

Frank Multrus, Sandra Majer

Methodenbericht zum 13. Studierendensurvey

Vergleich Papier-Onlinebefragung

- Werkstattbericht -

Universität Konstanz: Arbeitsgruppe Hochschulforschung

Hefte zur Bildungs- und Hochschulforschung 95



Frank Multrus, Sandra Majer

Methodenbericht zum 13. Studierendensurvey

Vergleich Papier-Onlinebefragung

Werkstattbericht

Hefte zur Bildungs- und Hochschulforschung (95)

Arbeitsgruppe Hochschulforschung, Universität Konstanz, September 2017

Herausgeber der Reihe „Hefte zur Bildungs- und Hochschulforschung“:

Arbeitsgruppe Hochschulforschung, Universität Konstanz,
Fachbereich Geschichte und Soziologie, 78457 Konstanz
Tel. 07531/88-2896

Die AG Hochschulforschung im Internet:
<http://www.uni-konstanz.de/ag-hochschulforschung>

ISSN 1616-0398

Inhalt

Der Studierendensurvey	5
Grundlagen: Konzepte und Themen	5
Leistungsmessung im Hochschulbereich als Dauerbeobachtung	5
Entwicklung und Abgrenzung des Studierendensurveys	7
Themenbereiche des Studierendensurveys	7
Methodenbericht zum 13. Studierendensurvey	9
1. Erhebung zum 13. Studierendensurvey	10
Design der 13. Erhebung	10
Untersuchungsfragen zum Erhebungsdesign	11
Weiteres Vorgehen: Analysen und Vergleiche	12
2. Datenbasis: Stichprobengröße im 13. Studierendensurvey	13
2.1 Rahmenbedingungen und Probleme	13
2.1.1 Repräsentativität	13
2.1.2 Probleme der Repräsentativität - Stichprobenfehler	16
3. Non-Response	21
3.1 Brutto-Rücklauf	22
3.1.1 Unit Non-Response	24
3.1.2 Netto-Rücklauf	24
3.1.3 Gründe für geringen Rücklauf	25
3.1.4 Besondere Probleme der Online-Erhebung	26
3.1.5 Probleme geringer Rücklaufquoten und Umgang mit Non-Response	26
3.2 Item Non-Response	27
3.2.1 Unterschiede zwischen randomisierter und selbstselektierter Gruppe	28
3.2.2 Item Non-Response mit restriktivem Unit Non-Response	28
3.2.3 Item Non-Response in Abhängigkeit von der Bearbeitungsdauer	29
3.2.4 Item Non-Response im Verlauf der Befragung	29
3.2.5 Hinweis auf problematische Fragen – Zahleneingabe	32
3.2.6 Unterschiede in der Darstellung zwischen Papier- und Onlinefragebogen	33
3.2.7 Letztes Item einer langen Itembatterie	34
3.3 Befragungsabbrüche	35
3.3.1 Completion Rates	37
3.3.2 Welche Faktoren begünstigen Abbrüche?	37
3.3.3 Fragebogenspezifische Faktoren	38

3.3.4	Abbruch und Antwortqualität	43
4.	Analyse von Paradata	47
4.1	Bearbeitungsdauer	47
4.1.1	Bearbeitungsdauer in den Gruppen der Online-Befragung.....	48
4.1.2	Item Non-Response und Bearbeitungsdauer	49
4.1.3	Dauer zwischen erstmaligem Zugriff und Abschluss/ Abbruch der Befragung .	50
4.2	Devices.....	51
4.2.1	Verwendete Geräte in der Online-Befragung.....	52
5.	Strukturvergleiche zwischen Untergruppen	56
5.1	Papier vs. Onlinegruppe (Mode-Effekt) bei Fremdselektion.....	57
5.2	Papier vs. Onlinegruppe (Mode-Effekt) bei Selbstselektion	60
5.2.1	Selektionseffekt bei Papierbefragung	60
5.2.2	Selektionseffekt bei Online-Befragung	67
5.3	Multivariater Vergleich: Einfluss von soziodemographischen Merkmalen.....	68
5.3.1	Effekte der Selbstselektion in der Papiergruppe	69
5.3.2	Mode-Effekte	70
6.	Strukturvergleiche mit Grundgesamtheit	72
6.1	Vergleich der Ergebnisse mit der 20. Sozialerhebung.....	78
7.	Vergleich der Designgruppen an inhaltlichen Indikatoren	82
7.1	Prüfung nominaler bzw. ordinaler Merkmale	82
7.2	Prüfung intervallskalierter Merkmale der Studiensituation	87
8.	Effekte von Designparameter	96
8.1	Motivationsseiten.....	96
8.2	Ankündigung und Erinnerung.....	101
9.	Herausforderungen bei Online-Befragungen	107
9.1	Teilnahmebereitschaft.....	107
9.2	Datenschutzbestimmungen.....	108
9.2.1	Personenbezogene Daten	109
9.2.2	Einwilligungsprinzip – Informed Consent.....	109
9.2.3	Technische und organisatorische Maßnahmen	109
9.2.4	Anonymisierung und Pseudonymisierung	110
9.2.5	Situation und Vorgehen im Studierendensurvey.....	110
9.3	Responsive Design	111
9.3.1	Device Effekte.....	112

9.3.2	Barrierefreiheit	114
9.4	Technische Anforderungen	114
10	Fazit und Ausblick.....	116
Anhang.....		129
	Anhänge zu Kapitel 3 – Unterschiede in der Darstellung zwischen Papier- und Onlinefragebogen.....	130
	Anhang zu Kapitel 4 – Tage bis zum Abschluss der Befragung	132
	Anhang zu Kapitel 5 – Ergebnisse der logistischen Regression (in STATA)	133

Der Studierendensurvey

Der Studierendensurvey ist eine Befragung von Studierenden an Universitäten und Fachhochschulen über ihre Erfahrungen im Studium. Das Instrument wurde von der Arbeitsgruppe Hochschulforschung Anfang der 80er Jahre an der Universität Konstanz entwickelt. Als Forschungsprojekt „Studiensituation und studentische Orientierungen“ wird es vom Bundesministerium für Bildung und Forschung gefördert und von einem wissenschaftlichen Beirat begleitet.

Mit regelmäßigen, standardisierten Befragungen werden die Erfahrungen und Bewertungen, Ansichten und Einstellungen von Studierenden im Sinne einer Dauerbeobachtung kontinuierlich und systematisch erfasst. Damit sollen verlässliche Befunde zur studentischen Situation und den studentischen Orientierungen gegenüber Studium, Beruf und Politik sowie zu deren Entwicklungen zusammengetragen werden. Seit dem Wintersemester 1982/83 wird alle zwei bis drei Jahre jeweils eine große Zahl deutscher Studierender aller Fachrichtungen zufällig ausgewählt und zu einem breiten Themenspektrum befragt. In den bislang 13 Erhebungen haben sich rund 106.000 Studierende an Universitäten und Fachhochschulen beteiligt.

Grundlagen: Konzepte und Themen

Im Folgenden werden grundlegende Informationen über den Studierendensurvey in aller Kürze dargestellt, um einen Eindruck über den Umfang und die thematische Breite des Instruments zu vermitteln.

Leistungsmessung im Hochschulbereich als Dauerbeobachtung

Ein Kern des Studierendensurveys ist seit Beginn der Erhebungen die Leistungsmessung im Hochschulbereich, um Grundlagen für die Hochschulpolitik und deren öffentliche Debatte bereit zu stellen. Dazu werden bewusst subjektive Indikatoren verwendet, indem Studierende nach ihren Erfahrungen und Bewertungen befragt werden.

Objektive Indikatoren beziehen sich auf die objektiv messbaren „Lebensumstände oder Dimensionen der Sozialstruktur“ (Noll, 2000, S. 4; Noll, 1989), wie zum Beispiel die Arbeitslosenquote, das Bruttoinlandsprodukt oder die Abiturientenquote. Es handelt sich um Kennzahlen, die Lebensbedingungen quantifizierbar machen sollen. Da sie nicht durch persönliche Einstellungen beeinflusst werden, stellen sie eine objektive Bewertung eines Sachverhalts dar.

Im Gegensatz dazu berücksichtigen subjektive Indikatoren die Sichtweise der Befragten, die ihre Einschätzung oder Einstellungen zu einem Thema artikulieren sollen. Dabei kann es sich um Aussagen zur Arbeitsplatzzufriedenheit, zum politischen Interesse oder zur Bewertung der eigenen Lebensqualität handeln. Hier wird der/die Befragte nicht nur als Träger von Informationen betrachtet, sondern um seine persönliche Bewertung gebeten.

Wo objektive Indikatoren lediglich eine Art Bestandsaufnahme der Studiensituation liefern können, ergänzen subjektive Indikatoren das Bild um die individuelle Wahrnehmung dieser Situation durch die Befragten. Identische Lebensbedingungen können von verschiedenen Personen ganz unterschiedlich wahrgenommen werden, weshalb die individuellen Ansichten der Studierenden berücksichtigt werden müssen, um zu einer umfassenden Beschreibung/Bewertung der Studiensituation zu gelangen. Als Rezipienten des Ausbildungssystems Hochschule erleben sie dieses direkt in all seinen unterschiedlichen Bereichen, Bedingungen und Ausgestaltungen. Ihre Antworten liefern ein subjektives Abbild des erfahrenen Systems, von dem sie aber nicht nur Teil sind, sondern dessen Berechtigung sie erst liefern. Als vorrangige Betroffene sind ihre Ansichten und Erfahrungen bedeutungsvoll

und notwendig für eine Analyse der Studienbedingungen, der Studiensituation und der studentischen Orientierungen.

Dieser Leistungsmessung unter Berücksichtigung der subjektiven Perspektive nimmt sich der Studierendensurvey an. Denn als Gegenstand solcher Leistungsmessungen im Hochschulbereich wurden frühzeitig Ende der 70er Jahre sechs Bereiche benannt und in das Erhebungsinstrument integriert.

- **„Effizienz“** des Studiums: z. B. Organisation der Lehre, Studienstrategien, Prüfungen (Prüfungsvorbereitung), Studiendauer und Verzögerungen, Studienabbruch; zur Diskussion über Nutzen und Erfolg.
- **„Qualifikation“** und Studierertrag: z. B. Praxisbezug, Forschungsbeteiligung, Erwerb von Fachwissen und allgemeinen Kompetenzen, Professionalisierung; zur Diskussion über Praxisbezug und Berufsvorbereitung.
- **„Evaluation“** und Studienqualität: z. B. Anforderungen im Fachstudium, Beurteilung der Lehre, Kontakte im Studium, Beratung und Betreuung durch Lehrende; zur Diskussion über Studienqualität und Studierbarkeit.
- **„Sozialisation“** und Orientierungen: z. B. Relevanz von Wissenschaft und Forschung, Werthaltungen und Ansprüche, persönliche Entwicklung, soziale Verantwortung; zur Diskussion über Bildung und Werte.
- **„Selektion“** und soziale Chancen: z. B. Hochschulzugang, Chancen von Studentinnen, Folgen sozialer Herkunft im Studium, Förderung des wissenschaftlichen Nachwuchses; zur Diskussion über Gerechtigkeit und Fairness.
- **„Platzierung“** der Absolventen: z. B. berufliche Aussichten, angestrebte Tätigkeitsfelder, Übergang auf den Arbeitsmarkt, Flexibilität und Identität, Berufsbefähigung; zur Diskussion über Chancen und Ungleichheit.

Das Konzept der **gesellschaftlichen Dauerbeobachtung**, dem auch der Studierendensurvey verpflichtet ist, wurde in den 70er Jahren entwickelt. Maßgeblich für dessen Entwicklung und Ausbau war die Sektion „Soziale Indikatoren“ der Deutschen Gesellschaft für Soziologie, in deren Rahmen Konzepte und Methoden des Studierendensurveys zur Diskussion gestellt wurden, sei es über die Erfassung von Bildungszielen, die gesellschaftlichen Vorstellungen Studierender oder den internationalen Vergleich (Bargel/Framhein 1976).

Die Kommission für wirtschaftlichen und sozialen Wandel (1977) hat später in ihrem Gutachten die Einbeziehung subjektiver Daten im Sinne von Erwartungen, Ansprüchen und Vorstellungen empfohlen (vgl. Peisert u.a. 1988). Diesen Empfehlungen folgend hat die Sozialberichterstattung in den 80er Jahren eine beachtliche Intensivierung erfahren, z.B. mit dem regelmäßigen Nationalen Sozialen Survey des Zentrums für Umfragen, Methoden und Analysen, ZUMA (vgl. Mayer/Schmidt 1984) sowie den Wohlfahrtsurveys (vgl. Glatzer/Zapf 1984) und dem als mehrjährige Wiederholungsbefragung angesetzten Sozio-ökonomischen Panel des SFB „Mikroanalytische Grundlagen der Gesellschaftspolitik“ (vgl. Hanefeld 1987).

Parallel dazu wurde in der Arbeitsgruppe Hochschulforschung an der Universität Konstanz das Konzept einer Dauerbeobachtung der Studiensituation entwickelt, das darauf ausgerichtet ist, in repräsentativem Umfang verschiedene Studierendengenerationen mit einem im Kern stabilen Befragungsinstrument sukzessive zu untersuchen.

Damit wurde der für den Bildungsbereich bereits in den 70er Jahren geforderten Notwendigkeit entsprochen, die statistische Berichterstattung durch subjektive Bildungsindikatoren zu ergänzen, die vom einzelnen und seinen Bedürfnissen ausgehen und die sozialen und psychischen Aspekte von Bildungserwartungen, Bildungsverhalten und Bildungsergebnissen in ihren Wandlungstendenzen aufhellen (vgl. Peisert u.a. 1988, BMBW 1976).

Vorbild für das Konstanzer Konzept der Dauerbeobachtung des Hochschulbereichs waren in formaler Hinsicht die Sozialerhebungen des Deutschen Studentenwerkes, die seit 1951 in regelmäßigen 3-Jahres-Abständen Informationen zur sozialen und wirtschaftlichen Lage der Studierendenschaft liefern (vgl. Kath 1952, Schnitzer 1986, Middendorff u.a. 2013), im Gegensatz zum Studierendensurvey aber viel stärker auf die Messung objektiver Indikatoren abzielen.

Entwicklung und Abgrenzung des Studierendensurveys

Die **Entwicklung** des Studierendensurveys Anfang der 80er Jahre konnte auf die Arbeiten aus dem Projekt „Hochschulsozialisation“ im Sonderforschungsbereich „Bildungsforschung“ (SFB 23) an der Universität Konstanz aufbauen, in dessen Rahmen Abiturienten, Studierende und Akademiker untersucht wurden (Peisert/Bargel/Framhein 1981, Framhein u.a. 1981). Eine wichtige Basis war ein *international vergleichendes Projekt zur Hochschulsozialisation* (1968 - 1982) - mit Forschergruppen aus England, Jugoslawien, Polen, der Tschechoslowakei und Schweden als Partner. Dabei konnten theoretische Perspektiven und geeignete Instrumentarien entwickelt werden (vgl. Bargel u.a. 1973, Framhein u.a. 1974).

Weitere Grundlagen waren Untersuchungen zur Situation und Sozialisation von Studierenden (Jenne u.a. 1969, Oehler u.a. 1976, Ipsen/Portele 1976), Studien zu Problemfeldern der Hochschulentwicklung und der studentischen Orientierungen (Habermas u.a. 1961, Apenburg u.a. 1977, Kaase 1971) sowie Ansätze zur Erstellung sozialer Indikatoren, die aus dem angloamerikanischen Raum übernommen wurden (vgl. Zapf 1976, Ballerstedt/Glatzer 1975).

Ebenfalls berücksichtigt wurden Untersuchungen zu Entwicklungsverläufen vom Studienzugang bis zum Übergang in den Beruf der Hochschul-Informationssystem GmbH (HIS 1980, Lewin/Schacher 1981, Reissert/Birk 1982), außerdem Arbeiten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) zu den Berufsaussichten und Berufswegen von Hochschulabsolventen (Saterdag/Stegmann 1980, Kaiser et.al. 1981, Tessaring/Werner 1981).

Des Weiteren wurden Untersuchungen über die allgemeinen gesellschaftlichen und politischen Haltungen der Studierenden herangezogen, etwa der Meinungsforschungsinstitute Infratest (1974, 1979, 1980, 1982) und IfD Allensbach (1967, 1968, 1978, 1982) sowie eine Reihe umfangreicher Einzeluntersuchungen zu studentischen Erfahrungen und Vorstellungen, Protesten und alternativen Haltungen (z.B. Kaase 1970, Oehler 1974, Does 1978, Kreutz 1979, Krause et al. 1980).

In Abgrenzung und in Ergänzung zu anderen Untersuchungen wie der Sozialerhebung oder Absolventenstudien wurde der Studierendensurvey gezielt auf studentische Motivation, Lernsituation, Studierverhalten sowie studienbezogene, gesellschaftliche und berufliche Orientierungen ausgerichtet. Er befasst sich somit mit der kulturellen und sozialen Dimension des studentischen Daseins.

