGEBNISSE DER EIGENEN UNTERSUCHUNG	67
4.1.1	<i>Abgleich der eigenen Prävalenzratenschätzungen mit denen der Literatur</i>	67
4.1.2	<i>Potentielle Gründe für die hohen PR-Schätzungen in der direkten Befragung</i>	68
4.1.3	<i>Rückbezugnahme auf die bisherige Studie zum BIBD</i>	70
4.1.4	<i>Interpretation der Fehlanpassung des multinomialen Grundmodells</i>	71
4.2	ASPEKTE DER EIGENEN STUDIE MIT OPTIMIERUNGSPOTENTIAL	72
4.2.1	<i>Verstärkung der Hinweise zur korrekten Anwendung von UCT und BIBD</i>	72
4.2.2	<i>Auswahl einer homogenen Stichprobe</i>	73
4.2.3	<i>Verringerung der inhaltlichen Salienz der kritischen Fragen</i>	73
4.2.4	<i>Explizitere Kennzeichnung des Bezugszeitrahmens der kritischen Fragen</i>	74
4.2.5	<i>Zusätzliche Erhebung der unkritischen Fragen in den direkten Bedingungen</i>	74
4.2.6	<i>Stärkere Einbindung von Dozenten in den Datenerhebungsprozess</i>	75
4.2.7	<i>Die Online-Umfrage als Alternative zur Papier und Bleistift-Befragung</i>	75
4.3	WEITERENTWICKLUNGEN DER INDIREKTEN BEFRAGUNGSTECHNIKEN	77
4.4	ANFORDERUNGEN AN DIE ZUKÜNFTIGE UMFRAGEFORSCHUNG	78
4.5	UMGANG MIT SENSIBLEN DATEN: BESTANDSAUFNAHME UND AUSBLICK	80
4.5.1	<i>Bedeutung und Hauptprinzipien des Datenschutzes</i>	81
4.5.2	<i>Datensparsamkeit als Schlüsselaspekt der indirekten Befragungsverfahren</i>	83
4.5.3	<i>Quo vadis, Datenschutz?</i>	83
	LITERATURVERZEICHNIS	85
	EIGENSTÄNDIGKEITSERKLÄRUNG	91
	ANHANG	92
	ANHANG A: FRAGEBOGEN DER BEDINGUNGEN 1 BIS 17	93
	ANHANG B: DATENFILE DER GESAMTAUSWERTUNG	104
	ANHANG C: GLEICHUNGSDATENFILE DER GESAMTAUSWERTUNG	105
	ANHANG D: MULTINOMIALE SIGNIFIKANZTESTS DER GESAMTAUSWERTUNG	109

Zusammenfassung

Antwortverzerrungen infolge sozialer Erwünschtheit bei Selbstauskünften stellen im Zuge der Erfragung sensibler Themen ein Problem dar. Um die Verfälschungsrate zu minimieren und somit näher an die „wahren“ Prävalenzraten kritischer Verhaltensweisen heranzukommen, ist die Verwendung indirekter Befragungsmethoden hilfreich, bei denen einer Person mehr Anonymität gewährt wird als bei der direkten Befragung. Zu den bisher eingesetzten indirekten Befragungsverfahren zählen die Randomized Response-Technik (RRT) und die klassische Unmatched Count-Technik (UCT). Ein wesentlicher Nachteil der klassischen UCT besteht jedoch darin, dass sie hohe Standardfehler generiert. In der vorliegenden Studie sollten direkte Befragung, die klassische Fassung der UCT sowie eine modifizierte Variante der UCT, ein sog. Balanced-Incomplete-Block-Design (BIBD), auf kritische Fragen, die den Bereich des Schummelns im schulisch-akademischen Bereich betreffen, angewendet werden. Die Parameterschätzungen erfolgten sowohl für die klassische UCT als auch für das BIBD über eine multinomiale Modellierung. Es wurde eine Papier und Bleistift-Befragung an einer vierstelligen Stichprobe der Heinrich-Heine-Universität Düsseldorf, der FH Düsseldorf bzw. der Bergischen Universität Wuppertal durchgeführt. Gezeigt werden konnte, dass das BIBD im Vergleich zur klassischen UCT die Standardfehler deutlich zu reduzieren imstande war. Zwar erbrachte die konventionelle Auswertung der klassischen UCT über einen Mittelwerts-Differenzen-Test signifikant höhere Basisratenschätzungen als die direkte Befragung. Dem gegenüber blieben jedoch im Zuge der multinomialen Auswertung sowohl die klassische UCT als auch das BIBD den Nachweis schuldig, höhere Prävalenzratenschätzungen für die sensiblen Fragen aufdecken zu können. Kontexteffekte, die das Setting der Befragung in den Mittelpunkt stellen, und die Notwendigkeit einer Heterogenitätsannahme für die multinomiale Modellierung zukünftiger BIBD-Anwendungen werden diskutiert.

1 Einleitung

Selbstauskünfte von Befragten stellen innerhalb der Sozialwissenschaften die dominierende bzw. oftmals sogar ausschließlich verwendete Datenquelle dar. Hierbei verfolgen Umfrageforscher grundsätzlich das Ziel, über Selbstauskünfte möglichst wahrheitsgetreue Information von den Befragten zu erlangen. Dies ist jedoch insbesondere dann nicht leicht zu erreichen, wenn der Befragungsinhalt durch sensible Merkmale gekennzeichnet ist. Unter sensiblen Merkmalen werden Verhaltensweisen verstanden, die als sozial unerwünscht gelten, und deren Zugeständnis die befragte Person in ein schlechtes Licht zu rücken drohen. Als sensitive Befragungsgebiete gelten zum Beispiel Drogenmissbrauch, Doping, sexuelle Belästigung oder Betrugsdelikte in der Wirtschaft¹. Corstange (2006, S. 1) verweist darauf, dass Menschen – abgesehen von der eher unbedeutenden Genugtuung, die aus der Befolgung der Norm „möglichst immer die Wahrheit zu sagen“ resultiert – kaum einen persönlichen Nutzen davontragen, wenn sie ehrlich antworten.

Nachdem lange Zeit versucht worden war, soziale Erwünschtheit als einheitliches Konstrukt zu erfassen, differenzierte erstmals Paulhus (1984) zwischen einer dispositionalen und einer situativen Komponente der sozialen Erwünschtheit. Die grundlegende Überlegung hierbei war es, soziale Erwünschtheit im Sinne einer Person-Situations-Interaktion aufzufassen. Das dispositionale “Self-Deceptive-Enhancement” kennzeichnete Paulhus (1984) als eine unbewusste, selbstwertdienliche Täuschung, die das Ziel verfolgt, das idealisierte Selbstbild zu wahren. Dem gegenüber steht die situative “Impression-Management”-Komponente, deren Einsatz bewusst erfolgt und welche die Intention aufweist, bei anderen Personen einen guten Eindruck zu hinterlassen. Mit dem sog. “Balanced Inventory of Desirable Responding” (BIDR) konzipierte Paulhus (1984) ein Instrument, welches imstande war, den Selbst- und den Fremdtäuschungsaspekt der sozialen Erwünschtheit getrennt zu erfassen. Musch, Brockhaus & Bröder (2002) entwickelten eine deutsche Version des BIDR, wobei sie die übersetzte Version des englischen Originals einer empirischen Itemselektion unterwarfen. Die aus der Itemselektion resultierenden Skalen für Selbst- und Fremdtäuschung, die jeweils 10 Items enthielten, konnten an drei unabhängigen Stichproben erfolgreich kreuzvalidiert werden. Wie Musch et al. (2002, S. 127) betonten, war es „mit Hilfe der vorgelegten Skalen erstmals im deutschsprachigen Raum möglich, Komponenten sozialer Erwünschtheit nicht nur konzeptuell, sondern auch empirisch zu unterscheiden“.

¹ Dalton, Daily & Wimbush (1997), S. 1049

1.1 Das Problem der Antwortverzerrungen bei Selbstauskünften

Antwortverzerrungen resultieren daraus, dass Fragen, die einen potentiell beschämenden, belastenden oder stigmatisierenden Charakter haben, mit höherer Wahrscheinlichkeit durch „unwahrheitsgemäße und ausweichende Antworten“ von den Befragten erwidert werden (Chaudhuri & Mukerjee, 1988, zitiert nach Dalton, Wimbush & Daily, 1994, S. 817). Da der wahre Wert, d.h. die Antworten ohne den Einfluss der sozialen Erwünschtheit, unbekannt ist, lässt sich der Effekt, der auf die Antwortverzerrungen zurückgeht, nur schwer erkennen. Auch die Tatsache, dass den Interviewten die Anonymität ihrer Angaben zugesichert wird, führt bei diesen nicht zu einer merklichen Abnahme ihres Misstrauens. Die klassische Umfrageforschung ermöglicht insofern nur eine Schätzung der unteren Grenze der Prävalenzrate eines sensiblen Merkmals.

1.2 Methoden der indirekten Befragung

Um zu besseren Prävalenzratenschätzungen für sensitive Verhaltensweisen zu gelangen, ist es sinnvoll, anstatt eines direkten Frageformats, auf indirekte Befragungsmethoden zurück zu greifen. Droitcour, Caspar, Hubbard, Parsley, Visscher & Ezzati (1991) verweisen darauf, dass dem Befragten durch eine indirekte Fragetechnik die Möglichkeit gegeben wird, „wahrheitsgemäß zu antworten, ohne eindeutig einem sozial unerwünschten Verhalten überführt werden zu können“ (S. 186). Insofern zeichnen sich indirekte Befragungstechniken dadurch aus, dass sie eine steigende Bereitschaft zu ehrlichen Antworten bewirken und zu valideren Selbstauskünften durch die Reduktion des systematischen Fehlers führen.

1.2.1 Die Randomized-Response-Technik

Bei der Randomized-Response-Technik entscheidet der Ausgang eines Zufallsexperimentes (z.B. ein Würfel), ob der Befragte ehrlich auf eine kritische „Ja-Nein-Frage“ antwortet oder ob er unabhängig vom Frageninhalt durch den Zufallsgenerator aufgefordert wird, die Frage zu bejahen. Hierbei behält der Befragte den Ausgang des Zufallsexperimentes für sich, so dass es dem Fragensteller nicht möglich ist, einzelne Individuen mit ihren Antworten in Verbindung zu bringen. Werden die Daten allerdings über viele Individuen hinweg aufsummiert, kann – vorausgesetzt, die Verteilung der Zufallsvariablen ist bekannt – dennoch ermittelt werden, wie viele Befragte im Schutze der Zufallsverschlüsselung das sensible Merkmal zugegeben haben.²

² Bei der ursprünglichen Version der Randomized-Response-Technik nach Warner (1965) entscheidet ein Zufallsgenerator, ob der Befragte die kritische Frage (z.B. „Haben Sie schon einmal Steuern hinterzogen?“) oder

1.2.1.1 Validierung der Randomized-Response-Technik

Eine erfolgreiche Validierung der Randomized-Response-Technik stammt von Wimbush & Dalton (1997), die eine Randomized-Response-Bedingung ($n = 196$) mit einer direkten Befragungsbedingung ($n = 210$) für das sensitive Themengebiet des Diebstahls am Arbeitsplatz kontrastierten. Hierbei erhielten die Befragten in der RR-Bedingung die folgende Instruktion:

If your coin flip is a head OR if you are / were involved in the theft from your employer of from \$5.00-\$9.99 in cash, supplies, or merchandise in a month, please put an "X" in the box to the right. Otherwise do not mark the box; just go to the next question (Wimbush & Dalton, 1997, S. 758).

Die Schätzung der Basisrate für ein begangenes Diebstahlverhalten erfolgte in der RR-Bedingung derart, dass ein Eintrag auf individueller Ebene keinen Rückschluss darüber zuließ, ob die jeweils befragte Person tatsächlich ein Diebstahlverhalten begangen hatte oder „bloß“ den Ausgang ihres Münzwurfes – mit dem Ereignis „Kopf“ – berichtete. Auf aggregativer Ebene konnten die Versuchsleiter die Summe der Einträge aller Befragten ermitteln. Ausgehend von der Wahrscheinlichkeit, dass 50 % der Münzwürfe das Ereignis „Kopf“ zum Ausgang haben, stellte nun gewissermaßen jeder – die 50 % überschreitende – zusätzliche Eintrag ein Eingeständnis begangenen Diebstahlverhaltens dar. Die Signifikanztestung erfolgte jeweils über einen Mittelwerts-Differenzen-Test. Während im Zuge der direkten Befragung nur 12,9 % der Befragten einen Diebstahl in der Größenordnung zwischen 5 und 9.99\$ einräumten, lag die Schätzung der Zugeständnisrate mit 26,1 % in der RR-Bedingung signifikant höher ($z = 3.25$; $p < .001$). Zahlreiche weitere Validierungsstudien zur RRT kamen zu vergleichbaren Ergebnissen wie Wimbush et al. (1997).

1.2.1.2 Verbesserung der RRT durch Verweigererdetektions-Modelle

Eine bedeutsame Modifikation der Randomized-Response-Technik wurde von Musch, Bröder & Klauer (2001) durch Implementierung eines multinomialen Modells zur Entdeckung von Verweigerern vorgenommen. Als Verweigerer galten hierbei solche Teilnehmer, die sich nicht an die Instruktionen der RRT hielten und trotz der Aufforderung, unabhängig vom Frageninhalt mit „Ja“ zu antworten, das Vorhandensein des kritischen Merkmals bestritten.

aber die verneinte Version der kritischen Frage (z.B. „Haben Sie noch nie Steuern hinterzogen?“) beantworten muss. In Abgrenzung zu Warners Originalmodell wird die oben dargestellte Modifikation der RRT in der Literatur zumeist als „Forced-Response“-Modell bezeichnet.

Die Detektion der Verweigererrate ist insofern bedeutsam, dass die RRT die Prävalenzrate des sensiblen Merkmals unterschätzt, sofern es sich bei den Regelverweigerern um Träger des kritischen Merkmals handelt. Vermutlich besteht ein bedeutsames Motiv der Verweigerung darin, dass eine vom Zufallsgenerator erzwungene Bejahung einer kritischen Frage auch Nichtmerkmalsträger in den Verdacht der Merkmalsträgerschaft führt. Indem der Befragte jedoch grundsätzlich verneint, kann er den Verdacht, Merkmalsträger zu sein, schon a priori von sich weisen. Ostapczuk (2008) versuchte die Verweigerungsdetektionsvariante der RRT von Musch et al. (2001) dadurch zu optimieren, dass vom Zufallsgenerator nun nicht mehr nur Ja-Antworten, sondern auch Nein-Antworten erzwungen wurden. Hierdurch konnte der Verweigereranteil wirksam reduziert werden, da sich Nicht-Merkmalsträger fortan weniger in der Gefahr sahen, unberechtigtweise des kritischen Merkmals überführt zu werden.

Dalton, Daily & Wimbush (1997) verweisen darauf, dass sich die Randomized-Response-Technik nicht für Befragungsarten eignet, bei denen der Interviewer nicht vor Ort ist und von daher nicht auf Verständnisprobleme im Zuge ihrer Umsetzung eingehen kann – dies ist z.B. bei postalischen oder Online-Befragungen der Fall. Einen weiteren Nachteil der RRT sehen Dalton et al. (1997) darin, dass die Befragten die wiederholte Durchführung eines Zufallsexperimentes (wie z.B. eines Münzwurfes) vor Beantwortung der Fragen möglicherweise als mühsam und anstrengend erleben. Der Preis, der für die höhere Anonymität infolge der Zufallsverschlüsselung bei der RRT zu zahlen ist, ist der, dass unweigerlich größere Stichprobenumfänge als bei der direkten Befragung vonnöten sind, um eine ähnlich niedrige Varianz des Schätzers und damit ein vergleichbares Konfidenzintervall für die Prävalenzratschätzung zu erhalten. Ostapczuk (2008, S. 13) relativiert letzteren Einwand jedoch damit, dass der erhöhte Probandenbedarf „von einer höheren Validität mehr als kompensiert werden kann“.

1.2.2 Die klassische Form der Unmatched-Count-Technik

Im Jahre 1984 stellte Miller mit der Unmatched-Count-Technik³ eine weitere Methode vor, mittels derer den Befragten eine höhere Anonymität bei der Beantwortung sensibler Fragen gewährt wird. Hierbei werden die Teilnehmer randomisiert auf zwei Gruppen aufgeteilt, wobei beide Gruppen einen Fragenblock erhalten; dieser besteht in der Kontrollgruppe aus fünf unkritischen Fragen und in der Experimentalgruppe aus denselben fünf Fragen zuzüglich einer kritischen Frage, die sich allesamt mit „Ja“ oder „Nein“ beantworten lassen. Die Aufgabe der Befragten besteht darin, die Fragen jeweils im Stillen für sich zu beantworten

³ In der Literatur findet sich diese häufig auch unter der Bezeichnung Randomized-List-Technik.

und im Anschluss daran lediglich Auskunft über die Summe ihrer Ja-Antworten innerhalb des Fragenblocks zu geben. Die Angabe über die Summe der Ja-Antworten liefert dem Fragensteller keine Information darüber, welche Frage(n) die teilnehmende Person bejaht hat. Obwohl die individuellen Antworten der Befragten auf die einzelnen Fragen geschützt bleiben, ist es auf aggregativer Ebene möglich, eine Schätzung der Basisrate für das sensitive Verhalten vorzunehmen, indem die Differenz aus dem Mittelwert der Experimentalgruppe und dem Mittelwert der Kontrollgruppe gebildet wird:

$$\hat{p} = MW_{EG} - MW_{KG}$$

EG = Experimentalgruppe (5 harmlose + *eine* kritische Frage) ; KG = Kontrollgruppe (5 harmlose Fragen)

Ein Beispiel für die klassische Form der Unmatched-Count-Technik nach Miller findet sich in Tabelle 1. Chaudhuri & Christofides (2007) weisen daraufhin, dass möglichst niedrig prävalente unkritische Fragen ausgewählt werden sollten, um zu vermeiden, dass eine Person der Experimentalgruppe in die Situation kommt, durch Angabe des Maximalscores – hier: sechs – die Bejahung der kritischen Frage unweigerlich zu offenbaren.

Tabelle 1

Ein Beispiel für die klassische Variante der Unmatched-Count-Technik nach Miller (1984)

<i>Experimentalbedingung</i>	<i>Kontrollbedingung</i>
A) Haben Sie in Ihrem Haushalt die Möglichkeit, eine Mikrowelle zu benutzen? B) Gehen Sie durchschnittlich im Monat mindestens einmal in ein Restaurant? C) Sind Sie im Dezember geboren? D) Waren Sie schon einmal in London? E) Können Sie von zu Hause aus ins Internet? F) Haben Sie schon einmal heimlich Emails oder SMS-Nachrichten gelesen, die für Ihre(n) Freund/ Freundin und nicht für Sie bestimmt waren?	A) Haben Sie in Ihrem Haushalt die Möglichkeit, eine Mikrowelle zu benutzen? B) Gehen Sie durchschnittlich im Monat mindestens einmal in ein Restaurant? C) Sind Sie im Dezember geboren? D) Waren Sie schon einmal in London? E) Können Sie von zu Hause aus ins Internet? ---
$SI = A+B+C+D+E+F$	$S2 = A+B+C+D+E$

Coutts & Jann (2008) konnten im Rahmen einer Online-Befragungsstudie zeigen, dass die Instruktionen der UCT von 93 % der Befragten im Gegensatz zu nur 85 % bei der RRT⁴ als verständlich eingestuft wurden. Die Ursache für die höhere Verständlichkeit der UCT sahen die Autoren in ihrer größeren Augenscheinvalidität begründet. Des Weiteren verwiesen Coutts & Jann (2008) darauf, dass die höhere Verständlichkeit der UCT einen Zeitvorteil mit sich brachte: so lag der Median für das Lesen und Umsetzen der UCT bei 116 Sekunden im Vergleich zu durchschnittlich 153 Sekunden im Falle der RRT⁵.

Um eine Entscheidung darüber zu treffen, inwieweit die UCT bezogen auf eine sensitive Frage eine höhere Basisratenschätzung als die direkte Befragung hervorzubringen imstande ist, ist es vonnöten, eine Signifikanztestung vorzunehmen. Wie im Zuge der Darstellung der bisherigen Validierungsstudien zur UCT gesehen werden kann, wurde die Signifikanztestung von den Autoren der jeweiligen Studien unterschiedlich angegangen. Die meisten Autoren favorisierten den „Difference of Proportion Test“, wobei dessen Anwendung insofern nicht ganz unbedenklich ist, dass die in ihn eingehenden Proportionen Schätzwerte (resultierend aus der Differenzbildung zweier Proportionen) und nicht „reine“ beobachtete Häufigkeiten sind. Dennoch hat er sich am ehesten für die Signifikanztestung im Zuge der UCT bewähren können.

1.2.2.1 Validierungsstudien zur Unmatched-Count-Technik

Einige der Validierungsstudien, die zur Unmatched-Count-Technik bisher durchgeführt worden sind, sollen in der Folge kurz skizziert werden: Dalton et al. (1994) führten eine Befragung von 240 professionellen Auktionären durch mit dem Ziel, den Anteil derjenigen Auktionäre zu bestimmen, die in der Vergangenheit im Zuge von Versteigerungen illegale Geschäftshandlungen verübt hatten. Hierbei wurden die Teilnehmer randomisiert einer direkten Befragungsbedingung beziehungsweise zwei verschiedenen Unmatched-Count-Bedingungen (jeweils mit einem Stichprobenumfang von $n = 80$) zugeteilt. In der direkten Befragungsbedingung mussten die Probanden sechs Fragen zu verschiedenen „kriminellen“ Auktionärshandlungen jeweils mit „Ja“ oder „Nein“ beantworten; zusätzlich wurden ihnen zwei unkritische Kontrollfragen – ebenfalls aus dem Auktionsbereich stammend – gestellt. In den beiden Unmatched-Count-Bedingungen wurden die Teilnehmer nacheinander mit acht verschiedenen Fragenblöcken konfrontiert, wobei beide Gruppen abwechselnd Blöcke, die

⁴ Coutts & Jan (2008) operationalisierten die RRT über fünf verschiedene Randomisierungsprozeduren (Münzwurf per Hand, elektronischer Münzwurf, Endziffern der eigenen Telefonnummer u.a.); die 85 % stellten das arithmetische Mittel der Verständlichkeitseinschätzungen dieser fünf Randomisierungsarten dar.

⁵ Der angegebene Median für die RRT entspricht hier wiederum dem arithmetischen Mittel der Mediane der fünf Randomisierungsprozeduren.

„nur“ fünf harmlose Fragen enthielten, beziehungsweise Blöcke, in denen zusätzlich jeweils eine kritische Frage enthalten war, vorgelegt bekamen. Auf diese Weise stellte jede der beiden UCT-Gruppen wechselweise Kontroll- bzw. Experimentalbedingung dar. Ebenso wie in der direkten Befragungsbedingung wurden auch im Zuge der UCT-Gruppen zwei Kontrollfragen eingesetzt, um zu dokumentieren, dass in diesem Fall direkte und UCT-Bedingung auf annähernd gleiche Prävalenzratschätzungen kommen würden. Dalton et al. (1994) konnten für fünf der sechs kritischen Verhaltensweisen im „Difference of Proportion Test“⁶ signifikant höhere Prävalenzratschätzungen im Falle der Verwendung der UCT im Vergleich zur „konventionellen“ direkten Befragung vorweisen. Beispielhaft sei hier die Basisratschätzung für die sensitive Frage „Haben Sie sich schon einmal ein fiktives Phantomgebot im Rahmen einer Auktion zunutze gemacht?“ angeführt, die im Falle der UCT bei 39,9 % (95 %-KI: 32,2 % - 47,6 %) im Vergleich zu nur 13,8 % bei der direkten Befragung lag ($z = 3.73$; $p < .001$). Wie von den Autoren vermutet worden war, ergaben sich für die Kontrollfragen keine signifikant verschiedenen Basisraten für die direkte bzw. die UCT-Bedingung. Für die Frage „Benutzen Sie Audio- oder Videoaufnahmen im Zuge der Versteigerungen?“ lag die Prävalenzratschätzung der UCT bei 48,9 % (95 %-KI: 41,1 % - 56,7 %) gegenüber 50,0 % in der direkten Bedingung ($z = -0.14$, n.s.).

Eine zweite Validierungsstudie zur Unmatched-Count-Technik findet sich bei LaBrie & Earleywine (2000), die die UCT auf die Thematik „ungeschütztes Sexualverhalten“ zur Anwendung brachten. Hierzu wurden 346 College-Schüler befragt, die randomisiert drei Gruppen zugeteilt wurden: zum einen einer direkten Befragungsbedingung ($n = 102$) und zum anderen zwei UCT-Bedingungen (jeweils $n = 122$), die wechselweise Kontroll- bzw. Experimentalgruppe darstellten. Jeder Fragenblock enthielt jeweils fünf bzw. in der Experimentalgruppe sechs Items, wobei nacheinander fünf verschiedene Fragenblöcke dargeboten wurden. Für die Mehrzahl der verwendeten kritischen Items konnten auch LaBrie et al. (2000) bei Anwendung des Binomialtests⁷ signifikant höhere Prävalenzschätzungen für die UCT gegenüber der direkten Befragung vorweisen. So gaben im Zuge der UCT 70 % der Befragten zu, bereits „Sex ohne Kondom gehabt zu haben“, während dies in der direkten Bedingung nur 59 % gestanden ($p < .05$). In Bezug auf das Item „Sex ohne Kondom nach

⁶ Der „Difference of Proportion Test“ basiert auf der folgenden Teststatistik:

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

⁷ Der Binomialtest testet zwei unabhängige Binomialwahrscheinlichkeiten auf ihre Gleichheit ($p_1 = p_2$).

vorherigem Alkoholkonsum gehabt zu haben“ zeigte sich im Binomialtest ein noch deutlicherer Unterschied zwischen direkter (36 %) und indirekter (65 %) Befragung ($p < .001$). Anzumerken bleibt an dieser Stelle, dass die Basisratenschätzung für das Item „Alkoholkonsum bis zum Vollrausch“ in der Studie von LaBrie et al. (2000) für die direkte Befragung höher ausfiel als für die UCT (77 % vs. 70 %; $p = .186$).

In einer dritten Validierungsstudie von Anderson, Simmons, Milnes & Earleywine (2007), die 454 College-Studenten im Hinblick auf potentielle essstörungsbedingte Handlungen befragten, zeigten sich ebenfalls überwiegend signifikante Unterschiede zwischen direkter Befragung und UCT-Bedingung⁸; diese fielen aber im Vergleich zu den vorherigen Studien wesentlich ambivalenter aus. So ergab sich zwar erwartungsgemäß eine signifikant höhere UCT-Basisratenschätzung in Bezug auf die Benutzung von Abführmitteln (11 % vs. 2 %, $X^2 = 12.4$, $p < .001$) für die weiblichen Befragten ($n = 279$) bzw. im Falle der männlichen Befragten ($n = 175$) eine höhere UCT-Prävalenzratenschätzung für die Umsetzung strikter Diäten (13 % vs. 6 %, $X^2 = 15.5$, $p < .001$). Andererseits aber führte die UCT bei beiden Geschlechtern zu signifikant niedrigeren Basisratenschätzungen als die direkte Befragung in Bezug auf die Angst vor Gewichtszunahme (Frauen: 53 % vs. 66 %, $X^2 = 12.4$, $p < .001$; Männer: 25 % vs. 33 %, $X^2 = 25.4$, $p < .001$) sowie die Idee, dass Andere ihrem Gewicht und ihrer Figur sehr viel Beachtung schenken würden (Frauen: 80 % vs. 86 %, $X^2 = 15.5$, $p < .001$; Männer: 43 % vs. 72 %, $X^2 = 38.1$, $p < .001$).

1.2.2.2 Die Problematik des hohen Standardfehlers

In Bezug auf alle bisherigen Validierungsstudien zur UCT bleibt zu konstatieren, dass die Prävalenzratenschätzungen der sensiblen Merkmale wegen der durch die Zusatzfragen erzeugten Varianz immer nur mit einem erheblichen Standardfehler angegeben werden konnten⁹. Die UCT ist somit zwar imstande den systematischen Fehler, nicht aber den Zufallsfehler zu reduzieren. Droitcour et al. (1991) demonstrierten, dass Probanden Fragenblöcke, in denen neben der kritischen Frage vier oder fünf unkritische Fragen eingesetzt wurden, gegenüber Fragenblöcken, in denen nur eine oder zwei „füllende“ harmlose Fragen vorhanden waren, präferierten. Der Kompromiss, den die UCT zu bewältigen hat, ist folglich der, sich auf der einen Seite nicht zu vieler unkritischer Fragen zu

⁸ Die Signifikanztestung erfolgte in dieser Studie über den Pearson'schen Chi-Quadrat-Test.

⁹ Es ist kritisch anzumerken, dass in den meisten UCT-Studien (z.B. Dalton et al., 1994; LaBrie et al., 2000) den erheblichen Standardfehlern, die die UCT mit sich bringt, nur begrenzte Aufmerksamkeit geschenkt wird. So werden zwar in den meisten Fällen Konfidenzintervalle angegeben, aber eine explizite Erwähnung der enormen zusätzlichen Varianz, die im Zuge der Anwendung dieser Technik entsteht, bleibt vielmals aus.

bedienen, da diese zusätzliche Varianz generieren und somit die Effizienz der Basisratenschätzung des sensiblen Merkmals verschlechtern, aber andererseits auch nicht zu wenige harmlose Fragen zu verwenden, da ansonsten die von Seiten der Befragten subjektiv wahrgenommene höhere Anonymität, die die UCT zu leisten imstande ist, in Mitleidenschaft gerät. Einen guten Kompromiss, der auch in der eigenen Untersuchung Anwendung finden soll, gehen Kane, Craig & Wald (2004) ein, die innerhalb ihrer Anwendung der Randomized List-Technik in der UCT-Kontrollgruppe vier bzw. in der UCT-Experimentalgruppe fünf Items verwenden.

1.2.2.3 Die Doppellistenteknik

Droitcour et al. (1991) zeigten mit der sog. Doppellisten-Version der UCT eine Möglichkeit auf, um die Effizienz der Prävalenzratenschätzungen zu verbessern. Bei ihrem Vorgehen wurden für zwei Teilstichproben verschiedene Listen eingesetzt, die unabhängig voneinander nach alt bewährtem Muster – die Kontrollgruppe erhielt nur die harmlosen Fragen, die Experimentalgruppe zusätzlich zu den harmlosen noch eine kritische Frage – zu separaten Basisratenschätzungen für das sensible Merkmal kamen. Hierbei ist hervorzuheben, dass die kritische Frage im Gegensatz zu den unkritischen Fragen in beiden Listen die gleiche war. Im Anschluss wurden die beiden Schätzer gemittelt, um so einen Einzelschätzer zu erhalten. Auf diese Weise konnte „die Varianz der Prävalenzratenschätzung halbiert werden, da indirekte Auskünfte über das sensitive Verhalten nun von doppelt so vielen Befragten vorlagen – nämlich von der ganzen Stichprobe und nicht bloß einer randomisierten Hälfte.“ (Droitcour et al., 1991, S. 189).

$$\hat{p} = \frac{1}{2} [(MW_{1.EG} - MW_{1.KG}) + (MW_{2.EG} - MW_{2.KG})]$$

1. und 2. Kontrollgruppe unterscheiden sich in den 5 harmlosen Fragen; dagegen ist die *eine* kritische Frage in 1. und 2. Experimentalgruppe gleich.

Diese Abänderung der ursprünglichen UCT-Variante beruht auf dem Sachverhalt, dass die Varianz des Schätzers eine Funktion der Stichprobengröße ist und die für die Schätzung der Prävalenzrate des sensiblen Merkmals zur Verfügung stehende Stichprobengröße bei der Doppel-Listen-Variante größer ist.

1.2.3 Das Balanced-Incomplete-Block-Design

Ein bedeutsamer Nachteil, der in den bisherigen Validierungsstudien zur UCT zum Tragen kam, bestand darin, dass für jede kritische Frage, die gestellt werden sollte, ein separater Fragenblock verwendet wurde. So mussten die Befragten in der Studie von Dalton et al.

(1994) beispielsweise 44 Fragen beantworten und insgesamt acht verschiedene Summenscores mitteilen. Dies erforderte von den Probanden einen nicht unerheblichen Zeit- und Rechenaufwand. Bei der Verwendung eines Balanced-Incomplete-Block-Designs, wie es erstmalig von Smith, Federer & Raghavarao (1974) vorgeschlagen wurde, müssen nicht mehr alle Fragen von allen Teilnehmern beantwortet werden. Vielmehr erhalten jeweils Teilmengen von Befragten verschiedene Fragenblöcke (Untersets), innerhalb derer sich sowohl sensitive als auch nicht-sensitive Fragen befinden (van der Linden, Veldkamp & Carlson, 2004). Raghavarao & Federer (1979) schränkten allerdings ein, dass kein Fragenblock ausschließlich sensitive Fragen beinhalten dürfe – denn so ginge die intendierte Uneindeutigkeit der Summenangabe der Befragten verloren. Wohl aber dürften Fragenblöcke mit ausschließlich unkritischen Fragen existieren, wie es für einzelne Untersets innerhalb bestimmter BIBDs der Fall ist. Die Anzahl der Fragenblöcke innerhalb eines BIBD unterscheidet sich in Abhängigkeit davon, welches spezielle BIBD eingesetzt wird.

Da im Falle von BIBDs nicht mehr nur zwei Gruppen – wie in der klassischen Variante der UCT – sondern mehrere Bedingungen unabdingbar sind, ist es im Zuge dieser Methode unweigerlich vonnöten, auf große Stichprobenumfänge zurückzugreifen. In der Studie von Smith et al. (1974) wurde dies ehemals beträchtlich vernachlässigt. So setzten die Autoren für sieben verschiedene Fragenblöcke insgesamt nur 84 Probanden – das heißt, gerade einmal 12 Probanden pro Unterset – ein.

1.2.3.1 Grundlegende Eigenschaften eines BIBDs

Smith & Street (2003) geben zunächst einen Einblick in die Notation eines Balanced-Incomplete-Block-Designs: ein (v, b, r, k, λ) -BIBD¹⁰ stellt ein Set von $v \geq 2$ Elementen (Fragen) und $b > 0$ Blöcken dar, wobei die folgenden Bedingungen erfüllt sein müssen:

1. jeder Block b beinhaltet k Fragen (wobei $v > k > 0$)
2. jede Frage tritt in exakt r Blöcken auf (wobei $r > 0$)
3. jedes Fragenpaar tritt simultan in exakt λ Blöcken auf (wobei $\lambda > 0$).

Die Unvollständigkeit des Designs resultiert daraus, dass die Blockgröße k kleiner ist als die Gesamtzahl der Fragen v , so dass also kein Block sämtliche Fragen beinhaltet. Die Balanciertheit des Designs ergibt sich aus der Konstanz des Parameters λ . Die ein BIBD spezifizierenden Parameter sind nicht grundsätzlich unabhängig voneinander; die zwei zentralen Gleichungen, die die Parameter verbinden, lauten:

¹⁰ In der Literatur findet sich für BIBDs häufig die kürzere Darstellung (v, k, λ) .

- a) $b_k = vr$
 b) $\lambda(v-1) = r(k-1)$

Aufgrund der symmetrischen Anforderungen, die ein BIBD aufzuweisen hat, ist die Anzahl überhaupt existierender BIBDs begrenzt. Wie van der Linden et al. (2004, S. 317) betonen, „stellt die Kalkulation eines optimalen BIBDs ein kombinationstheoretisches Problem dar“. Eine Übersicht über die verschiedenen, existierenden BIBDs soll im Zuge der eigenen Auswahlentscheidung für ein bestimmtes BIBD an späterer Stelle (Kap. 2.1.1) erfolgen.

1.2.3.2 Ein Beispiel für ein Balanced-Incomplete-Block-Design

Zur Veranschaulichung soll in der Folge ein konkretes BIBD betrachtet werden, nämlich das (6, 3, 2)-Design¹¹. Tabelle 2 gibt einen Überblick über die 10 verschiedenen Untersets dieses speziellen Designs. Von den insgesamt sechs verwendeten Fragen sind die Hälfte, also jeweils drei, sensitiv bzw. nicht-sensitiv. Jede Frage tritt in insgesamt fünf verschiedenen Blöcken auf, jedes Fragenpaar in jeweils zwei verschiedenen Blöcken. In jedem Fragenblock befinden sich maximal zwei sensitive Fragen; nur der letzte Fragenblock beinhaltet ausschließlich unkritische Fragen.