Themenbereiche des Studierendensurveys

Die **Themen**, die im Fragebogen des Studierendensurveys behandelt werden, sind breit gefächert und reichen von Tätigkeiten vor dem Studium über die Lehrsituation bis hin zur Studienqualität und dem Studierertrag. Zusätzlich sind die Probleme und Forderungen der Studierenden von Interesse, ihre Berufsorientierungen und –perspektiven sowie ihre gesellschaftlichen Vorstellungen. Das Spektrum an Themen umfasst folgende Bereiche:

1. Hochschulzugang, Fachwahl, Motive und Erwartungen
2. Studienstrategien, Studienverlauf und Qualifizierungen
3. Studienintensität, Zeitaufwand und Studiendauer
4. Studienanforderungen, Regelungen und Prüfungen

5. Kontakte und Kommunikation, soziales Klima, Beratung
6. Fachstudium, Situation der Lehre und Studienqualität
7. Studentische Lebensform, soziale Situation, Erwerbstätigkeit
8. Schwierigkeiten, Beeinträchtigungen und Belastungen
9. Hochschulpolitik: Partizipation und studentische Vertretung
10. Internationalität, Auslandsstudium und Europäischer Hochschulraum
11. Berufswahl, Berufswerte und Tätigkeitsbereiche
12. Berufsaussichten und Arbeitsmarktreaktionen
13. Demokratische Einstellungen und gesellschaftliche Vorstellungen
14. Politische Beteiligung, Standort und Ziele
15. Wünsche und Forderungen zur Hochschulentwicklung
16. Demographische und bildungsbiographische Daten

Die Fragen und Items ebenso wie die Antwortformate zu den verschiedenen Themen bleiben überwiegend unverändert, um die Dauerbeobachtung methodisch abzusichern. Dabei wird regelmäßig überprüft, ob das Verständnis der Fragen bei den Studierenden noch gewährleistet ist, zum Beispiel besonders intensiv, als die Studierenden aus den ‚neuen Ländern‘ ab 1992 in das Sample einbezogen wurden.

Durch die Bereiche der Leistungsmessung werden Indikatoren über inhaltliche Themen strukturiert, die in Verbindung zu einem übergeordneten Konzept stehen. Der Aufbau des Surveys ist daher mehrfach strukturiert:

- *Modular*: Die einzelnen Fragestellungen sind thematisch größeren Einheiten zugeordnet.
- *Temporär*: Der Studienverlauf wird nachgezeichnet, von der Zeit vor Studienbeginn bis zu der Zeit nach dem Studienabschluss.
- *Evaluativ*: Aufeinanderfolgende Komponenten sind einbezogen, womit der Differenzierung nach Input-Throughput-Output, oder nach Struktur-Prozess-Ergebnis entsprochen werden kann.
- *Funktional*: Es wird nach unterschiedlichen Ebenen differenziert, sowohl innerhalb des Systems Hochschule als auch im Handlungssystem. Für die Hochschule werden die einzelnen Ebenen unterschieden: Studierende, Lehrende, Veranstaltungen, Fach, Fachbereich, Studiengang, Hochschule, Hochschulsystem. Dem Handlungssystem wird durch die Differenzierung nach Individuum, Peer, Berufswelt, Hochschule, Gesellschaft entsprochen.

Die verwendeten Items basieren sowohl auf theoretischen Konzepten unterschiedlicher Reichweite, als auch auf Modellannahmen und empirischen Befundlagen (vgl. Multrus 2016). Viele davon können in Skalen oder Dimensionen integriert werden, andere lassen sich mit Modellen vereinbaren, für die sie als Prüfgrößen dienen können. Dabei sind umfassende Theorien (System und Funktion, Milieu und Habitus) ebenso vertreten wie Theorien mittlerer Reichweite (z.B. Coping, Dropout, Anomie, Partizipation, Berufswerte, Urteilsfähigkeit), bis hin zu skalaren Theorien und Modellbildungen (z.B. Anonymität, demokratische Einstellungen, forschendes Lernen, Lernkompetenz, Studienqualitätsbilanz, politische Aggressivität, alternative Orientierungen, Diskriminierung von Studentinnen), die zu Dimensionen und Indikatoren führen (Effizienzorientierung, Praxis- und Forschungsbezug, Beteiligung an Gruppen und Gremien, Prüfungsangst, Arbeitsmarktchancen, politisches Interesse, Anhängerschaft politischer Richtungen, Gesellschaftsbild, Studierenertrag).

Je nach Entwicklung im Hochschulbereich werden neue Themen aufgenommen, etwa neue Medien und Internetnutzung oder die zweistufige Studienstruktur mit ihren andersartigen Studienanforderungen. Außerdem können einzelne Fragen und Items aus den verschiedenen Themenbereichen in einem rotierenden Verfahren beibehalten oder in einzelnen Erhebungen weggelassen werden.

Neben der Berichterstattung zu inhaltlichen Themen des Studierendensurveys werden regelmäßig methodische Arbeiten zum Instrument durchgeführt, um Skalen, Strukturen, Indizes u.a. zu prüfen, zu verbessern oder zu entwickeln. Diese Arbeiten werden in einer eigenen Reihe „Hefte zur Bildungs- und Hochschulforschung“ publiziert.

Methodenbericht zum 13. Studierendensurvey

Der vorliegende Bericht beschäftigt sich mit der methodischen Prüfung der in der 13. Erhebung neu eingeführten Online-Befragung. Dabei geht es vorrangig darum zu klären, ob der bislang als Papierbefragung durchgeführte Studierendensurvey auch als Online-Befragung zu verlässlichen Befunden führt, damit die existierenden Zeitreihen methodisch gesichert weitergeführt werden können.

Vor dem Hintergrund, dass ab der 14. Erhebung nur noch online erhoben werden wird, ist gleichzeitig zu prüfen, inwieweit auftretende Unterschiede bzw. Veränderungen zwischen der Papier- und der Online-Befragung behandelt werden müssen. Das gilt für die Darstellung der Instrumente und der Erhebungsmodalitäten ebenso wie für die Analyse der Befunde und deren Interpretationen.

Eine Online-Befragung bietet zusätzliche Daten und Möglichkeiten von Designvariationen an, die in der 13. Erhebung genutzt wurden, um einerseits eingeführte Maßnahmen auf ihre Effekte zu prüfen, und andererseits um die miterhobenen Paradata zu analysieren und deren Nützlichkeit zu prüfen.

Vor diesem Hintergrund unterteilt sich der vorliegende Bericht in unterschiedliche Kapitel, die jeweils verschiedene Schwerpunkte beinhalten.

- Als Einstieg werden kurz Hintergrund und Konzeption des Studierendensurveys beschrieben.
- Im ersten Kapitel werden das Erhebungsdesign und die Datenlage beschrieben und die dazu anstehenden Untersuchungsfragen besprochen.
- Im zweiten Kapitel werden Rahmenbedingungen diskutiert, auf die bei der Interpretation der Befunde zu achten ist.
- Das dritte Kapitel beschäftigt sich mit dem Rücklauf und den Abbruchquoten.
- Im vierten Kapitel werden die vorhandenen Paradata analysiert und besprochen.
- Das fünfte Kapitel prüft die Stichprobenstrukturen als Ausgangsbasis für die Vergleiche zwischen den Designgruppen.
- Im sechsten Kapitel werden die Stichproben mit der Grundgesamtheit verglichen.
- Das siebte Kapitel beinhaltet die Vergleiche inhaltlicher Merkmale zwischen den Designgruppen.
- Im achten Kapitel werden verschiedene Maßnahmen auf ihre Effizienz geprüft.
- Das neunte Kapitel behandelt spezifische Herausforderungen bei Onlineuntersuchungen.
- Abschließend wird im zehnten Kapitel ein Fazit gezogen und Folgerungen als Ausblick angestellt.

In diesem Bericht können nicht alle Aspekte des Erhebungsdesigns des 13. Studierendensurveys eingearbeitet werden. Die Analysen zum Kurzfragebogen - eine um 40% gekürzte Version des Studierendensurveys - werden hier nicht aufgenommen, da dazu auch eine umfangreiche inhaltliche Analyse notwendig ist. Diese methodischen Bearbeitungen werden zu einem späteren Zeitpunkt durchgeführt werden. Des Weiteren wird auch die Kontrollgruppe mit E-Mail-Einladungen hier nicht in die Vergleiche einbezogen, sondern an anderer Stelle bearbeitet werden.

Der vorliegende Bericht soll keine Abhandlung über die Wissenschaftlichkeit von Methoden der Sozialwissenschaften sein, sondern praktische Belange bei der Studierendenbefragung erörtern. Daher wurde auch auf sehr komplexe Analyseverfahren verzichtet und mehr die praktische Bedeutung in den Vordergrund gerückt. Als Werkstattbericht richtet er sich mehr an Praktiker als an wissenschaftlich interessierte Methodiker.

1. Erhebung zum 13. Studierendensurvey

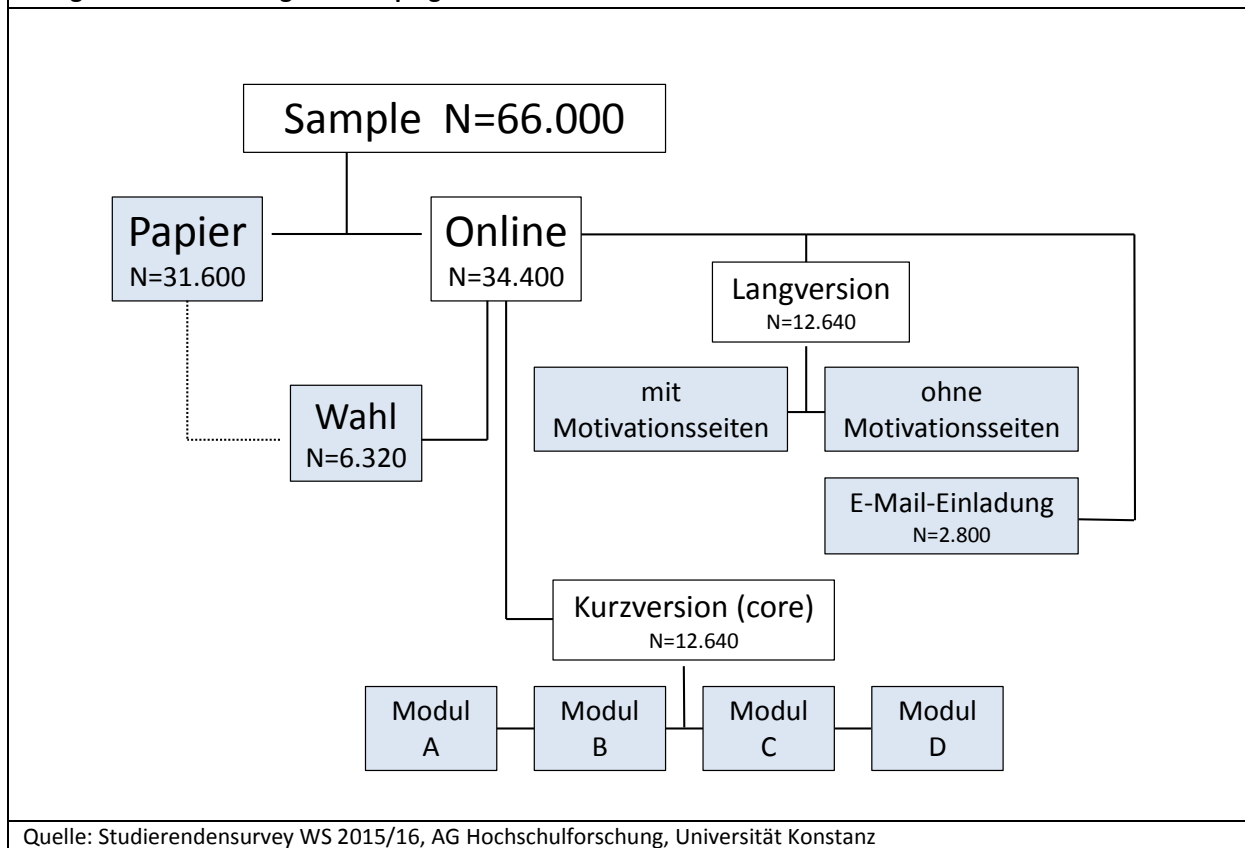
In den zwölf Erhebungen des Studierendensurveys von 1982 bis 2013 wurden nur Papierfragebogen an die Studierenden versendet. Zur 13. Erhebung im WS 2015/16 wurde das Design der Erhebung erheblich geändert, da neben der traditionellen Befragung über Papierfragebogen eine Online-Befragung neu hinzukam. Ziel dieser Kombination war es zu untersuchen und zu prüfen, inwieweit es möglich und methodisch haltbar ist, die Befragung zukünftig online durchzuführen. Gleichzeitig wurde die Situation genutzt, um weitere methodische und designtechnische Fragen zu Onlineuntersuchungen zu prüfen.

Design der 13. Erhebung

Um neben einem direkten Papier-Onlinevergleich auch weitere Designmöglichkeiten auszutesten, wurde ein komplexes Erhebungsdesign entwickelt, das insgesamt 10 Gruppen unterscheiden kann.

Abbildung 1.1

Design der 13. Erhebung und Samplegrößen



Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

- In der Gruppe mit Papierfragebogen wurden insgesamt 31.600 Studierende angeschrieben; sie erhielten den Fragebogen per Post (Gruppe 0).
- Studierende der Onlinegruppe erhielten einen Online-Zugang zum Fragebogen. Dazu wurden insgesamt 34.440 Studierende angeschrieben. Die Stichprobengrößen unterscheiden sich etwas zur Papiergruppe, weil eine Hochschule kurzfristig eine größere Stichprobe nur Online einbeziehen wollte.

- Die Onlinegruppe wurde in weitere Untergruppen aufgeteilt, um spezielle methodische Fragen analysieren zu können. Eine Untergruppe bildet die Wahlgruppe (N = 6.320), deren Studierende zusammen mit dem Online-Zugang auch einen schriftlichen Fragebogen per Post zugesendet bekamen. Diese Studierenden konnten also wählen, ob sie die Fragen auf Papier (Gruppe 1) oder online (Gruppe 2) beantworten wollen.
- Die übrigen Studierenden der Onlinegruppe erhielten nur einen Online-Zugang für die Befragung über das Internet. Der weitaus größte Teil dieser Studierenden (N=25.280) wurde per Post angeschrieben, um vergleichbare Bedingungen zur Papiergruppe zu erhalten. Diese Studierenden wurden zufällig in zwei gleichgroße Untergruppen aufgeteilt. Die eine Hälfte sollte den gesamten Onlinefragebogen (Langversion) beantworten, die andere Hälfte erhielt eine um ca. 40% gekürzte Version (Kurzversion). Für die Studierenden in der Langversion wurde zusätzlich eine weitere zufällige Aufteilung vorgenommen, wobei die eine Hälfte ein Version erhielt, in die drei Motivationsseiten eingespielt wurden (Gruppe 3), während die andere Hälfte diese Seiten nicht angezeigt bekam (Gruppe 4). Die Studierenden der Kurzversion wurden zufällig in vier weitere Untergruppen (Module) aufgeteilt, die sich darin unterscheiden, dass am Ende des Fragebogens jeweils unterschiedliche zusätzliche Fragen aufgenommen wurden (Gruppen 5-8). Die Aufteilungen der Lang- und Kurzversionen auf die jeweiligen Untergruppen wurden über eine Randomisierung beim Online-Zugang vorgenommen, sodass in etwa gleichgroße Untergruppen zustande kamen.
- Eine kleine Teilgruppe (N=2.800) der Online-Befragten wurde nur per E-Mail eingeladen, weshalb sie eine eigene Subgruppe im Design bildet (Gruppe 9).

Zusätzlich wurden alle Samplehochschulen angefragt, ob sie eine zweite Erinnerung per E-Mail nachlegen könnten, und ob sie einen kurzen Text auf ihre Website oder die sozialen Medien stellen würden, der die Untersuchung ankündigt.

Untersuchungsfragen zum Erhebungsdesign

Mit diesem Design lässt sich einerseits die Zeitreihe fortführen, da die Papiergruppe weiterhin die Datenbasis stellen kann; gleichzeitig lassen sich durch die verschiedenen Designgruppen andererseits auch wichtige methodische Fragen klären, womit insgesamt folgende Ansätze geprüft werden können:

- 1) Die Wahlgruppe gibt Auskunft, wie hoch der Anteil an Studierenden ist, die einen Online-Fragebogen einem Papierfragebogen vorziehen. Gleichzeitig kann das Profil der Studierenden mit unterschiedlichem Teilnahmemodus verglichen werden.
- 2) Der Online-Fragebogen bietet die Möglichkeit zu prüfen, ob die Erhebungsform Auswirkungen auf die Ergebnisse hat. Ob also Studierende online genauso antworten wie auf Papier. Darüber hinaus kann geprüft werden, ob es Unterschiede in der Zusammensetzung der Studierenden beider Erhebungsformen gibt oder ob sie vergleichbar sind. Diese Vergleiche sind von besonderer Bedeutung für die Frage der Aufrechterhaltung der Zeitreihen.
- 3) Die Vergleiche zwischen der Wahlgruppe und der Papier- bzw. der Onlinegruppe bieten die Möglichkeit zu prüfen, ob die Selbstselektion der Erhebungsart Unterschiede gegenüber der vorselektierten Erhebungsbedingung erbringt.
- 4) Die Variante mit Motivationsseiten erlaubt zu prüfen, ob solche Einschübe die Bearbeitungsmotivation verbessern können, und somit weniger Abbrüche zu beobachten sind.

- 5) Die gekürzte Online-Fragebogenversion gibt zum einen Auskunft, ob die Teilnahmebereitschaft bei einem kürzeren Instrument höher ist, und zum anderen ob häufiger der gesamte Fragebogen beantwortet wird. Darüber hinaus kann geprüft werden, ob der Umfang an Aussagen ausreicht, die Gesamtheit der Studiensituation und der studentischen Orientierungen abzubilden.
- 6) Die zusätzliche Gruppe 9 erlaubt zu prüfen, ob der Einlademodus (Online oder postalisch) einen Effekt aufweist. Dies ist wichtig, um zu prüfen ob der Medienbruch, der durch die postalische Einladung zu einer Online-Befragung akzeptiert wurde, nachteilige Effekte erzeugt.
- 7) Die unterschiedlichen Erinnerungsmodi erlauben zu prüfen, welchen Effekt eine 2. Online-Erinnerung auf die Beteiligung hat im Vergleich zur einmaligen postalischen Erinnerung. Und über die Gruppe 9 lassen sich online vs. postalisch als 1. Erinnerung gegenüberstellen.
- 8) Die verschiedentlich eingeführten Ankündigungen der Untersuchung auf den Websites oder sozialen Medien der Hochschulen erlauben eine Prüfung des Effekts solcher Maßnahmen auf die Beteiligung.

Weiteres Vorgehen: Analysen und Vergleiche

Die Hauptfrage betrifft die Vergleichbarkeit der Daten bei einem Umstieg von einer Papierbefragung zu einer Online-Befragung. Dahinter steht die Frage nach der Sicherung der Zeitreihen seit 1982.

Dazu muss zum einen geprüft werden, inwieweit Papier- und Onlinebefragte in ihrer Stichprobenstruktur vergleichbar sind und zum anderen, inwieweit die beiden Stichproben die gleichen Antwortmuster liefern. Sollten hierbei deutliche Unterschiede auftreten, ist zu prüfen, inwieweit die Befunde für den Zeitvergleich weiterhin verwertbar bleiben.

Um diese Frage zu beantworten, müssen Vergleiche zwischen den verschiedenen Designgruppen aufgestellt werden:

- Papier vs. Online (lang)
- Papier vs. Online in der Wahlgruppe
- Papier vs. Papier-Wahl
- Online vs. Online-Wahl

Mit diesen Vergleichen werden zum einen die Mode-Effekte und zum anderen die Selbst- vs. Fremdselektionseffekte untersucht.

Die zusätzlichen Designmerkmale bieten weitere Testungsmöglichkeiten:

- Online mit Motivationsseite vs. ohne Motivationsseite
- Online lang vs. online kurz
- E-mail-Einladung vs. postalische Einladung
- E-mail- Erinnerung vs. postalische Erinnerung
- 2. Erinnerung vs. nur eine Erinnerung
- Ankündigung im Internet (Website, Facebook) vs. keine Ankündigung

Für die methodisch-analytische Bearbeitung bedeutet dies jeweils Testungen zu:

- Strukturen der Stichprobenszusammensetzung
- Antwortverhalten in den Messvariablen
- Analyse der Nichtbeantwortungen

2. Datenbasis: Stichprobengröße im 13. Studierendensurvey

Als Untersuchungsgrundlage dienen die resultierenden Antworten der befragten Studierenden im 13. Studierendensurvey. An der Befragung haben sich insgesamt 8.549 Studierende beteiligt, was einem Rücklauf von insgesamt 13,2 % entspricht, wobei der Rücklauf in der Papiergruppe allerdings bei 16 % lag, während der Rücklauf bei der Onlinegruppe mit knapp 10 % deutlich schwächer ausfiel.

Das Erhebungsdesign erlaubt den Vergleich von 10 Untergruppen, die zu größeren Hauptgruppen zusammengefasst werden können. Für die anstehenden Vergleiche werden sowohl Haupt- als auch Untergruppen verwendet.

Tabelle 2.1				
Stichprobengröße 13. Studierendensurvey, nach Designgruppen und Hochschulart (WS 2015/16)				
(Angaben in absolut)				
	Design	Gesamt¹⁾	Uni	FH
Gruppe 0	Papier	4.959	3.885	1.074
Gruppe 1	Wahl Papier	661	508	153
Summe	Papier (Gruppe 0 + 1)	5.620	4.393	1.227
Gruppe 2	Wahl online	273	201	70
Gruppe 3	Online lang, A, mit Mot.seite	564	406	157
Gruppe 4	Online lang, B, ohne Mot.seite	596	441	153
Summe ¹⁾	Online Langversion (2-4)	1.433	1.048	380
Gruppe 5	Online, core, C, Zuwanderung	278	218	59
Gruppe 6	Online, core, D, Stile	283	217	66
Gruppe 7	Online, core, E, Bourdieu	301	227	74
Gruppe 8	Online, core, F, MKT	326	254	72
Summe	Online Kurzversion (5-8)	1.188	916	271
Gruppe 9	Online lang, B, E-mail, SB	308	308	-
Summe	Online gesamt (2-9)	2.929	2.272	651
Gesamt		8.549	6.665	1.878

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz
¹⁾ Durch fehlende Angaben zur Hochschulart stimmen Summen über Uni + FH nicht mit „Gesamt“ überein.

2.1 Rahmenbedingungen und Probleme

Vor den analytischen Bearbeitungen sind zuerst einige Rahmenbedingungen zu diskutieren, die sowohl für Studierendenbefragungen generell als auch für Online-Erhebungen im Speziellen zu berücksichtigen sind. Dies sind zum einen generelle Kritikpunkte an dieser Art von Befragungen, als auch spezifische Probleme, die durch einen Wechsel der Erhebungsmethode zum Tragen kommen können.

2.1.1 Repräsentativität

Eine Frage, die sowohl für die Papier- als auch die Online-Befragung von Bedeutung ist, ist die nach der Repräsentativität der Befunde bzw. der Stichprobe. Bilden die Antworten der Studierenden tatsächlich die Meinung aller Studierenden ab? Sind die ausgewählten Studierenden repräsentativ für alle Studierenden?

Um die Ansichten einer bestimmten Gruppe in der Gesellschaft zu erfahren, müsste diese in ihrer Ganzheit befragt und untersucht werden. Das wäre eine Totalerhebung aller ca. 2,7 Millionen Studierender. Damit könnten zwei wichtige Merkmale definiert werden: 1) die Zusammensetzung der Grundgesamtheit in wichtigen sozialen Indikatoren und 2) die Kennwerte ihrer Antworten.

Da Vollerhebungen aber nicht immer durchführbar sind, werden Stichproben genutzt, um zu den gleichen Aussagen zu gelangen. Die Forschungslogik besagt, dass eine zufällig ausgewählte Stichprobe aus der Grundgesamtheit Struktur und Kennwerte dieser Grundgesamtheit abbilden können sollte. Die Statistik liefert dazu die Werkzeuge, diese Annäherung zu prüfen und zahlenmäßig auszugeben.

Die wichtigste Aufgabe ist also sicher zu stellen, dass die ausgewählte Stichprobe repräsentativ für die Grundgesamtheit ist. Dazu ist die Zufallsauswahl eine Voraussetzung. Hier bestehen bei Studierendenbefragungen zwei mögliche Fehlerquellen.

- 1) Die Auswahl der Hochschulen
- 2) Die Auswahl der Studierenden an den Hochschulen

Da es keine gesamtdeutsche Studierendendatei gibt, auf die zugegriffen werden kann, führt der Weg zu den Studierenden über die Hochschulen. Diese müssen zustimmen, dass Studierende ihrer Hochschule befragt werden dürfen. Dann kann eine Zufallsauswahl an den zustimmenden Hochschulen getroffen werden.

Das erste methodische Problem ist also die Auswahl der Hochschulen, an denen Studierende befragt werden können. Zur Zeit gibt es 426 Hochschulen in Deutschland, an denen rund 2,7 Millionen Studierende eingeschrieben sind (Statistisches Bundesamt, 2016). Hier gilt es also eine Auswahl an Hochschulen zu treffen, die die Studierendenlandschaft in Deutschland zutreffend abbilden kann.

Das zweite Problem ist dann die Auswahl der Studierenden an diesen Hochschulen. Sie muss so gestaltet sein, dass die zu befragenden Studierenden alle wichtigen Merkmale der Studierenden in Deutschland abbilden können.

Zur Lösung dieser Probleme gibt es verschiedene Wege:

- 1) Versuch einer Vollerhebung: Es können alle Hochschulen angeschrieben werden und dann alle Studierenden der teilnehmenden Hochschulen eingeladen werden.
Das große Problem bei diesem Vorgehen wäre die Selbstselektion bzw. Selbsteliminierung von Hochschulen und Studierenden. Die letztlich teilnehmenden Studierenden wären an vorab selbstselektierten Hochschulen in unterschiedlichem Ausmaß verteilt. Inwieweit die Befunde repräsentativ wären, hinge davon ab, welche Hochschulen und welche Studierenden an diesen Hochschulen in der Stichprobe enthalten wären. Im ungünstigen Falle wären z.B. kleine Spezialhochschulen überrepräsentiert und traditionelle große Hochschulen unterrepräsentiert.
- 2) Theoretische Vorauswahl der Hochschulen: Die Selbstselektion der Hochschulen kann nicht vollständig ausgeschlossen werden. Aber es kann gewährleistet werden, dass die Hochschulen, an denen die Stichproben gezogen werden, bestmöglich die Hochschullandschaft abbilden, indem eine Vorauswahl solcher Hochschulen getroffen wird, die wichtige Kennwerte der Hochschullandschaft abbilden können. Solche Indikatoren können regionale Lage in Deutschland, Hochschulart, Größe der Hochschule, Fächerangebot etc. sein. Der Vorteil gegenüber einer zufälligen Vorauswahl an Hochschulen liegt in einer repräsentativeren Basis der zu befragenden Studierenden und damit eine größere Homogenität des Samples.

- 3) Bedingte Zufallsauswahl der Studierenden: An den teilnehmenden Hochschulen sollen die Studierenden dann zufällig ausgewählt werden. Um eine repräsentative Stichprobe zu erhalten, scheinen aber gewisse weitere Vorgaben nötig zu sein. Eine vorab definierte Anzahl an zufällig einzuladender Studierender hat den Vorteil, die Hochschulgröße einzubeziehen, so dass erstens vergleichbare und zweitens zur Hochschule passende Anteile der befragten Studierenden an den verschiedenen Hochschulen zustande kommen. Dies steigert die Homogenität und die Repräsentativität der Stichprobe.

Mit diesem Vorgehen kann zwar keine Repräsentativität garantiert werden, dennoch ist es als bestmöglicher Weg zur Herstellung von repräsentativen Stichproben zu bezeichnen. Dieser Weg wird bei den Erhebungen zum Studierendensurvey seit Beginn eingehalten.

Dennoch kann dieses Vorgehen die Selbstselektion bzw. -elimination nicht verhindern, die die Repräsentativität der Stichprobe bedrohen kann. Jene Hochschulen und Studierenden, die die Befragungen verweigern, könnten bestimmte Untergruppen der Grundgesamtheit ausschließen.

Dieser Kritikpunkt kann ohne Kenntnisse der Grundgesamtheit nicht wiederlegt werden. Aber es lässt sich prüfen, inwieweit Struktur und Kennwerte der Stichprobe validen Vergleichsdaten entsprechen. Die wichtigste Frage ist hierbei, was sind valide Vergleichsdaten?

a) Soziodemografische Struktur der Grundgesamtheit

Das Statistische Bundesamt (2016) liefert Daten zu den Studierenden in Deutschland, die auch einige soziodemografische Verteilungen enthalten. Dazu zählen Geschlecht, Hochschulart, Abschlussart, Fächer, Semester und Migrationshintergrund. Diese Daten beziehen sich auf die Gesamtheit der Studierenden in Deutschland und damit auf die Grundgesamtheit. Die Strukturen der Stichproben können nun anhand dieser Verteilungen verglichen werden.

b) Struktur früherer Untersuchungen

Ein Kritikpunkt gegenüber der Repräsentativität der Stichprobe bezieht sich auf den zunehmend schwächer werdenden Rücklauf bei Studierendenbefragungen (vgl. Ramm 2014). Die Annahme ist, dass je mehr Studierende sich solchen Befragungen verweigern, desto größer wird der Coverage-Fehler, die Annahme, dass bestimmte Teilgruppen in der Stichprobe nicht vertreten sind.

Ein Vergleich mit früheren Erhebungen des Studierendensurveys aus Zeiten, in denen ein Rücklauf von über 40 % erreicht werden konnte, bietet die Möglichkeit, weitere soziale Indikatoren einzubeziehen und in der Struktur zu vergleichen. Dabei muss allerdings eine zeitliche Entwicklung bzw. Veränderung in Betracht gezogen werden.

c) Strukturen anderer Untersuchungen

Ein Vergleich der Stichprobenstruktur mit Stichprobenstrukturen aus anderen, ähnlichen Untersuchungen bietet die Möglichkeit zu überprüfen, ob Abweichungen von der Grundgesamtheit typisch für bestimmte Erhebungs- oder Designarten sind.

d) Kennwerte der Stichproben

Für die Kennwerte der Stichprobe gilt zwar prinzipiell das gleiche Vorgehen wie bei den Strukturen, doch stehen hierbei meist keine Vergleichswerte aus der Grundgesamtheit zur Verfügung. Insofern bleiben nur Vergleiche zu anderen oder zu früheren Untersuchungen.

Stehen allerdings keine validen Vergleichsdaten zur Verfügung, bleiben nur die Daten der Erhebung selbst. Daran lassen sich theoretische Aussagenprüfungen oder Untergruppenvergleiche mit Rangfolgenbeziehungen durchführen.

2.1.2 Probleme der Repräsentativität - Stichprobenfehler

Die Repräsentativität der Befunde kann geprüft aber nicht gesichert werden. Unabhängig davon, wie gut eine Stichprobe die Grundgesamtheit in wichtigen Merkmalen abbildet, kann nicht bewiesen werden, dass die Stichprobe ein wirkliches Abbild der Grundgesamtheit darstellt, i.e. tatsächlich repräsentativ ist. Dies gelingt nur, wenn die Grundgesamtheitswerte bekannt sind. Ansonsten stellen Stichprobenbefunde immer nur eine Annäherung an die Grundgesamtheit dar, die höchstens statistisch (in Wahrscheinlichkeitsmaßen) beziffert werden können.

Werden die Stichproben nach wissenschaftlich-methodischen Ansätzen gezogen und bilden sie die Grundgesamtheit in wichtigen Strukturelementen sowie in prüfbaren Kennwerten ab, dann kann eine weitgehend repräsentative Abbildung angenommen werden.

Allerdings gibt es mehr als eine mögliche Fehlerquelle bei der Erhebung der Daten, die eine weitest gehende Repräsentativität beeinträchtigen können.

Auf der Ebene der Stichprobenziehung sind drei mögliche Fehlerquellen bekannt (Groves et al. 2004):

- Der Coverage-Fehler,
- der Sampling-Fehler und
- der Non-Response-Fehler

Der Coverage-Fehler besagt, dass nicht alle zur Grundgesamtheit gehörigen Einheiten in die Auswahl mit einbezogen werden. Der Fehler liegt also in den Kriterien, mit denen der Auswahlrahmen für die Grundgesamtheit definiert wird. Als möglicher Effekt werden bestimmte Gruppen nicht einbezogen.

Der Sampling-Fehler besagt, dass die Stichprobe nicht die Grundgesamtheit abbildet. Der Fehler liegt hier in der Ziehung der Stichprobe, bzw. in den Kriterien, nach denen die Ziehung vollzogen wird. Als möglicher Effekt werden auch hier bestimmte Gruppen nicht erreicht oder können die Kennwerte der Grundgesamtheit nicht repräsentativ abgebildet werden.

Der Non-Response-Fehler bezieht sich auf Ebene der Stichprobenziehung auf jene Einheiten, die nicht erreicht werden können, bzw. die nicht antworten. Auch dadurch kann möglicherweise ein bestimmter Teil der Grundgesamtheit nicht einbezogen werden.

Alle drei Fehler beziehen sich also darauf, dass bestimmte Untergruppen der Grundgesamtheit vielleicht nicht in der Befragung vertreten sind, womit die Befunde nicht wirklich repräsentativ wären.

Eine andere Klasse von Fehlern bezieht sich auf das Erhebungsinstrument und die Art der Fragen. Der Measurement-Fehler besagt, dass die Art der Fragen systematische Fehler (bias) verursachen können. Dabei sind Kontext- von Content-Effekten zu unterscheiden. Inhaltliche Effekte sind zum Beispiel die soziale Erwünschtheit, oder Fragen bzw. Antwortvorgaben, die bestimmte Antworten suggerieren. Kontext-Effekte sind Effekte, die auf das Design und die Art und Weise der Darstellung zurückgehen. Dazu gehört auch die Art der Erhebung (z.B.: Papier vs. Online).

Eine dritte Klasse von Fehlerquellen betrifft die Datenerfassung. Einerseits fallen darunter alle Methoden, die fehlende Daten oder Ausreißer behandeln, also auch die Datenbereinigung. Zum anderen ist hier die Weiterverarbeitung der Daten in den Analysen eine mögliche Fehlerquelle, wenn z.B. Gewichtungen einbezogen werden, oder auch auf eine Gewichtung verzichtet wird.

Alle angesprochenen Fehlerquellen können die Befunde verzerren, sodass ein Abbild der Grundgesamtheit zunehmend fraglich wird. Es gilt daher, diese Fehlerquellen zu minimieren, bzw. gezielt zu prüfen.

Den möglichen Fehlerquellen wird im Studierendensurvey in folgender Weise entsprochen:

Grundgesamtheit: Studierende an Universität und Fachhochschulen

Von den etwa 2,7 Millionen Studierenden, die 2016 an 426 Hochschulen in Deutschland studieren, werden nur jene als konzeptionelle Grundgesamtheit betrachtet, die an Universitäten und an Fachhochschulen eingeschrieben sind. Damit werden von den 426 Hochschulen insgesamt nur 107 Universitäten und 215 Fachhochschulen einbezogen. Die Studierendenanzahl verringert sich somit auf 2,331 Millionen deutsche Studierende, zu denen noch 88.763 Bildungsinländer kommen, die ebenfalls in die Befragung einbezogen wurden. Bildungsausländer und Studierende an pädagogischen oder theologischen Hochschulen, Kunsthochschulen und Verwaltungsfachhochschulen sowie private Hochschulen wurden nicht einbezogen. Die konzeptionelle Grundgesamtheit, für die der Studierenden-survey Aussagen machen will, ist somit eine Teilgruppe aller Studierenden in Deutschland, die trotz der vorgenommenen Einschränkungen 85 % aller Studierenden umfasst.