Tabelle 2

Operationalisierung eines (6, 3, 2)-Balanced-Incomplete-Block-Designs

Block									
Nr.1	Nr. 2	Nr. 3	Nr. 4	Nr. 5	Nr. 6	Nr. 7	Nr. 8	Nr. 9	Nr. 10
A	A	A	A	A	B	B	B	4	4
B	B	4	5	C	4	5	C	5	5
4	5	C	6	6	6	C	6	C	6

Zur besseren Übersicht sind die drei sensitiven Fragen durch Fettdruck hervorgehoben und mit Buchstaben gekennzeichnet, die drei nicht-sensitiven Fragen sind dagegen mit den Zahlen „4“ bis „6“ durchnummeriert.

Die Performance eines jeden Designs ist immer in Bezug auf die Blockgröße k sowie den Quotienten r/b zu bewerten; letzterer gibt den Anteil der Befragten wieder, die eine bestimmte Frage beantworten. Im Falle des (6, 3, 2)-Designs ergibt sich eine Blockgröße von $k = 3$ und ein Quotient von $r/b = 0,5$. Grundsätzlich gilt, wie Smith & Street (2003, S. 187) hervorheben:

¹¹ Ausführlich gekennzeichnet handelt es sich hierbei um das (6, 10, 5, 3, 2)-BIBD

As k increases so does the variance of the estimate. This appears reasonable as a larger number of attributes per questionnaire would increase the uncertainty attached to the estimation of any one particular attribute. As r/b increases, the variance of the estimate decreases. Again this appears reasonable. The more respondents that answer a particular question, the better the estimate of that attribute.

Ein Problem, welches sich augenscheinlich bei Betrachtung des Versuchsplans des (6, 3, 2)-BIBDs ergibt, besteht darin, dass die Angabe des Maximalscores „drei“ von Seiten eines / einer Befragten unweigerlich die Bekräftigung einer oder gar zweier sensitiver Verhaltensweisen offenlegt. Sind in einem Unterset zwei sensitive Fragen vorhanden, so hat neben des Maximalscores von „drei“ ebenso ein Summenscore von „zwei“ einen belastenden Charakter – weil auch dann klar ist, dass der Befragte zumindest eine der beiden kritischen Fragen bejaht haben muss.

Um das Problem des belastenden Gesamtscores zu umgehen, schlägt Smith (2005, S.2162) „eine Transformation“ vor, die der / die Befragte „im Vorfeld des Berichtens des Gesamtscores“ durchzuführen hat. Hierzu wird der Gesamtscore – unabhängig von dessen Wert – stets in der gleichen Weise modifiziert. Beispielsweise könnte der Ausgang eines Zufallsversuchs (eines Würfelwurfes) immer zum Gesamtscore addiert werden.¹² Auf diese Weise wäre das „Scoring“-Problem zwar nicht gänzlich beseitigt, aber zumindest minimiert und das Vertraulichkeitsempfinden bei den Teilnehmern insgesamt gestärkt. Ein nicht zu unterschätzender Nachteil des Hinzuziehens eines Zufallsexperimentes (hier: des Würfelwurfes) ist allerdings der, dass die Varianz des Erwartungswertes im Zuge dieser Transformation zunimmt. Wird anstatt des Würfelwurfes der Ausgang eines Münzwurfes (Kopf = „1“; Zahl = „2“) verwendet, so würde sich die zusätzliche Varianz zumindest partiell verringern¹³.

Es steht außer Frage, dass jede Möglichkeit, die tatsächliche und subjektive Vertraulichkeit der Befragten zu erhöhen von Vorteil ist. Allerdings räumt Smith (2005) selbst ein, dass das Vorgehen über die von ihr vorgeschlagenen Transformationen das Verfahren an sich erheblich verkompliziert. Weil zudem auch der substantielle Anstieg der Varianz im Zuge dieser Modifikation zu berücksichtigen ist, bleibt resümierend festzustellen, dass von derartigen Abänderungen der UCT im Zuge der eigenen Studie Abstand genommen

¹² Angenommen, der Ausgangsscore eines/ einer Befragten im (6,3,2)-BIBD ist „drei“, so würde nun hierzu der Ausgang eines Würfelwurfes (1, 2, 3, 4, 5, oder 6) hinzuaddiert; auf diese Weise würde nur ein Sechstel der Antwortenden, deren Ausgangsscore „drei“ war, identifiziert – nämlich nur diejenigen, die tatsächlich auf „neun“ (3+6) kommen.

¹³ Anstatt eines Sechstels würden aber im Falle des Münzwurfes 50 % der Probanden als Merkmalsträger des sensitiven Verhaltens demaskiert werden.

wird. An dieser Stelle dürfte bereits ersichtlich sein, dass der Auswahlentscheidung für ein spezifisches BIBD eine bedeutsame Rolle zukommt.

1.3 Prävalenzratenschätzungen mittels multinomialer Modelle

Da bei der Anwendung eines Balanced-Incomplete-Block-Designs nur die Information über den Summenscore der Befragten über mehrere Fragen hinweg zur Verfügung steht, ist es erforderlich, einen Weg aufzuzeigen, wie es trotz dieser begrenzten Information möglich ist, an Schätzungen für die Basisraten der sensitiven bzw. nicht-sensitiven Verhaltensweisen zu gelangen. Eine Auswertung über die Differenz der Mittelwerte, wie sie im Zuge der klassischen UCT durchgeführt werden konnte (vgl. Kap. 1.2.2), ist aufgrund der Teilverschachtelung der Fragen in den unterschiedlichen Bedingungen eines BIBDs nicht weiter möglich. Aus diesem Grund ist es notwendig sich der sog. multinomialen Modellierung zu bedienen, welche in der jüngeren Vergangenheit zunehmend stärker an Einfluss in der psychologischen Forschung, und zwar insbesondere innerhalb der Sozial- und Kognitionspsychologie, gewonnen hat (Stahl & Klauer, 2007).

1.3.1 Was sind multinomiale Modelle?

Multinomiale Modelle sind eine Familie formaler Modelle, die es erlauben, psychologische Prozesse zu messen und zu trennen, die nicht der direkten Beobachtung zugänglich sind. Hierbei ist es von zentraler Bedeutung, dass die psychologischen Prozesse ein begrenztes Set kategorialer (diskreter) Verarbeitungszustände umfassen (Riefer & Batchelder, 1988). Im Zuge der multinomialen Modellierung werden beobachtbare Kategoriewahrscheinlichkeiten p_1, p_2 etc. als Funktionen sog. „latenter“ Parameter $\theta_1, \theta_2, \dots, \theta_s$ und ihrer Komplemente ($1 - \theta_s$) dargestellt. Modelle mit diesen Eigenschaften werden als binäre probabilistische Ereignisbäume („kognitive Verarbeitungsbäume“) bezeichnet, wobei die Baumrepräsentation eine zeitliche Sequenz der latenten Ereignisse suggeriert (Erdfelder, 2000).

1.3.2 Ein Beispiel für ein einfaches multinomiales Modell

In der Folge soll noch einmal Bezug auf das (6, 3, 2)-BIBD genommen werden, wobei der Fokus nun auf den ersten Block dieses Designs gelegt wird (siehe Tabelle 2). Innerhalb des ersten Blocks gilt es für die Befragten die beiden sensitiven Fragen A und B sowie die nicht-sensitive Frage 4 zu beantworten und im Anschluss daran die Summe ihrer Ja-Antworten mitzuteilen. Wie in Abbildung 1 ersichtlich wird, gibt es hierbei insgesamt vier verschiedene (beobachtbare) Antwortkategorien, nämlich „keine“, „eine“, „zwei“ oder „drei“ Ja-Antwort(en).

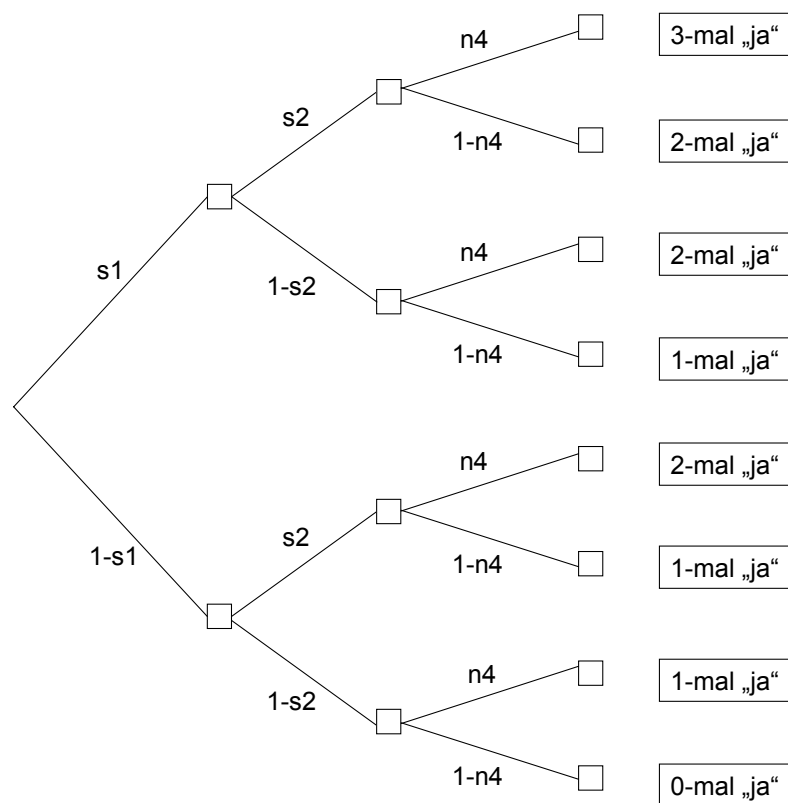


Abbildung 1: Übersicht über das einfache multinomiale Modell der Bedingung 1 im (6, 3, 2)-BIBD. Als Index für die Parameter der sensitiven Fragen A und B werden „s1“ bzw. „s2“, als Index für den Parameter der nicht-sensitiven Frage 4 wird „n4“ verwendet.

Während die Summen „null“ und „drei“ eindeutig darauf schließen lassen, dass die Probanden alle drei Fragen verneint bzw. alle drei Fragen bejaht haben, ergeben sich für die beobachtbaren Kategorien „eins“ und „zwei“ jeweils drei Möglichkeiten für das Zustandekommen der Gesamtsumme; diese sind in der folgenden Übersicht genauer spezifiziert:

$$\begin{aligned}
 1 &= s1 * (1 - s2) * (1 - n4) \\
 1 &= (1 - s1) * s2 * (1 - n4) \\
 1 &= (1 - s1) * (1 - s2) * n4
 \end{aligned}$$

$$\begin{aligned}
 2 &= s1 * s2 * (1 - n4) \\
 2 &= s1 * (1 - s2) * n4 \\
 2 &= (1 - s1) * s2 * n4
 \end{aligned}$$

- s1: Wahrscheinlichkeit, dass der oder die Befragte die sensitive Frage A bejaht
 (1-s1): Wahrscheinlichkeit, dass der oder die Befragte die sensitive Frage A verneint
 s2: W'keit, dass der oder die Befragte die sensitive Frage B bejaht
 (1-s2): W'keit, dass der oder die Befragte die sensitive Frage B verneint
 n4: W'keit, dass der oder die Befragte die nicht-sensitive Frage 4 bejaht
 (1-n4): W'keit, dass der oder die Befragte die nicht-sensitive Frage 4 verneint

1.3.3 Identifizierbarkeit eines multinomialen Modells

Die Voraussetzung für statistische Analysen auf dem Hintergrund eines bestimmten multinomialen Modells ist dessen Identifizierbarkeit. Wie Riefer et al. (1988, S.320) betonen, „ordnet ein substantielles multinomiales Modell jedem kognitiven Ereignis einen Parameterwert zu, der die Wahrscheinlichkeit für das Auftreten dieses Ereignisses repräsentiert“. Erdfelder (2000) verweist darauf, dass sich der Parameterraum Ω eines multinomialen Modells mit einem Parameter als das reelle Intervall $[0,1]$, für zwei Parameter als das Intervall $[0,1]^2$ bzw. allgemein als das Intervall $[0,1]^S$ kennzeichnen lässt. Die notwendige Bedingung für die globale Identifizierbarkeit eines binären multinomialen Modells ist die, dass es nicht mehr Parameter (S) als unabhängige Kategoriewahrscheinlichkeiten $[J(k)]$ aufweisen darf. Die hinreichende Bedingung für die globale Identifizierbarkeit eines binären multinomialen Modells gilt dann als erfüllt, „wenn alle Äste des Modells in unterschiedliche (empirische) Kategorien einmünden und deren Wahrscheinlichkeiten sämtlich positiv sind“ (Hu & Batchelder, 1994, zitiert nach Erdfelder, 2000, S. 94). Die Praxisrelevanz der hinreichenden Bedingung ist allerdings nicht so bedeutsam wie die der notwendigen Bedingung, da multinomiale Modelle mit beobachtbaren Ästen eher die Ausnahme als die Regel darstellen. Eine abgeschwächte Variante der Identifizierbarkeit, die lokale Identifizierbarkeit, ist dann gegeben, wenn ein multinomiales Modell in einer näheren Umgebung von θ_0 identifizierbar ist. Wie Erdfelder (2000) anmerkt, impliziert die globale Identifizierbarkeit eines multinomialen Modells unweigerlich auch seine lokale Identifizierbarkeit; umgekehrt ist dies nicht zwingend der Fall. So kann lokale Identifizierbarkeit zur statistischen Begründung der Parameterschätzung und der Modellprüfung ausreichen, wenn noch einige weitere Regularitätsbedingungen erfüllt sind (vgl. Erdfelder, 2000).

1.3.4 Parameterschätzung für ein multinomiales Modell

Das Ziel der multinomialen Modellierung besteht nun darin, einen Vektor von Parameterwerten $\vec{\theta} = (\theta_1, \dots, \theta_s, \dots, \theta_S)$ zu schätzen, der gewährleistet, eine vorgegebene Distanzfunktion zwischen den *beobachteten* Kategoriehäufigkeiten n_1, n_2, \dots, n_J und den *erwarteten* Kategoriehäufigkeiten bei Modellgültigkeit $N_{p1}(\theta), N_{p2}(\theta), \dots, N_{pJ}(\theta)$ zu minimieren. Häufig verwendete Distanzfunktionen sind die Likelihood-Quotienten-Statistik G^2 und Pearsons χ^2 .¹⁴ Die rechnerische Realisierung erfolgt über den Expectation-

¹⁴ Beide sind Spezialfälle von Read & Cressies' „Potenz-Divergenz-Statistik“ PD^λ ; die Bezeichnung „Potenz-Divergenz-Statistik“ leitet sich daraus ab, dass die Modell-Daten-Divergenz über eine gewichtete Summe von λ -

Maximization-Algorithmus, der den Vorteil hat, dass die geschätzten Parameter immer im Intervall $[0,1]^S$ liegen, jedoch den Nachteil aufweist, dass er nicht zwangsläufig ein globales Maximum findet. Der EM-Algorithmus lässt sich wie folgt skizzieren (angelehnt an Erdfelder, 2000, S. 108 / 109):

1. Wahl eines Startvektors $\vec{\theta}_i$ für die Parameter
2. Schätzung der erwarteten Häufigkeiten für die Äste des multinomialen Modells unter Zugrundelegung von $\vec{\theta}_i$ (Expectation-Schritt)
3. Iterationszähler-Erhöhung (setze $i = i + 1$); Berechnung eines neuen ML-Schätzvektors $\vec{\theta}_i$ unter Zugrundelegung der erwarteten Häufigkeiten aus 2 (Maximization-Schritt)
4. Wenn Änderungsbetrag $\text{Abs}(\vec{\theta}_i - \vec{\theta}_{i-1}) > (\text{Konvergenz-}) \text{ Kriterium } \epsilon$, erneut Schritt 2
5. Wenn Änderungsbetrag $\text{Abs}(\vec{\theta}_i - \vec{\theta}_{i-1}) < \epsilon$, Akzeptanz von θ_i als Parameterschätzer $\hat{\theta}$

Grundsätzlich sollten die resultierenden Schätzwerte erst dann als ML-Schätzungen akzeptiert werden, wenn sie mehrmals – von unterschiedlichen Zufallsstartwerten ausgehend – repliziert werden konnten. In jedem Fall ist es sinnvoll, die Parameterschätzwerte nicht bloß in Form von Punktschätzungen, sondern mittels Konfidenzintervallen anzugeben. Wie Riefer et al. (1988, S. 323) betonen, wird für die Bestimmung von Konfidenzintervallen für die Parameter die „Varianz-Kovarianz-Matrix der Schätzer, die sog. (erwartete) Fisher-Informationsmatrix $\Sigma(\theta)$ “ benötigt.

1.3.5 Anpassungsgüte eines multinomialen Modells

Sobald Parameterschätzungen für das multinomiale Modell errechnet wurden, stellt sich als nächstes die Frage, ob das Modell die (beobachteten) Daten angemessen zu erfassen imstande ist. Da die asymptotisch χ^2 -verteilte Likelihood-Quotienten-Statistik G^2 ohnehin als Nebenprodukt der ML-Parameterschätzung anfällt, sprechen nach Erdfelder (2000) alle Gründe für die Wahl von G^2 , wenn es um die Evaluation der Güte der Anpassung des multinomialen Modells an die erhobenen Daten geht. Ein guter “fit” des Modells wird hierbei dadurch indiziert, dass G^2 bei ausreichend großem N unterhalb des kritischen Wertes c_α der χ^2 -Verteilung liegt. Hierbei ist der kritische Wert c_α jeweils abhängig von der Anzahl der Freiheitsgrade des zugrunde liegenden multinomialen Modells. Eine Darstellung über die Bestimmung der Freiheitsgrade soll für das multinomiale Modell der eigenen Untersuchung an späterer Stelle (Kap. 3.2.1) erfolgen.

Potenzen der Quotienten beobachteter und erwarteter Kategoriehäufigkeiten quantifiziert wird (Read & Cressie, 1988, zitiert nach Erdfelder, 2000, S. 113).

1.3.6 Auswahl der Prüfstatistik für ein multinomiales Modell

Die Likelihood-Quotienten-Statistik G^2 eignet sich ebenfalls als Standard-Prüfstatistik für multinomiale Tests. So bietet sich die „Verfügbarkeit teststarker und einfach berechenbarer ΔG^2 -Differenzentests zur Prüfung spezieller Parameterhypothesen“ an (Erdfelder, 2000, S. 178). Probleme bei der Verwendung von G^2 ergeben sich nur im Falle zu kleiner Stichprobenumfänge, im Falle derer jedoch auf „Korrekturformeln“ zurückgegriffen werden kann. Kleine Stichprobenumfänge sollten aber allein schon deshalb vermieden werden, damit eine ausreichende Teststärke gewährleistet werden kann.

1.3.7 Teststärkeanalyse bei multinomialen Modellen

Nachdem der Teststärkenanalyse im Rahmen multinomialer Modelle lange Zeit keine Beachtung geschenkt wurde, waren Riefer & Batchelder (1990) die ersten, die die Notwendigkeit der Teststärkeanalyse für multinomiale Modelle postulierten. In der Folge wurde unter Rückgriff auf die Cohen'schen Effektstärkekonventionen¹⁵ oftmals eine „Kompromiss-Poweranalyse“ durchgeführt, deren Ziel es war, „eine angemessene Balance zwischen den beiden Fehlerwahrscheinlichkeiten erster und zweiter Art auf möglichst niedrigem Niveau herzustellen“ (z.B. Buchner, Steffens, Erdfelder & Rothkegel, 1997; Erdfelder & Buchner, 1998a; 1998b, zitiert nach Erdfelder, 2000, S. 150). Erdfelder (2000) verweist auf eine Vielzahl von Modifikationen der Teststärkeanalyse für multinomiale Modelle, die an dieser Stelle allerdings nicht weiter ausgeführt werden sollen. Als Richtgröße für die eigene empirische Erhebung wurden $n = 100$ Personen pro Bedingung angesetzt, um so eine Power gewährleisten zu können, die minimal bei 80 % lag.

1.3.8 Bedeutung der multinomialen Modellierung im Rahmen dieser Arbeit

Der Rückgriff auf multinomiale Modelle stellt eine profunde Methode dar, das „Informationsleck“ zu überwinden, welches sich aus der Anwendung indirekter Befragungstechniken ergibt. Es sollte klar geworden sein, dass mittels multinomialer Modellierung zwar nur Schätzungen für die „wahren“ Basisraten generiert werden können, dass aber diese Schätzungen weniger anfällig für Antwortverzerrungen als die Basisratenschätzungen der direkten Befragung sind. Für die eigene Untersuchung wird ein einfaches multinomiales Modell zugrunde gelegt, basierend auf der Annahme, dass die Parameter über verschiedene Personen hinweg *nicht* variieren. Diese Annahme ist nicht

¹⁵ Effektstärkekonventionen nach Cohen (1988), hier exemplarisch für Pearsons χ^2 : kleiner Effekt ($w = 0.10$), mittlerer Effekt ($w = 0.30$), großer Effekt ($w = 0.50$).

trivial, weshalb Stahl et al. (2007) im Rahmen des latenten Klassenansatzes verschiedene Parametervektoren für eine festgelegte Zahl C latenter Klassen zum Einsatz bringen.¹⁶ In der eigenen Untersuchung soll eine derartige Parameterheterogenität nicht berücksichtigt werden.

1.4 Validierung des Balanced-Incomplete-Block-Designs

Im Gegensatz zur klassischen UCT, zu der bereits einige Validierungsstudien vorliegen (vgl. Kap. 1.2.2.1), ist die Validierung eines Balanced-Incomplete-Block-Designs für die experimentelle Umfrageforschung ein bisher noch kaum ergründetes Gebiet. Gerade aus diesem Grund soll in der Folge Bezug auf eine aktuelle Studie von Benesch (2008) genommen werden, innerhalb der ein Vergleich der direkten Befragung, der Randomized Response-Technik sowie eines (6, 3, 2)-BIBDs (vgl. Kap. 1.2.3.2) auf das sensitive Themengebiet der „Missachtung der Privatsphäre des Partners / der Partnerin“ angewendet wurde. Die Darstellung soll in der Folge auf den Vergleich von direkter Befragung und BIBD begrenzt werden, da die RRT im Zuge der eigenen Studie bekanntlich keine Anwendung findet. In der Studie von Benesch (2008) wurden in der direkten Befragung sowie den zehn Bedingungen des (6, 3, 2)-BIBDs jeweils 500 Teilnehmerinnen und Teilnehmer befragt. Insofern gingen in jede BIBD-Bedingung die Antworten von 50 Personen ein. Durch einen Vortest konnte Benesch (2008) bereits im Vorfeld der Haupterhebung sicherstellen, dass sich die verwendeten kritischen Fragen in ihrem Sensitivitätsgrad deutlich unterschieden. Während der explizite Wortlaut der kritischen Fragen der Tabelle 3 zu entnehmen ist, sind die sensitiven Verhaltensweisen im folgenden Kasten schon einmal inhaltlich skizziert:

1. Verhalten mit dem niedrigsten Sensitivitätsgrad:

Das heimliche Ausfindigmachen des Buches, das der Partner / die Partnerin gerade liest

2. Verhalten mit einem mittleren Sensitivitätsgrad:

Das heimliche Lesen einer Postkarte, die der Partner / die Partnerin in seinem / ihrem Zimmer mit dem Foto nach vorne an der Wand aufgehängt hat

3. Verhalten mit dem höchsten Sensitivitätsgrad:

Das heimliche Lesen des vom Partner / der Partnerin an einem sicheren Ort aufbewahrten Tagebuches

Die Auswertung des Vergleichs zwischen direkter Befragung und BIBD erfolgte in dieser Studie über die in Kap. 1.3 dargestellte multinomiale Auswertungsprozedur auf der Grundlage „Maximum Likelihood“-gestützter Parameterschätzungen. Die Ergebnisse des Vergleichs von direkter Befragung und BIBD sind in Tabelle 3 unter Angabe der ΔG^2 -Werte zusammengefasst.

¹⁶ Stahl et al. (2007) sprechen in diesem Fall von „hierarchischen multinomialen Verarbeitungsbaummodellen“.

Tabelle 3

Überblick über die Ausgänge der G²-Differenzentests für die Gleichsetzung der direkten Parameter mit den BIBD-Parametern für jede der drei sensitiven Fragen

Bedingung	Kritische Frage 1		Kritische Frage 2		Kritische Frage 3	
	Würden Sie in Abwesenheit Ihres Partners / Ihrer Partnerin schauen, welches Buch er / sie gerade liest?		Würden Sie heimlich und ohne Erlaubnis eine Postkarte lesen, die Ihr/e Partner/in empfangen und in seinem / ihrem Zimmer mit dem Foto nach vorne an der Wand aufgehängt hat?		Würden Sie heimlich und ohne Erlaubnis das an einem sicheren Ort aufbewahrte Tagebuch Ihres Partners / Ihrer Partnerin lesen?	
Direkte Befragung (n = 500)	% „Ja“ (95 % KI)	79,8 % (76,3 - 83,3 %)	% „Ja“ (95 % KI)	44,4 % (40,0 - 48,8 %)	% „Ja“ (95 % KI)	5,8 % (3,8 - 7,8 %)
	% „Nein“ (95 % KI)	20,2 % (16,7 - 23,7 %)	% „Nein“ (95 % KI)	55,6 % (51,2 - 60,1 %)	% „Nein“ (95 % KI)	94,2 % (92,2 - 96,2 %)
Balanced-Inc-Block-Design (n = 500)	% „Ja“ (95 % KI)	71,6 % (63,7 - 79,5 %)	% „Ja“ (95 % KI)	59,8 % (51,1 - 68,5 %)	% „Ja“ (95 % KI)	22,2 % (13,4 - 30,9 %)
	% „Nein“ (95 % KI)	28,4 % (20,5 - 36,3 %)	% „Nein“ (95 % KI)	40,2 % (31,5 - 48,7 %)	% „Nein“ (95 % KI)	77,8 % (69,1 - 86,6 %)
G ² -Differenz (Veränderung der Modellpassung) ¹⁷	ΔG ² (1): d1 = s1		ΔG ² (1): d2 = s2		ΔG ² (1): d3 = s3	
	4.03 (n.s.)		9.97**		13.89**	

Für den G²-Differenzentest liegt der kritische χ^2 -Wert für eine signifikante Verschlechterung der Passung des Modells bei einem Freiheitsgrad und einem α von 1 % bei 6.63. Von daher ist ersichtlich, dass sich die Parameterschätzungen des BIBDs für die zweite und die dritte kritische Frage signifikant von den jeweiligen direkten Parameterschätzungen unterscheiden (**p < .01).

¹⁷ ΔG² gibt jeweils die Differenz aus dem im Zuge der spezifischen Restriktion erhaltenen G²'s und dem G² des unrestringierten Grundmodells an. Jede der hier aufgeführten Restriktionen ging mit der Minderung genau eines Freiheitsgrades im Vergleich zum Grundmodell einher. Ausführliche Erläuterungen diesbezüglich finden sich in Kap. 3.2.4.1.

Unerwarteter Weise fiel die Prävalenzratschätzung für die erste kritische Frage („Buch“) mit 79,1 % höher aus als die BIBD-Schätzung mit 71,6 %, wobei die Gleichsetzung der Parameter d_1 und s_1 die Signifikanz auf einem Fehlniveau von $\alpha = 1\%$ – verfehlte [$\Delta G^2(1) = 4.03$ (n.s.)]. Dieses Ergebnis ließ vermuten, dass das erfragte Verhalten ein nicht ausreichendes Maß an Sensitivität aufwies, um das indirekte Verfahren erfolgreich – im Sinne einer Aufklärung der Dunkelziffer – einsetzen zu können. Für die beiden weiteren kritischen Fragen lagen die Basisratschätzungen für das BIBD dagegen beträchtlich höher als in der direkten Befragungsbedingung. So erwies sich der BIBD-Schätzwert von 59,8 % für die zweite kritische Frage („Postkarte“) der direkten Bedingung mit 44,4 % ebenso überlegen wie die BIBD-Schätzung der dritten kritischen Frage („Tagebuch“) mit 22,2 % gegenüber nur 5,8 % in der direkten Befragung. Sowohl für die zweite als auch für die dritte kritische Frage ließen sich die höheren BIBD-Basisratschätzungen statistisch signifikant absichern [Gleichsetzung $d_2 = s_2$: $\Delta G^2(1) = 9.97$; $p < 0.01$; Gleichsetzung von d_3 und s_3 : $\Delta G^2(1) = 13.89$; $p < 0.01$].

1.5 Überblick über das Themengebiet „Akademische Unehrllichkeit“

Der experimentelle Vergleich der direkten und indirekten Befragungstechniken soll im Rahmen der eigenen Arbeit auf das Themengebiet des Schummelns im schulisch-akademischen Kontext¹⁸ angewendet werden – ein Problem, das in Zeiten von Leistungsdruck und schwierigen Arbeitsmarktverhältnissen aktuell ist und bleibt. In einem Zeitungsartikel der Frankfurter Allgemeinen Zeitung verweist Volk (2008) darauf, dass in Amerika viele Studenten einen „Eid“ ablegen, nicht zu betrügen. Für die Studentinnen des Sweet Briar College nahe Washington hat dieses Versprechen den folgenden Inhalt:

Sweet Briar women do not lie, cheat, steal or violate the right of others. Therefore, I pledge to uphold all standard of honourable conduct. I will report myself and others for any infraction of this pledge.

Ein derartiger Ehrenkodex hat sich bisher an deutschen Hochschulen noch nicht etablieren können. Allerdings soll in das Landeshochschulgesetz Baden-Württemberg zukünftig eine Vorschrift eingefügt werden, nach der Plagiarismus für die betreffenden Studenten zu hochschulrechtlichen Konsequenzen führen soll.¹⁹ Die Frage, ob ein solcher „Eid“ ein Mittel gegen Faulheit und Unmotivation ist, wird kontrovers diskutiert. Im Folgenden soll ein

¹⁸ In der englischen Fachliteratur hat sich in diesem Zusammenhang der Begriff der „Academic Dishonesty“ eingebürgert.

¹⁹ Frankfurter Allgemeine Zeitung vom 31.10.2008, Rubrik: Inland in Kürze, Titel: Plagiatoren droht Exmatrikulation (S. 4).

Überblick über die wichtigsten Fakten und Erkenntnisse in Bezug auf das Thema „akademische Unehrlichkeit“ gegeben werden.

1.5.1 Formen des Schummelns im schulisch-akademischen Kontext

Belcheir (2003) führte anhand einer Stichprobe von 1.500 Studenten der Boise State University eine qualitative und quantitative Analyse der verschiedenen Unterarten akademischer Täuschungsversuche durch. Mit einer Prävalenzratschätzung von 63 % erwies sich das Abschreiben von Hausaufgaben als die gängigste Form des Schummelns unter den befragten amerikanischen Studenten. Weitere von den Teilnehmern berichtete Arten von Täuschungsversuchen waren das gezielte Ablenken der Prüfungsaufsicht (5 %) und die Kommunikation mit Kommilitonen über eine Geheimsprache während der Prüfung (2 %). Baldwin, Daugherty, Rowley & Schwarz (1996) ermittelten eine Basisratschätzung von 2,9 % für das Absagen einer Prüfung unter Angabe einer nicht wahrheitsgemäßen Begründung. Hrabak, Vujaklija, Vodopivec, Hren, Marusic & Marusic (2004) führten eine Befragung von kroatischen Medizin-Studenten (n = 827) durch, wobei sich diese auf vier Clustern von betrügerischen Verhaltensweisen unterschiedlichen Schweregrades im schulisch-akademischen Kontext stützte. Zu den ernsthaftesten Täuschungsdelikten gehörten den Autoren zufolge das Ausnutzen privater Beziehungen zur Erlangung einer besseren Note (1,6 %) sowie die Bezahlung eines Prüfers, um eine Prüfung zu bestehen (0,7 %).

1.5.1.1 *Plagiarismus als Begleiterscheinung der ubiquitären Internetpräsenz*

Eine an Bedeutung zunehmende Komponente der akademischen Unehrlichkeit stellt in der jüngsten Vergangenheit der Plagiarismus²⁰ dar, der von Seiten des “Council of Writing Program Administrators” (2003) wie folgt definiert wird:

In an instructional setting, plagiarism occurs when a writer deliberately uses someone else’s language, ideas, or other original (not common-knowledge) material without acknowledging its source.

Im Rahmen einer Befragung von Studenten der University of Northumbria ermittelte Dordoy (2002) eine Prävalenz von 11,2 % für die Ausgabe eines aus dem Internet herunter geladenen Essays als das eigene Werk. Als Hauptmotive für ihr plagiatorisches Handeln gaben die Studentinnen und Studenten in der Studie von Dordoy (2002) das Anstreben einer besseren Note (59 %), ein problematisches Zeitmanagement (54 %) sowie die Leichtigkeit des Zugangs zu etwaigem Material über das Internet (40 %) an. Blackburn (1999), zitiert nach

²⁰ Plagiat (vom lat. Wort *plagium*, „Menschenraub“ abgeleitet)

Belcheir (2003, S.6), konnte im Rahmen seiner Studie eine „kurvilineare Beziehung zwischen den Fähigkeiten der Studenten und ihrer Wahrscheinlichkeit zu schummeln“ konstatieren: so war die Neigung zum Schummeln bei Unter- bzw. Überforderung der Studenten signifikant größer als bei adäquater Beanspruchung.

In Bezug auf den juristischen Hintergrund des Plagiarismus verweist Harke (2002), zitiert nach Weber-Wulff (2002, S. 67), darauf, dass unter anderem die ‚freie Benutzung‘ eines Werkes frei von der Zustimmung des Urhebers ist²¹. Diese setze voraus, dass „das Werk eines anderen lediglich als Anregung benutzt (werden dürfe), um ein eigenes selbstständiges Werk zu schaffen“. In jedem Fall müsse der Charakter des verwendeten Werkes „deutlich hinter den eigenständigen Zügen des neu geschaffenen Produktes“ zurücktreten. Ferner „(stelle) auch die Übersetzung eine Bearbeitung (dar), die nur mit Einwilligung des Urhebers veröffentlicht oder verwertet werden (dürfe)“²².

1.5.1.2 Interventionsmöglichkeiten in Bezug auf den Plagiarismus

Es erscheint angesichts der Ubiquität des Internets in der heutigen Zeit zunehmend notwendig, dass sich Lehrkräfte und Professoren Strategien aneignen, um potentielle Plagiatoren zu überführen. Weber-Wulff (2002) verweist in diesem Zusammenhang auf die Arbeitsgruppe an der Universität Berkeley um John Barrey, die eine Plagiarismus-Entdeckungs-Software entwickelt hat, mittels derer über unterschiedliche Algorithmen Dokumente mit im “World Wide Web” befindlichen Arbeiten verglichen werden können. Um plagiatorisches Handeln gar nicht erst aufkommen zu lassen, geben Hart & Friesner (2004) Lehrenden den Rat, Studentinnen und Studenten bezüglich der Bedeutsamkeit akademischer Loyalität zu unterweisen, regelmäßige Workshops zum Erlernen wissenschaftlichen Schreibens anzubieten sowie Aufgabenstellungen von Jahrgang zu Jahrgang zu variieren.

1.5.2 Geschlechtsbezogene Aspekte des Schummelns

Payne & Nantz (1994) konnten empirisch belegen, dass Männer signifikant häufiger als Frauen im akademischen Kontext betrügen. Die höhere Motivation der Männer zu schummeln ist in ihrem Bestreben, den eigenen sozioökonomischen Status zu verbessern, zu sehen. Wie Pashos (2001) betont, bewerten Frauen in Studien, in denen Partnerwahlpräferenzen untersucht werden, regelmäßig Statusmerkmale wie Einkommen und Beruf (und mit ihnen gekoppelte Eigenschaften des Statusstrebens wie Ehrgeiz und Fleiß) höher als Männer. Da gesellschaftliche Macht und Ressourcen traditionell überwiegend von Männern besessen

²¹ Rechtsgrundlage ist hier der § 24 des Urheberrechtsgesetzes (UrhG)

²² Rechtsgrundlage ist hier der § 23 des Urheberrechtsgesetzes (UrhG)

werden, stelle es für Frauen die Hauptmöglichkeit dar, sozial aufzusteigen, wenn sie Männer zum Partner wählen, die Eigenschaften besitzen, die mit hohem Status verknüpft sind (Caporael, 1989, zitiert nach Pashos, 2001).