Von den 322 Universitäten und Fachhochschulen in Deutschland werden je Erhebungswelle zwischen 25 und 30 Hochschulen nach theoretischen Vorgaben ausgewählt. Daher hat nicht jede Hochschule die gleiche Chance, in das Sample zu gelangen, sondern es werden nur solche einbezogen, die gewährleisten sollen, dass die Studierenden an diesen Hochschulen ein möglichst passendes Abbild der Studierendenschaft an Universitäten und Fachhochschulen in ganz Deutschland darstellen. Deshalb wird die Fächerstruktur, die Hochschulart, die Größe der Hochschule sowie die regionale Verteilung der Hochschulen in Deutschland in die Auswahl einbezogen.

Der Coverage-Fehler ist bei diesem Vorgehen, dass nur bestimmte Hochschulen ausgewählt werden und somit alle anderen Studierenden keine Chance haben, einbezogen zu werden. Der Kritikpunkt wäre also, dass die ausgewählten Hochschulen nicht die Studierendenschaft abbilden und die Befunde somit nur für dieses Hochschulsample Gültigkeit besitzen, was im Prinzip richtig ist. Allerdings soll die theoretische Vorauswahl der Hochschulen ja gerade ein besseres Abbild zulassen als eine zufällige Auswahl. Ein Beleg dafür wäre die hohe Homogenität der Studierenden. Der andere Grund für das theoretische Hochschulsample ist die Kontinuität der Befunde an den immer gleichen Hochschulen, sodass die Zeitreihe der Befunde bessere Gültigkeit erhält.

Studierende: Zufallsauswahl nach vordefinierter Stichprobengröße

Die Auswahl der Studierenden an den Hochschulen erfolgt nach dem Zufallsprinzip. Vorab wird allerdings die Größe der zu ziehenden Stichprobe definiert. Diese hängt zum einen von der Größe der Hochschule ab und zum anderen von dem zu erwartenden Rücklauf. Damit die Studierendenschaft passend abgebildet werden kann, müssen wichtige Strukturmerkmale auch erkennbar werden. D.h. die Stichproben müssen jeweils mindestens so groß sein, dass die Semester- und Fächerstruktur abgebildet werden kann. Dazu müssen Mindestgrößen erhoben werden, damit auch bei schwächer werdenden Rückläufen Vergleiche innerhalb dieser Merkmale möglich sind. Da die Hochschulen aber trotz theoretischer Vorauswahl unterschiedliche Studierendenzahlen aufweisen, gleichzeitig aber untereinander vergleichbar bleiben sollen, gilt es Stichproben zu ziehen, die beides berücksichtigen. Daher variieren die Stichproben zwischen jedem 6. und jedem 11. Studierenden an der entsprechenden Hochschule.

Der Sampling-Fehler tritt hier auf, da nicht alle Studierenden die gleichen Chancen haben, in die Stichprobe zu fallen. Denn die Zufallsauswahl geht normalerweise über die Auszählung der Studierendenkartei, die dann z.B. nach Namen geordnet wird. Inwieweit solch eine Auswahl tatsächlich zufällig ist, kann diskutiert werden.

Bei der Ziehung der Studierenden aus der Studierendenkartei kann es außerdem zu Coverage-Fehlern kommen; vor allem, wenn die Karteien nicht aktuell sind (vgl. Gabler & Häder, 2015, Häder, 2015). Ist ein Individuum mehrmals in der Kartei enthalten (bspw. zwei Adressen hinterlegt) liegt

Overcoverage vor: Elemente der Stichprobe sind nicht in der Grundgesamtheit vorhanden. Wenn Elemente in der Stichprobe fehlen, obwohl sie Teil der Grundgesamtheit sind (bspw. beurlaubte Studierende in separater Datei), liegt Undercoverage vor.

Die Studierendenkarteien der Hochschulen können als hinreichend vollständige Listen bezeichnet werden. Zumindest theoretisch hat damit jede/r Studierende die gleiche Möglichkeit, ausgewählt zu werden.

Bei diesem Vorgehen besteht eine weitere Fehlerquelle, die von Projektseite nicht behoben werden kann, nämlich die inkorrekte Zufallsauswahl durch die Hochschulen. Nur durch die Rückmeldung der Hochschulen, wie viele Einladungen tatsächlich versendet wurden, kann rückgeprüft werden, ob den Vorgaben entsprochen wurde, jedoch nicht, wie die Auswahl tatsächlich erfolgte.

Non-Response: Nicht erreichbare Studierende vs. Verweigerer

Um Studierende zur Befragung einladen zu können, müssen deren gültige Adressen vorliegen, sodass die Studierenden die Einladung auch erhalten. Sind die Adressdateien nicht auf dem neusten Stand, können die Befragungsunterlagen nicht zugestellt werden und Studierende müssen von der Befragung ausgeschlossen werden. Dies trat in der 13. Erhebungen des Studierenden surveys im WS 2015/16 in etwa zwei Prozent der Fälle auf.

Die Verweigerung der Befragung ist ein anderes Problem. Dies betrifft mittlerweile über 80 % der angeschriebenen Studierenden. Welche Untergruppen darunter vertreten sind und somit nicht in die Analysen eingehen, ist nicht zu beantworten. Die Kritik, dass damit kein realistisches Abbild der Studierenden mehr geschaffen werden kann, ist somit berechtigt. Allerdings gilt diese Kritik auch bei einem Rücklauf von 80 %. Und es gibt keine gültige Aussage, ab wieviel Prozent Befragte die Repräsentativität nicht mehr gewährleistet werden kann. Es gibt allerdings Aussagen zur Mindestgröße der Stichprobe, aufgrund derer Aussagen über die Grundgesamtheit gemacht werden können (vgl. Ramm 2014; Mossig, 2012). Diese Größenordnung wird in den Erhebungen des Studierenden surveys bislang immer überschritten. Dennoch ist der schwache Rücklauf bzw. der hohe Anteil an Verweigerern ein zunehmendes Problem der Studierendenbefragung.

Messfehler

Measurement-Fehler wurden seit Beginn des Projektes beachtet. Die Fragen und Items des Studierenden surveys sind methodisch geprüft und beinhalten theoretische Bezüge. Sie basieren zum großen Teil auf Arbeiten aus Forschungsprojekten zur Hochschulsozialisation, zur Hochschulentwicklung, zu studentischen Orientierungen, zu Entwicklungsverläufen vom Studienzugang bis zum Berufsübergang, zu Berufsaussichten und Berufswegen, zu gesellschaftlichen und politischen Haltungen und zur Erstellung sozialer Indikatoren.

Während der mittlerweile mehr als 30 Jahre, die der Studierenden survey im Feld ist, wurden viele methodische Arbeiten zu den Items, den zugehörigen Skalen und den zugrundeliegenden Dimensionen erstellt, die das Ziel hatten, die Zuverlässigkeit und Gültigkeit der Fragen und Items zu prüfen und zu verbessern. Eine Reihe von Publikationen in der Heftreihe „Hefte zur Bildungs- und Hochschulforschung“ geben hierrüber Auskunft.

Des Weiteren werden in den Veröffentlichungen zu den Befunden des Studierenden surveys stets andere Untersuchungen herangezogen, um Ergebnisse zu vergleichen oder zu replizieren, und so die Validität der Befunde zu prüfen.

Dennoch bleibt es Aufgabe in jeder neuen Erhebung, das Instrument stets auf solche möglichen Fehlerquellen zu untersuchen und weiter zu entwickeln. Im Zuge der Umstellung der Studierendenbefragung auf eine Online-Erhebung werden hier zusätzliche Arbeiten vollzogen werden müssen, die der Erhebungsmethode geschuldet sind.

Für die spezifischen Fehlerquellen, die auf die Art der Erhebungsmethode zurückgehen (Papier vs. Online), dient der Vergleich der beiden Gruppen, die in der 13. Erhebung des Studierenden surveys eingeführt wurden. Damit lässt sich untersuchen, ob die Items in einer Onlinedarstellung genauso beantwortet werden wie auf Papier.

Datenaufbereitung und -verarbeitung

Die Fehlerquellen, die auf die Datenaufbereitung zurückgehen können, wurden bislang ebenfalls sehr genau beachtet. Der Papierfragebogen in der 13. Erhebung wurde von einer Erfassungsfirma in die EDV übertragen. Die Güte der Übertragung wurde von der Arbeitsgruppe jeweils sehr genau kontrolliert. Die Fehlerquote lag bislang jeweils unter einem Prozent.

Bei der Online-Erhebung wurde ein Erfassungssystem des DZHWs verwendet, das die angewählten Skalenpunkte bzw. eingegebenen Zahlen automatisch in eine Datendatei überträgt. Auch hier wurden die Daten von der Arbeitsgruppe nach Erstellung der Datendatei auf Unstimmigkeiten überprüft und Fehler behoben.

Die erstellten Rohdateien (Papier und Online) werden dann an die Arbeitsgruppe weitergeleitet, die diese Daten in ein analysierbares Format überträgt. An dieser Stelle folgt die Abklärung der Behandlung von Ausreißern und Missing Values.

Ausreißer-Daten werden nur dann eliminiert, wenn sie eindeutig außerhalb der Messskala liegen, oder in einzelnen Fällen keinen Sinn ergeben (z.B. Anzahl Fachsemester von mehr als 60). Missing Values werden unterschiedlich codiert und behandelt, je nachdem ob es sich um einzelne Items handelt oder ob ganze Fragen bzw. Bereiche nicht beantwortet wurden. Für die Definition, ab wann eine Person aufgrund zu vieler fehlender Werte nicht mehr in die Analyse eingeht, wird in jeder Erhebung die konkrete Datenlage geprüft und dann ein Kriterium festgelegt. Dieses Kriterium kann sich bei Papier- und Online-Erhebungen unterscheiden.

Die denkbaren Fehlerquellen werden also so weit wie möglich beachtet. Dennoch bleiben sie bestehen und müssen in jeder neuen Erhebung überprüft werden.

Probleme des Methodenwechsels

Ein anderer Diskussionspunkt ergibt sich aus dem anstehenden Methodenwechsel der Erhebungsart, der Wechsel von Papierbefragung zur Online-Erhebung. Denn dieser Wechsel bringt einige neue Probleme mit sich.

Zum einen ist hier die Weiterführung der Zeitreihe zu nennen. Dauerbeobachtungen wie der Studierenden survey, der in seinen bislang 13 Erhebungswellen einen Zeitraum von 34 Jahren (von 1982 bis 2016) abdeckt, haben auch die Aufgabe, die Entwicklungen über die Zeit hinweg zu dokumentieren und zu analysieren. Dazu muss gewährleistet sein, dass die Daten soweit vergleichbar sind, dass auftretende Veränderungen auf die Zeit zurückgeführt werden können und nicht in Veränderungen des Instrumentes oder der Zusammensetzung des Samples begründet sind. Dazu wurde bislang versucht, das Hochschulsample konstant zu halten, sodass Studierende immer an den gleichen Hochschulen befragt wurden und das Instrument in großen Teilen identisch blieb. Äußere Einflüsse, die auf eine Veränderung der Hochschullandschaft zurückgehen, können diese Veränderungen zwar ebenfalls bedingen, lassen sich aber gezielt prüfen (z.B. Hinzunahme der neuen Länder in der Erhebung 1992 oder Einführung der neuen Abschlüsse durch die Bologna-Reform Ende des letzten Jahrhunderts).

Ein Wechsel der Erhebungsmethode kann diese Kontinuität gefährden, wenn dazu das Instrument angepasst werden muss. Dazu zählen die Veränderung von Frage- bzw. Item-Formulierungen, von Antwortvorgaben, von Skalenbreiten, von Anzahl bzw. Reihenfolge der Items oder vom Design.

Hier besteht also u.U. das Dilemma, entweder die Vergleichbarkeit zu früheren Erhebungen zu erhalten oder die Gestaltungsmöglichkeiten bzw. -notwendigkeiten der Online-Befragung zu nutzen. Da es aber wenig Sinn macht, eine neue Erhebungsart zu verwenden, deren Möglichkeiten dann aber nicht bestmöglich zu nutzen, wird die Entscheidung zu Gunsten der Online-Erhebung fallen, was bedeutet, dass die Zeitreihenvergleiche problematisch werden können. Es muss daher akzeptiert werden, dass manche Entwicklungen nicht mehr eindeutig nachzuzeichnen sind.

Der Wechsel in der Erhebungsform kann außerdem die Wahrscheinlichkeit für einen Coverage Bias erhöhen, wenn sich die beiden Stichproben (Papier und Online) hinsichtlich ihrer Zusammensetzung unterscheiden. Hinweise auf solche Unterschiede konnten in verschiedenen Untersuchungen nachgewiesen werden (vgl. Bandilla et. al, 2009; Bosnjak et. al, 2013; Mohorko, de Leeuw, Hox, 2013). Demnach finden sich bspw. Unterschiede im Alter, dem Bildungsstand und der Geschlechterverteilung zwischen Internetnutzern und Nicht-Nutzern. Wobei diese ausgeprägten Differenzen vor allem dadurch entstehen, dass Personen der Gruppen der Nicht-Nutzer keinen Zugang zum Internet haben. Unter Studierenden bzw. Personen in der gleichen Altersgruppe liegt die Internetnutzung laut einer Studie des statistischen Bundesamtes von 2016 bei 100%. Die heutigen Studierenden können damit zurecht als Generation der „digital natives“ bezeichnet werden.

Die Kontrolle der Sample-Zusammensetzung ist daher vor allem für die Wahlgruppe von Bedeutung, in der die Studierenden die Erhebungsform selbst bestimmen können.

Hinweise darauf, wie groß die Unterschiede bei einem Wechsel der Erhebungsmethode sind, können anhand der Vergleichsanalysen zwischen der Papier- und der Online-Erhebung in der 13. Welle des Studierendensurveys überprüft werden. Allerdings wurde dort darauf geachtet, dass keine inhaltlichen Unterschiede bestehen. Zur weiteren Prüfung solcher Unterschiede können dann Vergleichsuntersuchungen mit anderen Studien (z.B. dem Studienqualitätsmonitor) durchgeführt werden.

Ein Unterschied, der durch den Methodenwechsel herbeigeführt wird, ist die Veränderung des Teilnahmesamples, es nehmen weniger Frauen bzw. mehr Männer an einer Onlineuntersuchung teil im Vergleich zu einer Papierbefragung. Dies konnte sowohl im 13. Studierendensurvey als auch in den letzten drei Erhebungen der Sozialerhebung nachgewiesen werden. Diese veränderte Verteilung innerhalb der Samplestruktur kann zu veränderten Befunden führen. Allerdings lässt sich dies dadurch beheben, dass jene Untergruppen jeweils getrennt analysiert und miteinander verglichen werden.

3. Non-Response

Zentral für jede Art von Befragung ist der Rücklauf. Verlässliche Aussagen über akzeptable Rücklaufquoten für schriftliche und internetbasierte Befragungen sind allerdings selten und die Angaben sind stark von der Art, dem Umfang sowie dem Design der jeweiligen Untersuchung abhängig. Ohne weitere unterstützende Maßnahmen sind Rücklaufquoten von 20 % in schriftlichen Befragungen keine Seltenheit (Diekmann, 2006); jedoch lässt sich bspw. unter Zuhilfenahme besonderer Designelemente (Total Design Method/ Tailored Design Method) die Ausschöpfungsquote auf zum Teil 60 % bis 70 % deutlich steigern (Dillman, 1983, 2000).

Etwas moderater fällt die Vorgabe für Rücklaufquoten in schriftlichen Befragungen in einer Meta-Analyse von Shih und Fan (2008) aus: Hier liegt die durchschnittliche Rücklaufquote schriftlich-postalischer Befragungen bei etwa 40 %, für Online-Befragungen rund 10 % niedriger. Lozar Manfreda et. al kommen in ihrer Meta-Analyse von 45 Studien, die Unterschiede in den Rücklaufquoten von Online-Surveys und anderen Erhebungsmethoden untersucht, zu ähnlichen Ergebnissen, wonach die Rücklaufquote in Online-Befragungen durchschnittlich 11 % niedriger liegt als in anderen Befragungsmethoden (Lozar Manfreda et. al, 2008).

Für Online-Befragungen variieren die Angaben zu gängigen Rücklaufquoten noch stärker als für schriftlich-postalische Befragungen¹. In Studierendenbefragungen wurden teilweise Rücklaufquoten von 21,5 % (Sax, Gilmartin & Bryant, 2003), aber auch 41 % bis 58 % (Pötschke, 2009) erreicht. Die Sozialerhebung, die für den Studierenden-survey die beste Vergleichsstudie bildet, berichtet in ihrer Online-Befragung einen Netto-Rücklauf von 19,4 % (Poskowsky, Weiß, Kandulla, 2014).

Die Analyse der einschlägigen Literatur zeigt, dass es einerseits sehr unterschiedliche Befunde zu Rücklaufquoten gibt, und diese im Zeitverlauf abnehmen. Zudem bestehen deutliche Unterschiede nach den Erhebungsformen. Die Tendenz zur abnehmenden Teilnahmebereitschaft gilt für alle Erhebungsformen und ist ein zunehmendes globales Phänomen von Befragungen (vgl. Ramm 2014).