Hrabak et al. (2004) dokumentierten, dass Männer ferner bereitwilliger sind, ihr – den schulisch-akademischen Kontext betreffendes – betrügerisches Verhalten zu berichten. Evolutionspsychologisch lässt sich das Berichten begangener Täuschungsversuche als Imponiergehabe interpretieren: so versuchen Männer sich gegenüber Frauen mit ihrem Wagemut zu brüsten, um so ihren Fortpflanzungserfolg zu maximieren. Wie Belcheir (2003) konstatiert, haben Frauen tendenziell negativere Einstellungen, was das Schummeln anbetrifft. In jüngster Vergangenheit habe jedoch auch der Anteil weiblicher „Schummler“ zugenommen (McCabe, Trevino & Butterfield, 2001). Die Motivation zum Schummeln scheint geschlechtsbezogen unterschiedlich bedingt zu sein: so gab eine Vielzahl weiblicher Teilnehmerinnen als Beweggrund für ihr Schummeln „den Versuch, anderen helfen zu wollen“, an. Für die Mehrheit der männlichen Probanden liegt die Hauptursache für akademisches Täuschungsverhalten weniger in einem altruistischen Motiv, sondern viel mehr in einem „egoistischen, auf den eigenen Profit ausgerichteten Handeln“ (Calabrese & Cochran, 1990, zitiert nach Belcheir, 2003, S. 6).

1.6 Forschungshypothesen der vorliegenden Arbeit

Im Rahmen der eigenen empirischen Untersuchung sollen vier verschiedene Hypothesen geprüft werden. Die beiden ersten Hypothesen sind inhaltlich vergleichbar, da sie beide die Annahme zum Ausdruck bringen, dass die indirekten Befragungsmethoden imstande sind, höhere Prävalenzratenschätzungen als die (konventionelle) direkte Befragung zu erzielen. Hierbei stellt die erste Hypothese eine Kontrollhypothese dar, weil in zahlreichen Validierungsstudien zur klassischen Variante der UCT, bereits deren Eignung, höhere Basisratenschätzungen als die direkte Befragung aufzudecken, belegt werden konnte (siehe Abschnitt 1.2.2.1).

Hypothese 1:

Die Prävalenzratenschätzungen der klassischen Variante der Unmatched-Count-Technik fallen für die sensitiven Parameter höher aus als die Basisratenschätzungen der direkten Befragung.

Analog zur ersten Hypothese ergibt sich für das bislang noch kaum erprobte Balanced-Incomplete-Block-Design:

Hypothese 2:

Die Prävalenzratenschätzungen des Balanced-Incomplete-Block-Designs fallen für die sensitiven Parameter höher aus als die Basisratenschätzungen der direkten Befragung.

Die dritte Hypothese setzt die Standardfehler der indirekten Befragungsmethoden in Relation zu den Standardfehlern der direkten Befragungsbedingungen. Wie bereits in Kapitel 1.2.2.2 erwähnt, wurde den hohen Standardfehlern in den bisherigen (klassischen) UCT-Studien nur unzureichend Beachtung geschenkt. In der eigenen Untersuchung soll von daher dem quantitativen Ausmaß der Standardfehler in den indirekten Befragungsmethoden im Vergleich mit der direkten Befragung Rechnung getragen werden.

Hypothese 3:

Sowohl die Standardfehler der Parameterschätzwerte der klassischen Unmatched-Count-Technik als auch die Standardfehler der Parameterschätzwerte des Balanced-Incomplete-Block-Designs fallen höher aus als die Standardfehler der Basisratenschätzungen der direkten Befragung. Das BIBD weist zudem eine höhere Effizienz als die klassische UCT auf.

Da in zahlreichen Untersuchungen zur “Academic Dishonesty” Geschlechtsunterschiede hinsichtlich der Neigung zu Täuschungsversuchen vermutet und oftmals auch gefunden wurden (vgl. Kapitel 1.4.2), soll auch in der vorliegenden Studie eine geschlechtsspezifische Hypothese zur Testung kommen.

Hypothese 4:

Die Prävalenzratenschätzungen der sensitiven Verhaltensweisen fallen unabhängig davon, ob eine direkte oder indirekte Befragung durchgeführt wird, für die männlichen Teilnehmer höher als für die weiblichen Teilnehmerinnen aus.

Insgesamt spielt die vierte und letzte Hypothese eine eher untergeordnete Rolle und ist als Ergänzungshypothese zu verstehen. Die höhere Relevanz für den Gegenstand der eigenen Untersuchung weisen die ersten drei Hypothesen auf. Ihnen soll von daher auch die ausführlichere Darstellung im Rahmen der vorliegenden Arbeit gebühren.

2 Methodisches Vorgehen

Im folgenden Hauptabschnitt soll nun das methodische Vorgehen näher skizziert werden. Im Einzelnen werden die Ausgestaltung des Versuchsplans, die Auswahl der Stichprobe, die Versuchsdurchführung und die Datenauswertungs-Software aufgezeigt.

2.1 Versuchsplan

In den Versuchsplan der aktuellen Arbeit waren die drei zu kontrastierenden Umfragetechniken zu integrieren. Dies waren die direkte Befragungsbedingung, die „konservative“ Variante der UCT sowie die „innovative“ Variante der UCT, das BIBD.

2.1.1 Auswahl eines geeigneten Teilverschachtelungsdesigns

Smith et al. (2003) beschränkten sich im Rahmen ihres Versuchs, sich die Eigenschaften von BIBDs für Zwecke der indirekten Umfrageforschung zunutze zu machen, auf eine eingegrenzte Menge von Teilverschachtelungsdesigns. Tabelle 4 gibt einen Überblick über diejenigen BIBDs, die die folgenden Einschlusskriterien aufweisen: eine Gesamtanzahl von $6 \leq v \leq 13$ Fragen, eine Blockgröße von $3 \leq k \leq 5$ Fragen und eine Anzahl kritischer Fragen zwischen drei und sechs ($3 \leq n_s \leq 6$). Insgesamt sind innerhalb dieser Rahmenbedingungen die zehn in Tabelle 4 aufgeführten BIBDs existent.

Tabelle 4

Darstellung der entscheidungsrelevanten Kennwerte der von Smith et al. (2003) aufgeführten BIBDs

Design					r/b	n var (μ_i)			
v	b	r	k	λ		$n_s = 3$	$n_s = 4$	$n_s = 5$	$n_s = 6$
6	10	5	3	2	0.5	0.58	na ^x	na ^x	na ^x
7	7	3	3	1	0.43	0.64	0.59	na ^x	na ^x
7	14	6	3	2	0.43	0.64	0.59	na ^x	na ^x
9	12	4	3	1	0.33	0.78	0.74	na ^x	na ^x
7	7	4	4	2	0.57	0.88	0.81	na ^x	na ^x
8	14	7	4	3	0.5	0.91	0.85	na ^x	na ^x
10	15	6	4	2	0.4	1.03	0.97	0.91	na ^x
13	26	6	3	1	0.23	1.07	1.03	0.99	0.96
13	13	4	4	1	0.31	1.23	1.19	1.14	1.09
11	11	5	5	2	0.45	1.30	1.24	1.18	1.11

^x: „na“ bedeutet: für dieses spezielle Design nicht anwendbar, da ansonsten nicht gewährleistet ist, dass jeder Block mindestens eine nicht-sensitive Frage enthält. In der Übersicht ist p_s konstant 0.05 bzw. p_n konstant 0.1.

Die zehn von Smith et al. (2003) aufgeführten BIBDs waren nun in Bezug auf ihr Infragekommen für die Berücksichtigung innerhalb des eigenen Versuchsdesigns zu evaluieren. Die Auswahlentscheidung erfolgte hierbei nach den folgenden Kriterien:

1. Maximierung von r/b (denn jede [sensitive] Frage soll in möglichst vielen Blöcken vorkommen)
 - in Tab. 4 sind die das Kriterium 1 am Besten erfüllenden Designs jeweils durch hellgraue Schattierung gekennzeichnet
2. Minimierung von b (denn je größer die Anzahl der Blöcke b , desto größer wird der benötigte Stichprobenumfang)
 - in Tab. 4 sind die das Kriterium 2 am Besten erfüllenden Designs jeweils in Fett- und Kursivschrift hervorgehoben
3. Minimierung der Varianz (denn dies stellt den entscheidenden Vorteil der BIBDs gegenüber der „konservativen“ UCT dar)
 - in Tab. 4 sind die das Kriterium 3 am Besten erfüllenden Designs jeweils durch dunkelgraue Schattierung hervorgehoben

Wurden die zehn Teilverschachtelungsdesigns zunächst hinsichtlich des ersten Kriteriums bewertet, so wurde ersichtlich, dass das BIBD (7, 4, 2) mit einem r/b von 0.57 am Besten abschneidet; dahinter rangieren mit einem r/b von 0.5 die BIBDs (6, 3, 2) und (8, 4, 3). Gemäß dem zweiten Kriterium galt es diejenigen BIBDs ausfindig zu machen, deren Blockanzahl möglichst niedrig ist. Während die BIBDs (6, 3, 2) bzw. (8, 4, 3) mit 10 bzw. 14 Blöcken in Bezug auf das zweite Kriterium schlecht abschneiden, weisen die BIBDs (7, 3, 1) und (7, 4, 2) mit nur 7 Blöcken den geringsten Wert für b auf. Hinsichtlich der Minimierung der Varianz²³, die in Tabelle 4 jeweils für unterschiedliche n_s – variierend von 3 bis 6 sensitiven Fragen – angegeben ist, liegt das BIBD (6, 3, 2) bei einem n_s von 3 mit einer Varianz von 0.58 ganz vorne, gefolgt von den BIBDs (7, 3, 1) und (7, 3, 2), die bei einem n_s von 4 jeweils eine Varianz von 0.59 aufweisen. Legte man das Kriterium 3 für das BIBD (7, 4, 2) an, welches in Bezug auf die Kriterien 1 und 2 überlegen war, so lag dies mit einer Varianz von 0.88 für $n_s = 3$ bzw. 0.81 für $n_s = 4$ nur im Mittelfeld. Es lässt sich an dieser Stelle zusammenfassend sagen, dass sich das BIBD (7, 4, 2) – in Bezug auf die Kriterien 1 und 2 – bzw. das BIBD (7, 3, 1) – hinsichtlich der Kriterien 2 und 3 – von den zehn zur Disposition stehenden Designs am ehesten bewähren konnten. Von daher sollte die letztendliche Auswahlentscheidung zwischen diesen beiden Teilverschachtelungsdesigns fallen. Hierfür war es sinnvoll, die Versuchspläne der beiden BIBDs gegenüberzustellen. Eine derartige Gegenüberstellung findet sich in Tabelle 5a bzw. 5b.

²³ Aus Gründen der Vereinfachung wurde innerhalb des BIBD-Vergleichs bei Smith et al. (2003) die Wahrscheinlichkeit des Vorhandenseins eines kritischen Merkmals (p_s) konstant auf 0.05 und die Wahrscheinlichkeit des Vorhandenseins eines unkritischen Merkmals (p_n) konstant auf 0.10 angesetzt.

Es soll nunmehr ein viertes Kriterium eingeführt werden, welches die Vertraulichkeit, die die jeweiligen Teilverschachtelungsdesigns mit sich bringen, in den Fokus der Aufmerksamkeit rückt. In den Erläuterungen zu den BIBDs im Einführungsteil (siehe Kap. 1.2.3) wurde bereits angemerkt, dass in keinem Fragenblock innerhalb eines BIBDs ausschließlich sensitive Fragen enthalten sein dürfen, weil ansonsten die angestrebte Uneindeutigkeit des angegebenen Summenwertes eines Befragungsteilnehmers verloren geht.²⁴ Im Umkehrschluss lässt sich daraus folgern, dass ein BIBD einem anderen dann vorzuziehen ist, wenn es die sensitiven Fragen möglichst gut „abzupuffern“ imstande ist, d.h. in jedem Fragenblock mindestens eine, besser noch zwei oder mehr harmlose Füllfragen vorhanden sind. Da der Vertraulichkeitsverlust bei einer Anzahl von vier kritischen Fragen pro Design zu beträchtlich gewesen wäre, wurden für beide BIBDs nur die Konstellationen mit jeweils drei sensitiven Fragen ($n_s = 3$) in Erwägung gezogen, auch wenn dies mit einer geringfügigen Varianzzunahme einhergeht (vgl. Tabelle 4).

Tabelle 5a

Operationalisierung des Versuchsplanes für das BIBD (7, 3, 1) mit $n_s = 3$ sensitiven Fragen

Block innerhalb des BIBDs (7, 3, 1)						
Nr. 1	Nr. 2	Nr. 3	Nr. 4	Nr. 5	Nr. 6	Nr. 7
4	4	4	A	A	B	B
A	C	6	C	5	C	5
B	5	7	6	7	7	6

Tabelle 5b

Operationalisierung des Versuchsplanes für das BIBD (7, 4, 2) mit $n_s = 3$ sensitiven Fragen

Block innerhalb des BIBDs (7, 4, 2)						
Nr. 1	Nr. 2	Nr. 3	Nr. 4	Nr. 5	Nr. 6	Nr. 7
6	C	A	4	4	6	B
B	7	5	5	C	A	7
C	5	B	6	5	C	4
7	A	6	7	B	4	A

Die drei sensitiven Fragen sind jeweils mit Buchstaben versehen und innerhalb des Versuchsplans durch Fettdruck hervorgehoben; die vier nicht-sensitiven Fragen sind mit den Zahlen „4“ bis „7“ durchnummeriert

²⁴ Und genau diese Uneindeutigkeit ihres Antwortmusters darf nicht verloren gehen, da den Befragten durch sie im Gegensatz zur direkten Befragung eine methodisch bedingte höhere Anonymität erst ermöglicht wird.

Bei Betrachtung der Versuchspläne der beiden noch zur Disposition stehenden BIBDs fällt auf, dass innerhalb des BIBDs (7, 3, 1) in den Fragenblöcken 1, 4 und 6 jeweils zwei der drei Fragen sensitiv sind, so dass in diesen drei Blöcken jeweils nur eine Füllfrage existiert. Dagegen befinden sich innerhalb des BIBDs (7, 4, 2) in allen Fragenblöcken konsistent zwei Füllfragen, mit Ausnahme des Blockes 4, innerhalb dessen sogar ausschließlich unkritische Fragen auftreten. Wurde von daher als viertes Kriterium das Kriterium der maximierten Vertraulichkeit zugrunde gelegt, so war das BIBD (7, 4, 2) dem BIBD (7, 3, 1) vorzuziehen.

2.1.2 Auswahl der nicht-sensitiven Füll- bzw. Kontrollfragen

Ein weiterer, bedeutsamer Aspekt, der dazu beitrug, das subjektive Vertraulichkeitsempfinden innerhalb der BIBD-Fragenblöcke zu verbessern, bestand darin, niedrig prävalente, nicht-sensitive Fragen auszuwählen. Die Hintergrundüberlegung war hierbei die, dass niedrig prävalente „harmlose“ Fragen die Wahrscheinlichkeit der Angabe des Maximalscores durch eine(n) Befragte(n) innerhalb eines bestimmten Blockes noch weiter zu verringern imstande sind.²⁵ Auf der anderen Seite galt es aber ebenso zu vermeiden, dass die Prävalenzratenschätzung einer nicht-sensitiven Frage nahe bei 0 % lag, weil ansonsten Bodeneffekte drohten. Da – wie im vorigen Abschnitt dargestellt – die Entscheidung auf das BIBD (7, 4, 2) gefallen ist, mussten neben den drei sensitiven Fragen, deren Auswahl im nächsten Unterkapitel erläutert wird, insgesamt vier nicht-sensitive Fragen bestimmt werden. Eben diese vier nicht-sensitiven Fragen stellten (als Block 4 des BIBDs) gleichzeitig auch die Kontrollbedingung für die klassische UCT dar, die im Rahmen der eigenen Arbeit in den Experimentalbedingungen jeweils durch vier unkritische Fragen und eine kritische Frage²⁶ umgesetzt werden soll. Da für jede sensitive Frage eine eigene Experimentalbedingung geschaffen werden musste, die dann im Zuge der klassischen Auswertung mit der Kontrollbedingung zu vergleichen war, resultierten im Versuchsplan der eigenen Studie insgesamt drei Experimentalbedingungen für die klassische UCT.

Um bereits a priori an Schätzungen für die ungefähren Prävalenzraten der nicht-sensitiven Fragen zu gelangen, war eine diesbezügliche Datenbankrecherche sinnvoll. Hierbei war grundsätzlich darauf zu achten, dass sich die Prävalenzratenschätzungen jeweils auf eine studentische Population bezogen, da die eigene Datenerhebung in universitären

²⁵ Denn selbst wenn eine Befragte / ein Befragter innerhalb eines BIBD-Fragenblockes gleich zwei sensitive Fragen bejaht, kann die (ehrliche) Verneinung einer niedrig prävalenten harmlosen Frage dazu beitragen, dass er bzw. sie nicht in die Bedrängnis kommt, den Maximalscore angeben zu müssen.

²⁶ Die Entscheidung, im Rahmen der konservativen UCT-Variante neben der einen kritischen Frage „nur“ vier nicht-sensitive Fragen – und nicht fünf unkritische wie ursprünglich bei Miller (1984) – zu verwenden, wurde bereits im Abschnitt 1.2.2.2 begründet.

Veranstaltungen durchgeführt werden sollte. Als erste nicht-sensitive Frage, die den Anspruch erhob, niedrig prävalent auszufallen, wurde die Frage „Sind Sie bereits verheiratet?“ ausgewählt. So gaben bei einer Sozialerhebung des Hochschulinformationssystems im Jahre 2006, die vom Bundesministerium für Bildung und Forschung durchgeführt wurde, 5 % der insgesamt 17.000 befragten Studenten und Studentinnen an, schon verheiratet zu sein. Die zweite Frage unkritischer Art, die in das eigene Versuchsdesign aufgenommen wurde, lautete: „Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?“ Diese Frage war an eine Schüler- und Studentenbefragung des Statistischen Bundesamtes zum Thema „Informations- und Kommunikationstechnologien“ aus dem Jahre 2006 (mit einem Stichprobenumfang von $n = 2.269$) angelehnt, innerhalb derer 27 % angaben, diese Art des Informationsaustausches bereits in Anspruch genommen zu haben.²⁷ Wie schon die erste unkritische Frage stützte sich auch die dritte nicht-sensitive Frage „Wohnen Sie derzeit zur Untermiete?“ auf die Sozialerhebung des Hochschulinformationssystems aus dem Jahre 2006. Diese Frage wurde von 2 % der Befragten (bei einem n von 17.000) bejaht. Als vierte und letzte unkritische Frage erhielt die Frage „Haben Sie eine nicht entfernbare Tätowierung?“ Einzug in die eigene Studie. Im Rahmen einer Umfrage des Meinungsforschungsinstitutes Marplan GmbH an 19- bis 24-jährigen zur Thematik „Körperkult junger Erwachsener“ aus dem Jahre 2003 konnte auf die Frage „Haben Sie ein Piercing oder ein Tattoo?“ (bei einem Stichprobenumfang von $n = 2.507$) eine Merkmalsträger-Basisratenschätzung von 40 % ermittelt werden. Es durfte erwartet werden, dass die Prävalenzratenschätzung für das Aufweisen einer Tätowierung deutlich unter der durch die Marplan GmbH veröffentlichten Basisratenschätzung liegen würde, zumal innerhalb des eigenen Designs explizit von einer nicht entfernbaren Tätowierung gesprochen wurde, da „unechte“ (Natur-)Tattoos von den Befragten nicht mitgezählt werden sollten. Stirn (2007) berichtete für die deutsche Gesamtbevölkerung eine Prävalenzratenschätzung von 8,5 % für das Merkmal der Trägerschaft eines Tattoos, nahm aber keine altersspezifische Differenzierung vor.

2.1.3 Auswahl der sensitiven Fragen

Im Gegensatz zu den unkritischen Fragen, die als Füll- bzw. Kontrollgruppenfragen nur in den indirekten Befragungstechniken (klassische UCT sowie BIBD) zur Anwendung kamen, wurden die sensitiven Fragen sowohl in der direkten als auch in beiden indirekten

²⁷ Die Frage „Haben Sie schon einmal über das Internet telefoniert?“ wies im Vergleich zu den anderen nicht-sensitiven Fragen eine etwas höhere Prävalenzratenschätzung in der Fachliteratur auf. Die Berücksichtigung dieser Frage erfolgte in Anlehnung an Corstange (2006), da dieser empfiehlt, *ein* höher prävalentes nicht-sensitives Item einzubeziehen, um zu erreichen, dass möglichst wenige Probanden eine Summe von „null“ angeben und sich auf diese Weise unweigerlich als Nicht-Merkmalsträger offenbaren.

Befragungstechniken zum Einsatz gebracht. Hierbei sollten die kritischen Fragen derart ausgewählt werden, dass sie von den Befragten als in unterschiedlichem Maße sensitiv wahrgenommen wurden, das heißt auf unterschiedlichen Punkten eines Sensitivitäts-Kontinuums einzuordnen waren. Grundsätzlich sollten sich alle erfragten Täuschungsversuche sowohl auf die Schul- als auch die Studienzeit der Befragten beziehen, weshalb alle sensitiven Fragen mit der Formulierung „Haben sie schon einmal...“ begannen.²⁸ Als kritische Frage mit dem geringsten Sensitivitätsgrad diente die Frage „Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel zum Einsatz gebracht?“. Hrabak et al. (2004) ermittelten für das Benutzen eines Spickzettels während einer Prüfung eine Basisratenschätzung von 34,6 %. In einer Studie von Lin & Wen (2007) berichteten 27,9 % der befragten Studentinnen und Studenten, verbotenerweise Spickzettel-Notizen verwendet zu haben. Als mittelmäßig sensitiv wurde die Frage „Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür gar keinen ärztlichen Grund gab?“ in das eigene Design implementiert. Referenzquelle hierfür war eine Studie von Schmelkin, Kaufman & Liebling (2001), die für die Verhaltensweise, einen Prüfungstermin bzw. die Abgabefrist einer Hausarbeit durch Angabe eines fiktiven Grundes verzögert zu haben, eine Prävalenzratenschätzung von annähernd 9 % berichteten. Schließlich stellte die Frage „Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?“ die dritte und in höchstem Maße sensitive Frage des eigenen Versuchsdesigns dar. Zwar ließ sich in der Fachliteratur zur „Academic Dishonesty“ keine Basisratenschätzung finden, die explizit zur eigens ausgewählten Frage äquivalent war. Es sei jedoch an dieser Stelle neuerlich auf die Studie von Hrabak et al. (2004) verwiesen, die für das Item, die Unterschrift des Lehrers im Immatrikulationsbuch gefälscht zu haben, eine Prävalenzratenschätzung von 1,6 % ermittelten²⁹.

Es war grundsätzlich anzunehmen, dass im Zuge der direkten Befragung mit zunehmendem Sensitivitätsgrad der Fragen die Bereitwilligkeit, ein begangenes, sozial nicht erwünschtes Verhalten zuzugeben, abnehmen würde. Von daher stellen die mittels direkter Befragung erhaltenen Basisratenschätzungen für die kritischen Verhaltensweisen, die der Fachliteratur entstammten, vermutlich nur eine Untergrenze der „wahren“ Prävalenzraten dar.

²⁸ Es wurde somit die Lebenszeitprävalenz der jeweiligen sensitiven Verhaltensweisen erfragt.

²⁹ Es ist allerdings nicht ohne Weiteres davon auszugehen, dass ein Schüler / eine Schülerin, die die Unterschrift des Lehrers im Klassen- oder Immatrikulationsbuch zu fälschen bereit ist, unweigerlich auch die Unterschrift der Eltern zu fälschen gedenkt. Dennoch können beide Handlungen als vergleichbar sensitiv eingestuft werden.

2.1.4 Eine zusammenfassende Übersicht über das Untersuchungsdesign

Nachdem die Genese des Versuchsplans in den letzten Unterkapiteln ausführlich dargestellt worden ist, soll nun ein zusammenfassender Überblick über das eigene Untersuchungsdesign gegeben werden. Eine Veranschaulichung dessen erfolgt in Tabelle 6, innerhalb der auch noch einmal der genaue Wortlaut der verwendeten Fragen zu finden ist.

2.1.4.1 *Positionierung der sensitiven Fragen in den indirekten Bedingungen*

Die sieben Fragenblöcke des BIBDs (7, 4, 2) stellen die ersten sieben Bedingungen des eigenen Versuchsdesigns dar. Innerhalb der sechs Fragenblöcke mit den jeweils zwei kritischen Fragen, sollten diese stets die mittleren Positionen einnehmen, also jeweils als zweite und dritte Frage zum Einsatz kommen. Diese Positionierung erfolgte aus folgenden inhaltlichen Sequenzüberlegungen: wird eine kritische Frage direkt zu Beginn eines Fragenblockes gestellt oder treten gleich zwei kritische Fragen als Einstiegsfragen auf, so besteht die Gefahr einer frühzeitigen Abschreckung des bzw. der Befragten; um einen derartigen Abschreckeffekt zu verhindern, sollte der Einstieg durch eine „harmlose“ Frage erfolgen. Des Weiteren sollte eine kritische Frage niemals als letzte Frage innerhalb eines Fragenblockes auftreten, weil der Befragte hierdurch darauf gestoßen werden könnte, mit der ehrlichen Bejahung der letzten (kritischen) Frage den Maximalscore oder zumindest einen ebenfalls belastenden Score von „drei“ zu erreichen und aus diesem Grund das Vorhandensein des sensitiven Verhaltens womöglich leugnet. In den Bedingungen 8 bis 10, welche die konservative Anwendung der UCT – für jede der drei sensitiven Fragen – repräsentierten, sollte die kritische Frage aus den gleichen Überlegungen wie für die BIBD-Blöcke ebenfalls stets in der Mitte – hier dann als dritte von fünf Fragen – aufgeführt werden. Ein bedeutsamer Aspekt, der die Eignung des (7, 4, 2)-BIBDs für die eigene Untersuchung verdeutlicht, ist der, dass der BIBD-Fragenblock 4 gleichzeitig auch als Kontrollbedingung für die UCT-Experimentalbedingungen 8 bis 10 fungiert und von daher einen noch höheren personellen Aufwand im Zuge einer zusätzlichen Bedingung obsolet macht.

2.1.4.2 *Begründung der Formalisierung sieben direkter Bedingungen*

Es mag auf den ersten Blick ungewöhnlich erscheinen, dass die Bedingungen der direkten Befragung – 11 bis 17 – allesamt identisch sind. Die Unterteilung der direkten Befragung in sieben Untergruppen hatte in erster Linie den Zweck, eine analoge Darstellung zum BIBD zu wählen. Die Hintergrundüberlegung war hierbei die, dass die Parameterschätzungen der beiden Umfragetechniken (der direkten Befragung und des BIBDs) im Zuge der späteren Auswertung einander gleichgesetzt werden sollen. Auf diese Weise sollte die Testung

erfolgen, ob das BIBD als indirekte Befragungsmethode zu signifikant höheren Basisratenschätzungen für die sensitiven Fragen im Vergleich zur direkten Befragung führt. Resultieren die Basisratenschätzungen beider Befragungsarten aus den gleichen Stichprobenumfängen (nämlich $n = 700$), so lässt sich der potentielle Einwand, dass die Schätzungen einer unterschiedlichen Anzahl befragter Teilnehmerinnen und Teilnehmern entstammen, von vornherein entkräften

Tabelle 6

Überblick über die 17 Bedingungen des eigenen Untersuchungsdesigns

Bed.	BIBD							UCT			Direkte Befragung						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Fragen	6	7	5	4	4	6	7	4	4	4	A	A	A	A	A	A	A
	B	C	A	5	C	A	B	5	5	5	B	B	B	B	B	B	B
	C	A	B	6	B	C	A	A	B	C	C	C	C	C	C	C	C
	7	5	6	7	5	4	4	6	6	6							
								7	7	7							
Quote	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100	n = 100
	n = 700							n = 300			n = 700						

Sensitive Fragen:

A: Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel zum Einsatz gebracht?

B: Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen, schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür gar keinen ärztlichen Grund gab?

C: Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?

Nicht-sensitive Fragen:

4: Sind Sie verheiratet?

5: Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?

6: Wohnen Sie derzeit zur Untermiete?

7: Haben Sie eine nicht entfernbare Tätowierung?

Die drei sensitiven Fragen sind jeweils mit Buchstaben versehen und durch Fettdruck hervorgehoben, die vier nicht-sensitiven Fragen mit den Zahlen „4“ bis „7“ durchnummeriert. Die Bedingungen 1 bis 7 stellen die BIBD-Fragenblöcke, die Bedingungen 8 bis 10 die klassischen UCT-Bedingungen und die Bedingungen 11 bis 17 quasi sieben unabhängige Replikationen der direkten Befragung dar.

2.2 Wesentliche Aspekte der Fragebogengestaltung

Der Fragebogen der eigenen Studie hatte den Anspruch, aufgrund des benötigten vierstelligen Stichprobenumfangs schlicht gestaltet und verständlich zu sein, um Unklarheiten und Missverständnisse gar nicht erst aufkommen zu lassen. Er bestand von daher nur aus *einer* DIN-A4-Seite. Eine Übersicht über die Layouts der Fragebogen in den unterschiedlichen Bedingungen findet sich im Anhang A.

In der Kopfzeile des Fragebogens fand sich neben der Angabe über das betreuende Institut und den Rahmen der Arbeit³⁰ eine unauffällige Kodierung der Bedingung. So war jeweils im Anschluss an den Erhebungszeitpunkt nach einem Unterstrich die jeweilige Bedingung numerisch angegeben, um im Zuge der späteren Dateneingabe die Bedingungszuordnung zu erleichtern (z.B. Oktober 2007_1). Mit der Titulierung der Befragung als „Schnell-Umfrage“ wurde angestrebt, den Teilnehmern zu signalisieren, dass kein großer Zeitaufwand für die Teilnahme benötigt würde. Für die Bearbeitung des Fragebogens sollten drei Minuten bereits die Obergrenze darstellen. In einer kurzen Vorabinformation wurden die Befragten aufgefordert, den Fragebogen umgehend auszufüllen, da er noch vor Beginn der Lehrveranstaltung wieder eingesammelt werden sollte. Zudem wurde den Teilnehmenden vergewissert, dass die Fragebogen anonym für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit ausgewertet würden. Da für die Prüfung der Hypothesen nur die Kenntnis des Geschlechts von Interesse war, wurde nur diese eine demographische Variable erhoben.

Die Hauptinstruktion befand sich jeweils mittig auf der Seite und unterschied sich für die direkten und indirekten Befragungsbedingungen. Während sie in der direkten Befragung aus der Aufforderung „Wir bitten Sie um eine Antwort auf die folgenden Fragen!“ bestand, hatte sie für die indirekten Befragungstechniken den Wortlaut: „Ihnen wird im Folgenden ein Fragenblock à vier / fünf Fragen³¹ präsentiert. Alle Fragen sind derart gestellt, dass sich diese entweder mit „Ja“ oder mit „Nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der vier / fünf Fragen Sie bejaht haben!“. Der eigentliche Befragungsinhalt bestand im Falle der direkten Befragung aus den drei kritischen Fragen, wobei jeweils im Anschluss entweder das Kästchen für „Ja“ oder das Kästchen für „Nein“ anzukreuzen war. Die Abfolge der kritischen Fragen erfolgte in den direkten Bedingungen stets in der gleichen Reihenfolge, beginnend mit der Frage mit dem

³⁰ Intention der Angabe, dass es sich bei der Umfrage um eine Diplomarbeiten-bezogene Erhebung handelt, war der erhoffte Effekt, die Teilnahmebereitschaft der Studierenden hierdurch noch zu erhöhen.

³¹ In den BIBD-Bedingungen fand sich an dieser Stelle jeweils die Angabe der Zahl „vier“, in den klassischen UCT-Bedingungen die Angabe der Zahl „fünf“.

vermutlich geringsten Sensitivitätsgrad („Spickzettel“) und endend mit der Frage mit dem vermutlich höchsten Sensitivitätsgrad („gefälschte Unterschrift“). In den indirekten Bedingungen wurde der bzw. die Teilnehmende im Anschluss an die aufgelisteten Fragen explizit gefragt: „Wie viele dieser vier / fünf Fragen beantworten Sie mit „Ja“? Je nach Bedingung standen dann jeweils Kästchen für die Zahlen von „null“ bis „vier“ (im Falle der BIBD-Fragenblöcke) bzw. „null“ bis „fünf“ (im Falle der Fragenblöcke der klassischen UCT) zur Auswahl. Zur besseren Übersicht für die Versuchsteilnehmerinnen und -teilnehmer wurden die Fragen in allen Bedingungen jeweils durchnummeriert. Während in den direkten Bedingungen jede der drei kritischen Fragen an sich umrahmt war, wurde in den indirekten Bedingungen jeweils der gesamte Fragenblock durch einen Rahmen umgrenzt. Allen Befragungsbedingungen gemein war die abschließende Instruktion, in der ein Dank für die Teilnahme erfolgte bzw. darum gebeten wurde, die ausgefüllte Seite mittig zu falten und an den Rand der jeweiligen Stuhlreihe weiter zu geben.

2.3 Stichprobe

Aufgrund der Tatsache, dass mit dem Geschlecht insgesamt nur eine demographische Variable erhoben wurde, kann die Kennzeichnung der Stichprobe recht kompakt ausfallen. In der Folge sollen insbesondere die prozentualen Verteilungen der Befragten auf die Hochschulen und die Studienfächer im Vordergrund stehen.

2.3.1 Beschreibung der Stichprobe

Die Stichprobe der eigenen Untersuchung setzte sich zusammen aus 1.700 Studentinnen und Studenten dreier verschiedener Hochschulen des nördlichen Rheinlandes. Wie aus Abbildung 2 hervorgeht, bildeten die Studierenden der Heinrich-Heine-Universität Düsseldorf mit 80 % der Befragten hierbei den Schwerpunkt.

☐ Heinrich-Heine-Uni Düsseldorf ☒ Fachhochschule Düsseldorf ☐ Bergische Uni Wuppertal

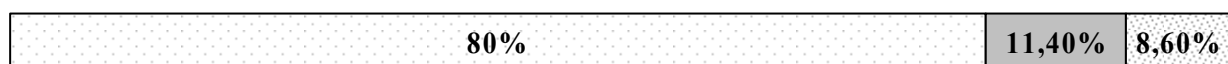


Abbildung 2: Prozentuale Verteilung der studentischen Teilnehmerinnen und Teilnehmer auf die drei Hochschulen Heinrich-Heine-Universität Düsseldorf, Fachhochschule Düsseldorf und Bergische Universität Wuppertal.

Da die Datenerhebung ausschließlich im universitären Rahmen durchgeführt wurde, erfüllten alle Befragten die Voraussetzung, auf mindestens 13 Jahre schulischer Ausbildung

zurückzublicken. Um später potentielle Geschlechtsunterschiede testen zu können, wurde intendiert, möglichst gleich viele Frauen und Männer in die Stichprobe einzubeziehen: insgesamt nahmen 811 Frauen (47,7 %) und 889 Männer (52,3 %) an der Studie teil, die sich annähernd gleichmäßig auf die unterschiedlichen Bedingungen verteilten.³² Wie in Abbildung 3 zu sehen ist, entstammten die Teilnehmerinnen und Teilnehmer sehr unterschiedlichen Fachbereichen, was die Stichprobe insgesamt repräsentativ für eine studentische Gesamtpopulation macht. Mit 17,4 % stellten die Sprachwissenschaftler den am stärksten vertretenen Fachbereich dar, dicht gefolgt von den Naturwissenschaftlern mit 14,6 % und den Medizinerinnen und Pharmakologen mit 14,2 %.