Einen anderen Ansatz liefert die Formel von Mossig, mit der sich der minimale Stichprobenumfang für eine endliche Grundgesamtheit berechnen lässt (Mossig, 2012, S.21). Mit der Formel lässt sich die Größe einer Stichprobe bestimmen, die mindestens nötig wäre, um über die deutsche Studierendenschaft gültige Aussagen treffen zu können:
$$n = \frac{N}{1 + \frac{(N-1) \times \epsilon^2}{z^2 \times P \times Q}}$$
,

wobei n = minimal erforderlicher Stichprobenumfang für eine endliche Grundgesamtheit,

N = Anzahl der Elemente der Grundgesamtheit,

ϵ = gewählter tolerierter Fehler,

z = aus der zentralen Wahrscheinlichkeit der Standardnormalverteilung berechneter Wert der gewählten Sicherheitswahrscheinlichkeit (S = 99 %; D(z) = 0,99 \rightarrow z = 2,575),

P = prozentualer Anteilswert an der Grundgesamtheit und

Q = 1-P ist (der P-Wert wird dabei so geschätzt, dass das Produkt P x Q den größtmöglichen Wert annimmt, um eine ausreichend große Stichprobengröße zu erhalten: P = 0,5 \rightarrow Q = 0,5).

¹ Die Gründe hierfür liegen in der vielfältigen Ausgestaltung webbasierter Befragungen und vor allem im Rekrutierungsprozess. Werden Besucher einer Internetseite über ein Werbebanner zu einer Befragung eingeladen, sind sehr niedrige Rücklaufquoten von 1-5 % vorstellbar. Führt hingegen ein Unternehmen auf dem firmeneigenen Server eine Mitarbeiterbefragung durch, ist eine Rücklaufquote von über 90 % möglich.

$$\text{Für den Studierendensurvey ergibt sich: } n \geq \frac{2.331.100}{1 + \frac{2.331.099 \times 0,05^2}{2,575^2 \times 0,5 \times 0,5}} = 662,874 \approx n \geq 663$$

Die Stichprobengröße muss also mindestens $n = 663$ ausmachen, damit die Ergebnisse der Grundgesamtheit von 2.331.100 Studierenden mit einer Sicherheitswahrscheinlichkeit von 99 % um nicht mehr als fünf Prozentpunkte vom tatsächlichen Ergebnis abweichen (vgl. Mossig, 2012, 8 ff. und insbesondere S.20 f.). 663 Fälle wären also statistisch gesehen ausreichend, um Aussagen über die deutschen Studierenden an Universitäten und Fachhochschulen machen zu können. Dabei handelt es sich allerdings nur um einen Zahlenwert. Wichtig ist vor allem die Übereinstimmung der Stichprobe mit wichtigen Merkmalen der Grundgesamtheit (vgl. Abschnitt 3.1.5 und Kapitel 6).

Im Folgenden wird der Rücklauf für die drei Untersuchungsgruppen (Gruppe 1 = Papierfragebogen, Gruppe 2 = Online-Fragebogen selbstselektiert, Gruppe 3 + 4 = Online-Fragebogen mit und ohne Motivationsseiten randomisiert) dargestellt und anschließend der Non-Response näher untersucht. Es wird dabei zwischen Unit Non-Response, dem Fehlen sämtlicher Daten für bestimmte Befragungseinheiten, und Item Non-Response, dem Fehlen einzelner Angaben einer Einheit, unterschieden.

Vorab sei darauf hingewiesen, dass die Brutto-Rücklaufquote in der Papierbefragung sehr viel weniger durch unvollständige Datensätze verzerrt ist als die Online-Befragung. Zwar gibt es sicherlich auch in der papierbasierten Untersuchung Personen, die den Fragebogen beginnen und nur unvollständig ausfüllen; in der Regel werden solche Fragebögen aber nicht abgeschickt, weshalb die Angaben in der Papierbefragung bemerkenswert vollständig sind. In der Online-Befragung hingegen ist sehr viel häufiger mit solchen unvollständigen oder lückenhaften Datensätzen zu rechnen. Diese unvollständigen Daten sind jedoch für das Auffinden problematischer Fragen und Items bzw. bei der Gestaltung und Verbesserung des Instruments (Länge) besonders nützlich.

3.1 Brutto-Rücklauf

Insgesamt wurden von den 28 beteiligten Hochschulen 50.120 Befragungsunterlagen für die Papiergruppe, die Langversion der Onlinegruppe sowie die Wahlgruppe verschickt² (vgl. Tabelle 3.1). Für die Berechnung der Ausschöpfungsquote werden die stichprobenneutralen Ausfälle abgezogen: 1.474 Sendungen konnten nicht zugestellt werden und 33 Fälle stellten sich als nicht zutreffend heraus (Student/ in befand sich zum Zeitpunkt der Befragung im Ausland/ Urlaubssemester, war nicht mehr immatrikuliert, etc.). Auffällig sind die hohen Ausfälle aufgrund nicht zugestellter Unterlagen. Ein Grund hierfür könnte die vergleichsweise hohe Mobilität der Studierenden sein: Ein Teil zieht während des Studiums mehrmals um, bspw. vom Wohnheim in eine WG oder in eine eigene Wohnung. Oftmals werden die Hochschulen nicht zeitnah über die neue Adresse informiert, weswegen die Anschriften in den Studierendenkarteien nicht immer aktuell sind.

Nach Abzug der stichprobenneutralen Ausfälle erhielten 30.789 Studierende den schriftlichen Fragebogen und 11.702 den Code zur Online-Befragung. 6.122 Studierende konnten zwischen dem schriftlichen Fragebogen und der Teilnahme an der Online-Befragung wählen. Die Feldphase für die Online-Befragung dauerte vom 1. Dezember 2015 bis zum 31. März 2016. In der Papierbefragung wurden aufgrund der Erfahrungen aus früheren Erhebungen noch bis Anfang Mai einzelne Fragebogen registriert. Zum Ende der Online-Befragung lag allerdings bereits 99 % des Rücklaufs aus der Papierbefragung vor. Von den 7.109 Rückläufen entfallen 4.959 auf den schriftlichen Fragebogen, 1.208 auf den

² In dieser Aufstellung sind die Gruppen, die eine Kurzversion des Fragebogens erhielten sowie die Gruppe mit den E-Mail-Einladungen nicht berücksichtigt: Diese beiden Gruppen erfordern eine gesonderte Betrachtung, die an dieser Stelle nicht durchgeführt werden kann. Im vorliegenden Bericht liegt der Fokus auf dem Methodenvergleich (Papier vs. Online) sowie dem Selbstselektionseffekt (Wahlgruppe), weshalb nur die Gruppen berücksichtigt werden, die auf dieselbe Weise (Post) rekrutiert wurden und die den langen Fragebogen erhalten haben.

Onlinefragebogen und 942 auf die Wahlgruppe. Die Brutto-Rücklaufquote über alle Gruppen hinweg beträgt damit nur 14,6 %, wobei sich die Quoten der einzelnen Befragungsgruppen deutlich voneinander unterscheiden. In der schriftlichen Befragung beläuft sich der Rücklauf auf 16,1 %, in der Online-Befragung nur noch auf 10,3 %. Gemessen an den zuvor genannten Richtwerten und Vergleichsstudien bleibt der Rücklauf des 13. Studierendensurveys deutlich hinter den Erwartungen zurück.

	Gesamt	Papier randomisiert	Online randomisiert	Wahl		Ge- samt
				Papier	On- line	
an Hochschulen verschickte Unterlagen	50.650	31.600	12.640		6.320	
von Hochschulen verschickte Unterlagen	50.120	31.549	12.264		6.307	
nicht zustellbar/ nicht zutref- fend ¹⁾	1.507	730/ 30	562		182/ 3	
bereinigte Stichprobe	48.613	30.789	11.702		6.122	
Brutto Rücklauf	7.109	4.959	1.208	661	281	942
Brutto Rücklaufquote	14,6 %	16,1 %	10,3 %	10,8 %	4,6 %	15,4 %
nicht verwertbar	56	-	48	-	8	8
Netto Rücklauf	7.053	4.959	1160	661	273	934
Netto Rücklaufquote	14,5 %	16,1 %	9,9 %	10,8 %	4,5 %	15,3 %

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz
¹⁾ unzustellbare Sendungen und Sendungen an Personen, die nicht Teil der Grundgesamtheit sind (Bildungsausländer/innen, beurlaubte Studierende, Exmatrikulierte, Studierende im Auslandsstudium)

Studierende, die zwischen Papier- und Onlinefragebogen wählen konnten, entschieden sich deutlich häufiger für den schriftlichen Fragebogen. Die Brutto-Rücklaufquote beträgt in dieser Gruppe 15,4 %, wobei 10,8 % sich für einen Papierfragebogen entschieden und 4,6 % für einen Onlinefragebogen. Damit entfallen 70,2 % des Rücklaufs in der Wahlgruppe auf Studierende, die den schriftlichen Fragebogen gewählt haben. Nur ein Drittel entschied sich für die Online-Befragung. Diese Dominanz der Papierbefragung ist sehr wahrscheinlich auch dem Design der Erhebung geschuldet, weil der schriftliche Fragebogen in der Wahlgruppe bereits der Einladung beilag. Um an der Online-Befragung teilnehmen zu können, mussten die Studierenden eine dem Einladungsschreiben beiliegende URL sowie ein Passwort an einem PC eingeben. Dieser Bruch in den Kommunikationskanälen erschien vielen Teilnehmern offenbar zu aufwendig, weshalb sie sich für den schriftlichen Fragebogen entschieden.

Die Ergebnisse der Rückläufe stellen heraus, dass die Möglichkeit der Wahl des Befragungsmodus insgesamt nicht zu einer höheren Teilnahmebereitschaft führt. Die beiden Rücklaufquoten in der Wahlgruppe liegen zwar jeweils niedriger als die der beiden randomisierten Gruppen, dürfen aber nicht direkt verglichen werden, da hier unterschiedliche Basisraten zugrunde liegen. Die Wahlgruppe umfasst beide Modi, daher müssten für einen Vergleich die beiden randomisierten Gruppen zusammengefasst werden. In diesem Falle würde ein Rücklauf von 14,5 % für die randomisierten Gruppen resultieren, der etwas niedriger liegt, als der Gesamtrücklauf mit 15,4 % in der Wahlgruppe.

3.1.1 Unit Non-Response

In sozialwissenschaftlichen Befragungen ist es meist nicht möglich, alle Personen zu einer Teilnahme zu motivieren. In groß angelegten Bevölkerungsumfragen kann meist ein Viertel bis die Hälfte der kontaktierten Personen nicht befragt werden (Schnell, Hill, Esser, 2011). Dieses als Non-Response bezeichnete Phänomen ist eine mögliche Fehlerquelle in Befragungen. Non-Response beschreibt eine Verzerrung der Ergebnisse, die dadurch zustande kommt, dass nicht alle angeschriebenen Personen an der Befragung teilnehmen (Bartsch, 2012). In der Literatur wird weiter danach unterschieden, ob für einen Befragten überhaupt keine Angaben vorliegen (Unit Non-Response), oder ob nur ein Teil der Angaben für einen Befragten vorliegen (Item Non-Response).

Im Falle der schriftlichen Befragungen liefert die Brutto-Rücklaufquote bereits ein relativ gutes Abbild der generellen Teilnahmebereitschaft innerhalb dieser Befragungsgruppe. Personen, die den Fragebogen abschicken, haben diesen in der Regel ernsthaft und vollständig bearbeitet. Das zeigt sich auch an der geringen Anzahl nicht verwertbarer Rückläufe. In der aktuellen Erhebung konnten nach Abzug der stichprobenneutralen Ausfälle alle Papier-Fragebogen für die Analyse verwendet werden.

In der Onlinegruppe gestaltet sich die Situation anders: Hier loggen sich zwar viele Befragte in die Online-Befragung ein, brechen dann aber direkt auf der Startseite oder im Verlauf der Befragung ab. Sinnvolle Angaben über die Teilnahmebereitschaft sind nicht möglich, und auch der Umfang unvollständiger Angaben ist in der Onlinegruppe sehr viel höher als in der papierbasierten Befragung. Hier muss eine Entscheidung getroffen werden, ab wann die Angaben der Befragten nicht mehr verwertbar sind.

Zur Bestimmung der verwertbaren Angaben wird meist der Item Non-Response herangezogen. Das Standardverfahren ist, alle Teilnehmer mit mehr als 50 % fehlender Angaben als nicht nutzbare Einheiten zu definieren (AAPOR, 2011). Diese werden dann von der Analyse ausgeschlossen. Alle anderen Befragten mit mehr oder weniger vollständigen Angaben gelten als „Partials“ und verbleiben im Datensatz. Denkbar ist auch eine restriktivere Einschränkung auf bspw. maximal 20 % oder 30 % fehlende Angaben. Alternativ schlagen Callegaro und Kollegen vor, eine Reihe an Variablen zu bestimmen, die für die Untersuchung von zentraler Bedeutung sind. Untersuchungseinheiten gelten nur dann als nutzbar, wenn für alle diese Variablen vollständige Angaben vorliegen (Callegaro, Lozar-Manfreda, Vehovar, 2015).

In Anlehnung an diese Vorgehensweise wurde in der aktuellen Erhebung eine eher weite Definition des Unit Non-Response gewählt: Alle Personen, die Angaben zu den ersten dreizehn Fragen gemacht hatten, wurden in die weiteren Analysen miteinbezogen. Der Grund hierfür liegt darin, dass bereits zu Beginn der Befragung zentrale Informationen zur Situation vor dem Studium und dem Übergang zwischen Schule und Hochschule erhoben werden. Der Nachteil dieser eher großzügigen Definition ist ein sehr hoher Item Non-Response, vor allem bei Fragen im fortgeschrittenen Fragebogenverlauf.

3.1.2 Netto-Rücklauf

Nach Abzug der nicht verwertbaren Angaben fällt der Rücklauf in der aktuellen Erhebung sehr gering aus. Während in den ersten sechs Erhebungen der klassischen Papierbefragung die Netto-Rücklaufquote noch über 40 % betrug, ging die Teilnahme seither zurück. Zur Jahrtausendwende belief sich die Rücklaufquote noch auf 34,8 %, im WS 2006/07 auf 32 % und in der vorigen Erhebung im WS 2012/13 nur noch auf 18,6 %. In der aktuellen Erhebung verzeichnet der Rücklauf erneut eine schwache Einbuße auf 16,1 %.

Wie bereits angesprochen ist der Anteil nicht verwertbarer Angaben in der Online-Befragung sehr viel höher; die Netto-Rücklaufquote beläuft sich auf 9,9 %. Sie liegt damit 6,2 Prozentpunkte niedriger als in der schriftlichen Befragung.

Wie in der randomisierten Befragungsgruppe konnten auch in der selbstselektierten Gruppe alle Rückläufe der Papierfragbogen verwertet werden, weshalb die Netto-Rücklaufquote ebenfalls 10,8 % beträgt. In der selbstselektierten Onlinegruppe lässt sich analog zur randomisierten Gruppe ein Rückgang der Netto-Rücklaufquote auf 4,5 % feststellen. Die kombinierte Netto-Rücklaufquote der Wahlgruppe liegt bei 15,3 % und damit 0,8 Prozentpunkte hinter der randomisierten Papierbefragung.

Auch wenn die Ausschöpfung insgesamt unbefriedigend ausfällt, zeigt sich besonders das häufig diskutierte Problem des geringeren Rücklaufs von webbasierten Befragungen.

3.1.3 Gründe für geringen Rücklauf

Seit den 1950er Jahren nimmt die Teilnahmeverweigerung in Bevölkerungsumfragen zu (Neller, 2005); ein Trend, der sich auch international beobachten lässt (de Leeuw & de Heer, 2002). Es verwundert daher nicht, dass auch andere Studierendenbefragungen und groß angelegte sozialwissenschaftliche Befragungen einen Rückgang der Teilnahmebereitschaft verzeichnen (Sozialerhebung, SQM, ALLBUS). Als einer der Hauptgründe für das generell nachlassende Teilnahmeinteresse an sozialwissenschaftlichen Befragungen wird häufig die „Überfischung“ der Befragten angeführt: Die Befragungsdichte hat so stark zugenommen, dass sich Befragte überfordert und demotiviert fühlen.

Für die allgemeine Situation in der Bevölkerung spielt sicherlich auch die Professionalisierung der Markt- und Meinungsforschung eine Rolle. Wo früher hauptsächlich Wissenschaftler mit ihren Instrumenten an die Befragten herantraten, hat sich parallel die Markt- und Meinungsforschung als eigener Wirtschaftszweig etabliert. Auf den Kauf eines Produkts bzw. nach der Inanspruchnahme einer Dienstleistung folgt heutzutage sofort eine Umfrage. Unternehmen befragen ihre Kunden, um deren Bedürfnisse besser zu kennen und sich einen Wettbewerbsvorteil zu sichern, und Meinungsforscher erfragen Trends und Einstellungen zu verschiedensten aktuellen Themen. Hinzu kommt eine stetig steigende Anzahl an Marktforschungsportalen, bei denen die Teilnahme an Befragungen sogar entlohnt wird. Die Konkurrenz durch andere, vor allem nichtwissenschaftliche Untersuchungen, hat also extrem zugenommen. Potentielle Befragte werden mit Einladungen zu Befragungen förmlich „überhäuft“.

An Hochschulen verschärft sich die Situation noch durch die gesetzlich verankerte Pflicht zur Evaluation der Lehre bzw. der Fachbereiche sowie durch weitere Erhebungen im Rahmen des hochschuleigenen Qualitätsmanagements. Zusätzliche interne (Qualifikations- und Forschungsarbeiten) sowie externe Untersuchungen belasten die Verwaltung der Hochschulen und die Studierenden zusätzlich.

Parallel stellen Menold und Züll in ihrer Auswertung der Kontaktprotokolle der ALLBUS-Erhebung 2008 auch eine allgemeine Verweigerungstendenz in der Bevölkerung fest. Diese allgemeine Verweigerung (generelle Verweigerung, kein Interesse, keine Zeit, Teilnahmeverbot) macht 80 % der Verweigerungsgründe aus (Menold & Züll, 2011). Weitere Gründe für eine Teilnahmeverweigerung sind negative Einstellungen gegenüber Umfragen (Umfragen nutzlos, zu viele Umfragen, schlechte Erfahrungen) sowie eher technische Bedenken (Datenschutz, Methodik), die in gleicher Weise auch auf Studierende zutreffen können.