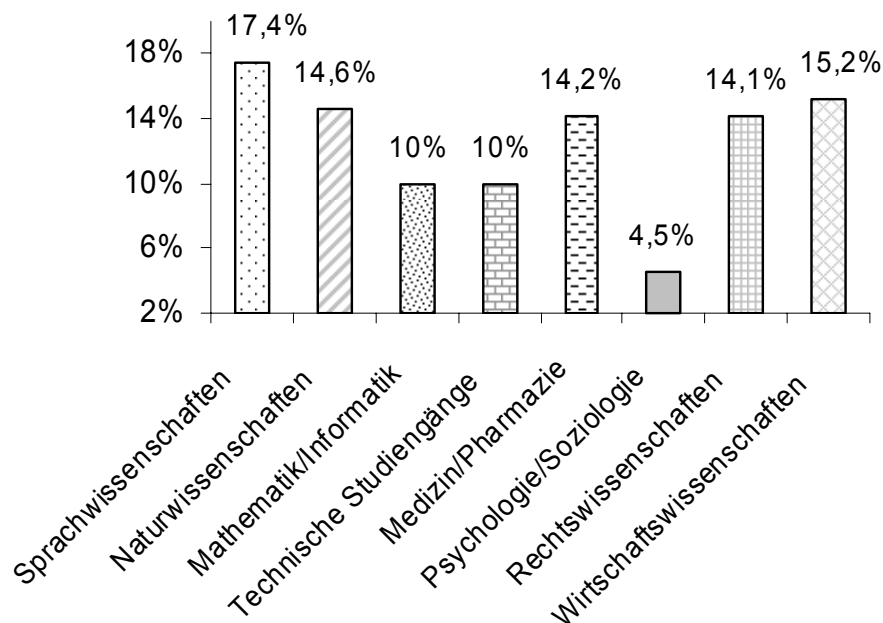


Abbildung 3: Übersicht über die anteilmäßige Aufteilung der befragten Studentinnen und Studenten auf die verschiedenen Studienfächer. Einige Fächer wurden wie folgt gebündelt: Sprachwissenschaften (Anglistik, Germanistik, Romanistik), Naturwissenschaften (Physik, Biologie, Chemie), technische Studiengänge (Elektrotechnik, Maschinenbau, Medientechnik).

Befragt wurden zahlreiche Studienanfänger (ca. 60 %), aber auch Studierende höherer Fachsemester. Die hohe Beteiligung von Erstsemestern erklärte sich insbesondere aus deren höherer Anwesenheit in den Veranstaltungen und deren tendenziell größerer Offenheit für die Teilnahme an der Befragung. Methodisch ist anzumerken, dass Studierende höherer Semester im Gegensatz zu Studienanfängern bereits im universitären Umfeld mit Prüfungssituationen

³² Das Minimum an teilnehmenden Frauen lag in den Bedingungen 2 und 4 bei 44 Befragten. Das Maximum an Männern wurde mit 56 Probanden in der Bedingung 2 erreicht.

konfrontiert waren und von daher potentiell mehr Gelegenheiten zu Täuschungsversuchen innerhalb ihrer Ausbildung hatten.

2.3.2 Planung der Stichprobengröße

Das Programm G*Power 3 wurde herangezogen, um den benötigten Stichprobenumfang zu ermitteln, der mit einer statistischen Power von 95 % einen kleinen bis mittleren Effekt (der Stärke $w = 0.15$) aufzudecken imstande war. Hierbei ergab sich auf dem konventionellen α -Niveau von 5 % die Notwendigkeit eines Stichprobenumfanges, der bei 1.636 lag. Aus diesem Grunde fiel die Entscheidung für einen Gesamtstichprobenumfang von $n = 1.700$, wobei jede der 17 Bedingungen mit $n = 100$ Personen besetzt wurde. Ein Vorteil, der sich aus der Ansetzung von $n = 100$ pro Bedingung ergab, war der, dass sich auf diese Weise die beobachteten Häufigkeiten der gegebenen Ja-Antworten in den unterschiedlichen Bedingungen mit minimalem Rechenaufwand in Prozentwerte umwandeln ließen.

2.4 Versuchsdurchführung

Die Datenerhebung erfolgte in Form einer Papier und Bleistift-Befragung überwiegend in Veranstaltungen an der HHU Düsseldorf. Um bereits im Vorhinein einen Überblick über den quantitativen Umfang der Studierendenschaft der verschiedenen Fächer zu gewinnen, wurde eine aktuelle Studierendenstatistik der Fachfälle von Seiten des Studierendensekretariates der Universität Düsseldorf herangezogen. Diese wies die Germanistik mit 4.285 Absolventen³³, gefolgt von der Medizin mit 2.617 Studenten und den Rechtswissenschaften mit 1.375 Immatrikulierten als die studierendenstärksten Fächer der HHU Düsseldorf aus.

Um Doppelerhebungen zu vermeiden, wurden innerhalb eines Faches bevorzugt Veranstaltungen ausgewählt, die aufeinander aufbauen (wie z.B. „Bürgerliches Recht I“ und „Bürgerliches Recht III“), da es als unwahrscheinlich erachtet werden kann, dass Studierende innerhalb eines Semesters beide Veranstaltungen belegen. Die gezielte Veranstaltungsauswahl wurde in der Regel über das elektronische Vorlesungsverzeichnis (EKVV) der Universität Düsseldorf durchgeführt, zumal in diesem Rahmen auch Änderungen vom Lehrplan in den Rubriken „Heutige Veranstaltungen“ bzw. „Heute ausfallende Veranstaltungen“ angekündigt wurden. Ziel war es, pro Veranstaltung möglichst viele Studentinnen und Studenten zu befragen, ohne jedoch die Dozenten und Dozentinnen im Zuge ihrer Lehrveranstaltungen zu stören. Als guter Indikator für die zu erwartende Teilnehmerzahl erwies sich ferner die Raumbelungsübersicht, da den größten Hörsälen der

³³ Von den insgesamt 4.285-Germanistik-Absolventen studierten dies allerdings nur 2.337 im Hauptfach.

HHU in der Regel auch die Veranstaltungen mit den höchsten Teilnehmerzahlen zugewiesen wurden.

2.4.1 Zeitpunkt der Datenerhebung

Als Zeitpunkt der Datenerhebung wurde der Beginn des Wintersemesters 2007 / 2008 angesetzt, da erfahrungsgemäß zu keinem anderen Zeitpunkt des Jahres mehr Studierende an den Düsseldorfer Hochschulen eingeschrieben sind. Dies liegt zum einen darin begründet, dass einige Studiengänge nur zu Beginn des Wintersemesters aufgenommen werden können, zum anderen auch darin, dass sich in der Folgezeit ein nicht unerheblicher Prozentsatz der Studienanfänger aus unterschiedlichen Gründen wieder exmatrikuliert. Ein weiterer Aspekt, der den Semesterbeginn als geeigneten Erhebungszeitraum bekräftigt, ist der, dass in den meisten Studiengängen in den ersten Semesterwochen die Teilnehmerzahlen der Veranstaltungen höher ausfallen als in den letzten Wochen des Semesters, was insbesondere darauf zurückzuführen ist, dass Prüfungen und Klausuren für die Studierenden mehrheitlich zum Ende der Vorlesungszeit anstehen.

2.4.2 Ablauf der Datenerhebung

Die Datenerhebung wurde jeweils von mindestens zwei Versuchsleitern durchgeführt, wobei sich diese ungefähr 20 bis 30 Minuten vor Beginn einer Lehrveranstaltung an den Eingängen der jeweiligen Hörsäle aufstellten. Um eine Randomisierung der Versuchsbedingungen zu gewährleisten, wurden die Fragebogen der 17 verschiedenen Bedingungen in der folgenden Weise angeordnet: 1; 2; 11; 3; 12; 4; 13; 5; 14; 6; 15; 7; 16; 8; 17; 9; 10 (die direkten Befragungsbedingungen sind jeweils durch Kursivdruck hervorgehoben)³⁴. Bei dieser Abfolge wechselten sich direkte und indirekte Befragungsbedingungen – mit Ausnahme der Bedingungskonstellationen 1 und 2 sowie 9 und 10 – einander ab. Auf diese Weise erhielten gleichzeitig ankommende und in der Folge oftmals nebeneinander sitzende Teilnehmerinnen und Teilnehmer nicht einen Fragebogen der gleichen Bedingung und zumindest in der Mehrheit der Fälle auch nicht die gleiche Umfragetechnik. Die Fragebogen wurden bis ca. 10 Minuten vor Beginn einer Lehrveranstaltung verteilt. Studierende, die zu einem späteren Zeitpunkt im Hörsaal eintrafen, erhielten keinen Fragebogen mehr, da ansonsten das rechtzeitige Einsammeln der ausgefüllten Bogen noch vor Beginn der Lehrveranstaltung gefährdet war. Wiesen Veranstaltungen eine Zäsur in Form einer kurzen Pause auf, konnte in Absprache mit der Veranstaltungsleiterin bzw. dem Veranstaltungsleiter zusätzlich die Pause

³⁴ Die angegebene Anordnung wurde bereits in der für den Ausdruck erstellten PDF-Datei der Fragebogen berücksichtigt.

für die Fortführung der Datenerhebung genutzt werden, was zumindest in Einzelfällen so gehandhabt wurde. Das Einsammeln der Fragebogen wurde so organisiert, dass jede Versuchsleiterin bzw. jeder Versuchsleiter jeweils für die ihr bzw. ihm zugeordnete Seite des Hörsaals zuständig war.

Aufgrund der Tatsache, dass im Zuge der fortgeschrittenen Datenerhebung die Gefahr von Doppelerhebungen zunahm³⁵, wurden ab dieser Phase gezielt Veranstaltungen anderer Hochschulen aufgesucht. Weil sich nach Abschluss der Erhebungen an der HHU ein deutlicher Überschuss an weiblichen Befragten herausgestellt hatte, wurden in der Endphase gezielt Veranstaltungen aufgesucht, die sich durch eine allgemein hohe Männerquote auszeichneten. Hierfür boten sich insbesondere die Veranstaltungen der technischen Studiengänge an der Fachhochschule Düsseldorf an.

2.4.3 Dateneingabe

Die Dateneingabe erfolgte begleitend zur Datenerhebung, damit kontinuierlich ein Überblick über die Anzahl der bereits befragten Studentinnen und Studenten in den einzelnen Versuchsbedingungen sowie über den aktuellen Geschlechteranteil gewährleistet war. Der beständigen Aktualisierung des Rohdatensatzes wurde insbesondere in der Endphase der Erhebung Sorge getragen, um so der leichten Asymmetrie in der Besetzung der verschiedenen Versuchsbedingungen³⁶ rechtzeitig entgegen zu wirken.

2.4.3.1 Aufbau der Datenmaske

Für die Eingabe der Daten in SPSS wurde eine auf den Fragebogen abgestimmte Datenmaske konzipiert. Konsistent für alle Bogen wurden die Angabe des Geschlechts sowie die bearbeitete Bedingung (die aus der Kopfzeile eines jeden Bogens ablesbar war; siehe Unterkapitel 2.2) eingegeben. In den indirekten Bedingungen (1 bis 10) wurde für die Kodierung des Fragenblocks lediglich eine Variable, bestehend aus der angekreuzten Summe der Ja-Antworten, eingegeben. Diese wurde mit dem Label „jasumme“ gekennzeichnet. Dem entgegen wurde in den direkten Befragungsbedingungen (11 bis 17) für jede sensitive Frage eine dichotom kodierte Variable – mit der Kodierung: die kritische Frage wurde bejaht vs. die kritische Frage wurde verneint –, eingeführt. Entsprechend der Frageninhalte wurden die drei Variablen der direkten Bedingungen mit den Labels „spick“, „attest“ und „untersch“

³⁵ Die erhöhte Gefahr von Doppelerhebungen resultierte daraus, dass Hauptfachstudenten verschiedener Studiengänge oftmals auch Veranstaltungen in Nebenfächern belegen.

³⁶ Eine leichte Asymmetrie der Besetzungszahlen in den verschiedenen Bedingungen blieb deshalb nicht aus, weil ein geringer Teil der Fragebogen nicht zurück an die Versuchsleiter gegeben wurde und dieser Nicht-Rücklauf immer erst im Zuge der Dateneingabe ersichtlich wurde.

versehen. In einer abschließenden String-Variablen, die wiederum allen Bedingungen gemein war, erfolgte eine Angabe über die Veranstaltung, in der die Datenerhebung stattgefunden hatte, wobei hier aus Gründen der Übersichtlichkeit nicht der genaue Veranstaltungstitel, sondern die Zuordnung zu einem Fachbereich und das Fachsemester der Studierenden festgehalten wurden (z.B. Anglistik, 1. Semester). In einer separaten Datei wurden zusätzliche Einträge in Bezug auf die jeweilige Veranstaltung (wie z.B. Dozent, Erhebungstag und -zeit, Anzahl angetroffener Studierender und Geschlechteranteil) aufgeführt.

2.4.3.2 Kriterien für den Ausschluss von Fragebogen

Im Zuge der Dateneingabe mussten einige Bogen aussortiert werden. Diese konnten für die spätere Auswertung nicht weiter berücksichtigt werden. Ein Ausschluss erfolgte immer dann, wenn ein Bogen nicht vollständig ausgefüllt worden war. Bogen galten dann als unvollständig ausgefüllt, wenn entweder keine Geschlechtsangabe getätigt wurde, in den direkten Bedingungen eine oder mehrere sensitive Fragen nicht beantwortet wurden oder in einer indirekten Befragungsbedingung die Angabe über die Summe der Ja-Antworten fehlte. Ein weiteres Ausschlusskriterium bestand in der fehlerhaften Anwendung der eine höhere Anonymität gewährenden indirekten Befragungsverfahren: so gaben ca. 9 % der in der Ersterhebung befragten Probanden³⁷ in einer klassischen UCT- bzw. BIBD-Bedingung nicht bloß die Summenangabe über ihre Ja-Antworten bekannt, sondern notierten zusätzlich hinter jeder Frage des Fragenblocks – gleich ob sensitiv oder nicht – ein „Ja“ bzw. „Nein“. Ein derartiges Vorgehen dokumentierte offenkundig, dass der oder die Befragte den Anonymitätsschutz des indirekten Verfahrens nicht erkannt oder wahrgenommen hatte. Im Falle der Berücksichtigung derart ausgefüllter Bogen hätte ein verzerrender Einfluss auf die Gesamtstatistik gedroht, der letztlich nur mit dem konsequenten Ausschluss der entsprechenden Bogen vermieden werden konnte. Es sei an dieser Stelle allerdings kritisch angemerkt, dass eine potentielle Motivation, die Fragen in den Frageblöcken der indirekten Bedingungen mit „Ja“ oder „Nein“ zu kennzeichnen, potentiell auch in einem bewussten Verzicht auf die erhöhte Anonymität des indirekten Verfahrens gelegen haben könnte, um gewissermaßen zu dokumentieren: „Ich habe nichts zu verbergen“. Es darf aber vermutet werden, dass eine bewusste Missachtung der Instruktion eher die Ausnahme darstellte.

³⁷ Aufgrund der fehlerhaften Anwendung eines indirekten Befragungsverfahrens (klassische UCT bzw. BIBD) mussten exakt 88 Fragebogen aussortiert werden.

2.4.3.3 Nacherhebungen

Um infolge des vorgenommenen Ausschlusses von Fragebogen nicht die angestrebte Stichprobengröße von $n = 100$ pro Bedingung einzubüßen, wurde für jeden aussortierten Fragebogen ein Bogen der gleichen Bedingung an einem anderen Probanden nachträglich erhoben. Auf diese Weise konnte sichergestellt werden, dass mit Abschluss der Nacherhebungen der angesetzte Stichprobenumfang von $n = 1.700$ auch wirklich erreicht werden konnte und keine Verschlechterung der statistischen Power zu verzeichnen war. Die Nacherhebungen fanden überwiegend in kleineren Veranstaltungen an der Heinrich-Heine-Universität Düsseldorf statt.

2.5 Die Datenauswertungs-Software HMMTree

Im Zuge der Auswertung wurde eine spezielle Software benötigt, die imstande war, auf Basis des zugrunde liegenden multinomialen Modells Prävalenzratschätzungen für die Parameter der indirekten Bedingungen zu generieren. Hierbei handelte es sich um die Software HMMTree³⁸, die über die Internetseiten des Instituts für Psychologie an der Albert-Ludwigs-Universität Freiburg kostenfrei herunter geladen werden kann.³⁹ Stahl & Klauer (2007) erläutern in ihren Ausführungen zu HMMTree, wie die genaue Formatierung der zu erstellenden Files umzusetzen ist. Hierbei unterscheiden sie zwei verschiedene Files: zum einen ein Gleichungsfile, innerhalb dessen sämtliche Äste des zugrunde liegenden multinomialen Modells spezifiziert sind und zum anderen ein Datenfile, das alle beobachtbaren Häufigkeiten der erhobenen Variablen beinhaltet. Erst wenn beide Files von HMMTree erfolgreich, d.h. ohne Fehlermeldung eingelesen worden sind, können Berechnungen der Parameterschätzungen einschließlich ihrer Konfidenzintervalle und Anpassungsgütetests vorgenommen werden.

³⁸ HMMTree ist allerdings nur für binäre multinomiale Modelle anwendbar; im Zuge der eigenen Anwendung fand ausschließlich ein binäres multinomiales Modell Verwendung.

³⁹ Der zugehörige Internet-Link lautet: <http://www.psychologie.uni-freiburg.de/Members/stahl/HMMTree>

3 Ergebnisse

Der Einstieg in den dritten Hauptteil der vorliegenden Arbeit rückt zunächst die für die multinomiale Auswertung relevanten, vorbereitenden Maßnahmen in den Vordergrund, und hierbei insbesondere die Syntax der zu programmierenden Files. Im Anschluss daran folgt die eigentliche Darstellung der Ergebnisse: zunächst für den Gesamtdatensatz und anschließend für die geschlechterspezifischen Datensätze.

3.1 Konzeption der Auswertungs-relevanten Files

Für die Auswertung mit HMMTree war es wichtig, bei der Erstellung der Files die von Stahl & Klauer (2007) beschriebenen Formatierungsrichtlinien zu berücksichtigen. Hierbei war grundsätzlich zu beachten, dass die Files in spezifischen Textdateiformaten abgespeichert werden mussten: für das Datenfile war dies das Format „mdt“, für das Gleichungsfile dagegen das Format „eqt“⁴⁰. In der Folge soll zunächst die Erstellung des Datenfile erläutert werden.

3.1.1 Generierung des Datenfile

Das Datenfile war das kürzere der beiden Files; es beinhaltete die Information über die empirisch gewonnenen Daten. Tabelle 7 gibt einen Überblick über die beobachteten Häufigkeiten der Ja-Antwortsummen in den indirekten Bedingungen sowie über die prozentuale Verteilung der „Ja“ und „Nein“-Antworten in den direkten Bedingungen für jede der drei sensitiven Fragen. Wie der Tabelle zu entnehmen ist, addierten sich die verschiedenen Ja-Antwortsummen in den indirekten Bedingungen jeweils zu 100 auf. In den direkten Bedingungen ergaben sich für jede (sensitive) Frage 700 Antworten.

Im Zuge der Syntaxkodierung für das Datenfile war es notwendig, jeder beobachteten Häufigkeit eine Kategorienummer zuzuordnen. In allen indirekten Bedingungen erhielt die geringste Ausprägung (Ja-Antwortsumme = 0) jeweils die niedrigste Kategorienummer, wohingegen der höchsten Ausprägung (maximale Ja-Antwortsumme) die höchste Kategorienummer zugeteilt wurde. Für jede BIBD-Bedingung wurden fünf Kategorienummern benötigt⁴¹, so dass insgesamt $7 * 5 = 35$ Kategorien für das BIBD vergeben wurden. Im Falle der drei klassischen UCT-Bedingungen fielen jeweils sechs Kategorienummern an, so dass für den UCT-Teil insgesamt $3 * 6 = 18$ Kategorien

40 Die MDT- und die EQT-Syntax sind kompatibel mit den MBT- (Hu, 1999) bzw. AppleTree- (Rothkegel, 1999) Programmen für die Analyse traditioneller multinomialer Verarbeitungsbaummodelle.

41 Die Kategorien 1 bis 5 deckten die erste BIBD-Bedingung, die Kategorien 6 bis 10 die zweite,...die Kategorien 31 bis 35 die siebte BIBD-Bedingung ab.

aufgewendet wurden⁴². Am sparsamsten bei der Vergabe der Kategorien schnitten die direkten Befragungsbedingungen ab, da diese für jede kritische Frage nur zwei Kategorienummern beanspruchten. Jeweils zuerst wurde der Anzahl der Ja-Antworten eine Kategorienummer zugewiesen, bevor die Anzahl der Nein-Antworten die um eins höhere Kategorienummer erhielt.⁴³ Für die direkten Bedingungen fielen somit insgesamt nur sechs Kategorienummern an. Wurden alle Bedingungen übergreifend betrachtet, so ergab sich eine Gesamtanzahl von $35 + 18 + 6 = 59$ Kategorienummern.

Tabelle 7

Beobachtete Häufigkeiten der Ja-Antwort-Summen in den indirekten Bedingungen und prozentuale Verteilung der Ja- / Nein-Antworthäufigkeiten der kritischen Fragen in den direkten Bedingungen

	Bedingung	Summe der Ja-Antworten	0	1	2	3	4	5	Gesamt
		Kategorie Nr.	1	2	3	4	5	6	
1	BIBD Bed 1	6,B,C,7	54	28	15	2	1		100
2	BIBD Bed 2	7,C,A,5	13	37	37	13	0		100
3	BIBD Bed 3	5,A,B,6	11	51	24	12	2		100
4	BIBD Bed 4+UCT Kontr.	4,5,6,7	42	42	13	3	0		100
5	BIBD Bed 5	4,C,B,5	27	48	19	4	2		100
6	BIBD Bed 6	6,A,C,4	31	40	27	1	1		100
7	BIBD Bed 7	7,B,A,4	37	36	21	5	1		100
8	UCT Experimental 1	4,5,A,6,7	19	34	36	8	2	1	100
9	UCT Experimental 2	4,5,B,6,7	26	44	21	8	1	0	100
10	UCT Experimental 3	4,5,C,6,7	32	48	14	2	1	3	100
11 bis 17	Direkte Befragung	A,B,C	Kritische Frage 1: ja: 351 (50,1 %)			Kritische Frage 1: nein: 349 (49,9 %)			700
			Kritische Frage 2: ja: 182 (26,0 %)			Kritische Frage 2: nein: 518 (74,0 %)			700
			Kritische Frage 3: ja: 116 (16,6 %)			Kritische Frage 3: nein: 584 (83,4 %)			700

Die drei sensitiven Fragen sind jeweils mit Buchstaben versehen und durch Fettdruck hervorgehoben, die vier nicht-sensitiven Fragen mit den Zahlen „4“ bis „7“ durchnummeriert. In den indirekten Bedingungen ist die Darstellung auf die absoluten Häufigkeiten beschränkt, da sich die Prozentwerte aufgrund des Stichprobenumfanges von $n = 100$ pro Bedingung unmittelbar ableiten lassen.

⁴² Die Kategorien 36 bis 41 deckten die erste UCT-Experimentalbedingung, die Kategorien 42 bis 47 die zweite und die Kategorien 48 bis 53 die dritte UCT-Experimentalbedingung ab.

⁴³ Der ersten sensitiven Frage wurden die Kategorien 54 und 55, der zweiten die Kategorien 56 und 57 und der dritten die Kategorien 58 und 59 zugewiesen.

Der Aufbau des Datenfile gestaltete sich derart, dass in der ersten Zeile die Angabe des Gesamtstichprobenumfanges erfolgte (in der eigenen Untersuchung $n = 1.700$). In den nachfolgenden Zeilen wurden die Kategoriennummern den beobachteten Häufigkeiten zugeordnet. Hierzu wurde am Zeilenanfang jeweils die spezifische Kategoriennummer aufgeführt, gefolgt von einem Leerzeichen und der anschließenden Angabe der jeweiligen absoluten Häufigkeit (siehe Anhang B).

3.1.2 Generierung des Gleichungsfile

Das Gleichungsfile war das umfangreichere der beiden Files, weil es für sämtliche Bedingungen alle möglichen Kombinationen hin zu einem bestimmten Output auflistete. In der ersten Zeile des Gleichungsfile wurde numerisch angegeben, wie viele Gleichungen insgesamt zu spezifizieren waren. Hierbei wurde jeder Ast des Baummodells durch exakt *eine* Gleichung repräsentiert. Alle Parameter, die innerhalb des der eigenen Studie zugrunde gelegten multinomialen Modells Verwendung fanden, waren binär in dem Sinne, dass jede Frage mit „Ja“ oder „Nein“ beantwortet werden konnte, weshalb sich deren Wahrscheinlichkeiten jeweils komplementär zu eins aufaddierten. Die Gesamtanzahl von Gleichungen (Ästen) belief auf 214.

Für jede (der sieben) BIBD-Bedingungen:	$2^4 = 16$ Gleichungen
Für jede (der drei) klassischen UCT-Bedingungen:	$2^5 = 32$ Gleichungen
Für jede (der drei) direkten Fragen:	$2^1 = 2$ Gleichungen
→ Gesamtanzahl: $7 * 16 + 3 * 32 + 3 * 2 = 112 + 96 + 6 = 214$ Gleichungen	

Beginnend mit der zweiten Zeile des Gleichungsfile wurden alle 214 zu spezifizierenden Äste untereinander aufgelistet. Jede Zeile enthielt drei Kodierungskomponenten, die jeweils durch ein Leerzeichen voneinander getrennt wurden. Der erste Wert, der aufgeführt wurde, war die Nummer des jeweiligen Teilbaumes. Für jede indirekte Bedingung wurde jeweils ein Teilbaum benötigt, im Falle der direkten Bedingungen war für jede kritische Frage ein Teilbaum vonnöten. Insgesamt ergaben sich somit 13 Teilbäume (10 für die indirekten und drei für die direkten Bedingungen). Die zweite Komponente war die einem spezifischen Ast zugehörige Kategoriennummer. Wie genau die Zuteilung der Kategoriennummern auf die verschiedenen Outputs (die möglichen Ja-Antwortsummen) erfolgte, wurde bereits im vorherigen Abschnitt erläutert. Der dritte und letzte Kodierungsbestandteil einer jeden Zeile bestand in der für jeden Ast individuellen Produktkombination der Parameter. Für die Kategorie 5 beispielsweise war diese das Produkt

aus n6, s1, s2 und n7. Sowohl das Datenfile als auch das Gleichungsfile wurden beendet, indem in einer abschließenden Zeile drei oder mehr Gleichheitszeichen aufgeführt wurden („==“). Das komplette Gleichungsfile ist im Anhang zu finden (Anhang C).

Abbildung 4 gibt einen Überblick über die 16 Äste des ersten Teilbaumes, der die erste Bedingung des (7, 4, 2)-BIBDs der eigenen Untersuchung spezifiziert. Dem einmaligen Ausgang „vier Ja-Antworten“ (alle Fragen wurden bejaht) wurde im ersten Teilbaum die Kategorienummer „fünf“ zugeordnet. Für die Outputs „drei Ja-Antworten“, „zwei Ja-Antworten“ und „eine Ja-Antwort“ existierten jeweils mehrere Gleichungen. Der erste Teilbaum schloß mit der wiederum nur einmaligen Kategorie „eins“ ab, da es nur eine Möglichkeit für die Antwortsumme „null“ gab.

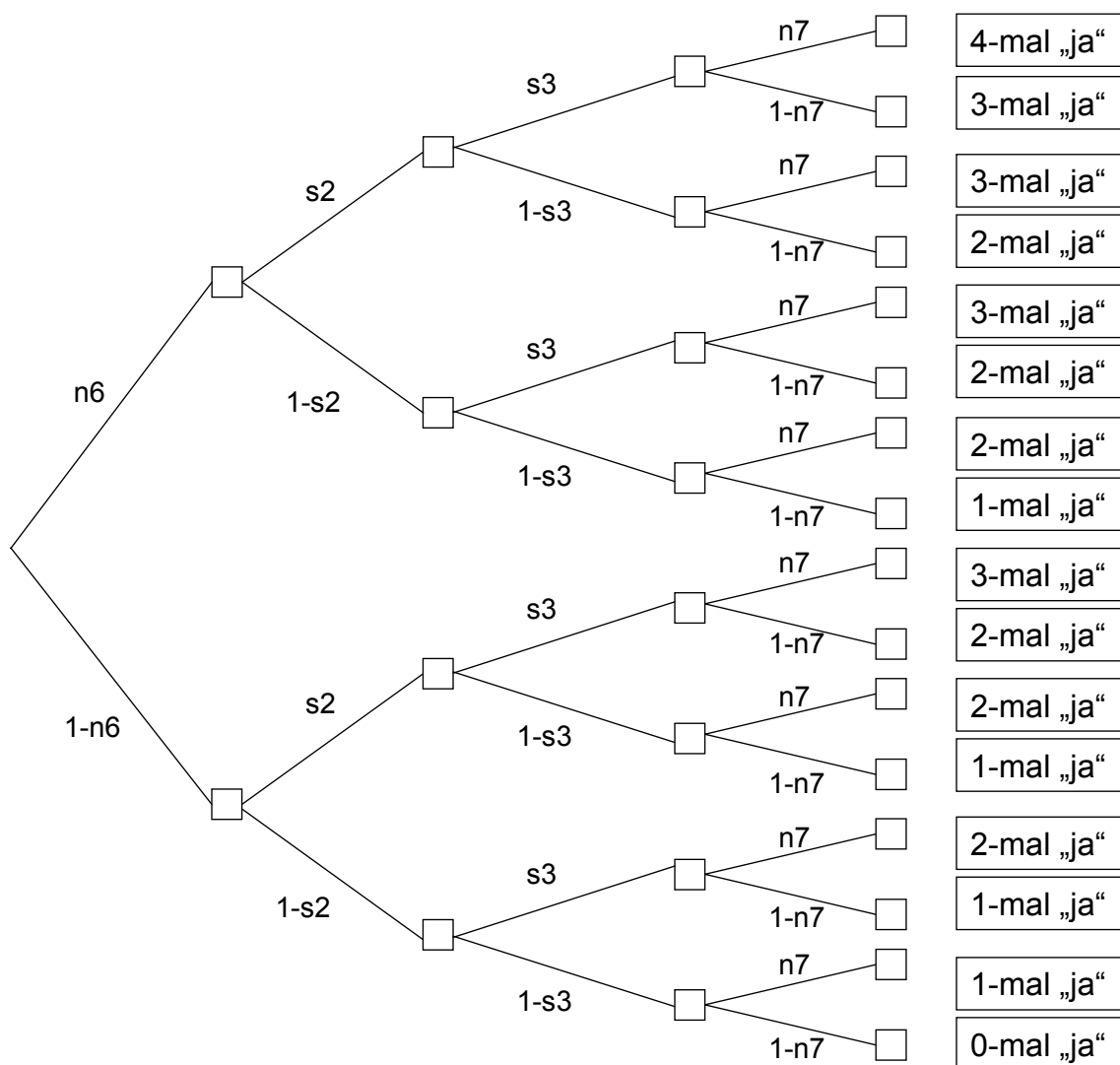


Abbildung 4: Übersicht über das multinomiale Modell der ersten Bedingung des (7, 4, 2)-BIBDs der eigenen Untersuchung. Die kategorialen Zuordnungen der Outputs waren wie folgt: Kat. 5 (viermal „ja“), Kat. 4 (dreimal „ja“), Kat. 3 (zweimal „ja“), Kat. 2 (einmal „ja“), Kat. 1 (keinmal „ja“). Die sensitiven Parameter sind mittels „s“, die nicht-sensitiven Parameter mittels „n“ gekennzeichnet.

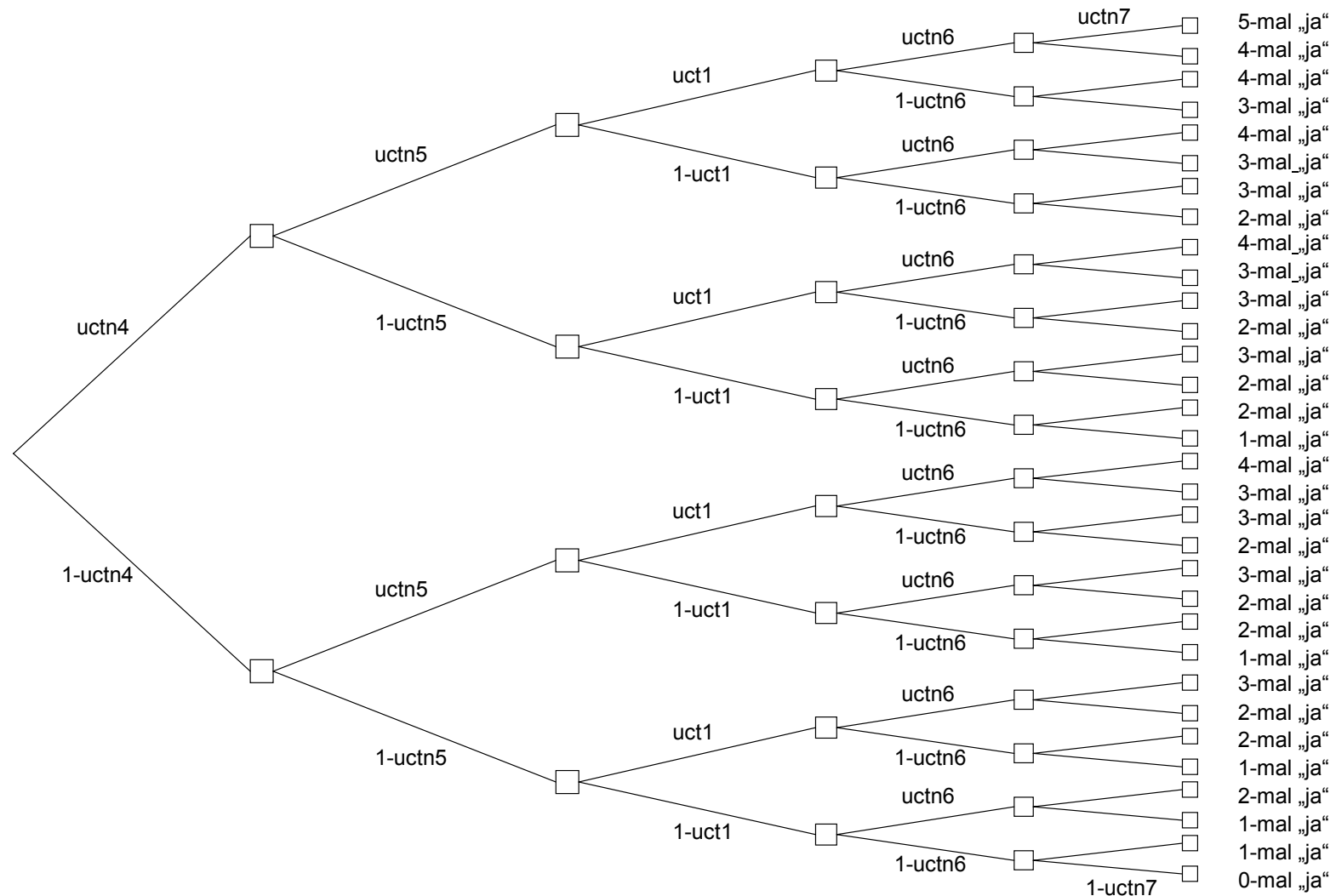


Abbildung 5: Übersicht über das multinomiale Modell der Bedingung 8 (erste klassische UCT-Experimentalbedingung) des eigenen Untersuchungsdesigns. Die kategorialen Zuordnungen der Outputs waren wie folgt: Kat. 41 (fünfmal „ja“), Kat. 40 (viermal „ja“), Kat. 39 (dreimal „ja“), Kat. 38 (zweimal „ja“), Kat. 37 (einmal „ja“) und Kat. 36 (keinmal „ja“). Die sensitiven Parameter sind mittels „uct“, die nicht-sensitiven Parameter mittels „uctn“ bezeichnet.