Während die allgemeinen Gründe nur bedingt kontrolliert werden können (z.B. kein Versand der Umfrage in der Prüfungszeit, Themenauswahl), müssen vor allem die negativen Einstellungen und Bedenken der Studierenden thematisiert werden. Insbesondere der Nutzen einer Teilnahme an der Befragung muss deutlicher herausgestellt werden, bilden die Ergebnisse des Studierendensurveys doch

häufig die Grundlage für Diskussionen und Änderungen in der Hochschulpolitik, im hochschuleigenen Qualitätsmanagement sowie in den Medien und der interessierten Öffentlichkeit. Weitere Maßnahmen zur Erhöhung der Teilnahmebereitschaft der Studierenden, aber auch der Hochschulen, werden in Kapitel 9 „Herausforderungen bei Online-Befragungen“ präsentiert.

3.1.4 Besondere Probleme der Online-Erhebung

Für den geringen Rücklauf der Online-Befragung können darüber hinaus weitere Gründe angeführt werden. Die Einladungen zur Befragung wurden unabhängig vom Erhebungsmodus (Papier vs. Online) per Post versandt. Dieses Vorgehen wurde gewählt, um den Aufwand der beteiligten Hochschulen möglichst gering zu halten. Teilnehmer der Online-Befragung erhielten mit der Einladung einen URL-Code zum Online-Fragebogen sowie einen persönlichen Zugangscode. Der Wechsel von der postalischen Einladung zur Online-Befragung stellt allerdings einen Medienbruch dar: Bei der Eingabe des URL-Codes bzw. des individuellen Passworts können Eingabefehler passieren und der Wechsel zwischen Anschreiben und PC ist allgemein auch aufwendiger als die direkte Bearbeitung des beiliegenden Fragebogens in der schriftlichen Befragungsgruppe. Durch den Versand der Einladung per E-Mail kann der Medienbruch verhindert werden. Zudem besteht zwischen Studierenden und ihren Hochschulen bereits eine Beziehung, weshalb ein Versand der Einladung über die hochschuleigene E-Mail-Adresse eine höhere Legitimität aufweist als eine postalische Einladung einer fremden Einrichtung.

Im Gegensatz zur Papiergruppe wurde den Teilnehmern der Online-Befragung außerdem auf der Startseite des Onlinefragebogens die voraussichtliche Bearbeitungsdauer des Fragebogens angezeigt. Sie wurde auf eine Stunde geschätzt. Wie die Auswertung der Abbrüche zeigt, wirkt diese lange Bearbeitungsdauer zumindest auf einige Befragte abschreckend: Von 1.247 Personen, die auf die Online-Befragung zugegriffen haben, brachen 39 noch auf der Startseite ab (3,2 %). In der Wahlgruppe stiegen sieben von 288 Personen noch auf der Startseite aus (2,5 %).

3.1.5 Probleme geringer Rücklaufquoten und Umgang mit Non-Response

Der hohe Anteil an nicht erreichten bzw. unwilligen Teilnehmern stellt ein Problem für die Güte der erhobenen Daten dar (Schnell, Hill & Esser, 2011; Koch & Blohm, 2015). Zahlreiche Ausfälle erhöhen den Stichprobenfehler, weil die realisierte Stichprobe der aktuellen Untersuchung u.U. deutlich von der ursprünglich geplanten Stichprobe abweicht. Zudem besteht die Gefahr, dass die Aussagen auf Basis der gewonnenen Daten verzerrt sind, weil die Ausfälle nicht zufällig sind, sondern ein spezifischer Selektionsmechanismus vorliegt (Non-Response Bias): Teilnehmer und Nicht-Teilnehmer könnten sich in ihrem Antwortverhalten substantiell voneinander unterscheiden.

Die Rücklaufquote allein ist jedoch noch kein hinreichendes Kriterium für die Datengüte und ein hoher Anteil an Non-Response kann, aber muss nicht zu Verzerrungen führen. Koch und Blohm (Koch & Blohm, 2015) verweisen auf eine Zusatzbefragung des ALLBUS 2008, in der nur geringe Abweichungen zu den Ergebnissen des regulären ALLBUS festgestellt wurden, obwohl eine deutlich höhere Rücklaufquote erzielt werden konnte (Blohm & Koch, 2009, 2015). Solange die Teilnahmeverweigerung (Unit Non-Response) nicht systematisch ist und die Fallzahlen nicht zu klein werden, ist nicht mit bedeutenden Verzerrungen zu rechnen. Unterscheiden sich jedoch Teilnehmer von Verweigerern, dann erhöht sich das Risiko für verzerrte Ergebnisse durch eine niedrige Rücklaufquote.

Wichtiger ist es daher, den Umfang des Non-Response Bias zu bestimmen und die Ergebnisse der Befragung anhand anderer vergleichbarer Umfragedaten zu quantifizieren. Im Falle von Studierendenbefragungen wie dem Studierendensurvey ist die Situation relativ günstig, da mit den offiziellen Da-

ten der Hochschulstatistik des Statistischen Bundesamtes eine sehr gute Datengrundlage für den Vergleich besteht. Die Gegenüberstellung beschränkt sich auf die Verteilung zentraler soziodemographischer Merkmale in beiden Untersuchungen. Die Voraussetzung für diesen Vergleich ist, dass die Merkmale in vergleichbarer Weise erhoben wurden. In Kapitel 6 wird die Stichprobe des Studierendensurveys eingehend mit den Daten des Statistischen Bundesamtes abgeglichen sowie Abweichungen identifiziert und diskutiert.

3.2 Item Non-Response

Neben Befragungsausfällen aufgrund von Verweigerung oder Nichterreichbarkeit (Unit Non-Response) sind auch partielle Ausfälle bei den Befragten möglich: Einzelne Items werden nicht beantwortet. Häufig ist das bei sensiblen Fragen (Einkommen, sozial erwünschtes oder widerrechtliches Verhalten) oder Wissensfragen der Fall. Davon zu unterscheiden sind Antwortausfälle aufgrund des Designs des Befragungsinstruments (Frage/Item überlesen, Antwort unkenntlich) oder der Länge („Durchklicken“ bis zum Ende). Die Analyse von Item Non-Response (IN) kann wichtige Hinweise auf schwierige Items oder ein problematisches Design liefern. Zudem kann die Kombination von Unit Non-Response und Item Non-Response zu einer weiteren starken Absenkung des Rücklaufs bei einzelnen Fragen führen. Bei einer Netto-Rücklaufquote von zehn Prozent müssten alle Personen ein Item beantworten, damit die Rücklaufquote auf Itemebene ebenfalls gerade einmal zehn Prozent beträgt (Diekmann, 2006). Antworten nur 70 % der Teilnehmer, beträgt die Rücklaufquote auf Itemebene nur noch sieben Prozent, bei 50 % nur noch fünf Prozent usw.

Im Folgenden werden nur jene Befragten berücksichtigt, die verwertbare Angaben gemacht haben. Weil der Unit Non-Response sehr weit gefasst wurde (vgl. Abschnitt 3.1.1), ist mit einer Verzerrung der Ergebnisse durch fehlende Angaben und Befragungsabbrüche zu rechnen. Auch sind mit der vorliegenden Definition des Unit Non-Response noch zahlreiche Lurker im Datensatz vorhanden. Diese sind dadurch gekennzeichnet, dass sie den Fragebogen bis zum Ende durchklicken, ohne (sinnvolle) Angaben zu machen. Zum Vergleich werden die Berechnungen noch mit einer alternativen Definition des Unit Non-Response durchgeführt: Alle Befragten, die weniger als 50 % der erwarteten Angaben gemacht haben, gelten als nicht verwertbar (Definition 2).

Um den durchschnittlichen Item Non-Response pro Befragten (unit-level item non-response, UIN) zu berechnen, wird für jede/n Befragte/n die Anzahl der nichtbeantworteten Items ermittelt und dieser Wert durch die Anzahl der zu beantworteten Items geteilt. Der Fragebogen enthielt 413 Items, bei 372 Items wurden von den Teilnehmern Antworten erwartet (vgl. Tabelle 3.2).

	Papier rand.		Wahl Papier		Online rand.		Wahl Online	
	absolut	in %	absolut	in %	absolut	in %	absolut	in %
Item Non-Response	5,9	1,6	6,1	1,6	73	19,6	55,2	14,8
Standardabweichung	14	3,8	14,8	4	117,4	31,6	105,8	28,4
Median	2	0,5	2	0,5	5	1,3	4	1,1

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

Die Angaben der Studierenden in der randomisierten bzw. selbstselektierten schriftlichen Befragung sind auffallend vollständig. Hier fehlen durchschnittlich nur sechs Angaben. Im Gegensatz dazu sind die Ergebnisse für die online Befragten wie erwartet stark verzerrt. Die Verteilung ist stark rechtschief/ linkssteil und bimodal mit einem zweiten niedrigeren Gipfel am rechten Ende der Verteilung,

was auf Befragungsabbrüche im Befragungsverlauf zurückzuführen ist, die den Item Non-Response im fortgeschrittenen Fragebogen erhöhen. 50 % der Befragten haben lediglich vier (Wahlgruppe Online) bzw. fünf (Online randomisiert) fehlende Angaben, bearbeiten den Fragebogen also sehr ernsthaft. Die oberen 50 % streuen hingegen sehr stark, was auf Abbrüche im Befragungsverlauf sowie Ermüdungseffekte („Durchklicken“ bis zum Ende) zurückzuführen ist.

3.2.1 Unterschiede zwischen randomisierter und selbstselektierter Gruppe

Die Unterschiede in den fehlenden Werten in den Papiergruppen sind nur gering und erreichen keine statistische Signifikanz. In den Onlinegruppen ist die Differenz zwischen randomisierten und selbstselektierten Befragten hingegen sehr deutlich ausgeprägt: Der t-test ergibt einen signifikanten Effekt bezüglich der Gruppeneinteilung, $t(1431) = 2,30, p < .05$, wobei die Teilnehmer der Wahlgruppe deutlich weniger fehlende Angaben aufweisen als Teilnehmer der randomisierten Befragungsgruppe.

Die randomisierte Onlinegruppe ließe sich noch weiter unterteilen in jene Personen, die während der Befragung Motivationsseiten eingeblendet bekamen (Gruppe 3) und jene Befragten ohne solche unterstützenden Maßnahmen (Gruppe 4). Diese sollten die Befragten dazu ermutigen, die Befragung vollständig zu bearbeiten und so einen vorzeitigen Abbruch verhindern. Ist die Maßnahme erfolgreich, sollten die Angaben der Gruppe mit Motivationsseiten weniger stark durch Befragungsabbrüche verzerrt sein und damit auch der Item Non-Response niedriger sein als in der Gruppe ohne solche Maßnahmen. Der t-test zeigt allerdings keinen Hinweis auf signifikante Unterschiede zwischen den beiden Gruppen: $t(11589) = -0,084, p = .933$.

Um für Ermüdungseffekte und zumindest teilweise für Lurker zu kontrollieren, wird im Folgenden die alternative Definition des Unit Non-Response verwendet: Es werden nur Befragte berücksichtigt, die mindestens 50 % der vorgesehenen Angaben gemacht haben.

3.2.2 Item Non-Response mit restriktivem Unit Non-Response

Das Bild wird einheitlicher, wenn die zweite Definition des Unit Non-Response verwendet wird (vgl. Tabelle 3.3). An den Ergebnissen der schriftlich Befragten ändert die Einschränkung auf Personen, die mehr als 50 % der Angaben gemacht haben, wenig. Nach wie vor fehlen pro Befragtem durchschnittlich sechs Angaben. Bei den online Befragten führt die Einschränkung hingegen zu einer deutlichen Reduzierung des durchschnittlichen Item Non-Response. In der randomisierten Befragungsgruppe fehlen nun noch durchschnittlich 18 Angaben, in der Wahlgruppe 16.

Tabelle 3.3
Durchschnittlicher Item Non-Response pro Person nach Befragungsgruppe – Nur Befragte mit weniger als 50 % fehlenden Angaben
 (Angaben in absolut und in Prozent)

	Papier rand.		Wahl Papier		Online rand.		Wahl Online	
	absolut	in %	absolut	in %	absolut	in %	absolut	in %
Item Non-Response	5,8	1,6	5,9	1,6	18,6	4,9	15,9	4,3
Standardabweichung	13,2	3,6	13	3,5	42,3	11,4	35,6	9,6
Median	2	0,5	2	0,5	3	0,8	3	0,8

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

Dennoch streuen die Daten stark: 25 % der Befragten haben neun und mehr fehlende Angaben. Das zeigt, dass auch nach der Einschränkung die Ergebnisse der Onlinegruppe unvollständiger sind als die

der Papiergruppe. Die größere Streuung durch Befragungsabbrüche bzw. Ermüdungseffekte im Befragungsverlauf lassen sich nur in den Onlinegruppen nachzeichnen.

Mit der Verwendung der restriktiveren Definition des Unit Non-Response ist der Unterschied zwischen randomisierter und selbstselektierter Onlinegruppe nicht länger signifikant ($t(1163) = 0.90$, $p = .366$). Dieser Effekt ist aber hauptsächlich auf Personen der randomisierten Onlinegruppe zurückzuführen, die gleichzeitig Motivationsseiten eingeblendet bekamen: Wird die Anzahl fehlender Werte der Personen mit und ohne Motivationsseite verglichen, so weist die erste Gruppe deutlich weniger Item Non-Response auf (13 zu 24 fehlenden Angaben), $t(927) = 4,132$, $p < .000$. Die Ergebnisse deuten darauf hin, dass die Angaben der Teilnehmer mit Motivationsseiten weniger durch Item Non-Response verzerrt sind. Sie schließen deshalb aber nicht häufiger die Befragung ab: Von den Personen in der Motivationsbedingung beenden 72 % erfolgreich den Fragebogen, in der Gruppe ohne Motivationsseiten sind es 71 % ($p = 0,772$). Das bedeutet, dass die Motivationsseiten in der Summe nicht zu einer Verringerung der Befragungsabbrüche führen, aber die Vollständigkeit der Angaben nachhaltig positiv beeinflussen.

3.2.3 Item Non-Response in Abhängigkeit von der Bearbeitungsdauer

Darüber hinaus ist die Antwortqualität auch abhängig von der Dauer der Befragung. Die Informationen über die Bearbeitungsdauer liegen nur für die Online-Befragung vor. Hier besteht zwischen der Anzahl fehlender Angaben und der Bearbeitungsdauer ein mittlerer, signifikanter, negativer Zusammenhang ($r_s = -0.5332$, $p = 0,000$)³. Das bedeutet, dass mit steigender Bearbeitungsdauer der Anteil nicht beantworteter Fragen sinkt. Personen, die zwischen 20 und 30 Minuten im Web-Survey verbracht haben, weisen einen höheren Item Non-Response auf als Teilnehmer, die länger für die Beantwortung benötigten. Ihnen fehlen durchschnittlich 19 Angaben. Zwischen 50 und 60 Minuten Bearbeitungszeit ist der Item Non-Response am geringsten (sechs fehlende Angaben), danach steigt er wieder an.

3.2.4 Item Non-Response im Verlauf der Befragung

Die Analyse des Item Non-Response im Verlauf der Befragung zeigt, ob es bei den Teilnehmern zu Ermüdungseffekten in Form von unmotivierten Antworten („Durchklicken“) kommt. Von Interesse sind hier insbesondere die Unterschiede zwischen den Befragten der Papier- und der Onlinegruppe: Die Bearbeitung des Fragebogens nimmt vergleichsweise viel Zeit in Anspruch und liegt deutlich über den gängigen Empfehlungen zur Bearbeitungsdauer für Onlineumfragen. Zudem ist in Online-Befragungen mit einer kürzeren Aufmerksamkeitsspanne, Ablenkungen oder Multitasking zu rechnen. Daher ist überwiegend in der Gruppe der online Befragten mit besagten Ermüdungseffekten zu rechnen. Um diese Punkte näher zu untersuchen, wird im Folgenden der Item Non-Response für jene Items betrachtet, bei denen von allen Befragten eine Antwort erwartet wurde (vgl. Abb. 3.1).

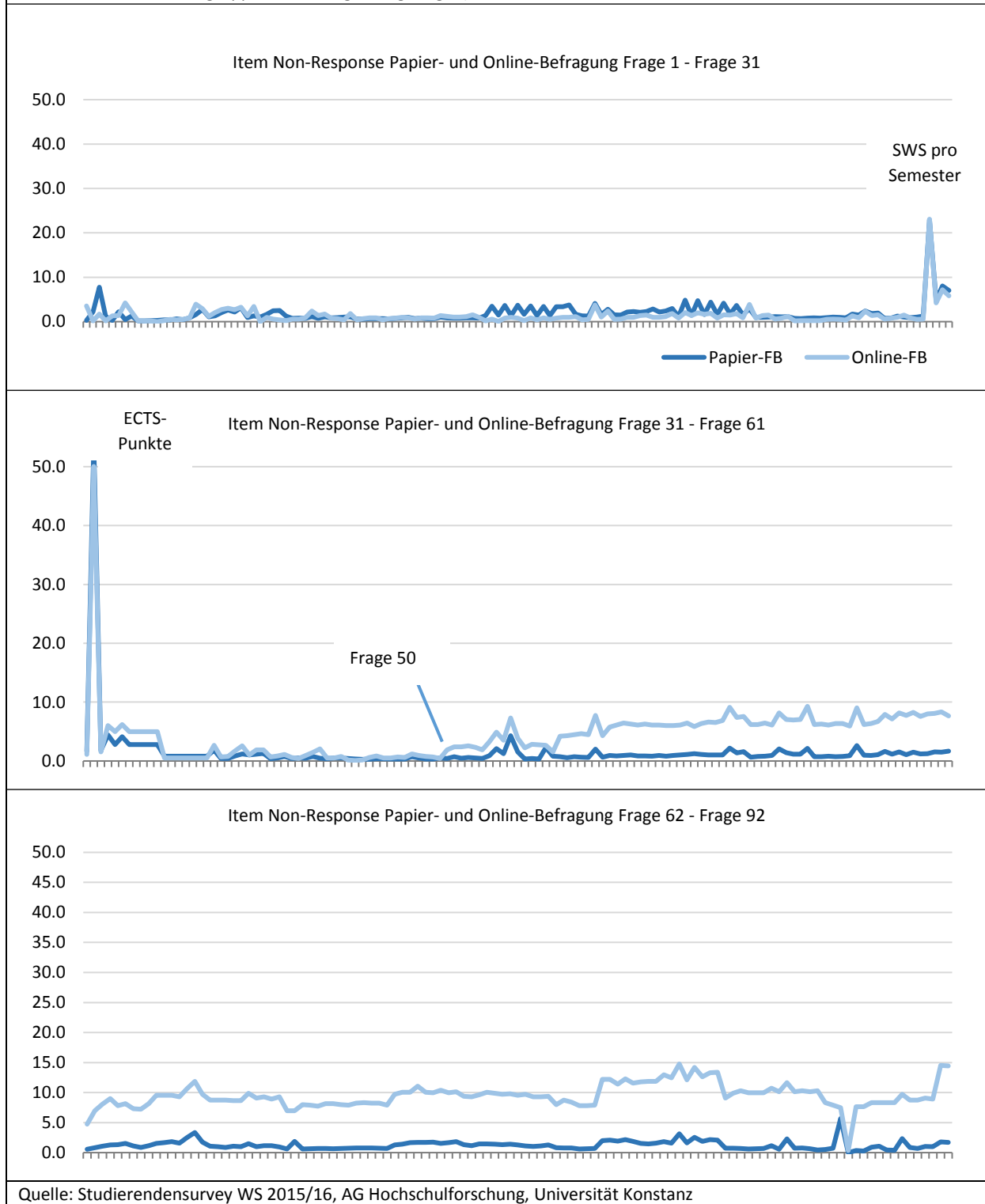
Bis zur Hälfte der Befragung zeigen sich noch keine auffälligen Unterschiede zwischen der Papier- und der Onlinegruppe; erst danach (ca. Frage 50) steigt der Item Non-Response in der Gruppe der online Befragten stetig an. In der Gruppe, die den schriftlichen Fragebogen bearbeitet hat, bleibt der Anteil fehlender Angaben hingegen bis zuletzt gering.

³ Betrachtet werden nur Befragte mit mehr als 50 % der Angaben, die die Befragung am selben Tag abgeschlossen/abgebrochen haben.

Abbildung 3.1

Item Non-Response nach Erhebungsart in Prozent (min. 50 % der Angaben)

(Papier-FB = Papierfragebogen + Wahlgruppe Papierfragebogen; Online-FB = Online langer Fragebogen mit Motivation, ohne Motivation + Wahlgruppe Online langer Fragebogen)



Um die Entwicklung des Item Non-Response im Zeitverlauf genauer untersuchen zu können, wird der Fragebogen in Abschnitte unterteilt. Als Anker zwischen den Abschnitten dienen die Motivationsseiten, die in Gruppe 4 zur Anwendung kamen. Diese wurden nach Frage 36, Frage 53 und Frage 62 eingeblenet. Folglich werden vier Abschnitte betrachtet: von Beginn bis zur ersten Motivationsseite,

zwischen erster und zweiter Motivation, zwischen zweiter und dritter Motivation sowie schließlich zwischen dritter Motivation und der letzten Seite des Fragebogens (vgl. Tabelle 3.4).

Die Ergebnisse eines Kruskal-Wallis Tests bestätigen, was die Grafik bereits vermuten lässt: Der Item Non-Response steigt nur in der Online-Befragung im Verlauf des Fragebogens an. Die Unterschiede zwischen den einzelnen Gruppen sind jedoch in jedem Abschnitt signifikant. Zu Beginn weisen die Teilnehmer der schriftlichen Befragung sogar mehr Item Non-Response auf als die Teilnehmer des Web-Surveys ($\chi^2(3) = 80.713$, $p < 0,0001$). Die Unterschiede liegen dabei klar zwischen Papier- und Onlinegruppe: Sowohl die randomisierte als auch die selbstselektierte Papiergruppe unterscheidet sich signifikant von den Teilnehmern der Online-Befragung.

Tabelle 3.4					
Durchschnittlicher Item Non-Response im Verlauf des Fragebogens nach Befragungsgruppe (WS 2015/16)					
(Nur Befragte mit mindestens 50 % der Antworten; Angaben in Häufigkeiten)					
	Papier rand.	Papier selbsts.	Online rand.	Online selbsts.	Kruskal-Wallis
1. Abschnitt	3	3,1	2,2	2,4	$\chi^2(3) = 80.713$, $p < 0,0001$
2. Abschnitt	0,7	0,6	1	1,3	$\chi^2(3) = 536.851$ $P < 0,0001$
3. Abschnitt	0,7	0,6	3,1	4	$\chi^2(3) = 692.097$ $P < 0,0001$
4. Abschnitt	1,4	1,6	9,6	11	$\chi^2(3) = 686.248$ $P < 0,0001$

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

Ein Problem stellen die durch Abbrüche verursachten fehlenden Angaben dar, die die Ergebnisse eventuell verzerren. Um diesen Fehler zu kontrollieren, werden in einer gesonderten Analyse nur Personen betrachtet, die bei der letzten Frage eine Angabe gemacht haben. Damit soll sichergestellt werden, dass die Teilnehmer den Fragebogen bis zum Ende ernsthaft ausgefüllt haben und keine Abbrecher oder Lurker⁴ die Ergebnisse verfälschen (vgl. Tabelle 3.5).

Tabelle 3.5					
Durchschnittlicher Item Non-Response im Verlauf des Fragebogens nach Befragungsgruppe (WS 2015/16)					
(Nur Befragte, die die letzte Frage beantwortet haben; Angaben in Häufigkeiten)					
	Papier rand.	Papier selbsts.	Online rand.	Online selbsts.	Kruskal-Wallis
1. Abschnitt	3	3,2	1,9	2,4	$\chi^2(3) = 9.026$, $p = 0,0289$
2. Abschnitt	0,7	0,6	1	1,1	$\chi^2(3) = 55.096$ $P = 0,0001$
3. Abschnitt	0,7	0,5	1,4	1,4	$\chi^2(3) = 80.963$ $P = 0,0001$
4. Abschnitt	1,2	1,3	2,5	2,9	$\chi^2(3) = 79.324$ $P = 0,0001$

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

⁴ Dabei wird unterstellt, dass Lurker sich bis zum Ende des Fragebogens durchklicken, ohne inhaltliche Angaben zu machen. Es ist unwahrscheinlich, dass diese Personen ausgerechnet die letzte Frage wieder ernsthaft beantworten, zumal sie sich nicht sicher sein können, dass es sich tatsächlich um die letzte Frage handelt.

Die Unterschiede zwischen den Gruppen bleiben bestehen, allerdings weisen die Teilnehmer der Online-Befragung nun deutlich weniger fehlende Angaben auf. Bis auf den ersten Abschnitt, in dem die schriftlich Befragten wieder den höheren Item Non-Response aufweisen, liegen die fehlenden Angaben in der Online-Befragung durchweg höher.

Es existieren demnach Hinweise auf eine etwas höhere Erschöpfung in der Online-Befragung. Obwohl die Teilnehmer darauf hingewiesen wurden, dass sie die Befragung jederzeit unterbrechen können um sich später erneut einzuloggen, nutzen nur wenige diese Möglichkeit. Es ist in der Praxis auch einfacher, einen Papierfragebogen nach einer Unterbrechung weiterzubearbeiten als sich wiederholt in ein Web-Survey einzuloggen. Darüber hinaus bietet der Papierfragebogen einen Anhaltspunkt, wie lange das Ausfüllen noch dauern wird. In der Online-Befragung wurde hingegen auf die Darstellung eines Fortschrittsbalkens verzichtet, da die Ergebnisse hierzu in der einschlägigen Literatur gemischt sind und in langen Befragungen von einem Feedback über den Stand der Befragung eher abgeraten wird (Crawford et al, 2001; Yan et. al, 2010; Conrad et. al, 2010). Diese Faktoren können für den höheren Item Non-Response der Online-Befragung mitverantwortlich sein.

3.2.5 Hinweis auf problematische Fragen – Zahleneingabe

Die Item Non-Response-Analyse kann auch Aufschluss über problematische Items liefern, denn eine hohe Anzahl fehlender Werte kann auf Probleme in der Darstellung, der Formulierung oder generell mit dem Inhalt der Frage hindeuten. Spitzen im Item Non-Response zeigen sich dabei in beiden Erhebungsmodi gleichermaßen (vgl. Abb. 3.1), was darauf hindeutet, dass einzelne Items, unabhängig von der Art der Erhebung, fehlende Angaben provozieren. Insbesondere die Zahleneingabe erscheint problematisch. Das zeigt sich bspw. bei der Frage nach den im Fachsemester vorgesehenen Wochenstunden oder bei der Angabe der bisher erworbenen ECTS-Punkte. Sowohl in der Papier- als auch in der Onlinegruppe steigt der Item Non-Response hier sehr stark an; bei der Schätzung der bisher erworbenen ECTS-Punkte auf bis zu 54 % in der schriftlichen Befragung (zu 50 % Online). Die Frage nach den ECTS-Punkten lassen Teilnehmer der schriftlichen Erhebung etwas häufiger offen ($\chi^2 = 7.2134$, $p = 0.007$). Bei der Angabe zu den Semesterwochenstunden existieren hingegen keine Unterschiede zwischen den Erhebungsmodi⁵.

Die offene Abfrage exakter Häufigkeitsangaben ist, losgelöst vom jeweiligen Befragungsmedium, mit Schwierigkeiten verbunden (Porst, 2014). Die Teilnehmer müssen aus der Befragungssituation heraus die benötigte Information aus dem Gedächtnis abrufen. Wie anspruchsvoll diese Aufgabe ist, hängt von der Verfügbarkeit und der Salienz der jeweiligen Information ab: Einprägsame Ereignisse, im Studienalltag bspw. Prüfungen, werden besser erinnert, da mit dem jeweiligen Ereignis bspw. starke Emotionen (Aufregung, Angst, Erleichterung) verbunden sind. Ebenso sind Informationen leichter verfügbar, die zeitlich nicht weit zurückliegen oder von besonderer Relevanz für den Befragten sind. Der niedrigere Item Non-Response im Falle der Semesterwochenstunden kann dann darauf zurückgeführt werden, dass die Studierenden diese Information leichter abrufen können: Da es sich um das laufende Semester handelt, ist die gesuchte Information zeitlich nah und relevant. Jedoch besucht kaum ein Studierender heutzutage noch Seminare gezielt nach der Anzahl der Credit Points, die er/sie für dieses erlangt. Die Studienverlaufspläne zeichnen Inhalt und Aufbau des Studiums bereits so genau vor, dass sich Studierende nicht aktiv um den Erwerb der nötigen ECTS-Punkte kümmern müssen (studienbegleitende Leistungen), weshalb die Schätzung der ECTS-Punkte in der Befragungssituation schwieriger ist.

⁵ Auch nach der Einschränkung des Samples auf Personen mit mindestens 50 % der Angaben oder auf Teilnehmer, die die letzte Frage beantwortet haben, zeigen sich keine signifikanten Effekte.

3.2.6 Unterschiede in der Darstellung zwischen Papier- und Onlinefragebogen

In dieser Phase der Umstellung des Instruments auf ein Web-Survey orientierte sich die Online-Befragung noch stark an der schriftlichen Vorlage. Auch wenn mit dieser Vorgehensweise noch nicht die Gestaltungsmöglichkeiten der Online-Befragung vollumfänglich genutzt werden konnten bzw. Regeln zur Gestaltung von Online-Befragungen teilweise missachtet wurden, war es doch notwendig, um potentiell auftretende Unterschiede weitestgehend auf das Befragungsmedium zurückführen zu können.

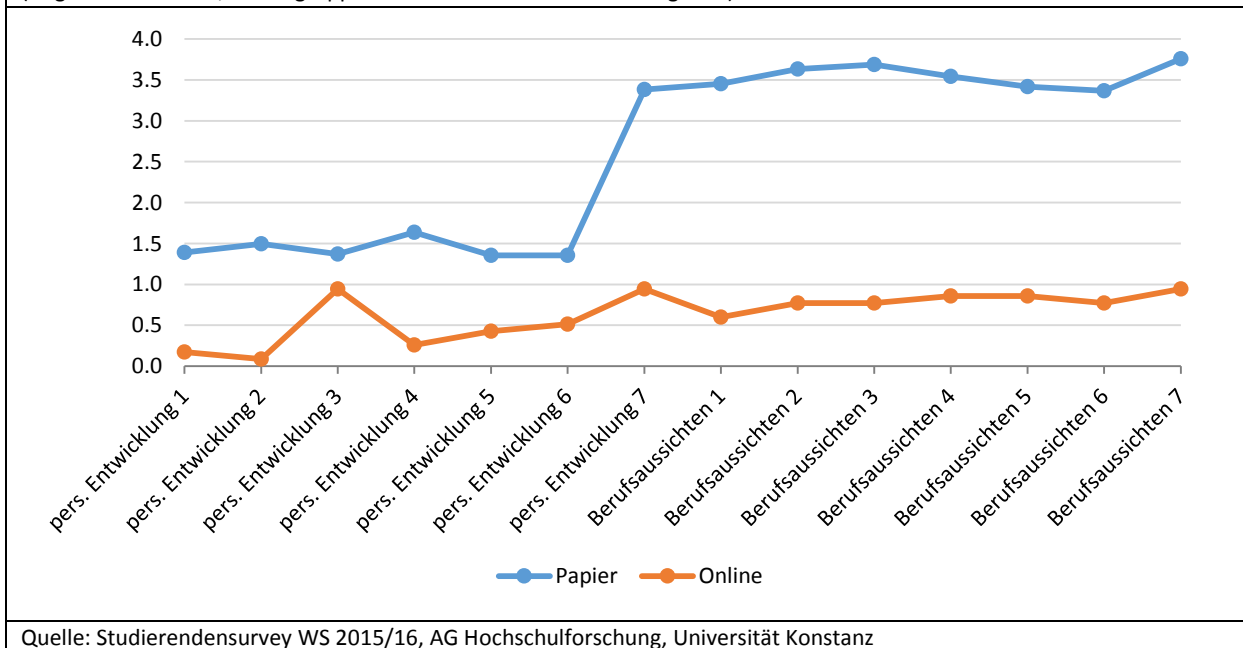
An drei Stellen musste in der Online-Befragung trotzdem vom ursprünglichen Design abgewichen und eine Anpassung vorgenommen werden. In der schriftlichen Befragung werden an diesen Stellen zwei Fragen nebeneinander dargestellt (vgl. Anhang, Abb. A1 und A2). Da dieses Design im Web-Survey nicht umgesetzt werden konnte, wurden die Fragen auf derselben Fragebogenseite untereinander abgebildet. Bei der Analyse inhaltlicher Unterschiede zwischen den Befragungsgruppen konnten Differenzen in der Verteilung zwischen Online-Befragung (ohne Wahlgruppe) und der schriftlichen Befragung festgestellt werden (vgl. Kapitel 7.1). Um die zweite Frage lesen zu können, müssen einige Befragte auf dem Bildschirm nach unten scrollen, was einen zusätzlichen Aufwand bedeutet, der auch eine Antwortverweigerung begünstigen könnte.

Frage 16 bestand aus zwei Teilfragen mit je sieben Items: In der ersten Teilfrage sollten die Teilnehmer angeben, wie wichtig ihnen verschiedene Aktivitäten für ihre persönliche Entwicklung erscheinen. Anschließend sollten sie den Nutzen derselben Aktivitäten für ihre Berufsaussichten angeben. In Abbildung 3.2 wird der Item Non-Response (IN) jedes Items für beide Teilfragen gegenübergestellt. Tatsächlich zeigt sich ein Effekt der Darstellung der beiden Fragen: Entgegen der Erwartungen steigt der IN jedoch in der Gruppe mit Papierfragebogen stärker an als in der Onlinegruppe⁶.

Abbildung 3.2

Item Non-Response Frage 16 – Darstellung zweier Teilfragen neben- bzw. untereinander

(Angaben in Prozent, Onlinegruppe: nur Personen mit 50 % der Angaben)



Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

⁶ Die parallele Darstellung zweier Fragen erfolgte an drei Stellen im Fragebogen. Im ersten Fall, bei der Frage nach Angeboten zum Studienbeginn, wurde jedoch nicht von allen Studierenden eine Antwort bei der zweiten Frage (Bewertung des Nutzens der Angebote) erwartet, weshalb nur ein Teil der Befragten hier Angaben machen konnte. Darum wurden diese beiden parallel dargestellten Fragen nicht berücksichtigt.

Während der Item Non-Response zwar auch in der Onlinegruppe in der zweiten Teilfrage ansteigt, sind die Ausschläge in der Papiergruppe sehr viel größer: Hier steigt der IN um durchschnittlich 1,8 Prozentpunkte an (Anstieg zwischen 0,4 – 2,3 Prozentpunkte). In der Onlinegruppe beträgt die Differenz im IN zwischen beiden Teilfragen durchschnittlich nur 0,3 Prozentpunkte (Veränderung zwischen -0,2 und 0,7 Prozentpunkte). Das gleiche Muster wiederholt sich bei Frage 19 (Möglichkeiten zur Weiterbildung, siehe Anhang, Abb. A3): Auch hier weisen Teilnehmer der schriftlichen Befragung einen höheren IN bei den Items der zweiten Teilfrage auf (1,9 – 3,7 Prozentpunkte) als Teilnehmer der Online-Befragung (0,3 – 1,3).