3.1.3 Auftretende Modellparameter in der eigenen Studie

Während die Spezifizierung der Äste in den BIBD-Bedingungen über die bereits eingangs eingeführten Parameter s_1 , s_2 und s_3 für die sensitiven Fragen bzw. die Parameter n_4 , n_5 , n_6 und n_7 für die nicht-sensitiven Fragen erfolgen konnte, war es vonnöten für die Gleichungen der klassischen UCT-Bedingungen bzw. der direkten Bedingungen zusätzliche Parameter-Indizes einzuführen. In den UCT-Bedingungen sollten uct_1 , uct_2 und uct_3 die Parameter der kritischen Fragen und $uctn_4$, $uctn_5$, $uctn_6$ und $uctn_7$ die Parameter der unkritischen Fragen kennzeichnen (vgl. Abbildung 5). In den direkten Bedingungen kamen d_1 , d_2 und d_3 zum Einsatz, wobei diese die Bejahung der jeweiligen kritischen Frage repräsentierten.

Während es in den indirekten Bedingungen im Zuge der Auswertung einer späteren Parameterschätzung bedurfte, um an die jeweiligen Prävalenzratschätzungen der sensitiven bzw. nicht-sensitiven Merkmale zu gelangen, waren die Schätzwerte für die direkten Befragungsbedingungen praktisch schon bekannt. Da die dritte Kodierungskomponente in den direkten Bedingungen – wie in Abbildung 6 zu sehen ist – nicht aus einer Produktkombination von Parametern, sondern lediglich aus einem Einzelparameter bzw. dessen Komplement bestand, stellten die in Tabelle 7 aufgeführten prozentualen Anteile der Ja-Antworten bereits die Basisratschätzungen für die sensitiven Verhaltensweisen dar.

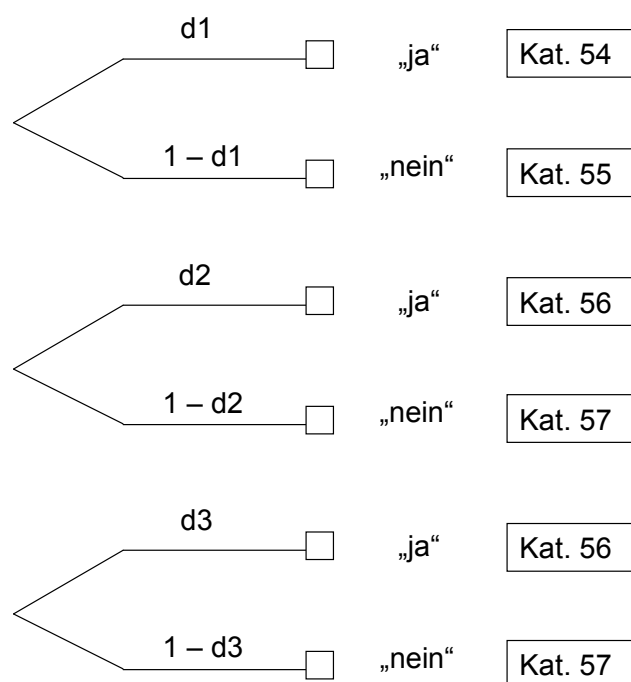


Abbildung 6: Übersicht über die multinomialen Teilbäume 11 bis 13, denen die Kategorien 54 bis 59 zugeordnet wurden. Als Index für die Parameter wurde „d“ verwendet, wobei die kritische Frage A durch d_1 , die kritische Frage B durch d_2 und die kritische Frage C durch d_3 repräsentiert wurde.

3.2 Vorgehen bei der Auswertung mit HMMTree

Nachdem die Files in vollständiger und korrekt formatierter Form vorlagen, konnten beide in HMMTree eingeladen werden. Hierzu war im Menü-Unterpunkt “File” über die Instruktion “Load equations” zuerst das Gleichungsfile aufzurufen. HMMTree verfügt über vier verschiedene Oberflächen, die mit den Labels “Parameters”, “Output”, “Model” und “Data” gekennzeichnet sind. Nach erfolgreicher Ladung des Gleichungsfile wurden im Menü “Parameters” alle Modellparameter in nicht restringierter Form aufgeführt. Die Parameter konnten im weiteren Verlauf als konstant festgelegt oder mit anderen Modellparametern gleichgesetzt werden. Über die Oberfläche “Model” ließ sich das gesamte Gleichungsfile einsehen. Nachdem in einem zweiten Schritt über die Anweisung “Load data” auch das Datenfile aufgerufen wurde, war unter der Oberfläche “Data” eine Zuordnung der beobachteten Häufigkeiten zu den Kategorienummern vorzufinden. Erst nachdem die Information beider Files erfolgreich integriert worden war, gab HMMTree auf der rechten Seite das Menü für die Berechnungen und Anpassungsgütetests frei. In selbigem Menü ließen sich nun ferner die Kennwerte des einbezogenen multinomialen Modells einsehen.

3.2.1 Kennwerte des zugrunde liegenden multinomialen Modells

Von den insgesamt fünf aufgeführten Modellkennwerten haben die meisten im Zuge der bisherigen Darstellung bereits Erwähnung gefunden. So wurde die Anzahl der Kategorien auf 59, die Anzahl der zu schätzenden Parameter auf 17 und die Anzahl der (Teil-) Bäume auf 13 beziffert. Der vierte Kennwert gab die maximale Anzahl von Ästen pro spezifischer Kategorie an. Innerhalb des zugrunde liegenden Modells der eigenen Untersuchung lag die maximale Anzahl von Ästen pro spezifischer Kategorie bei 10. Diese Maximalanzahl von Ästen wurde im Zuge der größten Teilbäume des Modells in den klassischen UCT-Experimentalbedingungen erreicht; so traten beispielsweise die Kategorien 38 (zweimal „Ja“) und 39 (dreimal „Ja“) des achten Teilbaums (vgl. Abbildung 5) insgesamt jeweils zehnmal auf. Der letzte von HMMTree ausgewiesene Modellkennwert war die Anzahl der freien Kategorienhäufigkeiten. Diese lässt sich wie folgt ermitteln:

$\text{Anzahl freier Kategorienhäufigkeiten} = \text{Anzahl der Kategorien} - \text{Anzahl der (Teil-)Bäume}$

Die Anzahl der freien Kategorienhäufigkeiten wurde von HMMTree auf 46 beziffert ($59 - 13$). Um die Anzahl der Freiheitsgrade des zugrunde liegenden multinomialen Modells zu bestimmen, musste die Anzahl der freien Kategorienhäufigkeiten um die Anzahl der zu

schätzenden Parameter vermindert werden; das nicht restringierte Modell wies demnach 29 Freiheitsgrade ($46 - 17$) auf.

3.2.2 Parameterrestriktionen des Grundmodells

Wie bereits im vorherigen Abschnitt (3.2.1) erwähnt, führte HMMTree alle Modellparameter zunächst in nicht restringierter Form auf. Bevor mit der Hypothesen-gerichteten Auswertung begonnen werden konnte, galt es noch eine wichtige Grundüberlegung mit in die Parameterdeklaration einzubeziehen. Es sollte nämlich vorausgesetzt werden, dass sich die nicht-sensitiven UCT-Parameter von den nicht-sensitiven BIBD-Parametern *nicht* unterscheiden. Diese Annahme begründete sich darin, dass die Hemmschwelle für die Bejahung einer unkritischen Frage derart niedrig ist, dass sie die Verwendung einer indirekten Befragungstechnik – gleich ob klassische UCT oder BIBD – obsolet macht. Die Daseinsberechtigung eines indirekten Befragungsverfahrens besteht darin, den Befragten bei sozial unerwünschten Items eine höhere Anonymität zu gewährleisten. Im Falle einer harmlosen Thematik muss der höhere Schutz, den ein indirektes Verfahren mit sich bringt, jedoch nicht in Anspruch genommen werden. Die Prävalenzratenschätzungen sollten sich von daher bei harmlosen Fragen für verschiedene indirekte Befragungsverfahren nicht unterscheiden.⁴⁴

Wurden die jeweils inhaltlich korrespondierenden, harmlosen Parameter der UCT- bzw. BIBD-Bedingungen gleichgesetzt ($uctn4 = n4$, $uctn5 = n5$, $uctn6 = n6$ und $uctn7 = n7$), so hatte HMMTree fortan „nur“ noch 13 Parameter ($17 - 4$) zu schätzen. Bedingt durch diese Modifikation wies das Grundmodell jetzt insgesamt 33 ($46 - 13$) Freiheitsgrade auf. Über den Button „Run“ konnte diese erstmalige Parameterschätzung ausgeführt werden. Über die „Output“-Oberfläche wurden die Ergebnisse angezeigt. Der Output beinhaltete die Namen von Modellgleichungs- und Datenfile sowie den Wert der asymptotisch χ^2 -verteilten log likelihood-Statistik G^2 für das berechnete Modell. Des Weiteren erfolgte eine Übersicht über die Basisratenschätzungen der einzelnen Parameter zuzüglich der Angabe ihrer Konfidenzintervalle.

Tabelle 8 gibt einen Überblick darüber, wie die Werte der einzelnen Parameter für das Grundmodell ausfielen. In den direkten Bedingungen stimmten die ML-Schätzer mit den bereits bekannten, beobachteten Häufigkeiten überein (vgl. Tabelle 7). Die angegebenen

⁴⁴ Zwar ist die Gleichsetzung der nicht-sensitiven Parameter von UCT und BIBD inhaltlich plausibel, aber es muss darauf hingewiesen werden, dass HMMTree das multinomiale Modell ohne diese Gleichsetzungen (mit dann 17 zu schätzenden Parametern) als nicht identifizierbar auswies („The core model ist not identified.“). Die Nicht-Identifizierbarkeit des Modells in seiner nicht restringierten Form soll im Zuge der späteren Diskussion noch einmal aufgegriffen werden.

Parameterschätzwerte d1 bis d3 lieferten von daher keine neue Information, sondern bestätigten vielmehr die ohnehin schon dokumentierten, beobachteten Häufigkeiten in den direkten Befragungsbedingungen. Bereits auf den ersten Blick konnte konstatiert werden, dass die Prävalenzratenschätzungen des Grundmodells in den direkten Bedingungen unerwartet hoch ausfielen und sich augenscheinlich von den Schätzwerten der indirekten Bedingungen nur geringfügig unterschieden.

Tabelle 8

Überblick über die Schätzungen der 13 Parameter sowie ihrer Konfidenzintervalle und Standardfehler für das multinomiale Grundmodell der eigenen Untersuchung

Bedingung		Parameterwerte	Unter- und Obergrenzen des 95 %-KIs ^x	Standardfehler
Direkte Befragung (n = 700)	1	d1 = 50,1 %	46,4 % - 53,8 %	0.019
	2	d2 = 26,0 %	22,8 % - 29,2 %	0.016
	3	d3 = 16,6 %	13,8 % - 19,3 %	0.014
Balanced-Incomplete-Block-Design (n = 700)	4	s1 = 52,0 %	42,6 % - 61,2 %	0.047
	5	s2 = 21,7 %	11,3 % - 32,0 %	0.053
	6	s3 = 24,3 %	13,8 % - 34,7 %	0.053
	7	n4 = 13,9 %	4,1 % - 23,8 %	0.051
	8	n5 = 44,8 %	35,7 % - 53,9 %	0.046
	9	n6 = 14,4 %	4,5 % - 24,4 %	0.051
	10	n7 = 14,8 %	4,8 % - 24,8 %	0.051
Unmatched-Count-Technik (n = 300)	11	uct1 = 52,1 %	31,0 % - 73,2 %	0.108 ^y
	12	uct2 = 25,7 %	3,5 % - 47,9 %	0.113 ^y
	13	uct3 = 15,0 %	-6,2 % - 36,3 %	0.109 ^y
	14	uctn4 gleich n4		
	15	uctn5 gleich n5		
	16	uctn6 gleich n6		
	17	uctn7 gleich n7		

^x: Die Abkürzung „KI“ steht für Konfidenzintervall. d1 bis d3 repräsentieren die Parameterwerte der direkten Befragungsbedingung, s1 bis s3 sowie n4 bis n7 bilden die Parameterschätzungen für das BIBD ab und uct1 bis uctn7 kennzeichnen die Parameterschätzungen für die klassische UCT.

^y: Im Gegensatz zu den direkten und den BIBD-Bedingungen mit einem Stichprobenumfang von jeweils n = 700, basieren die hier angegebenen UCT-Konfidenzintervalle und -Standardfehler auf einem Stichprobenumfang von n = 300. Von daher lassen sie sich *nicht* ohne Weiteres in Bezug zu den Standardfehlern der direkten Befragung und des BIBDs setzen.

Der G²-Wert des Grundmodells belief sich auf 83.98 und lag damit deutlich über dem kritischen Wert c_α der χ^2 -Verteilung bei 33 Freiheitsgraden von 47.40. Damit kam es zu einem

signifikanten “miss-fit” des Grundmodells. Aufgrund der Tatsache, dass ein Modell die realen Verhältnisse nur ungenau abzubilden imstande ist, sind Ungenauigkeiten im Grundmodell praktisch kaum zu vermeiden. Diese Ungenauigkeiten können sich bei großen Stichprobenumfängen nach Riefer & Batchelder (1988, S. 325) wie folgt auswirken:

A fourth problem is that the goodness-of-fit test...is frequently rejected if N is very large. This is because most substantive models are only approximations to reality and therefore are technically false. Because the power of the goodness-of-fit test increases with N, rejection of a good, but technically incorrect, model becomes very likely with very large N...Until some of this work becomes standard in the psychological modelling area, it seems reasonable to continue to analyze data with technically incorrect models if they explain much of the variance in the data.

Es ist sinnvoll, an dieser Stelle auch darauf hinzuweisen, dass eine wichtige Voraussetzung der multinomialen Modellierung in der Unabhängigkeitsannahme der Parameter besteht. Diese Annahme muss in multinomialen Modellen immer getroffen werden, obgleich deren Einhaltung kritisch zu sehen ist. So muss beispielsweise in der eigenen Untersuchung modelltheoretisch angenommen werden, dass es keinen kausalen Zusammenhang zwischen dem Status des Verheiratetseins und der Wohnform der Untermiete gibt (beide Fragen sind nicht-sensitive Kontrollfragen im eigenen Versuchsdesign), wenngleich deren Unabhängigkeit insofern bezweifelt werden kann, dass jemand der bereits verheiratet ist, wohl kaum zur Untermiete wohnen dürfte. Gemäß der Modelltheorie sollten zudem Reihenfolgeeffekte für die Beantwortung der Fragen keine Rolle spielen. In der Realität aber dürfte jemand, der vor der Beantwortung der letzten Frage des Fragenblockes steht, sehr wohl durch seine / ihre Antworten auf die vorhergehenden Fragen beeinflusst sein und somit *nicht* unabhängig antworten. Mit diesem kurzen Exkurs sollte keinesfalls der Wert der multinomialen Modellierung in Zweifel gestellt werden. Im Zuge der Interpretation der Ergebnisse ist es aber sinnvoll, sich der Voraussetzungen und Grenzen dieses Verfahrens bewusst zu sein.

3.2.3 Betrachtung der Standardfehler für die Parameter des Grundmodells

Es soll in der Folge nun zunächst auf die dritte Hypothese der eigenen Arbeit eingegangen werden, die vorhersagte, dass sowohl die Standardfehler der Parameterschätzungen der klassischen Unmatched-Count-Technik als auch die Standardfehler der Parameterschätzwerte des Balanced-Incomplete-Block-Designs höher ausfallen sollten als die Standardfehler der Prävalenzratenschätzungen der direkten Befragung. Des Weiteren wurde vermutet, dass das BIBD eine höhere relative Effizienz als die klassische UCT aufweist.

An dieser Stelle ist es zweckmäßig, die vierte Spalte der Tabelle 8 heranzuziehen, innerhalb der jeweils die Standardfehler der unterschiedlichen Parameterschätzwerte angegeben sind. Bei der Betrachtung der Standardfehler war es wichtig, nur diejenigen Standardfehler zueinander in Bezug zu setzen, die auf dem gleichen Stichprobenumfang basierten. Der durchschnittliche Standardfehler betrug für die direkten Parameterschätzwerte bei einem Stichprobenumfang von $n = 700$ circa 0.016, während der durchschnittliche Standardfehler der BIBD-Parameterschätzungen bei gleichem Stichprobenumfang etwa 0.050 betrug. Damit waren die Standardfehler in den BIBD-Bedingungen mehr als dreimal (präziser: 3,07-mal) so groß wie in den direkten Befragungsbedingungen, was einen deutlichen Effizienzverlust des BIBD gegenüber der direkten Befragung bedeutete, wie es in der dritten Hypothese auch vorhergesagt worden war.

Um den durchschnittlichen Standardfehler der klassischen UCT-Bedingungen in Relation zum durchschnittlichen Standardfehler der direkten Bedingungen setzen zu können, war es notwendig, diesen Vergleich auf Grundlage der Standardfehler gleicher Stichprobenumfänge durchzuführen, die jedoch im eigenen Versuchsdesign nicht gegeben waren: direkte Befragung ($n = 700$) und UCT ($n = 300$) wiesen hier nämlich heterogene Stichprobenumfänge auf. Da es sich im Zuge der direkten Bedingungen jedoch um sieben unabhängige Replikationen der gleichen Bedingung handelte, war es problemlos möglich, für die Gegenüberstellung der Standardfehler von UCT und direkter Befragung, nur drei der unabhängigen Replikationen der direkten Befragung heranzuziehen.⁴⁵ Der durchschnittliche Standardfehler betrug für die direkten Parameterschätzwerte bei einem reduzierten Stichprobenumfang von $n = 300$ ca. 0.025, während der durchschnittliche Standardfehler der klassischen UCT bei gleichem Stichprobenumfang etwa 0.110 ausmachte. Damit waren die Standardfehler in den UCT-Bedingungen mehr als viermal (präziser: 4,40-mal) so groß wie in den direkten Befragungsbedingungen. Somit konnte die von der dritten Hypothese vorhergesagte größere Ineffizienz der klassischen UCT gegenüber der direkten Befragung auch für die zweite indirekte Befragungstechnik bestätigt werden. Ferner ließen die

⁴⁵ Die hier aufgeführte Tabelle gibt einen Überblick über die für einen Stichprobenumfang von $n = 300$ angepassten Schätzwerte der drei Parameter der direkten Befragung zuzüglich ihrer Konfidenzintervalle und Standardfehler. Hierzu wurden per Zufallsauswahl drei der sieben direkten Bedingungen ausgewählt (nämlich die Bedingungen 11, 14 und 17). Die hier angegebenen Standardfehler der direkten Befragung konnten in dieser Form nun ins Verhältnis zu den Standardfehlern der UCT-Parameter (siehe Tabelle 8) gesetzt werden.

Bedingung		Parameterwerte	95 %-Konfidenzintervall	Standardfehler
Direkte Befragung ($n = 300$)	11, 14 und 17	d1 = 49,0 %	43,3 % - 54,7 %	0.029
		d2 = 23,7 %	18,9 % - 28,5 %	0.024
		d3 = 17,7 %	13,4 % - 22,0 %	0.022

Ergebnisse den Schluss zu, dass das BIBD eine höhere relative Effizienz als die klassische UCT aufweist⁴⁶.

3.2.4 Hypothesenkonforme Parameterrestriktionen

Nachdem die dritte Hypothese bestätigt werden konnte, soll es nun darum gehen, die erste und die zweite Hypothese zu prüfen. Inhaltlich besagten die beiden ersten Hypothesen, dass die Prävalenzratenschätzungen der indirekten Verfahren (in Hypothese 1 wird Bezug auf die klassische Variante der UCT und in Hypothese 2 auf das BIBD genommen) für die sensitiven Parameter höher ausfallen sollten als die Prävalenzratenschätzungen der direkten Befragung.

3.2.4.1 Grundsätzliches Vorgehen bei der Parameterrestriktion

Zur Prüfung der ersten und der zweiten Hypothese war es notwendig, jeden direkten Parameter einerseits mit seinem korrespondierenden (inhaltsgleichen) Parameter der UCT sowie andererseits mit seinem korrespondierenden (inhaltsgleichen) Parameter des BIBDs gleichzusetzen. Insgesamt ergaben sich auf diese Weise sechs Restriktionen, die jeweils sequentiell und unabhängig voneinander durchzuführen waren (erst wurde d1 mit uct1 gleichgesetzt, dann d2 mit uct2, danach d3 mit uct3, anschließend d1 mit s1, daraufhin d2 mit s2 und schlussendlich d3 mit s3). Bei jeder Restriktion wurde das Grundmodell um genau *einen* Parameter restringiert, so dass nun nicht mehr 13, sondern nur noch 12 Parameter zu schätzen waren. HMMTree berechnete für jede vorgenommene Restriktion ein neues G^2 , wobei dessen Wert immer in Bezug zum G^2 des Grundmodells (83.98) zu setzen war. Über die Differenz aus dem im Zuge der spezifischen Restriktion erhaltenen neuen G^2 's und dem G^2 des Grundmodells konnte jeweils eine Aussage darüber getroffen werden, ob sich infolge der Gleichsetzung der Parameter die Passung des Modells signifikant verschlechtert hatte. Da jede vorgenommene Restriktion mit der Minderung genau *eines* Freiheitsgrades im Vergleich zum Grundmodell einherging, war der kritische G^2 -Differenzenwert, der für eine signifikante Verschlechterung der Modellpassung überschritten werden musste – bei einem Signifikanzniveau von $\alpha = 5\%$ – der Wert 3.84. Zusätzlich zur Angabe des neu berechneten G^2 's gab HMMTree im Zuge jeder Restriktion die Werte der 12 zu schätzenden Parameter sowie der dazugehörigen Konfidenzintervalle an.

⁴⁶ Ein Vergleich der Standardfehler der klassischen UCT und des BIBDs war aufgrund der heterogenen Stichprobenumfänge (UCT: $n = 300$; BIBD: $n = 700$) nicht direkt möglich. Da die Standardfehler der beiden indirekten Befragungstechniken jedoch unabhängig voneinander zu den Standardfehlern der direkten Befragung ins Verhältnis gesetzt wurden und der Effizienzverlust gegenüber der direkten Befragung für die UCT (Faktor: 4,40) höher als für das BIBD (Faktor: 3,07) ausfiel, konnte hieraus die höhere (relative) Effizienz des BIBDs abgeleitet werden.

3.2.4.2 Die Ergebnisse der sechs Parameterrestriktionen im Einzelnen

Es sollen nun zuerst die Gleichsetzungen der direkten Parameter mit den korrespondierenden UCT-Parametern thematisiert werden, um eine Aussage über die Hypothese 1 treffen zu können. Im Anschluss daran folgen die Gleichsetzungen der direkten Parameter mit den korrespondierenden BIBD-Parametern, um den Wahrheitsgehalt der Hypothese 2 zu prüfen. Tabelle 9 gibt einen Überblick über den Ausgang aller sechs Parameterrestriktionen.

Wurden die Parameterschätzwerte der UCT-Parameter $uct1$, $uct2$ und $uct3$ zunächst auf deskriptiver Ebene mit den Parameterschätzungen der direkten Parameter $d1$, $d2$ und $d3$ verglichen, so fiel auf, dass sich die Schätzungen der inhaltsgleichen Parameter nur geringfügig voneinander unterschieden. Aus diesem Grund konnte bereits auf Basis der deskriptiven Werte vermutet werden, dass durch die Gleichsetzungen keine nennenswerten Änderungen der Passung des Modells zu verzeichnen sein dürften. Die Ergebnisse der drei Restriktionen bestätigten diese Vermutung. Sowohl für die Gleichsetzung von $d1$ und $uct1$ [$\Delta G^2(1) = 0.04$ n.s.] als auch für die Gleichsetzungen von $d2$ und $uct2$ [$\Delta G^2(1) = 0.001$ n.s.] sowie $d3$ und $uct3$ [$\Delta G^2(1) = 0.05$ n.s.] wurde der kritische Wert (3.84) jeweils deutlich verfehlt. Bemerkenswert war, dass die G^2 -Differenz für die zweite kritische Frage erst durch Angabe der dritten Nachkommastelle ersichtlich wurde. Von daher muss an dieser Stelle konstatiert werden, dass die erste Hypothese eindeutig nicht bestätigt werden konnte. Im Gegensatz zu den UCT-Validierungsstudien, die in Unterkapitel 1.2.2.1 vorgestellt wurden, bestätigten die Ergebnisse der eigenen Untersuchung die vorausgesagte Überlegenheit der Unmatched-Count-Methode, gegenüber der direkten Befragung höhere Basisratenschätzungen zu erzielen, somit *nicht*. Ähnlich wie für die UCT-Parameter soll nun auch für die BIBD-Parameter zunächst ein deskriptiver Blick im Vorfeld der eigentlichen Testung erfolgen. Während sich für die erste kritische Frage („Spickzettel“) ähnlich wie bei der UCT nur eine geringfügig höhere Prävalenzratenschätzung des BIBD gegenüber der direkten Befragung (52,0 % vs. 50,1 %) ergab, zeigten sich für die zweite und dritte sensitive Frage etwas deutlichere Unterschiede. Unerwartungsgemäß fiel die Basisratenschätzung des BIBD für die zweite kritische Frage („ärztliches Attest“) mit 21,7 % im Vergleich zur direkten Befragung (26,0 %) unterlegen aus. Dementgegen fand sich im Falle der dritten kritischen Frage („gefälschte Unterschrift“) für das BIBD mit 24,3 % eine höhere Basisratenschätzung im Vergleich zur direkten Befragung (16,6 %). Obwohl die Abweichungen der BIBD-Parameterschätzwerte von den direkten Prävalenzratenschätzungen höher ausfielen als die Abweichungen der UCT-Parameterschätzwerte von den direkten Schätzungen, konnten auch über die Gleichsetzungen der BIBD-Parameter mit den direkten Parametern keine signi-

Tabelle 9

Überblick über die Ausgänge der G^2 -Differenzentests für die Gleichsetzung der direkten Parameter mit den UCT- bzw. den BIBD-Parametern für jede der drei sensitiven Fragen

Bedingung	Kritische Frage 1		Kritische Frage 2		Kritische Frage 3	
	Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel zum Einsatz gebracht?		Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen, schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür keinen ärztlichen Grund gab?		Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?	
Direkte Befragung						
	Ja	50,1 %	Ja	26,0 %	Ja	16,6 %
	Nein	49,9 %	Nein	74,0 %	Nein	83,4 %
Unmatched-Count-Technik						
	Ja	52,1 %	Ja	25,7 %	Ja	15,0 %
	Nein	47,9 %	Nein	74,3 %	Nein	85,0 %
Balanced-Incomplete-Block-Design						
	Ja	52,0 %	Ja	21,7 %	Ja	24,3 %
	Nein	48,0 %	Nein	78,3 %	Nein	75,7 %
G²-Differenz (Veränderung der Modellpassung)						
	ΔG² (1): d1 = uct1	0.04 (n.s.)	ΔG² (1): d2 = uct2	0.001 (n.s.)	ΔG² (1): d3 = uct3	0.05 (n.s.)
	ΔG² (1): d1 = s1	0.14 (n.s.)	ΔG² (1): d2 = s2	0.89 (n.s.)	ΔG² (1): d3 = s3	2.57 (n.s.)

Für den G^2 -Differenzentest liegt der kritische χ^2 -Wert für eine signifikante Verschlechterung der Passung des Modells bei einem Freiheitsgrad und einem α von 5 % bei 3.84. Von daher ist ersichtlich, dass sich weder die Parameterschätzungen der UCT noch die Parameterschätzungen des BIBD zufallstechnisch von den direkten Parameterschätzungen unterscheiden. Zog man im Falle der Gleichsetzung der Parameter d3 und s3 eine einseitige Testung heran, so verfehlte die G^2 -Differenz 2.57 den kritischen Wert (2.71) nur marginal.

fikanten Verschlechterungen der Passung des Modells erzielt werden. Allerdings zeigte sich hier eine deutliche Zunahme der G^2 -Differenzwerte mit steigendem Grad der Sensitivität der Fragen⁴⁷. So ergab sich im Zuge der Gleichsetzung von d1 und s1 ein ΔG^2 von 0.14 (n.s.), für d2 = s2 betrug $\Delta G^2 = 0.89$ (n.s.) und im Falle der Gleichsetzung von d3 und s3 ein ΔG^2 von 2.57 (n.s.). In Bezug auf das letzte Ergebnis bot sich die Umformulierung der ursprünglich zweiseitigen Hypothese in eine gerichtete Hypothese an. Jedoch verfehlte das ΔG^2 von 2.57 den kritischen Wert [$\chi^2(1) = 2.71$; $\alpha = 10\%$] auch im Falle der einseitigen Testung, wenn auch nur marginal. Es bleibt an dieser Stelle zu vermerken, dass die letzte Restriktion (d3 = s3) den einzigen zufallskritisch eher unwahrscheinlichen Effekt der eigenen Studie aufzudecken imstande war. Denn die Wahrscheinlichkeit, allein aufgrund des Zufalls einen Unterschied zwischen den Basisratenschätzungen 16,6 % (direkt) und 24,3 % (BIBD) zu finden, lag bei lediglich 10,8 %. Tatsächlich ergab die einseitige Testung einen signifikanten Effekt für die dritte kritische Frage, wenn in das multinomiale Modell lediglich die direkten und die BIBD-Bedingungen einbezogen wurden, nicht aber die klassischen UCT-Bedingungen. In diesem Fall ergab sich für die Gleichsetzung von d3 und s3 ein ΔG^2 von 2.85, welches den kritischen Wert [$\chi^2(1) = 2.71$; $\alpha = 10\%$] überschritt. Dieses Ergebnis ist insofern berichtenswert, weil aufgrund der Nicht-Berücksichtigung der UCT-Bedingungen in diesem Fall keine Zusatzannahmen vonnöten waren, um das Modell zu identifizieren. Nichts desto trotz konnte auch die zweite Hypothese nicht bestätigt werden: das BIBD war ebenso wenig wie die klassische UCT imstande, höhere Prävalenzratenschätzungen als die direkte Befragung zu erzeugen.

3.3 Klassische Auswertung der Unmatched Count-Bedingungen

Die Tatsache, dass die vierte BIBD-Bedingung des eigenen Versuchsplans gleichzeitig auch die Kontrollgruppe für die UCT-Experimentalgruppen darstellte, machte es ohne größeren Aufwand möglich, die Basisratenschätzungen der UCT-Parameter nun auch einmal über die klassische Vorgehensweise zu ermitteln (vgl. Kap. 1.2.2). Hierzu war es notwendig, den Ja-Summen-Mittelwert der UCT-Kontrollgruppe (Bedingung 4) von den jeweiligen Ja-Summen-Mittelwerten der drei UCT-Experimentalgruppen (Bedingungen 8 bis 10) zu subtrahieren. Die auf diese Weise erhaltenen Basisratenschätzungen sind im folgenden Kasten illustriert, wobei

⁴⁷ Obwohl die G^2 -Differenzwerte mit steigendem Grad der Sensitivität der Fragen tatsächlich zunahmen, bleibt aber zu bedenken, dass die Abweichung der Prävalenzratenschätzungen zwischen direkter und BIBD-Bedingung für die zweite kritische Frage entgegen der Hypothese ausfiel; die direkte Prävalenzratenschätzung war hier nämlich höher als die BIBD-Schätzung (26,0 % vs. 21,7 %).

aus Gründen der Vergleichbarkeit auch die Basisratenschätzungen der multinomialen Auswertung aufgeführt wurden (vgl. Kap. 3.2.4.2).

a) Bestimmung der Mittelwerte der UCT-Bedingungen:

$$MW_{\text{Bed.4}}(\text{UCT-KG}): 0.42 * 0 + 0.42 * 1 + 0.13 * 2 + 0.03 * 3 + 0.00 * 4 = 0.77$$

$$MW_{\text{Bed.8}}(\text{UCT-EG}_1): 0.19 * 0 + 0.34 * 1 + 0.36 * 2 + 0.21 * 3 + 0.05 * 4 + 0.01 * 5 = 1.43$$

$$MW_{\text{Bed.9}}(\text{UCT-EG}_2): 0.26 * 0 + 0.44 * 1 + 0.21 * 2 + 0.08 * 3 + 0.01 * 4 + 0.00 * 5 = 1.14$$

$$MW_{\text{Bed.10}}(\text{UCT-EG}_3): 0.32 * 0 + 0.48 * 1 + 0.14 * 2 + 0.02 * 3 + 0.01 * 4 + 0.03 * 5 = 1.01$$

b) Ermittlung der Mittelwert-Differenzen:

$$uct1_{\text{klassisch}} = MW_{\text{Bed.8}} - MW_{\text{Bed.4}} = 1.43 - 0.77 = 0.66 \quad (uct1_{\text{multinomial}} = 0.52)$$

$$uct2_{\text{klassisch}} = MW_{\text{Bed.9}} - MW_{\text{Bed.4}} = 1.14 - 0.77 = 0.37 \quad (uct2_{\text{multinomial}} = 0.26)$$

$$uct3_{\text{klassisch}} = MW_{\text{Bed.10}} - MW_{\text{Bed.4}} = 1.01 - 0.77 = 0.24 \quad (uct3_{\text{multinomial}} = 0.15)$$

Wie der Vergleich der mittels unterschiedlicher Auswertungsformen erhaltenen Schätzungen erkennen lässt, fielen die UCT-Schätzwerte im Zuge der klassischen Auswertung deutlich höher als die UCT-Schätzwerte bei der multinomialen Auswertung aus. Aus diesem Grund war es sinnvoll, mittels des "Difference of Proportion"-Tests eine Testung auf Signifikanz durchzuführen. Anhand Tabelle 10 kann eingesehen werden, dass alle drei UCT-Parameterschätzungen im Falle der klassischen Auswertung auf einem Fehlerniveau von 5 % statistisch signifikant höher als in der direkten Befragung ausfielen.

Tabelle 10

Vergleich der Schätzungen der Zugeständnisraten für die kritischen Verhaltensweisen von direkter Befragung bzw. klassisch ausgewerteter Unmatched-Count-Technik

Messung	Direkte Befragung	Unmatched-Count-Technik	Z ^a
Nutzung eines präparierten Spickzettels	53,5 % ^b	66,0 %	2.57*
Unberechtigt ausgestelltes ärztliches Attest	27,5 % ^b	37,0 %	2.04*
Fälschung der elterlichen Unterschrift	14,5 % ^b	24,0 %	2.49*

^a: Der Z-Score repräsentiert den Ausgang des "Difference of Proportion Tests" und indiziert die statistische Signifikanz: *p < .05. Alle drei nicht-sensitiven UCT-Parameterschätzungen waren statistisch signifikant höher als die Schätzungen der direkten Befragung.

^b: Die hier aufgeführten Basisratenschätzungen der direkten Befragung resultieren aus einem Stichprobenumfang von n = 200 und weichen insofern von den Schätzwerten aller direkten Bedingungen (mit einem n von 700) ab [vgl. Tab. 8]. Der Stichprobenumfang von n = 200 kam in diesem Fall durch die Zufallsauswahl der direkten Bedingungen 13 und 16 zustande.

Die Ergebnisse der klassischen Vorgehensweise stimmten also nicht mit den multinomial erzielten Resultaten überein. Es muss zwar an dieser Stelle bedacht werden, dass die Ergebnisse der klassischen Auswertung auf einem deutlich geringeren Stichprobenumfang als die der multinomialen Auswertung basieren und insofern weniger repräsentativ sind. Nichts desto trotz sind die Ergebnisse deshalb berichtenswert, weil im Falle der klassischen Auswertung im Gegensatz zum multinomialen Modell *keine* Zusatzannahmen getroffen werden mussten.⁴⁸

3.4 Geschlechtsspezifische Auswertung

Aufgrund der Tatsache, dass innerhalb der Stichprobe ein relativ ausgeglichenes Geschlechterverhältnis erzielt werden konnte (52,3 % Männer vs. 47,7 % Frauen), sollen abschließend die Ergebnisse der geschlechtsspezifischen Auswertung aufgeführt werden. Ziel der geschlechtsspezifischen Auswertung war es, zu prüfen, inwieweit sich die vierte Hypothese, derzufolge Männer sensitive Verhaltensweisen signifikant häufiger als Frauen berichten – und dies unabhängig davon, ob der Erhebung eine direkte oder indirekte Befragung zugrunde liegt –, durch die eigene Erhebung bestätigen ließ.

3.4.1 Deskriptive Betrachtung der geschlechtsspezifischen Parameterschätzwerte

Die Darstellung der Ergebnisse der Parameterschätzwerte für Männer und Frauen soll zunächst in deskriptiver Weise erfolgen. Im Gegensatz zum Gesamtmodell mussten im Zuge des Grundmodells der geschlechtsspezifischen Auswertung doppelt so viele Parameter (insgesamt 26) geschätzt werden. Die Gleichsetzung der nicht-sensitiven UCT-Parameter $uctn_4$ bis $uctn_7$ mit den nicht-sensitiven BIBD-Parametern n_4 bis n_7 wurde wie schon für das Grundmodell der Gesamtauswertung vorgenommen, weil sich auch hier in der völlig unrestringierten Form das Modell nicht identifizieren ließ. Tabelle 11 gibt einen Überblick über die Parameterschätzungen der Basisraten der Frauen und Männer. Bei Betrachtung der Konfidenzintervalle fällt auf, dass diese im Rahmen der geschlechtsspezifischen Auswertung erheblich breiter als die Konfidenzintervalle in der Gesamtauswertung ausfielen, was sich dadurch erklärte, dass durch die Geschlechtsauftrennung nur jeweils halb so viele Personen pro Bedingung vorlagen ($n \approx 50$ statt $n = 100$).

⁴⁸ Leider ließ sich ein multinomiales Modell, in welches nur die direkten und die UCT-Bedingungen (Bed. 4 als UCT-Kontrolle, 8-10 als UCT-Experimentalgruppen), nicht aber die übrigen BIBD-Bedingungen (1-3 und 5-7) eingingen, deswegen nicht identifizieren, weil keine Erhebung der vier nicht-sensitiven Fragen im direkten Befragungsformat stattgefunden hatte und somit keine „Ankerwerte“ für die Schätzungen der nicht-sensitiven UCT-Parameter zur Verfügung standen.

Tabelle 11

Überblick über die Schätzungen der insgesamt 26 Parameter sowie ihrer Konfidenzintervalle, differenziert für das Grundmodell der Männer und Frauen

Bedingung		Parameterwerte (Männer)	95 %-KI ^x (Männer)	Parameterwerte (Frauen)	95 %-KI ^x (Frauen)
Direkt	1	d1 = 50,5 %	45,0 - 55,8 %	d1f = 49,9 %	44,8 - 55,0 %
	2	d2 = 28,7 %	23,8 - 33,6 %	d2f = 23,6 %	19,2 - 27,9 %
	3	d3 = 18,1 %	14,0 - 22,2 %	d3f = 15,2 %	11,5 - 18,8 %
BIBD	4	s1 = 49,1 %	35,6 - 62,7 %	s1f = 53,9 %	41,1 - 66,6 %
	5	s2 = 19,1 %	4,5 - 33,7 %	s2f = 23,3 %	8,8 - 37,8 %
	6	s3 = 25,4 %	10,5 - 40,2 %	s3f = 23,0 %	8,5 - 37,5 %
	7	n4 = 7,3 %	-5,3 - 19,8 %	n4f = 7,3 %	4,3 - 33,0 %
	8	n5 = 53,7 %	42,6 - 64,9 %	n5f = 37,9 %	24,1 - 51,6 %
	9	n6 = 17,2 %	2,9 - 31,5 %	n6f = 13,4 %	0,0 - 27,1 %
	10	n7 = 15,1 %	0,0 - 29,3 %	n7f = 16,5 %	2,5 - 30,6 %
UCT	11	uct1 = 57,1 %	28,7 - 85,6 %	uct1f = 41,4 %	9,8 - 73,0 %
	12	uct2 = 18,7 %	-10,0 - 47,4 %	uct2f = 31,0 %	-1,6 - 63,5 %
	13	uct3 = 3,6 %	-20,8 - 28,1 %	uct3f = 17,0 %	-14,4 - 48,4 %
	14	uctn4 gleich n4		uctn4f gleich n4f	
	15	uctn5 gleich n5		uctn5f gleich n5f	
	16	uctn6 gleich n6		uctn6f gleich n6f	
	17	uctn7 gleich n7		uctn7f gleich n7f	

^x: Die Abkürzung „KI“ steht für Konfidenzintervall. Die Parameterbezeichnungen der Männer entsprechen den Parametersignaturen des Gesamtmodells. Für die Frauen wurde jeweils an die üblichen Parameterbezeichnungen der Buchstabe „f“ angehängt.

Den Parameterschätzwerten der unterschiedlichen Befragungstechniken lagen die folgenden Stichprobenumfänge zugrunde: direkte Befragung (369 Männer, 331 Frauen), BIBD (376 Männer, 324 Frauen) und UCT (144 Männer, 156 Frauen)

Die Prävalenzratenschätzungen der Frauen und Männer wichen im Falle der direkten sowie der sensitiven BIBD-Parameter nicht erheblich voneinander ab. Während die Basisratenschätzungen der Männer im Falle der direkten Parameter zumindest hypothesenkonform ausfielen, indem die Männer leicht erhöhte Raten gegenüber den Frauen verzeichnen konnten, setzte sich diese Tendenz im Falle der sensitiven BIBD-Parameterschätzungen nicht fort. Vielmehr konnten die Frauen im Falle der ersten beiden kritischen Fragen um jeweils 4 % höhere Prävalenzratenschätzungen als die Männer vorweisen (53,9 % vs. 49,1 % im Falle der „Spickzettel“-Frage bzw. 23,3 % vs. 19,1 % bei der Frage nach dem „ärztlichen Attest“). Nur die dritte sensitive BIBD-Parameterschätzung

(„gefälschte Unterschrift“) fiel bei den Männern mit 25,4 % leicht höher als bei den Frauen (23,0 %) aus. Quantitativ deutlichere Unterschiede in den Schätzungen der Prävalenzraten fanden sich dagegen für die drei kritischen UCT-Parameter $uct1$ bis $uct3$ im Vergleich zu $uct1f$ bis $uct3f$. Diese beachtlichen Differenzen der UCT-Schätzwerte für Frauen und Männer fielen allerdings neuerlich uneinheitlich aus. Konnten die Männer im Zuge der ersten kritischen Frage („Spickzettel“) mit 57,1 % eine deutlich höhere Basisratenschätzung als die Frauen (mit 41,4 %) verbuchen, wiesen die Frauen im Falle der zweiten und dritten kritischen Frage die höheren Prävalenzratenschätzungen auf. Bei der Frage nach dem „unberechtigterweise ausgestellten ärztlichen Attest“ dominierte die Basisratenschätzung der Frauen mit 31,0 % (gegenüber nur 18,7 % bei den Männern) ebenso wie bei der Frage nach der „gefälschten Unterschrift der Eltern“ (17,0 % der Frauen gegenüber 3,6 % der Männer).

3.4.2 Interferenzstatistische Absicherung der auffälligen Geschlechterdifferenzen

Während aufgrund der nur geringfügigen geschlechtsspezifischen Unterschiede für die direkten und die sensitiven BIBD-Parameter auf eine interferenzstatistische Absicherung verzichtet werden konnte, wurde eine solche für die deskriptiv auffälligen Differenzen der drei sensitiven UCT-Parameter für sinnvoll erachtet. Hierzu war es notwendig, ein neues Gleichungs- und ein neues Datenfile zu generieren. Das geschlechtsbezogene Gleichungsfile war nun doppelt so lang wie das Gleichungsfile der Gesamtauswertung, weil es jeweils separate Gleichungen für Männer und Frauen enthielt (pro Geschlecht 214, insgesamt also 428). In diesem Zusammenhang war es vonnöten, die jeweiligen Parameter für Frauen und Männer unterschiedlich zu benennen; wie schon in Tabelle 11 berücksichtigt, wurden die Parameter der Frauen jeweils durch ein an den ursprünglichen Parameternamen angehängtes „f“ von den Parametersignaturen der Männer unterschieden. Ebenso wie das Gleichungsfile stellte sich auch das Datenfile für die Geschlechterauswertung als doppelt so lang heraus, weil es nacheinander erst die beobachteten Häufigkeiten bei den Männern und dann die beobachteten Häufigkeiten bei den Frauen auflistete. Die Anzahl der benötigten Kategorien verdoppelte sich insofern von ursprünglich 59 auf 118. Wurden die für die Geschlechterauswertung modifizierten Files nun in HMMTree eingeladen und die a priori-Parameterrestriktionen des Grundmodells vorgenommen ($uctn4 = n4$, $uctn4f = n4f$ etc.), so registrierte HMMTree 26 Teilbäume, 92 freie Kategoriehäufigkeiten und maximal zehn Äste pro Kategorie. Für das G^2 des geschlechtsspezifischen Grundmodells ergab sich ein Wert von 146.23, der den kritischen Wert c_α von 85.97 (bei nunmehr 66 Freiheitsgraden) wie schon in Gesamtauswertung deutlich überschritt, so dass auch im Falle der geschlechtsspezifischen Auswertung ein signifikanter „miss-fit“ des Grundmodells konstatiert werden musste.

Um zu prüfen, inwieweit sich der Modellfit durch Gleichsetzung bestimmter Parameter von Frauen und Männern verschlechtern würde, wurden nacheinander in voneinander unabhängigen Durchgängen die Parameter $uct1$ und $uct1f$, $uct2$ und $uct2f$ sowie $uct3$ und $uct3f$ gleichgesetzt. Trotz der deskriptiv durchaus beachtlichen Differenzen in den Basisratenschätzungen für die Frauen und die Männer (für die erste kritische Frage ca. 16 %, für die zweite kritische Frage etwa 12 % und für die dritte kritische Frage ungefähr 13 %), wurde im jeweiligen G^2 -Differenztest für keinen der drei Vergleiche der kritische χ^2 -Wert für einen Freiheitsgrad von 3.84 (bei einem α von 5 %) überschritten. So ergab sich im Zuge der Gleichsetzung von $uct1$ und $uct1f$ ein ΔG^2 von 0.631 (n.s.), für $uct2 = uct2f$ ein ΔG^2 von 0.376 (n.s.) und im Falle der Gleichsetzung von $uct3$ und $uct3f$ ein $\Delta G^2 = 0.536$ (n.s.). Eine plausible Erklärung für die deutliche Nicht-Signifikanz der drei Vergleiche ist mit Sicherheit in den merklich größeren Standardfehlern, die aus den geringeren Stichprobengrößen pro Bedingung ($n \approx 50$) im Vergleich zur Gesamtauswertung resultierten, zu sehen. Es ist von daher sinnvoll, kritisch anzumerken, dass das eigene Versuchsdesign für die Aufdeckung eines geschlechtsspezifischen Effektes nicht geeignet war, da durch die Geschlechtsauftrennung die Stichprobengrößen pro Bedingung zu klein ausfielen.

Es soll nun abschließend noch einmal Bezug auf die vierte Hypothese genommen werden. Aufgrund der Tatsache, dass die Prävalenzratenschätzungen für Frauen und Männer im Falle der direkten und der sensitiven BIBD-Parameter tendenziell indifferent und im Falle der sensitiven UCT-Parameter heterogen ausfielen, muss die vierte Hypothese der eigenen Erhebung als widerlegt angesehen werden. Die Zugaberaten der Männer für die kritischen Verhaltensweisen waren überwiegend nicht höher als die der Frauen, in einigen Fällen fielen sie sogar niedriger aus.

4 Diskussion

Im abschließenden Teil sollen die Ergebnisse der eigenen Untersuchung nun vor dem Hintergrund der theoretischen Fragestellung bewertet und kritisch reflektiert werden. Auf Grundlage dessen sollen Implikationen für die künftige Umfrageforschung getroffen und ein Ausblick auf potentielle Weiterentwicklungen der vorgestellten indirekten Befragungstechniken gegeben werden. Der Diskussionsteil endet mit einer Debatte über die Bedeutsamkeit des Datenschutzes sensibler Daten in der heutigen Zeit.

4.1 Evaluation der Ergebnisse der eigenen Untersuchung

Die eigene Untersuchung hatte den Anspruch, Verfahren aufzuzeigen, mit deren Hilfe sich die Basisratenschätzungen sensibler Befragungsinhalte valider erfassen ließen, indem den Befragten eine höhere Anonymität bei der Beantwortung der kritischen Fragen zugestanden wurde. Die Ergebnisse der eigenen Studie lassen letzten Endes keine andere Interpretation zu als die, dass sich keines der beiden indirekten Verfahren – also weder die klassische UCT noch das innovative BIBD – im Hinblick auf eine höhere Validität der Prävalenzerfassung sensibler Merkmale bewähren konnte. Trotz der verfahrensbedingt gewährten höheren Anonymität fielen die Basisratenschätzungen für beide indirekten Befragungstechniken im Zuge der multinomialen Auswertung nicht signifikant höher als die der direkten Befragung aus. Im Falle der klassischen UCT bleibt allerdings zu konstatieren, dass die konventionelle Auswertung über die Differenzbildung der Mittelwerte von Experimental- und Kontrollgruppe signifikante Ergebnisse beisteuerte. Des Weiteren konnte gezeigt werden, dass der Effizienzverlust des BIBDs gegenüber der direkten Befragung im Vergleich zur klassischen UCT geringer ausfiel und insofern eine Verbesserung der Varianzproblematik erzielt werden konnte. Alles in allem kann aber der Effizienzverlust eines indirekten Verfahrens nur dann in Kauf genommen werden, wenn das Primärziel – nämlich die Generierung von Prävalenzratenschätzungen höherer Validität im Zuge einer stärkeren Ausschöpfung der Dunkelziffer – erreicht werden kann. Da die Ergebnisse der eigenen Untersuchung diesem Kriterium nicht gerecht wurden, bleibt zu hinterfragen, worin die Ursachen für diese unzufrieden stellenden Resultate zu sehen sind.

4.1.1 Abgleich der eigenen Prävalenzratenschätzungen mit denen der Literatur

An dieser Stelle soll zunächst ein Vergleich der eigenen, für die einzelnen kritischen Verhaltensweisen erzielten Basisratenschätzungen mit denen der vorherigen Literaturrecherche vorgenommen werden.

Für die Frage mit dem geringsten Sensitivitätsgrad („Haben Sie schon einmal in einer Prüfung einen präparierten Spickzettel benutzt?“) ermittelten Hrabak et al. (2004) eine Basisratenschätzung von 34,6 % bzw. Lin & Wen (2007) eine Schätzung von 27,6 %. In der eigenen Untersuchung wurden diese Prozentwerte deutlich übertroffen. So ergaben sich für alle drei Befragungstechniken Prävalenzratenschätzungen von knapp über 50 %. Ein ähnliches Bild zeigte sich auch im Hinblick auf die mittelmäßig sensitive Frage nach dem „unrechtmäßigen Ausstellenlassens eines ärztlichen Attestes“. Während Schmelkin et al. (2001) eine Basisratenschätzung von knapp 9 % für die Verhaltensweise, einen Prüfungstermin bzw. die Abgabefrist einer Hausarbeit durch Angabe eines fiktiven Grundes verzögert zu haben, berichteten, lagen die Schätzwerte in der eigenen Untersuchung für die direkte Befragung und die klassische UCT nahe 25 % bzw. für das BIBD bei ca. 21 %. Auch wenn die Schätzwerte der zweiten kritischen Frage aufgrund der Tatsache, dass sich die Fragen in Literatur und eigener Studie inhaltlich geringfügig voneinander unterscheiden, nur bedingt vergleichbar sind, übertrafen die eigens ermittelten Prävalenzratenschätzungen auch in diesem Fall den Schätzwert aus der Literatur deutlich. Da für die höchst-sensitive Frage der eigenen Untersuchung („Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?“) in der Literatur keine inhaltlich annähernd äquivalente Basisratenschätzung gefunden werden konnte, ließ sich ein quantitativer Prozentwertvergleich für die dritte kritische Frage nicht durchführen.

Nichts desto trotz lassen die Ergebnisse des Abgleichs der eigens ermittelten Prävalenzratenschätzungen mit denen der Literatur eindeutig die Schlussfolgerung zu, dass sich die befragten Studentinnen und Studenten in der eigenen Untersuchung sehr bereitwillig zu den kritischen Verhaltensweisen bekannten, und dies auch in Abwesenheit einer höheren Gewähr von Anonymität in den direkten Befragungsbedingungen. Es bleibt zu ergründen, worin die Ursachen für die bereitwilligen Zugaberraten bereits im Zuge der direkten Befragung zu sehen sind.

4.1.2 Potentielle Gründe für die hohen PR-Schätzungen in der direkten Befragung

Es liegt an dieser Stelle nahe, den folgenden Einwand kritisch zu hinterfragen: Lassen sich die hohen Prävalenzratenschätzungen in den direkten Befragungsbedingungen dadurch erklären, dass die für sensitiv erachteten Fragen in der Wahrnehmung der Befragten nicht sensitiv genug waren, um sie in ihrem Antwortverhalten ernsthaft in Bedrängnis zu bringen? Denn wie bereits eingangs betont wurde, kann eine indirekte Befragungstechnik ihre Qualität, höhere und damit validere Basisratenschätzungen im Vergleich zu einer direkten Befragung

zu erzielen, nur dann sicherstellen, wenn die Befragungsinhalte ein ausreichendes Maß an Sensitivität aufweisen.

Um zu prüfen, inwieweit die für die eigene Untersuchung ausgewählten kritischen Fragen tatsächlich sensitiv genug sind, wäre es hilfreich gewesen, im Vorfeld der eigenen Erhebung einen Prätest durchzuführen, auf Grundlage dessen ein Vorab-Urteil über den Grad der Sensitivität der einzelnen Fragen möglich gewesen wäre. Zwar existiert kein formales Kriterium darüber, ab welcher Prävalenzrate – oder besser: Prävalenzratenschätzung – ein Verhalten als ausreichend sensitiv zu klassifizieren ist. Aber grundsätzlich dürfte eine Frage (wie die nach der „Benutzung eines präparierten Spickzettels“ in der eigenen Studie), die bereits jeder zweite in der direkten Befragungsbedingung bekräftigte, eine zu hohe Prävalenz für den Versuch, die potentielle Überlegenheit eines indirekten Befragungsverfahrens nachzuweisen, offenbaren.⁴⁹ Für solch eine Vorab-Einschätzung wäre ein Prätest sicherlich dienlich gewesen.

Um ein zügiges Ausfüllen des Fragebogens zu gewährleisten, dem aufgrund des großen Stichprobenumfanges, der benötigt wurde, eine hohe Bedeutung zukam, wurde die Erhebung von demographischen Daten in der eigenen Studie auf die Geschlechtsvariable begrenzt. Ein potentieller Einwand in diesem Zusammenhang besteht darin, dass durch den Verzicht auf die Erhebung weiterer demographischer Variablen (wie Alter, Studiengang, Semesterzahl und Veranstaltungstitel) eine höhere subjektiv empfundene Anonymität *unabhängig* von der eigentlichen Befragungsbedingung geschaffen wurde.⁵⁰ Ein ebenfalls wichtiger Aspekt, den es in diesem Zusammenhang zu bedenken gilt, ist die Tatsache, dass den Teilnehmerinnen und Teilnehmern durch die Vorabinformation beim Ausfüllen des Bogens bereits bewusst war, für welchen Zweck die Daten erhoben wurden. Das Bedrohlichkeitspotential der gleichen Umfrage hätte einen anderen Charakter gehabt, wenn nicht ein Kommilitone im Zuge seiner Diplomarbeitserhebung, sondern ein politisch oder juristisch agierender Hochschulausschuss durchführendes Organ der Befragung gewesen wäre. Das Zugeständnis sozial unerwünschter Verhaltensweisen war im Zuge der eigenen Erhebung offenkundig nicht mit negativen Konsequenzen verbunden. Es darf vermutet werden, dass die Zugaberaten in den direkten Befragungsbedingungen nicht vergleichbar hoch ausgefallen wären, wenn von den Teilnehmern ernsthafte Sanktionen befürchtet worden wären.

⁴⁹ Hintergrundüberlegung ist hier die, dass ein sozial unerwünschtes Verhalten grundsätzlich eher niedrig prävalent ausfallen sollte; die zu hohe Verbreitung eines Merkmals „gefährdet“ seine Sensitivität.

⁵⁰ Denn die Wahrscheinlichkeit, dass ein in einer Großveranstaltung ausgefüllter Fragebogen allein aufgrund der Angabe des Geschlechts von einem Versuchsleiter / einer Versuchsleiterin mit der eigenen Person in Verbindung gebracht werden kann, dürfte von den Befragten als niedriger eingestuft werden, als wenn dem Versuchsleiter / der Versuchsleiterin weitere soziodemographische Daten über die eigene Person vorliegen würden.

4.1.3 Rückbezugnahme auf die bisherige Studie zum BIBD

Es soll an dieser Stelle noch einmal Rückbezug auf die in Kap. 1.4 dargestellte BIBD-Studie von Benesch (2008) genommen werden, die im Gegensatz zur eigenen Untersuchung ein alternatives BIBD, nämlich das (6, 3, 2)-Design in Bezug auf die Thematik „Missachtung der Privatsphäre des Partners / der Partnerin“ anwendete und für zwei der drei kritischen Fragen auf einem α -Niveau von 1 % signifikant höhere Basisratenschätzungen im Vergleich zur direkten Befragung aufzeigen konnte. Zweifelsohne rücken die Ergebnisse der Studie von Benesch (2008) die Resultate der eigenen Untersuchung in ein anderes Licht. Es scheint sich die Vermutung zu erhärten, dass in der eigenen Untersuchung das Scheitern der indirekten Befragungstechniken, signifikant höhere Basisratenschätzungen als die direkte Befragung zu erzielen, nicht auf die Untauglichkeit der verwendeten indirekten Verfahren zurückzuführen ist, sondern dass vielmehr der Befragungskontext von den Teilnehmern als nicht ausreichend sensitiv empfunden wurde. Letztendlich scheint die Auswahl des Themengebiets der „Missachtung der Privatsphäre des Partners / der Partnerin“ eine günstigere gewesen zu sein, die die Befragten in scheinbar stärkere Gewissenskonflikte als die Auseinandersetzung mit ihren „schulisch-akademischen Täuschungsversuchen“ geführt hatte. Ein Aspekt, der allerdings in diesem Zusammenhang nicht unberücksichtigt bleiben darf, ist der, dass in der Studie von Benesch (2008) die Fragen im Konjunktiv formuliert wurden („Würden Sie ...?“), während in der eigenen Untersuchung eine Formulierung in der Vergangenheitsform („Haben Sie bereits...?“) gewählt worden war. Diese Unterscheidung ist deswegen nicht unerheblich, weil jemand, der ein kritisches Verhalten im Konjunktiv bekräftigt („Ja, ich würde...“) im Gegensatz zu jemandem, der ein Verhalten der Vergangenheit zugibt, im Nachhinein immer noch die Möglichkeit hat, zu behaupten, dieses Verhalten noch nie in die Tat umgesetzt zu haben. Insofern lässt eine Aussage im Konjunktiv einer befragten Person einen größeren Spielraum als eine Aussage, die in der Vergangenheitsform formuliert ist. Es muss also unter Validitätsaspekten angemerkt werden, dass in beiden Studien nicht dasselbe Konstrukt gemessen wurde, sondern einmal ein bereits begangenes und zum anderen ein potentiell kritisches Verhalten, welches nicht notwendigerweise bereits begangen wurde.

Für die weitere Evaluation der beiden BIBD-Studien bietet es sich an dieser Stelle an, noch einmal Rückbezug auf die in Kap. 2.1.1 aufgeführten Auswahlkriterien bei der Entscheidung für ein bestimmtes Teilverschachtelungsdesign heranzuziehen. Ein grundlegendes Merkmal des von Benesch (2008) verwendeten (6, 3, 2)-BIBDs im Vergleich zum (7, 4, 2)-BIBD der eigenen Untersuchung bestand darin, dass pro Fragenblock nur drei (und nicht vier) Fragen gestellt wurden. Die Maximierung der Vertraulichkeit war von daher

im (6, 3, 2)-Design eine schlechtere, da in einigen Blöcken nur *eine* nicht-sensitive „Pufferfrage“ vorhanden war; im (7, 4, 2)-Design waren es dagegen mindestens *zwei*. Werden das (6, 3, 2)- und das (7, 4, 2)-BIBD unter Effizienzgesichtspunkten einander gegenüber gestellt, so fällt auf, dass die BIBD-Standardfehler bei Benesch (2008) im Vergleich zur eigenen Studie etwas geringer ausfielen, da hier aufgrund der Beschränkung auf drei Fragen pro Block weniger Zusatz-Varianz erzeugt wurde (vgl. Tab. 4). Waren die BIBD-Standardfehler in der eigenen Untersuchung im Schnitt 3,07-mal größer als die Standardfehler der direkten Befragung, so fielen diese im Falle des (6, 3, 2)-BIBDs zumindest für die ersten beiden kritischen Fragen deutlich geringer aus (2,24- bzw. 2,02-mal größer als in der direkten Befragung). Allerdings wies die dritte kritische Frage bei Benesch (2008) eine wesentlich schlechtere Effizienz auf: so waren die Standardfehler hier 4,4-mal größer als die der direkten Befragung. Ein solcher Wert war in der eigenen Untersuchung nur für die klassische UCT konstatiert worden, die in punkto Effizienz ja noch deutlich schlechter als das BIBD abgeschnitten war (vgl. Kap. 3.2.3). Im Hinblick auf die statistische Power ist die Anwendung des (6, 3, 2)-Designs Benesch (2008) insofern zu kritisieren, dass mit einem n von 50 pro BIBD-Bedingung die Besetzungszahl im Vergleich zur eigenen (7, 4, 2)-Anwendung mit $n = 100$ pro BIBD-Block deutlich niedriger ausfiel.

4.1.4 Interpretation der Fehlanpassung des multinomialen Grundmodells

Wie an früherer Stelle bereits vermerkt wurde, kam es sowohl für das multinomiale Gesamtmodell als auch für das multinomiale geschlechtsspezifische Modell zu einem signifikanten „miss-fit“ des Grundmodells. Auch wenn unter Bezugnahme auf Riefer & Batchelder (1988) das Scheitern der Modellanpassung insofern relativiert wurde, dass mit wachsendem Stichprobenumfang und damit einhergehender steigender Power die Ablehnung des Grundmodells zunehmend wahrscheinlicher wird, darf dieser „miss-fit“ keinesfalls bagatellisiert werden. Wie bereits in Kap. 1.3.8 erläutert wurde, lag der eigenen Untersuchung ein einfaches multinomiales Modell, basierend auf der Annahme, dass die Parameter über verschiedene Personen hinweg *nicht* variieren, zugrunde.

Stahl & Klauer (2007) verweisen darauf, dass die Ursachen einer Modell-Fehlanpassung unterschiedlich sein können. Ein schlechter „fit“ kann einerseits daraus resultieren, dass die Modellstruktur die Realität (im Falle des eigenen Designs also die Antwortprozesse der Partizipanten) nicht gut genug zu beschreiben vermag. Der schlechte „fit“ kann aber ebenfalls seine Ursache darin haben, dass das Modell sehr wohl die Struktur der Antwortprozesse einer Person zu beschreiben imstande ist, aber die Parameter über verschiedene Personen hinweg variieren. Die traditionellen multinomialen Modelle sind nicht

in der Lage, zwischen diesen beiden Situationen zu differenzieren: bei vorhandener Parameterheterogenität kommt es bei einfachen multinomialen Modellen vielfach zu einer Fehlanpassung, obwohl sie die Antwortstruktur eines jeden Individuums akkurat zu beschreiben imstande sind. Gerade im Hinblick darauf, dass sich die Stichprobe der eigenen Untersuchung aus einer sehr breiten Anzahl von Studierenden verschiedener Fachbereiche zusammensetzte (vgl. Kap. 2.3.1), scheinen Zweifel an der Gültigkeit der Homogenitätsannahme nicht unberechtigt zu sein.

4.2 Aspekte der eigenen Studie mit Optimierungspotential

Im Folgenden werden einige Aspekte der eigenen Untersuchung im Hinblick auf ihr Verbesserungspotential für Folgestudien betrachtet. In diesem Zusammenhang sollen eine Modifikation der Instruktion zur Anwendung der indirekten Verfahren, eine homogenere Zusammensetzung der Stichprobe, die Verringerung der inhaltlichen Salienz, die explizitere Angabe des Bezugszeitpunktes der kritischen Fragen, die zusätzliche Erhebung der unkritischen Fragen in den direkten Bedingungen, die stärkere Einbeziehung von Dozenten in die Datenerhebung bzw. die Anwendung einer onlinegestützten Befragung diskutiert werden.

4.2.1 Verstärkung der Hinweise zur korrekten Anwendung von UCT und BIBD

Wie bereits in Kapitel 2.4.3.2 thematisiert wurde, war es bei ca. 9 % der in der Ersterhebung befragten Probanden der indirekten Befragungsbedingungen notwendig, ihre Bogen aufgrund der fehlerhaften Anwendung der klassischen UCT bzw. des BIBDs wieder auszusortieren. Die fehlerhafte Anwendung schlug sich darin nieder, dass die Befragten nicht bloß eine Angabe über die Summe der Ja-Antworten innerhalb ihres Fragenblocks – wie von der Versuchsinstruktion impliziert – tätigten, sondern für jede einzelne der vier bzw. fünf Fragen ein „Ja“ bzw. „Nein“ niederschrieben. Es ist davon auszugehen, dass die Ursache einer solchen Fehlanwendung der indirekten Verfahren in erster Linie darin zu suchen ist, dass die Betroffenen den vorherigen Instruktionstext nur flüchtig überflogen oder gar überhaupt nicht lasen. Um die Ausschlussrate und damit den Nacherhebungs-Aufwand zu reduzieren, sollte bei der Fragebogenkonzeption künftiger Anwendungen dieser Methoden darauf geachtet werden, dass innerhalb des eigentlichen Fragenblocks, der durch seine Einrahmung in der eigenen Studie eine Art „Blickfang“ darstellte (siehe Anhang A), unmittelbar vor der Auflistung der Fragen noch einmal explizit darauf verwiesen wird, dass *keine* Antworten auf die Einzelfragen aufzuführen sind.

4.2.2 Auswahl einer homogenen Stichprobe

Der Blick auf die Zusammensetzung der Stichprobe der eigenen Untersuchung lässt erkennen, dass sich diese aus einer Vielzahl von Studentinnen und Studenten unterschiedlicher Fachrichtungen zusammensetzte. Auf diese Weise war es möglich, die Repräsentativität der Stichprobe zu erhöhen, da so eine Generalisierung der Ergebnisse auf die Population der Studentinnen und Studenten und nicht bloß auf eine eingeschränkte studentische Population, wie z.B. die der Germanistik- und Medizinstudenten – sofern nur diese befragt worden wären – erzielbar war.⁵¹ Grundsätzlich muss jedoch kritisch angemerkt werden, dass die Studierenden eines bestimmten Fachbereichs immer auch dadurch zusätzliche Varianz erzeugen, dass sie sich von den Studierenden anderer Fächer unterscheiden. Insofern hätte eine homogenere Stichprobe, deren Daten in wenigen Großveranstaltungen der studierendenstärksten Fächer zusammen getragen worden wären, den Vorteil gehabt, die interdisziplinäre Zusatzvarianz zu reduzieren. Gerade im Hinblick auf die ohnehin erhebliche Varianz, die die indirekten Befragungsverfahren im Vergleich zur direkten Befragung erzeugen, ist jede Möglichkeit der Varianzverminderung als erstrebenswert anzusehen. Für künftige Datenerhebungen ähnlicher Größenordnung sollte daher der Gang an andere Hochschulen nicht gescheut werden, um dazu beizutragen, mit relativ wenigen Erhebungseinheiten den gewünschten Stichprobenumfang zu erzielen.

4.2.3 Verringerung der inhaltlichen Salienz der kritischen Fragen

In einigen Fällen ließ sich im Zuge der Versuchsdurchführung beobachten, dass Teilnehmerinnen und Teilnehmer den Versuchsleitern rückmeldeten, die Fragen innerhalb der Fragenblöcke der indirekten Bedingungen würden inhaltlich nicht zueinander passen. Einige Befragte fingen an, über potentielle Zusammenhänge zwischen den einzelnen Fragen zu reflektieren. Die übermäßige Auseinandersetzung mit der mangelnden Kohärenz der Fragen innerhalb der Fragenblöcke war durchaus ein Störfaktor, weil die Befragten so von der Konzentration auf die eigentliche Aufgabe abgelenkt wurden. Aus diesem Grunde ist darüber nachzudenken, inwieweit es nicht sinnvoll ist, die Thematik der nicht-sensitiven Fragen inhaltlich dem Gegenstand der sensitiven Fragen anzupassen anstatt sie willkürlich auszuwählen. Als potentielle Kontrollfrage im Zuge der eigenen Untersuchung wäre z.B. die Frage „Haben Sie innerhalb ihrer Schul- oder Studiumszeit in einem Schul- oder Universitätschor mitgesungen?“ denkbar gewesen.

⁵¹ Germanistik und Medizin stellten die zum Erhebungszeitpunkt (WS 2007/2008) studierendenstärksten Fächer an der HHU Düsseldorf dar.

4.2.4 Explizitere Kennzeichnung des Bezugszeitrahmens der kritischen Fragen

Wie innerhalb des Kapitels 2.1.3 erläutert wurde, zeichneten sich alle kritischen Fragen durch den gleichen Aufbau aus, indem sie jeweils durch die Formulierung „Haben Sie bereits...?“ nach der Lebenszeitprävalenz der jeweiligen sensitiven Verhaltensweisen fragten. Vielen Probanden schien aber nicht unmittelbar klar zu sein, inwieweit sie die Fragen lediglich auf ihre Studienzeit oder auch auf ihre Schulzeit beziehen sollten, so dass sich einige Studierenden diesbezüglich bei den Versuchsleitern erkundigten. Um dieser Unklarheit vorzubeugen, hätte entweder im Rahmen der einführenden Instruktion oder direkt in die Frage eingebunden der explizite Hinweis erfolgen können, dass der Bezugszeitrahmen sowohl die Schul- als auch die Studienzeit der Befragten sein soll. Es ist nämlich davon auszugehen, dass nicht jeder Teilnehmende die Möglichkeit hatte, sich aufgrund dieser Unklarheit an einen der Versuchsleiterinnen und Versuchsleiter zu wenden und womöglich aus diesem Grund einen falschen Bezugszeitrahmen bei der Beantwortung zugrunde legte. Der Nachteil einer expliziteren Kennzeichnung des Bezugszeitrahmens wäre allerdings der, dass die Fragen bzw. die Eingangsinstruktion hierdurch noch umfangreicher gewesen wären.

4.2.5 Zusätzliche Erhebung der unkritischen Fragen in den direkten Bedingungen

Wie sich nach Abschluss der multinomialen Auswertung herausgestellt hatte, wäre es günstig gewesen, wenn die vier nicht-sensitiven Fragen auch mittels direkter Befragung erhoben worden wären. Da nämlich das multinomiale Gesamtmodell nur unter der Zusatzannahme, dass die nicht-sensitiven UCT- und BIBD-Parameter gleichgesetzt wurden, identifiziert werden konnte, hätte ein reduziertes multinomiales Modell, in welches nur die direkten und die UCT-Bedingungen eingegangen wären, wichtige zusätzliche Erkenntnisse gebracht. Während ein solches, reduziertes (multinomiales) Modell zwar für die direkten und die BIBD-Bedingungen identifiziert werden konnte (mit dem Resultat eines signifikanten Befundes für die dritte, kritische Frage bei einseitiger Testung; vgl. Kap. 3.2.4.2), blieb diese Möglichkeit für die direkten und die UCT-Bedingungen verwehrt, weil keine Ankerwerte für die Schätzungen der nicht-sensitiven Parameter zur Verfügung standen. Eben dieses Problem hätte nicht bestanden, wenn die nicht-sensitiven Fragen in den direkten Bedingungen erhoben worden wären⁵². So aber blieb schließlich „nur“ die Erkenntnis, dass die klassische UCT-Auswertung entgegen der multinomialen Auswertung signifikant höhere UCT-

⁵² Dies hätte allerdings zur Folge gehabt, dass die Bearbeitungszeiten für die Fragebögen der direkten Bedingungen etwas größer gewesen wären.

Basisratenschätzungen im Vergleich zu den Prävalenzratenschätzungen der direkten Befragung aufzudecken imstande war.

4.2.6 Stärkere Einbindung von Dozenten in den Datenerhebungsprozess

Im Zuge der eigenen Datenerhebung ist auf die Einbindung von Dozenten in die Datenerhebung überwiegend verzichtet worden. Die Befragung wurde weitestgehend im Vorfeld einer Lehrveranstaltung durchgeführt und so der Beginn der Veranstaltung nicht gestört. Dies hatte zur Folge, dass all diejenigen Studentinnen und Studenten, die „just-in-time“ den Hörsaal betraten, nicht mehr an der Erhebung teilnehmen konnten. Gerade aufgrund des im Vergleich zu anderen Fragebogen überschaubaren Umfangs eines wie in der eigenen Untersuchung verwendeten Kurzumfrage-Bogens sollte die stärkere Einbeziehung von Dozenten in den Datenerhebungsprozess in Erwägung gezogen werden. Auf diese Weise ließe sich die erstrebenswerte höhere Homogenität der Stichprobe vermutlich leichter in die Tat umsetzen. Zudem ist denkbar, dass die Ankündigung der Befragung durch einen Veranstaltungsdozenten – gerade bei einem Befragungsinhalt wie dem der „schulisch-akademischen Unehrllichkeit“ – einen stärker einschüchternden Einfluss auf das Antwortverhalten der Befragten haben dürfte und womöglich die Bereitswilligkeit zur ehrlichen Beantwortung in der direkten Befragung nicht unerheblich reduzieren würde. Insbesondere wenn die stärkere Einbindung der Dozenten in Kopplung mit einer Erhebung zusätzlicher demographischer Daten erfolgt worden wäre (vgl. Kap. 4.1.2), ist anzunehmen, dass die direkt Befragten auf die kritischen Fragen nicht vergleichbar bereitwillig wie in der eigenen Untersuchung geantwortet hätten. Würden mehrere demographische Variablen erfragt, sollten diese – anders als in der eigenen Untersuchung – jedoch nicht zu Beginn des Fragebogens, sondern möglichst erst im Schlussteil erfolgen. Wie Alreck & Settle (2004) erläutern, bergen demographische Items nämlich immer die Gefahr einer Verweigerungshaltung der Befragten. Tritt diese Haltung erst zum Ende des Fragebogens auf, so sind die Antworten innerhalb des Hauptteils weniger verzerrt, als wenn sich die verweigernde Haltung bereits zu Beginn der Befragung etabliert.

4.2.7 Die Online-Umfrage als Alternative zur Papier und Bleistift-Befragung

Es ist nicht zu bestreiten, dass ein erheblicher Nachteil der vorgestellten indirekten Befragungstechniken darin besteht, dass ein hoher Stichprobenumfang unumgänglich gewährleistet werden muss, um eine möglichst niedrige Varianz der Parameterschätzungen zu erhalten. In der eigenen Untersuchung wurde im Rahmen der geschlechtsspezifischen Auswertung deutlich, dass durch die Trennung der Daten in zwei Geschlechts-Datensätze

erheblich größere Standardfehler resultierten; diese waren letztlich Ausdruck dessen, dass das Design für die Prüfung auf einen Geschlechtseffekt nicht hinreichend geeignet war. Abhilfe hätte hier nur ein noch erheblich größerer Stichprobenumfang geschaffen.

Mit den Mitteln einer Papier und Bleistift-Befragung sind weit im vierstelligen Bereich liegende Stichprobenumfänge schwierig zu erreichen. Von daher stellt die Online-Befragung aufgrund ihres geringeren zeitlichen und finanziellen Aufwandes eine ernsthaft in Erwägung zu ziehende Alternative dar, zumal – wie Bandilla (2002, zitiert in Ehling & Merz, 2002) betont – die Qualität der online erhobenen Daten den auf traditionellem Wege gewonnenen Daten vergleichbar, wenn nicht sogar höher einzuschätzen ist. Bortz & Döring (2002) verweisen darauf, dass es im Zuge der Online-Befragung wichtig sei, den Fragebogen nicht einfach der gesamten Netzöffentlichkeit zu präsentieren⁵³, sondern ihn per Email nur den gezielt in das Sample einbezogenen Individuen oder Gruppen zugänglich zu machen. Hierbei komme allerdings erschwerend hinzu, dass viele Internet-Nutzer „über mehrere E-Mail-Adressen verfügen (bzw.) derartige Adressen aufgrund eines Providerwechsels schnell veralten können“ (Bandilla, 2002, S. 50). Ein wesentlicher Grund, warum sich ein Fragebogen wie der in der eigenen Untersuchung verwendete, für eine Online-Befragung eignen dürfte, besteht darin, dass die Kürze des Bogens im Vergleich zu anderen Umfragen den kontaktierten Personen unmittelbar suggeriert, dass für dessen Bearbeitung nicht viel Zeit in Anspruch genommen wird. Insofern kann die Gefahr eines zu niedrigen Rücklaufs hier als eher unwahrscheinlich angesehen werden. Um die Rücklaufquote dennoch zu verbessern, dürfte die Einbeziehung Umfrage begleitender Maßnahmen, wie Vorankündigungen oder Erinnerungsschreiben bei Nicht-Teilnahme, laut Bandilla (2002) weithin förderlich sein.

Gerade mit Rückblick auf einige der in Kapitel 4.2 thematisierten, optimierbaren Aspekte der eigenen Papier und Bleistift-Befragung ist auch im Zuge der Online-Befragung zu bedenken, welche Rolle der Hervorhebung relevanter Textelemente, der expliziten Benennung potentiell missverständlicher Aspekte und der übersichtlichen Gestaltung der Website zukommt. Gräf (2002, zitiert in Ehling et al., 2002, S. 57) weist darauf hin, dass Nutzer „mit einem Blick das Wichtigste auf einer Seite erfassen wollen (und eher) einen scannenden Blick auf Webpages werfen“ anstatt vollständig zu lesen. In Bezug auf das Layout der Befragung rät Gräf (2002) zu einem neutralen, aber ansprechenden Design, welches seiner Funktion nachkommt, den Fragebogeninhalt zum Befragten zu transportieren. Überflüssige und schmückende Elemente sollten möglichst unterlassen werden. Des Weiteren sei es sinnvoll, rudimentäre Daten über Abbrecher zu erzielen. So empfiehlt Gräf (2002), die

⁵³ Für WWW-Fragebogen ergibt sich nämlich das methodische Problem der Selbstselektion der Rekrutierten.

Probanden, die die Befragung vor Erreichung des Umfrageendes verlassen, in einem sich neu öffnenden Fenster abschließend zu den Gründen ihres Abbruchs zu befragen. In der Praxis habe sich gezeigt, dass bei dieser letzten Frage noch relativ hohe Ausschöpfungsquoten erreicht werden konnten.

4.3 Weiterentwicklungen der indirekten Befragungstechniken

Der Anspruch der bisher vorgestellten indirekten Befragungsmethoden (RRT, klassische UCT und BIBD) war es, nicht bloß – wie in der direkten Befragung – eine Schätzung der unteren Grenze der Prävalenz eines sensitiven Merkmals zu erfassen, sondern zu einer valideren Schätzung der tatsächlichen Prävalenz zu kommen. Eine validere Schätzung muss neben der Korrektur der Unterschätzung der wahren Prävalenz eines sensitiven Merkmals aber möglichst auch einer Überschätzung der wahren Prävalenzrate entgegenwirken (Walsh & Braithwaite, 2008). Aus diesem Grund brachten Zimmerman & Langer (1995) die Abfrage der so genannten „selbst berichteten Ehrlichkeit“ („self-reported honesty“) ins Spiel. In ihrer Studie wurde eine vierstellige Stichprobe von Oberschülern ($n = 1.514$) in einer direkten Bedingung bzw. zwei UCT-Bedingungen (einer Kontrollgruppe á 5 Items und einer Experimentalgruppe á 5 + 1 Items) mit sensitiven Fragen, die ihr Sexualverhalten betrafen, konfrontiert. Beispielhaft sei hier die Frage „Hatten Sie schon einmal Sex mit einem gleichgeschlechtlichen Partner?“ angeführt. Wurden die Basisratenschätzungen gemäß der klassischen Auswertungsmethode (über die Mittelwerts-Differenzen) ermittelt, so ergaben sich – bezogen auf die angegebene Frage – für die direkte Befragung eine Schätzung von 8,7 % und im Falle der UCT eine Schätzung von 15,9 %. Im Anschluss an jeden UCT-Experimentalgruppen-Fragenblock wurden die betroffenen Schülerinnen und Schüler zusätzlich gefragt, inwieweit sie die jeweilige sensitive Frage ehrlich beantwortet hätten. Hierzu mussten sie sich auf einer vierstufigen Skala mit den Ausprägungen „kaum ehrlich antwortend – mäßig ehrlich antwortend – überwiegend ehrlich antwortend – absolut ehrlich antwortend“ einstufen. Daraufhin wurden die UCT-Prävalenzratenschätzungen über eine Gewichtung mittels der angegebenen selbst berichteten Ehrlichkeit nach unten korrigiert. Die Gewichtungsprozedur erfolgte derart, dass die unterste Kategorie der Skala („kaum ehrlich antwortend“) mit 0.00, die beiden mittleren Kategorien mit 0.33 bzw. 0.67 und die oberste Kategorie („absolut ehrlich antwortend“) mit 1.00 multiplikativ verrechnet wurden. Für das sensitive Merkmal des „Geschlechtsverkehrs mit einem gleichgeschlechtlichen Partner“ ergab sich auf diese Weise eine korrigierte Basisratenschätzung von nur noch 13,7 % (zuvor: 15,9

%).⁵⁴ Auch wenn die Notwendigkeit, die Untergrenze einer wahren Prävalenz nach oben hin zu korrigieren im Vergleich zur Korrektur der Überschätzung der tatsächlichen Prävalenz nach unten hin die bedeutendere darstellen dürfte, ist die Modifikation der UCT von Zimmerman et al. (1995) unter Validitätsgesichtspunkten eine sinnvolle. Beide Autoren bemängelten aber ebenfalls die problematisch hohen Standardfehler, die die Füllfragen der klassischen UCT mit sich brachten. Die Möglichkeit, anstelle der klassischen UCT nunmehr das BIBD mit der Erfassung der „selbst berichteten Ehrlichkeit“ zu kombinieren, ist aufgrund der hierbei niedriger ausfallenden Standardfehler in Erwägung zu ziehen.

Corstange (2006) änderte die klassische UCT insofern ab, dass die Befragten in der Kontrollgruppe jede Frage direkt beantworten mussten, während die Probanden der Experimentalgruppe unverändert die Ja-Antwortsumme des um das sensitive Item erweiterten Fragenblockes zu berichten hatten. Durch diese Modifikation ergab sich der von Corstange (2006, S. 7 / 8) wie folgt geschilderte Vorzug einer Verankerung der sensitiven Basisratenschätzung⁵⁵:

We may estimate the individual probabilities associated with the non-sensitive items from the control group and use these estimates to anchor the sensitive probability in the treatment group, reducing the number of unknowns to one. Thus, for example, if the average probability across four list items were again .50, but this time we estimated that the non-sensitive probabilities were .20, .75 and .80, we would know that the only possible value the sensitive probability could take would be .25, since that is the only value that could produce the average probability of .50 given the estimates.

Die von Corstange (2006) in diesem Rahmen thematisierte Abänderung der klassischen UCT, deren Validierung allerdings noch aussteht, stellt eine interessante Alternative zum bisherigen Vorgehen der Mittelwert-Differenzbildung von UCT-Kontroll- und Experimentalgruppe dar.

4.4 Anforderungen an die zukünftige Umfrageforschung

Die Umfrageforschung muss in Zukunft der potentiell größer werdenden Skepsis der Befragten in Bezug auf die Sicherheit ihrer Daten mit innovativen Techniken entgegenreten. Anwendungen wie das BIBD bzw. die Erfassung der selbst berichteten Ehrlichkeit stellen methodische Ansätze dar, deren Forschungsbedarf noch längst nicht ausgeschöpft ist. Die

⁵⁴ Die von Zimmerman & Langer (1995) durchgeführte Befragung erfolgte über mehrere Messzeitpunkte, so dass sie auf Grundlage dessen auch eine Inkonsistenzanalyse (der Antworten) durchführen konnten. Auf die Darstellung der Ergebnisse dessen soll aus Gründen der Übersichtlichkeit aber an dieser Stelle verzichtet werden.

⁵⁵ Corstanges Ausführungen bezogen sich hier auf ein klassisches UCT-Design, welches in der Kontrollgruppe drei nicht-sensitive Fragen und in der Experimentalgruppe dieselben drei nicht-sensitiven Fragen plus eine sensitive Frage umfasste. Durch die Reduktion der nicht-sensitiven Items auf eine Anzahl von drei versuchte er die zusätzlich erzeugte Varianz in Grenzen zu halten.

Befunde der eigenen BIBD-Anwendung sollten insbesondere im Hinblick auf die Ergebnisse Benesch (2008) nicht überbewertet werden. Wie bereits in Tabelle 4 ersichtlich wurde, stehen weitere BIBDs zur Verfügung, die bisher in Kontexten der Umfrageforschung noch nicht erprobt worden sind. Insbesondere das (7, 3, 1)-BIBD könnte in Folgestudien interessant sein, da es bereits in der eigenen Studie in der näheren Auswahl stand (vgl. Kap. 2.1.1). In jedem Fall scheint es aber notwendig, in zukünftigen BIBD-Anwendungen von den bisher zugrunde gelegten einfachen multinomialen Modellen abzuweichen und stattdessen eine Heterogenitätsannahme einzubeziehen. Allerdings hätte die Berücksichtigung des Heterogenitätsansatzes zur Konsequenz, dass der benötigte Stichprobenumfang noch deutlich höher als in den bisherigen BIBD-Studien ausfallen müsste. Dies ist unter Kosten-Nutzen-Abwägungen sicherlich nicht unproblematisch.

Gegen eine Anwendung der „klassischen“ UCT-Variante im Zuge der künftigen Umfrageforschung dürften zwei wesentliche Einwände deutlich geworden sein: zum einen ist sie im Gegensatz zum BIBD immer nur imstande, *eine* kritische Frage – statt mehrerer – einzubeziehen. Zum anderen schneidet sie unter Effizienzgesichtspunkten im Vergleich zum BIBD schlechter ab. Von daher ist von ihrer Verwendung in kommenden Umfrageforschungs-Projekten eher abzusehen, wenngleich die Modifikation der klassischen UCT, wie sie zuletzt durch Corstange (2006) erfolgte, nicht vorzuverurteilen ist, da ihre Evaluation noch aussteht. Die Randomized Response-Technik hat seit ihrer Konzeption durch Warner (1965) eine lange Geschichte mit vielen Modifikationen hinter sich. Wie Ostapczuk (2008) zuletzt zeigen konnte, stieg die Validität der Basisratenschätzungen der RRT, wenn es gelang, den Anteil der Verweigerer möglichst gering zu halten. Dies konnte Ostapczuk (2008) über die erstmalige Zulassung von Zufallsgenerator-bedingten „Nein“-Antworten wirksam in die Tat umsetzen (vgl. Kap. 1.2.1.2). Von daher sollte auch die RRT als ein probates Instrument zur Kontrolle von Antwortverzerrungen bei Selbstauskünften nicht aus dem Auge verloren werden.⁵⁶

Wie die fehlerhafte Anwendung der indirekten Befragungstechniken (UCT bzw. BIBD) im Zuge der eigenen Untersuchung gezeigt hat, muss ein Leitmotiv bei der Konzeption neuer bzw. erweiterter indirekter Befragungsmethoden in der leichten Handhabbarkeit des / der Verfahren liegen. Wie Alreck et al. (2004) hervorheben, dürfen die Motivation und die Konzentration der Befragten bei der Bearbeitung eines Fragebogens keineswegs überschätzt werden. Insbesondere ist zu bedenken, dass in der eigenen Untersuchung eine studentische

⁵⁶ Es bleibt allerdings nachzutragen, dass die RRT im Gegensatz zum (6, 3, 2)-BIBD in der Studie von Benesch (2008) nicht imstande war, signifikant höhere Prävalenzratenschätzungen als die direkte Befragung zu erzielen.

Stichprobe für die Erprobung der indirekten Befragungstechniken herangezogen wurde. In Folgestudien sollte das BIBD auch für Personen eines niedrigeren Bildungsstatus Erprobung finden. Aktuelle Arbeiten kamen diesbezüglich zu dem Ergebnis, dass die Verständlichkeit und die Einhaltung der Instruktionen von RRT bzw. UCT bei Befragten mit niedrigerem Bildungsstand abnimmt (Tsuchiya, Hirai & One, 2007, zitiert nach Coutts & Jann, 2008).

Einen für die zukünftige Umfrageforschung ebenfalls interessanten Aspekt könnten die Forschungsarbeiten von Ahart & Sackett (2004) darstellen, die differentialpsychologische Konstrukte in die Erfragung sensibler Inhalte mit einbezogen. So implementierten sie Testverfahren für die beiden Persönlichkeitsdimensionen Integrität und Gewissenhaftigkeit als Prädiktoren für kontraproduktive Verhaltensweisen (wie z.B. das Begehen eines Diebstahls oder die Tätiung einer Krankmeldung beim Arbeitgeber, ohne wirklich erkrankt zu sein). In ihrer Untersuchung wurden die Basisratenschätzungen für sensitive Verhaltensweisen separat für verschiedene Persönlichkeitssubgruppen (weniger integere, eher integere, weniger gewissenhafte, eher gewissenhafte Personen) zum einen direkt und zum anderen indirekt (mittels der klassischen UCT) erhoben. Zwar konnten Ahart & Sackett (2004) nicht konsistent höhere UCT-Basisratenschätzungen im Vergleich zur direkten Befragung aufzeigen, aber zumindest deutliche Differenzen zwischen den verschiedenen Persönlichkeits-Subgruppen untermauern. Auch die Befunde Aharts & Sacketts (2004) weisen somit daraufhin, dass es sinnvoll sein dürfte, eine Heterogenitätsannahme für die Parameter, wie sie in Kapitel 4.1.4 für BIBD-Anwendungen vorgeschlagen wurde, in die zukünftige Modellplanung zu implementieren, zumal es inhaltlich plausibel erscheint, dass sich Menschen mit unterschiedlichen Persönlichkeitseigenschaften bei der Beantwortung sensibler Fragen in ihrem Antwortverhalten unterscheiden. So ist beispielsweise davon auszugehen, dass sich gewissenhafte Menschen grundsätzlich schwerer als weniger Gewissenhafte damit tun, im direkten Befragungskontext ein begangenes, kritisches Verhalten zu leugnen, weil sie im Zuge dessen vermutlich mit intensiveren Gewissenskonflikten zu kämpfen haben als eher sorglose Personen⁵⁷.

4.5 Umgang mit sensiblen Daten: Bestandsaufnahme und Ausblick

Um den Forschungsbedarf zu innovativen Methoden in der Umfrageforschung auch noch aus einer anderen Perspektive deutlich zu machen, soll im nun folgenden, abschließenden Teilabschnitt ein gesellschaftspolitischer Blick auf die Chancen und Grenzen des Umgangs

⁵⁷ Insofern dürfte eine gewissenhafte Person von einer indirekten Befragungstechnik stärker profitieren als eine weniger gewissenhafte Person, da es sie emotional weitaus stärker erleichtern dürfte, ein kritisches Verhalten zugeben zu können, ohne „überführbar“ zu sein.

mit sensiblen Daten geworfen werden. In diesem Zusammenhang sollen insbesondere juristische Aspekte des Datenschutzes und das potentielle Missbrauchspotential sensibler Daten zur Sprache kommen.

4.5.1 Bedeutung und Hauptprinzipien des Datenschutzes

In der modernen Informationsgesellschaft nimmt der Schutz von personenbezogenen Daten⁵⁸ eine zunehmend wichtigere Rolle ein. Westeuropa und Nordamerika entwickeln sich seit etwa 1970 von einer Industrie- zu einer Informationsgesellschaft. Dadurch werden Informationen und damit auch personenbezogene Daten zu einem immer bedeutender werdenden Wirtschaftsgut. Dassler (2007) verweist in diesem Zusammenhang darauf, dass Datenerfassung, Datenweitergabe und Datenanalyse in einer „Welt der Digitalisierung“ durch zahlreiche Interessengruppen stetig vorangetrieben werden. Die neuen Möglichkeiten der elektronischen Datenverarbeitung bringen ein nicht zu unterschätzendes Missbrauchspotential sowohl des Staates als auch durch Private mit sich, die „ein die Freiheitsrechte des Einzelnen beeinträchtigendes Interventionswissen erzeugen“ (Kaase, 1999, S. 82). Ein jüngstes Beispiel stellt die vielerorts kritisch betrachtete, für 2009 angestrebte Einführung der elektronischen Gesundheitskarte dar. Wie Schaar (zitiert nach Kalinowski, 2008) diesbezüglich in einem Zeitungsartikel der Westdeutschen Zeitung bemerkt, droht die angestrebte verbesserte medizinische Versorgung der Patienten mit einem Verlust an Datenschutz einherzugehen. Sie eröffnet die Gefahr eines „gläsernen Patienten“. Datenschutzorganisationen warnen davor, dass Lebensversicherer unlauter erworbene Gesundheitsdaten nutzen könnten, um Menschen mit Gesundheitsrisiken aus der Versicherung auszuschließen oder Arbeitgeber die Einstellung von Mitarbeitern von genetischen Dispositionen abhängig machen könnten.

Im Rahmen einer internationalen Datenschutzkonferenz im Jahre 2005 wurden die wichtigsten Hauptprinzipien des Datenschutzes in Form der „Erklärung von Montreux“ niedergelegt. Diese sollen in der Folge erläutert und in Abbildung 7 veranschaulicht werden. Allen vier Datenschutzprinzipien ist gemein, dass sie die Betroffenen vor dem Missbrauch ihrer personenbezogenen Daten schützen sollen. Grundlage dieser Zielsetzung ist die Wahrung des Rechts der Betroffenen auf informationelle Selbstbestimmung⁵⁹. Jeder Mensch

⁵⁸ Personenbezogene Daten sind gemäß § 3 Absatz 1 Bundesdatenschutzgesetz Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer Person. Bestimmbar sind Personen zum Beispiel durch Telefonnummern, E-Mail-Adressen oder IP-Adressen. In Abgrenzung dazu stehen anonyme Daten, bei denen die Person unbekannt und somit nicht bestimmbar ist.

⁵⁹ Amtliche Sammlung der Entscheidung des Bundesverfassungsgerichts, Band 65, S. 1, vom 15.12.1983

solle grundsätzlich selbst entscheiden, wem zu welchem Zeitpunkt seine persönlichen Daten zugänglich gemacht werden.

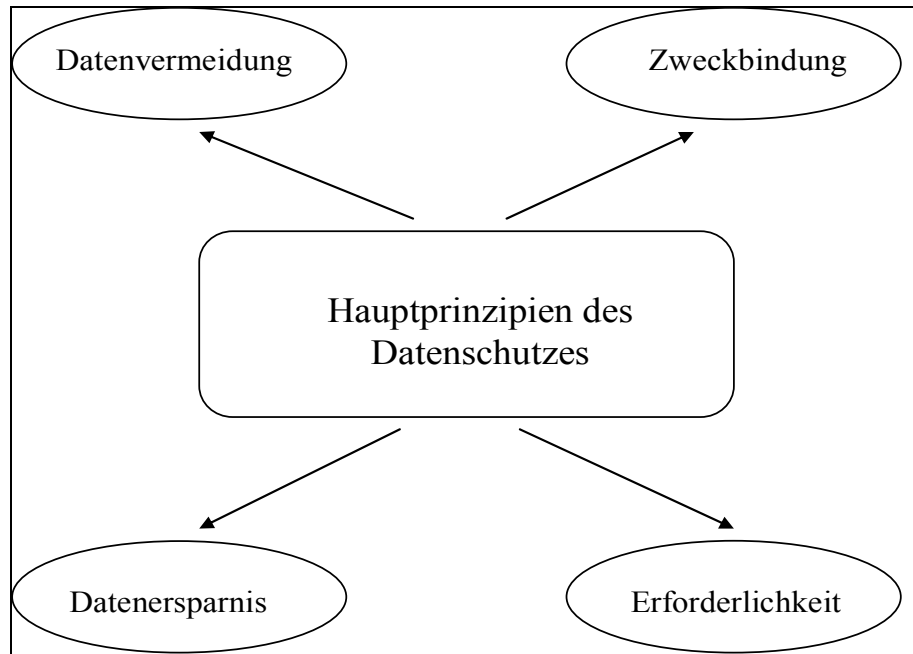


Abbildung 7: Überblick über die Hauptprinzipien des Datenschutzes in Anlehnung an die „Erklärung von Montreux“ im Rahmen der internationalen Datenschutzkonferenz 2005, zitiert nach Dassler (2007); eigene Darstellung.

Das Ziel des Prinzips der Datensparsamkeit ist es, Informationen möglichst sorgsam zu erheben. Gleichzeitig appelliert dieses Prinzip an die Befragten, personenbezogene (sensible) Daten nur zurückhaltend kundzutun. Darüber hinaus sollen, im Zuge der Datenvermeidung, nicht mehr benötigte Daten umgehend gelöscht werden, damit sie einem potentiell späteren Missbrauch nicht zur Verfügung stehen. Datensparsamkeit und Datenvermeidung stehen in einem engen Zusammenhang zum Prinzip der Erforderlichkeit, demzufolge nur diejenigen personenbezogenen Daten erhoben werden dürfen, die für die Aufklärung der spezifischen Fragestellung notwendig sind. Zu guter Letzt besagt das Prinzip der Zweckgebundenheit, dass eine Verarbeitung von personenbezogenen Daten ausschließlich zu dem Zweck erfolgen darf, zu dem sie erhoben wurden. So dürfen zum Beispiel die „Videodaten bei der LKW-Maut ausschließlich für das Mautsystem und nicht zu Fahndungszwecken“ verwendet werden (Dassner, 2007, S. 19).

4.5.2 Datensparsamkeit als Schlüsselaspekt der indirekten Befragungsverfahren

Nachdem die Hauptprinzipien des Datenschutzes soeben dargestellt wurden, ist es zweckmäßig, diese nun zu den indirekten Befragungstechniken in Beziehung zu setzen. Hierbei soll das Prinzip der Datensparsamkeit in den Fokus der Aufmerksamkeit gerückt werden. Das Kernziel der Datensparsamkeit ist es, die Erhebung personenbezogener Daten auf eine minimale Datenmenge zu reduzieren, die für die Klärung der zugrunde liegenden Fragestellung unumgänglich ist. Wie Kaase (1999) betont, gilt das Interesse der Durchführenden in vielen Befragungskontexten lediglich den Basisratenschätzungen für eine ganze Gruppe. Um z.B. Entscheidungsträgern Kennzahlen für die Notwendigkeit von Interventionsmaßnahmen darbieten zu können, ist der konkrete Personenbezug des Einzelnen oftmals uninteressant; was zählt sind letztlich die aggregierten Daten. Und exakt an dieser Stelle erweisen sich die indirekten Befragungstechniken als vorteilhaft: so lassen sich diese nämlich immer dann, wenn es um die Prävalenz von Gruppencharakteristika geht, Erfolg versprechend einsetzen. Die klassische UCT bzw. das BIBD bieten gar nicht erst die Möglichkeit, personenbezogene Daten im Nachhinein zu missbrauchen, weil sich eine direkte Überführung von Einzelpersonen mittels dieser Verfahren gar nicht realisieren lässt. Ebenso bekräftigt Dassler (2007), dass es immer besser ist, personenbezogene Daten gar nicht erst zu erheben und zu speichern, als sie anschließend mit aufwendigen Gesetzen und Kontrollen schützen zu müssen.

4.5.3 Quo vadis, Datenschutz?

Aufgrund der vielfältigen Nutzung des Internets erfolgt die Datenverarbeitung zunehmend international und entwickelt sich immer mehr zu einer globalen Herausforderung. Die strikte Einhaltung des Prinzips der Zweckbindung wird auf eine harte Probe gestellt, da insbesondere Wirtschaftsunternehmen an einer Zweckentfremdung personenbezogener Daten Interesse haben. Wie Dassner (2007) betont, sind Unternehmen darauf angewiesen, möglichst genaue Kundenprofile zu erstellen, um Zielgruppen-orientiertes Marketing zu betreiben. Dazu nehmen sie zum Beispiel am System "Payback" teil. Zwar erhält der Kunde für einen Einkauf bei einem Mitgliedsunternehmen einen Bonus, dafür werden aber die Daten über seinen Kauf vom Unternehmen verarbeitet. Dieser Tausch von personenbezogenen Daten gegen relativ unerhebliche wirtschaftliche Vorteile scheint angesichts der großen Anzahl der teilnehmenden Kunden gesellschaftlich akzeptiert zu sein. Und nicht zuletzt in Zeiten der anhaltenden Unruhen auf den globalen Finanzmärkten, wünschen viele Firmen genaue und aktuelle Daten über die Kreditwürdigkeit ihrer Kunden. Bereits vor knapp zehn Jahren sah Kaase (1999) die

Gefahr, dass Daten personenbezogener Form zu Akquisitionszwecken verwendet werden und dass sich eine „Kultur des Datenhandels“ etabliert. Diese Gefahr hat sich mittlerweile – leider – realisiert, wie man an den jüngsten Skandalen der Deutschen Telekom AG bezüglich der Behandlung der personenbezogenen Daten ihrer Kunden sieht. Eine Novellierung des Bundesdatenschutzgesetzes im Hinblick auf den Schutz personenbezogener Daten nicht nur vor dem Staat, sondern auch vor Privaten ist deshalb erforderlich und steht auch schon auf der politischen Agenda: so fordern vor allem Vertreter der Partei „Bündnis 90/ Grüne“ die Einführung eines „Grundrechts auf Datenschutz“, aber zumindest die Ergänzung des Datenschutzgesetzes. Jedoch sind nicht nur die Rechtswissenschaften und die Politik, sondern auch technische Lösungsmöglichkeiten gefragt. So fordert der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit, Schaar, in einem Interview mit Dassler (2007, S. 127), dass „der Datenschutz schon in Produkte und Dienste eingebaut werden sollte und bereits in der Entwurfsphase daran gedacht werden müsste, dass personenbezogene Daten zu schützen sind“. Hätte die Deutsche Telekom AG schon im Vorfeld die einfachen und nahe liegenden Möglichkeiten der Technik genutzt, um zu verhindern, dass Dritte unbefugt auf die personenbezogenen Daten der Kunden zugreifen können, wäre es gar nicht erst zum Skandal gekommen. Im Übrigen ist davon auszugehen, dass wegen der dadurch vergrößerten Datenerparnis auch die indirekte Datenerhebung zur Vermeidung des oben genannten Datenmissbrauchs beigetragen hätte.

Literaturverzeichnis

- Ahart, A.M. & Sackett, P.R. (2004). A new method of examining relationships between individual difference measures and sensitive behaviour criteria: Evaluating the Unmatched Count Technique. *Organizational Research Methods*, 7 (1), 101-114.
- Alreck, P.L. & Settle, R.B. (2004). *The survey research handbook*. McGraw-Hill Irwin (3rd edition). Boston, Massachusetts.
- Anderson, D.A., Simmons, A.M., Milnes, S.M. & Earleywine, M. (2007). Effect of response format on endorsements of eating disordered attitudes and behaviours. *International Journal of Eating Disorders*, 40 (1), 90-93.
- Baldwin, D.C., Daugherty, S.R., Rowley, B.D. & Schwarz, M.R. (1996). Cheating in medical school: a survey of second-years students at 31 schools. *Academic medicine*, 71 (3), 267-273.
- Bandilla, W. (2002). Chancen und Probleme einer internetbasierten Datenerhebung. In: Ehling, M. & Merz, J. (2002). *Neue Technologien in der Umfrageforschung – Anwendungen bei der Erhebung von Zeitverwendung*. *Schriften des Forschungsinstitutes Freie Berufe*. Baden-Baden: Nomos, 45-54.
- Belcheir, M.J. (2003). Academic dishonesty at Boise State University. *Research Report 2003-2005*.
- Benesch, C. (2008). *Ein experimenteller Vergleich von Verfahren zur Kontrolle von Antwortverzerrungen infolge sozialer Erwünschtheit*. Düsseldorf, Universität. Unveröffentlichte Diplomarbeit.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. Auflage). Berlin, Heidelberg, New York: Springer.
- Chaudhuri, A. & Christofides, T.C. (2007). Item Count Technique in estimating the proportion of people with a sensitive feature. *Journal of Statistical Planning and Inference*, 137, 589-593.

Cohen (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Erlbaum.

Constange (2006). *Sensitive questions, truthful answers – Modelling the List Experiment multivariately with LISTIT*. University of Michigan, Ann Arbor.

Council of Writing Program Administrators (2003). *Defining and avoiding plagiarism: The WPA statement on best practices* (<http://www.wpacouncil.org>).

Coutts, E. & Jann, B. (2008). *Sensitive questions in online surveys: Experimental results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)*. Sociology Working paper No. 3. Swiss Federal Institute of Technology, Zurich.

Dalton, D., Daily, C.M. & Wimbush, J.C. (1997). Collecting “sensitive” data in Business Ethics Research: A case for the Unmatched Count Technique (UCT). *Journal of Business Ethics*, 16, 1049-1057.

Dalton, D., Wimbush, J.C. & Daily, C.M. (1994). Using the Unmatched Count Technique (UCT) to estimate base rates for sensitive behaviour. *Personnel Psychology*, 47, 817-828.

Dassner, S. (2007). *Datenschutz in der modernen Informationsgesellschaft*. Berlin: Wissenschaftlicher Verlag.

Dordoy, A. (2002). *Cheating and plagiarism: staff and student perceptions at Northumbria*. Northumbria University Conference.

Droitcour, J., Caspar, M.L., Parsley, T.L., Visscher, W. & Ezzati, T.M. (1991). The Item Count Technique as a method of indirect questioning: a review of its development and a case study application. *Measurement Errors in Surveys*, 11, 185-210.

Erdfelder, E. (2000). *Multinomiale Modelle in der kognitiven Psychologie*. Habilitationsschrift zur Erlangung der Lehrbefugnis im Fach Psychologie an der Philosophischen Fakultät der Rheinischen Friedrich-Wilhelm-Universität in Bonn.

- Faul, F., Erdfelder, E., Lang, A.G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural and biomedical sciences: *Behavior Research Methods*, 33, 175-191.
- Gräf, L. (2002). WWW-Erhebungsinstrumente: Anforderungen, Gestaltung und Datenqualität. In: Ehling, M. & Merz, J. (2002). Neue Technologien in der Umfrageforschung – Anwendungen bei der Erhebung von Zeitverwendung. *Schriften des Forschungsinstitutes Freie Berufe*. Baden-Baden: Nomos, 55-63.
- Hart, M. & Friesner, T. (2004). Plagiarism and poor academic practice – A threat to the extension of e-learning in higher education? *Electronic Journal on e-Learning*, 2 (1), 89-96.
- Hrabak, M., Vujaklija, A., Vodopivec, I., Hren, D., Marusic, M. & Marusic, A. (2004). Academic misconduct among medical students in a post-communist country. *Medical Education*, 38 (3), 276-285.
- Kaase, M. (1999). *Qualitätskriterien der Umfrageforschung*. Deutsche Forschungsgemeinschaft. Berlin: Akademischer Verlag.
- Kalinowski, S. (2008). Die elektronische Gesundheitskarte. *Westdeutsche Zeitung*. Anzeigen-Sonderveröffentlichung: Krankenkassen – Das Gesundheitsmagazin (09.09.2008).
- Kane, J.G., Craig, S.C. & Wald, K.D. (2004). Religion and presidential politics in Florida: a list experiment. *Social Science Quarterly*, 85 (2), 281-294.
- LaBrie, J.W. & Earleywine, M. (2000). Sexual risk behaviours and alcohol: higher base rates revealed using the Unmatched-Count Technique. *Journal of Sex Research*, 37 (4), 321-327.
- Lin, C.S. & Wen, L.M. (2006). Academic dishonesty in higher education – a nationwide study in Taiwan. *Higher Education*, 54, 85-97.

- McCabe, D.L., Trevino, L.K. & Butterfield, K.D. (2001). Cheating in academic institutions: a decade of research. *Ethics & Behavior*, 11, 219-232.
- Musch, J., Brockhaus, R. & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica*, 48, 121-129.
- Musch, J., Bröder, A. & Klauer, K.C. (2001). Improving survey research on the World-Wide Web using the randomized response technique. In: U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science*, 179-192. Lengerich: Pabst.
- Ostapczuk, M.S. (2008). Experimentelle Umfrageforschung mit der Randomized Response Technik. Düsseldorf, Universität. Unveröffentlichte Dissertation.
- Pashos, A. (2001). Über die Rolle von Status, physischer Attraktivität und Taktiken in der menschlichen Partnerwahl. Hamburg, Universität. Dissertation.
- Paulhus, D.L. (1984). Two component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Payne, S.L. & Nantz, K.S. "Social accounts and metaphors about cheating". *College Teaching*, 42 (3), 90-96.
- Raghavarao, D. & Federer, W.T. (1979). Block Total Response as an alternative to the Randomized Response Method in surveys. *Journal of the Royal Statistical Society: Series B*, 41 (1), 40-45.
- Riefer, D.M. & Batchelder, W.H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95, 318-339.
- Schmelkin, L.P., Kaufman, A.M. & Liebling, D.E. (2001). Faculty assessments of the clarity and prevalence of academic dishonesty. Annual meeting of the American Psychological Association, San Francisco, C.A.

- Smith, L.L., Federer, W.T. & Raghavarao, D. (1974). A comparison of three techniques for eliciting truthful answers to sensitive questions. *Proceedings of the Social Statistics Section, American Statistical Association*. Baltimore, 447-452.
- Smith, N.F. & Street, D.J. (2003). The use of balanced incomplete block designs in designing randomized response surveys. *Australian and New Zealand Journal of Statistics*, 45 (2), 181-194.
- Smith, N.F. (2005). The design of scoring schemes for surveys using the Block Total Response Method. *Communication in Statistics – Theory and Methods*, 34, 2157-2168.
- Stahl, C. & Klauer, K.C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, 39, 267-273.
- Stirn, A. (2007). “Mein Körper gehört mir.” – Kulturgeschichte und Psychologie des Piercing und Tattoo. *Therapeutische Rundschau*, 64, 115-119. Bern: Hans Huber-Verlag.
- Van der Linden, W.J., Veldkamp, B.P. & Carlson, J.E. (2004). *Applied Psychological Measurement*, 28 (5), 317-331.
- Volk, P. (2008). Ehrenwort vor dem ersten Semester. *Frankfurter Allgemeine Zeitung*, Beruf & Chance – Campus (09./10.08.2008)
- Walsh, J.A. & Braithwaite, J. (2008). Self-reported alcohol consumption and sexual behaviour in males and females: using the unmatched-count technique to examine reporting practices of socially sensitive subjects in a sample of university students. *Journal of Alcohol & Drug Education*, 52 (2), 49-72.
- Weber-Wulff, D. (2002). *Schummeln mit dem Internet? Report Literaturrecherche*. Hannover: Heinz Heise-Verlag, Heft 1, 64-69.
- Wimbush, J.C. & Dalton, D.R. (1997). Base rates for employee theft: Convergence of multiple methods. *Journal of Applied Psychology*, 82, 756-763.

Zimmerman, R.S. & Langer, L.M. (1995). Improving estimates of prevalence rates of sensitive behaviors: The Randomized List Technique and consideration of self-reported honesty. *The Journal of Sex Research*, 32 (2), 107-117.

Eigenständigkeitserklärung



Hiermit versichere ich, dass die Diplomarbeit zum Thema

„Die Kontrolle sozialer Erwünschtheit bei Selbstauskünften: ein experimenteller Vergleich von direkter Befragung, Unmatched-Count-Technik und Balanced-Incomplete-Block-Design“

von mir selbstständig verfasst wurde.

Ich versichere zudem, dass

- dabei keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie
- alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen als solche kenntlich gemacht wurden.

Dies gilt für gedruckte Texte ebenso wie für elektronische Ressourcen (z.B. www-Dokumente).

Die Diplomarbeit ist in der vorliegenden oder einer modifizierten Form noch nicht in einer anderen Lehrveranstaltung als Studienleistung vorgelegt worden. Sie enthält keine Teile (Textpassagen, Grafiken etc.) von Hausarbeiten anderer Personen.

Datum: 11.11.2008

Name: Thomas Seppelfricke

Unterschrift:

Anhang

Anhang A: Fragebogen der Bedingungen 1 bis 17

(hierbei sind die identischen Bedingungen 11 bis 17 nur einfach aufgeführt)

Anhang B: Datenfile der Gesamtauswertung

Anhang C: Gleichungsfile der Gesamtauswertung

Anhang D: Multinomiale Signifikanztests der Gesamtauswertung

Anhang A: Fragebogen der Bedingungen 1 bis 17

Institut für Experimentelle Psychologie
Abteilung für Diagnostik und Differentielle Psychologie
Prof. Dr. Jochen Musch

Heinrich-Heine-Universität Düsseldorf
Diplomarbeit Thomas Seppelfricke
Okt./ Nov. 2007_1

SCHNELL-UMFRAGE

Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á vier Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der vier Fragen Sie insgesamt bejaht haben!

1. Wohnen Sie derzeit zur Untermiete?
2. Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen, schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür gar keinen ärztlichen Grund gab?
3. Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?
4. Haben Sie eine nicht entfernbare Tätowierung?

Wie viele dieser vier Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á vier Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der vier Fragen Sie insgesamt bejaht haben!

1. Haben Sie eine nicht entfernbare Tätowierung?
2. Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?
3. Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel verwendet?
4. Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?

Wie viele dieser vier Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á vier Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der vier Fragen Sie insgesamt bejaht haben!

1. Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?
2. Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel verwendet?
3. Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen, schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür gar keinen ärztlichen Grund gab?
4. Wohnen Sie derzeit zur Untermiete?

Wie viele dieser vier Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

Institut für Experimentelle Psychologie
Abteilung für Diagnostik und Differentielle Psychologie
Prof. Dr. Jochen Musch

Heinrich-Heine-Universität Düsseldorf
Diplomarbeit Thomas Seppelfricke
Okt./ Nov. 2007_4

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á vier Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der vier Fragen Sie insgesamt bejaht haben!

1. Sind Sie verheiratet?
2. Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?
3. Wohnen Sie derzeit zur Untermiete?
4. Haben Sie eine nicht entfernbare Tätowierung?

Wie viele dieser vier Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á vier Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der vier Fragen Sie insgesamt bejaht haben!

1. Sind Sie verheiratet?
2. Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?
3. Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen, schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür keinen ärztlichen Grund gab?
4. Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?

Wie viele dieser vier Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

Institut für Experimentelle Psychologie
Abteilung für Diagnostik und Differentielle Psychologie
Prof. Dr. Jochen Musch

Heinrich-Heine-Universität Düsseldorf
Diplomarbeit Thomas Seppelfricke
Okt./ Nov. 2007_6

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á vier Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der vier Fragen Sie insgesamt bejaht haben!

1. Wohnen Sie derzeit zur Untermiete?
2. Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel zum Einsatz gebracht?
3. Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?
4. Sind Sie verheiratet?

Wie viele dieser vier Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á vier Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der vier Fragen Sie insgesamt bejaht haben!

1. Haben Sie eine nicht entfernbare Tätowierung?
2. Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen, schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür keinen ärztlichen Grund gab?
3. Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel zum Einsatz gebracht?
4. Sind Sie verheiratet?

Wie viele dieser vier Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

Institut für Experimentelle Psychologie
Abteilung für Diagnostik und Differentielle Psychologie
Prof. Dr. Jochen Musch

Heinrich-Heine-Universität Düsseldorf
Diplomarbeit Thomas Seppelfricke
Okt./ Nov. 2007_8

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á fünf Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der fünf Fragen Sie insgesamt bejaht haben!

1. Sind Sie verheiratet?
2. Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?
3. Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel zum Einsatz gebracht?
4. Wohnen Sie derzeit zur Untermiete?
5. Haben Sie eine nicht entfernbare Tätowierung?

Wie viele dieser fünf Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier) ☐ 5 (fünf)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á fünf Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der fünf Fragen Sie insgesamt bejaht haben!

1. Sind Sie verheiratet?
2. Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?
3. Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen, schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür gar keinen ärztlichen Grund gab?
4. Wohnen Sie derzeit zur Untermiete?
5. Haben Sie eine nicht entfernbare Tätowierung?

Wie viele dieser fünf Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier) ☐ 5 (fünf)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

Institut für Experimentelle Psychologie
Abteilung für Diagnostik und Differentielle Psychologie
Prof. Dr. Jochen Musch

Heinrich-Heine-Universität Düsseldorf
Diplomarbeit Thomas Seppelfricke
Okt./ Nov. 2007_10

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der Umfrageforschung bitte ich Sie, den vorliegenden Kurz-Fragebogen jetzt gleich auszufüllen; ich werde ihn in 2 Minuten, noch vor Beginn der Veranstaltung, wieder einsammeln.

Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird ausschließlich von mir, für den alleinigen Zweck einer wissenschaftlichen Abschlussarbeit vollkommen anonym ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Ihnen wird im Folgenden ein Fragenblock á fünf Fragen präsentiert.

Alle Fragen sind derart gestellt, dass sich diese entweder mit „ja“ oder mit „nein“ beantworten lassen. Beantworten Sie die Fragen im Stillen für sich und kreuzen Sie am Ende des Blockes an, wie viele der fünf Fragen Sie insgesamt bejaht haben!

1. Sind Sie verheiratet?
2. Haben Sie schon einmal über das Internet telefoniert (z.B. mit Skype oder Voice-over-IP)?
3. Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?
4. Wohnen Sie derzeit zur Untermiete?
5. Haben Sie eine nicht entfernbare Tätowierung?

Wie viele dieser fünf Fragen beantworten Sie mit „Ja“?

☐ 0 (keine) ☐ 1 (eine) ☐ 2 (zwei) ☐ 3 (drei) ☐ 4 (vier) ☐ 5 (fünf)

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

Institut für Experimentelle Psychologie
Abteilung für Diagnostik und Differentielle Psychologie
Prof. Dr. Jochen Musch

Heinrich-Heine-Universität Düsseldorf
Diplomarbeit Thomas Seppelfricke
Okt./ Nov. 2007_11-17

SCHNELL-UMFRAGE



Sehr geehrter Studierender/ Sehr geehrte Studierende,

für meine Diplomarbeit zum Vergleich unterschiedlicher Methoden der experimentellen Umfrageforschung möchte ich Dich bitten, den vorliegenden Kurz-Fragebogen umgehend auszufüllen; er soll möglichst noch vor Beginn der Vorlesung wieder eingesammelt werden. Der Fragebogen steht nicht im Zusammenhang mit der Lehrveranstaltung und wird vollkommen anonym, ausschließlich von mir und dem meine Diplomarbeit betreuenden Professor, vertraulich ausgewertet.

Ihr Geschlecht: ☐ männlich ☐ weiblich

Wir bitten Sie um eine Antwort auf die folgenden Fragen!

1. Haben Sie schon einmal in einer Prüfung einen vorher präparierten Spickzettel zum Einsatz gebracht?

- ☐ ja
☐ nein

2. Haben Sie, um ein Fernbleiben vom Unterricht, einer Lehrveranstaltung oder einer Prüfung zu rechtfertigen, schon einmal ein ärztliches Attest ausstellen lassen, obwohl es hierfür gar keinen ärztlichen Grund gab?

- ☐ ja
☐ nein

3. Haben Sie schon einmal eine gefälschte Unterschrift verwendet, um eine schlechte Note vor ihren Eltern zu verheimlichen?

- ☐ ja
☐ nein

Herzlichen Dank für Ihre Unterstützung! Bitte falten Sie das Blatt und geben Sie es an den Rand der Stuhlreihe weiter, wo ich es einsammeln werde.

Anhang B: Datenfile der Gesamtauswertung

Data file (n=1700)

```
1 54
2 28
3 15
4 2
5 1
6 13
7 37
8 37
9 13
10 0
11 11
12 51
13 24
14 12
15 2
16 42
17 42
18 13
19 3
20 0
21 27
22 48
23 19
24 4
25 2
26 31
27 40
28 27
29 1
30 1
31 37
32 36
33 21
34 5
35 1
36 19
37 34
38 36
39 8
40 2
41 1
42 26
43 44
44 21
45 8
46 1
47 0
48 32
49 48
50 14
51 2
52 1
53 3
54 351
55 349
56 182
57 518
58 116
59 584
```

====

Anhang C: Gleichungsfile der Gesamtauswertung

```

214
1 5 n6*s2*s3*n7
1 4 n6*s2*s3*(1-n7)
1 4 n6*s2*(1-s3)*n7
1 3 n6*s2*(1-s3)*(1-n7)
1 4 n6*(1-s2)*s3*n7
1 3 n6*(1-s2)*s3*(1-n7)
1 3 n6*(1-s2)*(1-s3)*n7
1 2 n6*(1-s2)*(1-s3)*(1-n7)
1 4 (1-n6)*s2*s3*n7
1 3 (1-n6)*s2*s3*(1-n7)
1 3 (1-n6)*s2*(1-s3)*n7
1 2 (1-n6)*s2*(1-s3)*(1-n7)
1 3 (1-n6)*(1-s2)*s3*n7
1 2 (1-n6)*(1-s2)*s3*(1-n7)
1 2 (1-n6)*(1-s2)*(1-s3)*n7
1 1 (1-n6)*(1-s2)*(1-s3)*(1-n7)
2 10 n7*s3*s1*n5
2 9 n7*s3*s1*(1-n5)
2 9 n7*s3*(1-s1)*n5
2 8 n7*s3*(1-s1)*(1-n5)
2 9 n7*(1-s3)*s1*n5
2 8 n7*(1-s3)*s1*(1-n5)
2 8 n7*(1-s3)*(1-s1)*n5
2 7 n7*(1-s3)*(1-s1)*(1-n5)
2 9 (1-n7)*s3*s1*n5
2 8 (1-n7)*s3*s1*(1-n5)
2 8 (1-n7)*s3*(1-s1)*n5
2 7 (1-n7)*s3*(1-s1)*(1-n5)
2 8 (1-n7)*(1-s3)*s1*n5
2 7 (1-n7)*(1-s3)*s1*(1-n5)
2 7 (1-n7)*(1-s3)*(1-s1)*n5
2 6 (1-n7)*(1-s3)*(1-s1)*(1-n5)
3 15 n5*s1*s2*n6
3 14 n5*s1*s2*(1-n6)
3 14 n5*s1*(1-s2)*n6
3 13 n5*s1*(1-s2)*(1-n6)
3 14 n5*(1-s1)*s2*n6
3 13 n5*(1-s1)*s2*(1-n6)
3 13 n5*(1-s1)*(1-s2)*n6
3 12 n5*(1-s1)*(1-s2)*(1-n6)
3 14 (1-n5)*s1*s2*n6
3 13 (1-n5)*s1*s2*(1-n6)
3 13 (1-n5)*s1*(1-s2)*n6
3 12 (1-n5)*s1*(1-s2)*(1-n6)
3 13 (1-n5)*(1-s1)*s2*n6
3 12 (1-n5)*(1-s1)*s2*(1-n6)
3 12 (1-n5)*(1-s1)*(1-s2)*n6
3 11 (1-n5)*(1-s1)*(1-s2)*(1-n6)
4 20 n4*n5*n6*n7
4 19 n4*n5*n6*(1-n7)
4 19 n4*n5*(1-n6)*n7
4 18 n4*n5*(1-n6)*(1-n7)
4 19 n4*(1-n5)*n6*n7
4 18 n4*(1-n5)*n6*(1-n7)
4 18 n4*(1-n5)*(1-n6)*n7
4 17 n4*(1-n5)*(1-n6)*(1-n7)
4 19 (1-n4)*n5*n6*n7
4 18 (1-n4)*n5*n6*(1-n7)

```

```

4 18 (1-n4)*n5*(1-n6)*n7
4 17 (1-n4)*n5*(1-n6)*(1-n7)
4 18 (1-n4)*(1-n5)*n6*n7
4 17 (1-n4)*(1-n5)*n6*(1-n7)
4 17 (1-n4)*(1-n5)*(1-n6)*n7
4 16 (1-n4)*(1-n5)*(1-n6)*(1-n7)
5 25 n4*s3*s2*n5
5 24 n4*s3*s2*(1-n5)
5 24 n4*s3*(1-s2)*n5
5 23 n4*s3*(1-s2)*(1-n5)
5 24 n4*(1-s3)*s2*n5
5 23 n4*(1-s3)*s2*(1-n5)
5 23 n4*(1-s3)*(1-s2)*n5
5 22 n4*(1-s3)*(1-s2)*(1-n5)
5 24 (1-n4)*s3*s2*n5
5 23 (1-n4)*s3*s2*(1-n5)
5 23 (1-n4)*s3*(1-s2)*n5
5 22 (1-n4)*s3*(1-s2)*(1-n5)
5 23 (1-n4)*(1-s3)*s2*n5
5 22 (1-n4)*(1-s3)*s2*(1-n5)
5 22 (1-n4)*(1-s3)*(1-s2)*n5
5 21 (1-n4)*(1-s3)*(1-s2)*(1-n5)
6 30 n6*s1*s3*n4
6 29 n6*s1*s3*(1-n4)
6 29 n6*s1*(1-s3)*n4
6 28 n6*s1*(1-s3)*(1-n4)
6 29 n6*(1-s1)*s3*n4
6 28 n6*(1-s1)*s3*(1-n4)
6 28 n6*(1-s1)*(1-s3)*n4
6 27 n6*(1-s1)*(1-s3)*(1-n4)
6 29 (1-n6)*s1*s3*n4
6 28 (1-n6)*s1*s3*(1-n4)
6 28 (1-n6)*s1*(1-s3)*n4
6 27 (1-n6)*s1*(1-s3)*(1-n4)
6 28 (1-n6)*(1-s1)*s3*n4
6 27 (1-n6)*(1-s1)*s3*(1-n4)
6 27 (1-n6)*(1-s1)*(1-s3)*n4
6 26 (1-n6)*(1-s1)*(1-s3)*(1-n4)
7 35 n7*s2*s1*n4
7 34 n7*s2*s1*(1-n4)
7 34 n7*s2*(1-s1)*n4
7 33 n7*s2*(1-s1)*(1-n4)
7 34 n7*(1-s2)*s1*n4
7 33 n7*(1-s2)*s1*(1-n4)
7 33 n7*(1-s2)*(1-s1)*n4
7 32 n7*(1-s2)*(1-s1)*(1-n4)
7 34 (1-n7)*s2*s1*n4
7 33 (1-n7)*s2*s1*(1-n4)
7 33 (1-n7)*s2*(1-s1)*n4
7 32 (1-n7)*s2*(1-s1)*(1-n4)
7 33 (1-n7)*(1-s2)*s1*n4
7 32 (1-n7)*(1-s2)*s1*(1-n4)
7 32 (1-n7)*(1-s2)*(1-s1)*n4
7 31 (1-n7)*(1-s2)*(1-s1)*(1-n4)
8 41 uctn4*uctn5*uct1*uctn6*uctn7
8 40 uctn4*uctn5*uct1*uctn6*(1-uctn7)
8 40 uctn4*uctn5*uct1*(1-uctn6)*uctn7
8 39 uctn4*uctn5*uct1*(1-uctn6)*(1-uctn7)
8 40 uctn4*uctn5*(1-uct1)*uctn6*uctn7
8 39 uctn4*uctn5*(1-uct1)*uctn6*(1-uctn7)
8 39 uctn4*uctn5*(1-uct1)*(1-uctn6)*uctn7
8 38 uctn4*uctn5*(1-uct1)*(1-uctn6)*(1-uctn7)
8 40 uctn4*(1-uctn5)*uct1*uctn6*uctn7

```

```

8 39 uctn4*(1-uctn5)*uct1*uctn6*(1-uctn7)
8 39 uctn4*(1-uctn5)*uct1*(1-uctn6)*uctn7
8 38 uctn4*(1-uctn5)*uct1*(1-uctn6)*(1-uctn7)
8 39 uctn4*(1-uctn5)*(1-uct1)*uctn6*uctn7
8 38 uctn4*(1-uctn5)*(1-uct1)*uctn6*(1-uctn7)
8 38 uctn4*(1-uctn5)*(1-uct1)*(1-uctn6)*uctn7
8 37 uctn4*(1-uctn5)*(1-uct1)*(1-uctn6)*(1-uctn7)
8 40 (1-uctn4)*uctn5*uct1*uctn6*uctn7
8 39 (1-uctn4)*uctn5*uct1*uctn6*(1-uctn7)
8 39 (1-uctn4)*uctn5*uct1*(1-uctn6)*uctn7
8 38 (1-uctn4)*uctn5*uct1*(1-uctn6)*(1-uctn7)
8 39 (1-uctn4)*uctn5*(1-uct1)*uctn6*uctn7
8 38 (1-uctn4)*uctn5*(1-uct1)*uctn6*(1-uctn7)
8 38 (1-uctn4)*uctn5*(1-uct1)*(1-uctn6)*uctn7
8 37 (1-uctn4)*uctn5*(1-uct1)*(1-uctn6)*(1-uctn7)
8 39 (1-uctn4)*(1-uctn5)*uct1*uctn6*uctn7
8 38 (1-uctn4)*(1-uctn5)*uct1*uctn6*(1-uctn7)
8 38 (1-uctn4)*(1-uctn5)*uct1*(1-uctn6)*uctn7
8 37 (1-uctn4)*(1-uctn5)*uct1*(1-uctn6)*(1-uctn7)
8 38 (1-uctn4)*(1-uctn5)*(1-uct1)*uctn6*uctn7
8 37 (1-uctn4)*(1-uctn5)*(1-uct1)*uctn6*(1-uctn7)
8 37 (1-uctn4)*(1-uctn5)*(1-uct1)*(1-uctn6)*uctn7
8 36 (1-uctn4)*(1-uctn5)*(1-uct1)*(1-uctn6)*(1-uctn7)
9 47 uctn4*uctn5*uct2*uctn6*uctn7
9 46 uctn4*uctn5*uct2*uctn6*(1-uctn7)
9 46 uctn4*uctn5*uct2*(1-uctn6)*uctn7
9 45 uctn4*uctn5*uct2*(1-uctn6)*(1-uctn7)
9 46 uctn4*uctn5*(1-uct2)*uctn6*uctn7
9 45 uctn4*uctn5*(1-uct2)*uctn6*(1-uctn7)
9 45 uctn4*uctn5*(1-uct2)*(1-uctn6)*uctn7
9 44 uctn4*uctn5*(1-uct2)*(1-uctn6)*(1-uctn7)
9 46 uctn4*(1-uctn5)*uct2*uctn6*uctn7
9 45 uctn4*(1-uctn5)*uct2*uctn6*(1-uctn7)
9 45 uctn4*(1-uctn5)*uct2*(1-uctn6)*uctn7
9 44 uctn4*(1-uctn5)*uct2*(1-uctn6)*(1-uctn7)
9 45 uctn4*(1-uctn5)*(1-uct2)*uctn6*uctn7
9 44 uctn4*(1-uctn5)*(1-uct2)*uctn6*(1-uctn7)
9 44 uctn4*(1-uctn5)*(1-uct2)*(1-uctn6)*uctn7
9 43 uctn4*(1-uctn5)*(1-uct2)*(1-uctn6)*(1-uctn7)
9 46 (1-uctn4)*uctn5*uct2*uctn6*uctn7
9 45 (1-uctn4)*uctn5*uct2*uctn6*(1-uctn7)
9 45 (1-uctn4)*uctn5*uct2*(1-uctn6)*uctn7
9 44 (1-uctn4)*uctn5*uct2*(1-uctn6)*(1-uctn7)
9 45 (1-uctn4)*uctn5*(1-uct2)*uctn6*uctn7
9 44 (1-uctn4)*uctn5*(1-uct2)*uctn6*(1-uctn7)
9 44 (1-uctn4)*uctn5*(1-uct2)*(1-uctn6)*uctn7
9 43 (1-uctn4)*uctn5*(1-uct2)*(1-uctn6)*(1-uctn7)
9 45 (1-uctn4)*(1-uctn5)*uct2*uctn6*uctn7
9 44 (1-uctn4)*(1-uctn5)*uct2*uctn6*(1-uctn7)
9 44 (1-uctn4)*(1-uctn5)*uct2*(1-uctn6)*uctn7
9 43 (1-uctn4)*(1-uctn5)*uct2*(1-uctn6)*(1-uctn7)
9 44 (1-uctn4)*(1-uctn5)*(1-uct2)*uctn6*uctn7
9 43 (1-uctn4)*(1-uctn5)*(1-uct2)*uctn6*(1-uctn7)
9 43 (1-uctn4)*(1-uctn5)*(1-uct2)*(1-uctn6)*uctn7
9 42 (1-uctn4)*(1-uctn5)*(1-uct2)*(1-uctn6)*(1-uctn7)
10 53 uctn4*uctn5*uct3*uctn6*uctn7
10 52 uctn4*uctn5*uct3*uctn6*(1-uctn7)
10 52 uctn4*uctn5*uct3*(1-uctn6)*uctn7
10 51 uctn4*uctn5*uct3*(1-uctn6)*(1-uctn7)
10 52 uctn4*uctn5*(1-uct3)*uctn6*uctn7
10 51 uctn4*uctn5*(1-uct3)*uctn6*(1-uctn7)
10 51 uctn4*uctn5*(1-uct3)*(1-uctn6)*uctn7
10 50 uctn4*uctn5*(1-uct3)*(1-uctn6)*(1-uctn7)

```

```

10 52 uctn4*(1-uctn5)*uct3*uctn6*uctn7
10 51 uctn4*(1-uctn5)*uct3*uctn6*(1-uctn7)
10 51 uctn4*(1-uctn5)*uct3*(1-uctn6)*uctn7
10 50 uctn4*(1-uctn5)*uct3*(1-uctn6)*(1-uctn7)
10 51 uctn4*(1-uctn5)*(1-uct3)*uctn6*uctn7
10 50 uctn4*(1-uctn5)*(1-uct3)*uctn6*(1-uctn7)
10 50 uctn4*(1-uctn5)*(1-uct3)*(1-uctn6)*uctn7
10 49 uctn4*(1-uctn5)*(1-uct3)*(1-uctn6)*(1-uctn7)
10 52 (1-uctn4)*uctn5*uct3*uctn6*uctn7
10 51 (1-uctn4)*uctn5*uct3*uctn6*(1-uctn7)
10 51 (1-uctn4)*uctn5*uct3*(1-uctn6)*uctn7
10 50 (1-uctn4)*uctn5*uct3*(1-uctn6)*(1-uctn7)
10 51 (1-uctn4)*uctn5*(1-uct3)*uctn6*uctn7
10 50 (1-uctn4)*uctn5*(1-uct3)*uctn6*(1-uctn7)
10 50 (1-uctn4)*uctn5*(1-uct3)*(1-uctn6)*uctn7
10 49 (1-uctn4)*uctn5*(1-uct3)*(1-uctn6)*(1-uctn7)
10 51 (1-uctn4)*(1-uctn5)*uct3*uctn6*uctn7
10 50 (1-uctn4)*(1-uctn5)*uct3*uctn6*(1-uctn7)
10 50 (1-uctn4)*(1-uctn5)*uct3*(1-uctn6)*uctn7
10 49 (1-uctn4)*(1-uctn5)*uct3*(1-uctn6)*(1-uctn7)
10 50 (1-uctn4)*(1-uctn5)*(1-uct3)*uctn6*uctn7
10 49 (1-uctn4)*(1-uctn5)*(1-uct3)*uctn6*(1-uctn7)
10 49 (1-uctn4)*(1-uctn5)*(1-uct3)*(1-uctn6)*uctn7
10 48 (1-uctn4)*(1-uctn5)*(1-uct3)*(1-uctn6)*(1-uctn7)
11 54 d1
11 55 (1-d1)
12 56 d2
12 57 (1-d2)
13 58 d3
13 59 (1-d3)
====

```

Anhang D: Multinomiale Signifikanztests der Gesamtauswertung

Model: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Equation-file.eqn'

Data: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Datafile.mdt'

No. parameters estimated:	13
-2*log(likelihood):	4939.903020
AIC:	4965.903020
BIC:	5044.412066
G ² :	83.983834

Parameter values [lower and upper limit of 95% CI]

1	d1 = 0.501429	[0.464388 0.538469]
2	d2 = 0.260000	[0.227506 0.292494]
3	d3 = 0.165714	[0.138169 0.193259]
4	n4 = 0.139369	[0.040818 0.237919]
5	n5 = 0.447724	[0.356500 0.538947]
6	n6 = 0.144204	[0.044801 0.243607]
7	n7 = 0.148219	[0.048309 0.248129]
8	s1 = 0.519535	[0.426090 0.612980]
9	s2 = 0.216748	[0.113037 0.320459]
10	s3 = 0.242541	[0.138300 0.346782]
11	uct1 = 0.520956	[0.309593 0.732320]
12	uct2 = 0.256779	[0.035047 0.478510]
13	uct3 = 0.150361	[-0.062250 0.362972]
14	uctn4 = equal to n4	
15	uctn5 = equal to n5	
16	uctn6 = equal to n6	
17	uctn7 = equal to n7	

Note: Confidence intervals are based on expected Fisher information.

Model: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Equation-file.eqn'

Data: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Datafile.mdt'

No. parameters estimated:	12
-2*log(likelihood):	4939.944203
AIC:	4963.944203
BIC:	5036.414092
G ² :	84.025018

Parameter values [lower and upper limit of 95% CI]

1	d1 = equal to uct1
2	d2 = 0.260000 [0.227506 0.292494]
3	d3 = 0.165714 [0.138169 0.193259]
4	n4 = 0.140178 [0.042597 0.237760]
5	n5 = 0.449793 [0.363232 0.536354]
6	n6 = 0.145116 [0.046905 0.243327]
7	n7 = 0.149130 [0.050464 0.247796]
8	s1 = 0.518563 [0.426099 0.611027]
9	s2 = 0.215784 [0.114024 0.317544]
10	s3 = 0.241241 [0.138878 0.343604]
11	uct1 = 0.502177 [0.465678 0.538676]
12	uct2 = 0.252450 [0.043578 0.461323]
13	uct3 = 0.148264 [-0.051597 0.348124]
14	uctn4 = equal to n4
15	uctn5 = equal to n5
16	uctn6 = equal to n6
17	uctn7 = equal to n7

Note: Confidence intervals are based on expected Fisher information.

Model: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Equation-file.eqn'

Data: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Datafile.mdt'

No. parameters estimated:	12
-2*log(likelihood):	4939.904020
AIC:	4963.904020
BIC:	5036.373908
G ² :	83.984834

Parameter values [lower and upper limit of 95% CI]

1	d1 = 0.501429	[0.464388 0.538469]
2	d2 = equal to uct2	
3	d3 = 0.165714	[0.138169 0.193259]
4	n4 = 0.139223	[0.042280 0.236165]
5	n5 = 0.447367	[0.359883 0.534851]
6	n6 = 0.144036	[0.046525 0.241546]
7	n7 = 0.148055	[0.050080 0.246030]
8	s1 = 0.519706	[0.427494 0.611918]
9	s2 = 0.216920	[0.115673 0.318167]
10	s3 = 0.242773	[0.140892 0.344654]
11	uct1 = 0.521586	[0.322843 0.720329]
12	uct2 = 0.259915	[0.227767 0.292062]
13	uct3 = 0.150738	[-0.046323 0.347800]
14	uctn4 = equal to n4	
15	uctn5 = equal to n5	
16	uctn6 = equal to n6	
17	uctn7 = equal to n7	

Note: Confidence intervals are based on expected Fisher information.

Model: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Equation-file.eqn'

Data: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Datafile.mdt'

No. parameters estimated:	12
-2*log(likelihood):	4939.949783
AIC:	4963.949783
BIC:	5036.419672
G ² :	84.030597

Parameter values [lower and upper limit of 95% CI]

1	d1 = 0.501429	[0.464388 0.538469]
2	d2 = 0.260000	[0.227506 0.292494]
3	d3 = equal to uct3	
4	n4 = 0.138732	[0.042071 0.235394]
5	n5 = 0.445671	[0.357360 0.533982]
6	n6 = 0.143472	[0.046258 0.240686]
7	n7 = 0.147498	[0.049805 0.245190]
8	s1 = 0.520412	[0.428271 0.612554]
9	s2 = 0.217585	[0.116486 0.318684]
10	s3 = 0.243676	[0.141929 0.345423]
11	uct1 = 0.524079	[0.326168 0.721990]
12	uct2 = 0.260606	[0.055145 0.466067]
13	uct3 = 0.165126	[0.137844 0.192408]
14	uctn4 = equal to n4	
15	uctn5 = equal to n5	
16	uctn6 = equal to n6	
17	uctn7 = equal to n7	

Note: Confidence intervals are based on expected Fisher information.

Model: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Equation-file.eqn'

Data: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Datafile.mdt'

No. parameters estimated:	12
-2*log(likelihood):	4940.039586
AIC:	4964.039586
BIC:	5036.509475
G ² :	84.120401

Parameter values [lower and upper limit of 95% CI]

1	d1 = equal to s1
2	d2 = 0.260000 [0.227506 0.292494]
3	d3 = 0.165714 [0.138169 0.193259]
4	n4 = 0.139835 [0.041407 0.238262]
5	n5 = 0.450742 [0.361153 0.540331]
6	n6 = 0.144817 [0.045122 0.244511]
7	n7 = 0.148654 [0.048482 0.248827]
8	s1 = 0.504107 [0.469606 0.538608]
9	s2 = 0.218799 [0.115568 0.322030]
10	s3 = 0.245734 [0.142081 0.349388]
11	uct1 = 0.517481 [0.307466 0.727496]
12	uct2 = 0.252600 [0.033325 0.471875]
13	uct3 = 0.148275 [-0.061829 0.358378]
14	uctn4 = equal to n4
15	uctn5 = equal to n5
16	uctn6 = equal to n6
17	uctn7 = equal to n7

Note: Confidence intervals are based on expected Fisher information.

Model: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Equation-file.eqn'

Data: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Datafile.mdt'

No. parameters estimated:	12
-2*log(likelihood):	4940.794933
AIC:	4964.794933
BIC:	5037.264822
G ² :	84.875748

Parameter values [lower and upper limit of 95% CI]

1	d1 = 0.501429	[0.464388 0.538469]
2	d2 = equal to s2	
3	d3 = 0.165714	[0.138169 0.193259]
4	n4 = 0.136850	[0.038570 0.235129]
5	n5 = 0.443823	[0.352370 0.535276]
6	n6 = 0.141166	[0.042138 0.240195]
7	n7 = 0.145339	[0.046888 0.243789]
8	s1 = 0.513574	[0.420012 0.607137]
9	s2 = 0.254483	[0.223641 0.285325]
10	s3 = 0.235484	[0.133491 0.337476]
11	uct1 = 0.530087	[0.324708 0.735466]
12	uct2 = 0.268153	[0.051870 0.484436]
13	uct3 = 0.155969	[-0.051449 0.363387]
14	uctn4 = equal to n4	
15	uctn5 = equal to n5	
16	uctn6 = equal to n6	
17	uctn7 = equal to n7	

Note: Confidence intervals are based on expected Fisher information.

Model: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Equation-file.eqn'

Data: 'C:\Eigene Dateien\Diplomarbeit Seppelfricke\Auswertung\Datafile.mdt'

No. parameters estimated:	12
-2*log(likelihood):	4942.472567
AIC:	4966.472567
BIC:	5038.942456
G ² :	86.553382

Parameter values [lower and upper limit of 95% CI]

1	d1 = 0.501429	[0.464388 0.538469]
2	d2 = 0.260000	[0.227506 0.292494]
3	d3 = equal to s3	
4	n4 = 0.145391	[0.047852 0.242930]
5	n5 = 0.454422	[0.365075 0.543768]
6	n6 = 0.150894	[0.053031 0.248758]
7	n7 = 0.156578	[0.057542 0.255613]
8	s1 = 0.531566	[0.440851 0.622282]
9	s2 = 0.228218	[0.127584 0.328851]
10	s3 = 0.172721	[0.145734 0.199708]
11	uct1 = 0.499776	[0.287332 0.712219]
12	uct2 = 0.231284	[0.013385 0.449184]
13	uct3 = 0.138881	[-0.069478 0.347239]
14	uctn4 = equal to n4	
15	uctn5 = equal to n5	
16	uctn6 = equal to n6	
17	uctn7 = equal to n7	

Note: Confidence intervals are based on expected Fisher information.