Eine Erklärung für dieses kontraintuitive Ergebnis könnte die Fragenreihenfolge sein: Bei der ersten parallel dargestellten Frage sollten die Studierenden Auskunft über Angebote beim Übergang an die Hochschule geben und diese in der zweiten Teilfrage anschließend bewerten. Nur jene Personen, die die Angebote genutzt haben, können logischerweise eine Bewertung abgeben. Es ist möglich, dass bei der schriftlichen Befragung dieses Antwortmuster bei der Bearbeitung der folgenden parallel dargestellten Fragen noch aktiv war. In der Online-Befragung mussten die Teilnehmer um die zweite Teilfrage komplett zu sehen auf dem Bildschirm nach unten scrollen und wurden so dazu gezwungen, auch den Fragetext noch einmal zu lesen.

Die Unterschiede im Antwortverhalten (inhaltliche Antwort vs. keine Antwort/Item Non-Response) bei der zweiten Teilfrage sind häufig signifikant, aber anders gelagert als erwartet: Die Teilnehmer der Online-Befragung geben häufiger inhaltliche Antworten als Teilnehmer der schriftlichen Befragung. Da die Differenz zwischen den Gruppen aber nie mehr als 3 Prozentpunkte beträgt, können die Unterschiede vernachlässigt werden. Wird die Onlinegruppe außerdem noch stärker an die Papiergruppe angepasst (nur 4 % fehlende Angaben⁷) verschwinden die Effekte fast vollständig, und nur vereinzelt sind noch Unterschiede (1 Prozentpunkt) zwischen beiden Gruppen festzustellen. Trotz unterschiedlicher Darstellung der Fragen in den beiden Erhebungsmodi lassen sich keine größeren Effekte nachweisen.

3.2.7 Letztes Item einer langen Itematterie

Lange Itematterien stellen insbesondere in Online-Befragungen ein Problem dar. Auch hier hängt es von der Darstellung des Fragebogens am Bildschirm und der Bildschirmauflösung ab, ob die Teilnehmer auf der Seite navigieren müssen, um alle Items zu sehen. Items, die in einer langen Itematterie weiter unten gelistet sind, werden daher unter Umständen übersehen. Als Folge sind Differenzen zwischen Papier- und Online-Befragung möglich: Teilnehmer der Online-Befragung beantworten die Items unten auf dem Bildschirm eventuell seltener.

Um dies zu überprüfen wurden Fragebatterien mit zehn und mehr Items näher untersucht. Die Angaben der Befragten wurden dichotomisiert (inhaltliche Antwort vs. keine Antwort) und auf Unterschiede zwischen den beiden Erhebungsmodi hin untersucht. Außerdem wurde der IN für die gesamte Fragebatterie berechnet und zwischen den beiden Befragungsgruppen verglichen.

Bei der Analyse von Itematterien in der zweiten Hälfte des Fragebogens ist Vorsicht geboten, weil fehlende Werte aufgrund von Befragungsabbrüchen die Ergebnisse verzerren können (vgl. auch Abb. 3.1). Um diese Unschärfe einzugrenzen, werden wieder nur Personen berücksichtigt, die weniger als vier Prozent fehlende Angaben gemacht haben (= durchschnittlicher IN der Personen, die die Online-Befragung erfolgreich abgeschlossen haben). Damit wird der IN durch Lurker und Abbrüche weitestgehend kontrolliert; Effekte können somit auf das Befragungsinstrument zurückgeführt werden. Insgesamt zehn Itematterien des Fragebogens bestanden aus zehn und mehr Items. Ein Mode-Effekt in

⁷ Durchschnittlicher IN der Onlinegruppe für Personen, die die Befragung erfolgreich abgeschlossen haben. Das Sample verringert sich dadurch noch einmal deutlich: Papiergruppe N=4.987; Onlinegruppe N=904.

der vorhergesagten Richtung besteht, wenn Teilnehmer der Onlinegruppe seltener inhaltliche Angaben machen bzw. der IN über die gesamte Fragebatterie höher ist als in der Papiergruppe.

In sechs der zehn Itembatterien (Frage 13, Frage 60, Frage 61, Frage 68, Frage 77, Frage 79) weist die Onlinegruppe einen höheren IN auf als die Papiergruppe ($p=0.05$; $p=0.01$). Auf Itemebene existieren vereinzelt signifikante Unterschiede im Verhältnis inhaltlicher zu nicht inhaltlichen Angaben zwischen den beiden Gruppen, jedoch sind diese äußerst gering ausgeprägt (maximal 1-2 %) und zeigen sich auch nicht in der vorhergesagten Weise: Der Item Non-Response steigt in der Onlinegruppe nicht wie erwartet bei den weiter unten auf dem Bildschirm angezeigten Items an, sondern folgt keinem spezifischen Muster. Das bedeutet, dass Teilnehmer der schriftlichen Befragung etwas häufiger vollständige bzw. inhaltliche Angaben in den langen Itembatterien machen, Teilnehmer der Onlinegruppe wiesen hingegen dementsprechend häufiger fehlende Werte auf. Diese Unterschiede zeigen sich aber 1) nicht in der vorhergesagten Weise (IN steigt über die Itembatterie an) und sind 2) zu gering ausgeprägt, so dass ein Mode-Effekt nicht bestätigt werden kann.

3.3 Befragungsabbrüche

Sowohl in Onlinesurveys als auch in schriftlichen Befragungen stehen keine Interviewer/innen zur Verfügung, die die Befragung steuern, Teilnehmer motivieren oder Missverständnisse klären könnten, weshalb Abbrüche, also das vorzeitige Beenden der Befragung, ein besonderes Problem darstellen. Im Gegensatz zur schriftlichen Befragung liegen für die Online-Erhebung des Studierendensurveys jedoch Informationen über den Zeitpunkt des Abbruchs sowie über die entsprechende Ausstiegsseite vor. Die Analyse dieser Informationen kann Aufschluss über problematische Fragen und andere Qualitätsprobleme des Instruments liefern. Zudem können sich Personen, die die Befragung vorzeitig abbrechen, hinsichtlich untersuchungsrelevanter Merkmale von Antwortenden unterscheiden, was eine verzerrte Stichprobe zur Folge hätte (Knapp & Heidingsfelder, 1999; Peytchev, 2009). Die aus der Analyse der Abbruchstatistik gewonnenen Ergebnisse liefern außerdem wichtige Hinweise für die weitere Ausgestaltung und Überarbeitung des Instruments.

Befragungsabbrüche beeinflussen darüber hinaus auch den Item Non-Response bzw. den Unit Non-Response: Je nachdem an welcher Stelle im Fragebogen der Abbruch erfolgt, kann der Item Non-Response für bestimmte problematische Fragen ansteigen. Gehen die Abbrüche hauptsächlich auf die Länge des Fragebogens zurück, steigt der Item Non-Response im Verlauf der Befragung an. Fragen im fortgeschrittenen Teil des Fragebogens weisen dann einen höheren Item Non-Response auf. Erfolgt ein Abbruch bereits sehr früh im Fragebogen oder werden bestimmte Fragen nicht beantwortet (vgl. Abschnitt 3.1.1), müssen die betreffenden Fälle aussortiert werden, was wiederum die für Analysen verwertbare Stichprobe weiter einschränkt. Eine eingehendere Analyse der Abbrüche ist damit unerlässlich.

Der Klassifizierung von Callegaro, Lozar Manfreda und Vehovar (2015) folgend können Befragungsabbrüche noch weiter in *introduction breakoffs*, *questionnaire breakoffs* und *total breakoff* unterschieden werden. In Tabelle 3.6 werden für die Gruppen der Online-Befragung die Anteile der jeweiligen Abbrüche berichtet.

Die Introduction Breakoff Rate (IBR) gibt an, wie viele der Personen, die die Online-Befragung besucht haben, auf der Startseite abbrechen. In der Meta-Analyse von Vehovar und Čehovin machten Abbrüche auf der Startseite 85 % aller Abbrüche aus. Im Gegensatz dazu ist die IBR in der vorliegenden Studie vergleichsweise gering: In der Wahlgruppe beträgt sie 2,49 %, in der randomisierten Onlinegruppe 3,23 %. Allerdings handelte es sich in der Meta-Analyse von Vehovar und Kollegen um klassische „Do-It-Yourself“-Web-Surveys, bei denen Teilnehmer über Werbebanner, Pop-Up Fenster, etc. rekrutiert werden. Sehr hohe Abbruchraten auf der Startseite sind in diesen Online-Befragungen

durchaus üblich. Dennoch ist der Befund zur IBR bemerkenswert, wurde doch auf der Startseite bereits die voraussichtliche Bearbeitungsdauer der Befragung mit 1h angegeben. Trotz der oft angeführten „Befragungsmüdigkeit“ der Studierenden scheint die lange Bearbeitungsdauer nicht das Hauptproblem zu sein.

Tabelle 3.6			
Abbruchraten in den beiden Gruppen der Online-Befragung (WS 2015/16)			
(Angaben in Prozent)			
	Introduction Breakoff Rate	Questionnaire Breakoff Rate	Total Breakoff Rate
Online Wahlgruppe	$7/281 * 100 = 2,49 \%$	$60/(281-7) * 100 = 21,90 \%$	$(7 + 60)/281 * 100 = 23,84 \%$
Online randomisiert	$39/1208 * 100 = 3,23 \%$	$341/(1208-39) * 100 = 29,17 \%$	$(39 + 341)/1208 * 100 = 31,46 \%$
Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz			

Die Questionnaire Breakoff Rate (QBR) gibt Aufschluss darüber, welcher Anteil der Personen, die über die Startseite hinaus gekommen sind, vor Abschluss des Fragebogens die Befragung abgebrochen haben. In der Wahlgruppe haben knapp 22 % vorzeitig abgebrochen, in der randomisierten Online-Befragung brechen die Teilnehmer mit 29 % etwas häufiger die Befragung ab. Offenbar sind die Teilnehmer der selbstselektierten Onlinegruppe etwas motivierter bzw. weisen eine höhere Ausdauer auf.

Die Total Breakoff Rate (TBR) summiert die Abbrüche auf der Startseite und im Fragebogenverlauf auf und setzt sie ins Verhältnis mit der Anzahl an Personen, die die Befragung begonnen haben. Die TBR wird häufig auch als Overall Breakoff Rate, also die Quote für den gesamten Abbruch bezeichnet. Sie liegt in der randomisierten Onlinegruppe deutlich höher als in der selbstselektierten Gruppe (31 % zu 24 %).

Der Vergleich zwischen randomisierter und selbstselektierter Onlinegruppe zeigt, dass eine Wahlmöglichkeit die Befragten nachhaltig motivieren kann. Mit der bewussten Entscheidung für die Online-Befragung haben die Teilnehmer bereits Zeit bzw. „Aufwand“ investiert. Sie hätten den beiliegenden Papierfragebogen ausfüllen können, entschieden sich aber für die Online-Befragung und nahmen damit die umständliche Eingabe der URL sowie des Passworts an einem PC in Kauf. Menschen neigen dazu, an Handlungsalternativen festzuhalten, wenn bereits Zeit, Geld oder Mühe in diese Alternative investiert wurde. In der Psychologie spricht man von „Sunk-Cost-Effects“: Anstatt die Situation neu zu bewerten und eine andere Handlungsalternative zu wählen, wird die bereits gewählte Taktik weiterverfolgt, weil damit Kosten einhergingen (Kahneman & Tversky, 1979; Arkes & Ayton, 1999). Die Ergebnisse legen nahe, dass in der Wahlgruppe eventuell Sunk-Cost-Effekte wirksam geworden sein könnten.

Das drängendere Problem in beiden Gruppen ist die hohe QBR in Kombination mit den sehr geringen Rücklaufquoten: Weil die Rücklaufquoten in Online-Erhebungen generell niedriger sind als in schriftlichen Befragungen, was auch auf den 13. Studierendensurvey zutrifft, ist es umso wichtiger, möglichst viele Teilnehmer in der Befragung zu halten, denn nur so kann eine hohe Datenqualität sichergestellt werden. Das Hauptproblem ist zwar der hohe Unit Non-Response (online randomisiert: 90 %), dennoch müssen die wenigen Teilnehmer, die das Web-Survey starten, bis zum erfolgreichen Abschluss der Befragung motiviert und in der Befragung gehalten werden.

3.3.1 Completion Rates

Gemessen am Umfang des Fragebogens und der zuvor in Pretests ermittelten durchschnittlichen bzw. tatsächlichen Bearbeitungsdauer (ca. 60 Minuten, vgl. Abschnitt 4.2), sind die Completion Rates erstaunlich hoch (vgl. Tabelle 3.7). Die Completion Rate (CR) ist die Anzahl bis zum Ende ausgefüllter und abgeschickter Fragebogen geteilt durch die Anzahl an Befragten, die auf die Befragung zugegriffen haben. Es werden also nur jene Personen berücksichtigt, die auf den Web-Survey zugegriffen haben, nicht auf alle, die kontaktiert wurden. In der selbstselektierten Online-Befragung liegt die CR bei 78,6 %, in den beiden randomisierten Gruppen etwas niedriger bei 72,2 % (mit Motivationsseiten) und 71,4 %.

Natürlich gilt auch für die CR: je höher, desto besser. Eine CR von 78 % wird in verschiedenen Online-Communities aber bereits als guter Standard angesehen (FluidSurveys, 2014; MeasuringU, 2011). In einer aktuellen Auswertung von 25.080 Web Surveys berichten Liu und Wronski (2017) von einer durchschnittlichen CR von 87 %. Ihre Daten stammen aus einem online nonprobability access panel des Software Anbieters Survey Monkey. Die relativ hohe durchschnittliche CR führen die Autoren auf eine Besonderheit der Survey Monkey Umfragen zurück, wonach Befragungen, die eine CR von weniger als 60 % aufweisen, pausiert und eingehend auf Fehler überprüft werden. Ein Mechanismus, der durchaus sinnvoll erscheint. Die CR lässt sich jederzeit während der Feldphase bestimmen. Ein aktives Monitoring könnte logische oder technische Fehler aufdecken und wesentlich zur Erhöhung des Rücklaufs beitragen.

Tabelle 3.7
Completion Rates in den Gruppen der Online-Befragung (WS 2015/16)
(Angaben in Prozent)

	Completion Rate
Wahlgruppe Online	221/ 281 *100 = 78,6
Online mit Motivationsseiten	425/ 589 * 100 = 72,2
Online ohne Motivationsseiten	442/ 619 * 100 = 71,4

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

Die meisten Aussagen zu CR oder auch zum Abbruch beziehen sich auf Web-Surveys aus dem Bereich der Markt- und Meinungsforschung, die Teilnehmer auf verschiedene Weise (Banner, Links, Kunden-E-Mails, etc.) und teilweise aus speziellen Online Panels rekrutieren. Damit liegen andere Voraussetzungen vor als für den Studierendensurvey. Das schränkt die Vergleichbarkeit ein. Vehovar und Čehovin (2014) berichten über Ergebnisse einer Meta-Analyse von 5.752 Online-Befragungen, die mehrheitlich von Studierenden bearbeitet wurden. Hier lag die CR bei 60 % (zitiert nach Callegaro, Lozar Manfreda und Vehovar, 2015). Gemeinsam ist allen Aussagen, dass sie nur als Richtlinie dienen können. In Anbetracht der hohen Teilnahmeverweigerung ist jedoch eine konservative Vorgabe zur anvisierten Completion Rate geboten.

3.3.2 Welche Faktoren begünstigen Abbrüche?

Für gewöhnlich werden drei Faktorenbündel bei der Analyse von Befragungsabbrüchen berücksichtigt: 1) Merkmale der Befragten, 2) Merkmale der Befragung und 3) Merkmale der Fragen/Seiten der Online-Befragung. Merkmale der Befragten umfassen soziodemographische Aspekte, die Kompetenzen der Befragten sowie ihre Einstellungen und Motive (El-Menouar, Blasius, 2005; Peytchev, 2009). Befragte können mit der Teilnahme außerdem einen bestimmten Nutzen für sich selbst oder andere

verbinden (Ergebnisse werden für Verbesserungen an der Hochschule genutzt, Ergebnisse fließen in die Hochschulpolitik ein, Meinung kundtun, materielle und immaterielle Incentives).

Befragungsspezifische Faktoren beinhalten bspw. die Art und den Inhalt der Einladung zur Befragung, die Dauer, die Anzahl der Erinnerungen, den Zeitpunkt der Befragung bzw. des Versands der Erinnerungen, das Thema und die generellen Gestaltungsmerkmale wie zum Beispiel das Vorhandensein eines Fortschrittsbalkens (Peytchev, 2009).

Merkmale der Fragen/der Seiten umfassen das Frageformat (offene vs. geschlossene Fragen, Einfach- oder Mehrfachantworten, Matrixfragen, etc.), die tatsächliche Länge des Fragebogens, technische Probleme oder spezielle Softwarevoraussetzungen, Inhalt der Fragen (sensible Fragen, Einstellungsfragen, Fragen die das Abrufen von Informationen aus dem Gedächtnis erfordern), Anzahl der Fragen etc. (Ebd.).

Über den Unterschied zwischen Antwortenden und Abbrechern können nur bedingt Aussagen getroffen werden, da für Letztere nur begrenzte Informationen vorliegen. 50 % der Abbrüche passieren bereits bis Frage 17 (von 92) im Fragebogen (ohne Abbruch auf der Startseite), weshalb für einen Vergleich potentiell bedeutende Informationen (Demographie) über diese Gruppe nicht vorliegen. Eine Follow-Up-Befragung der Abbrecher zur Ermittlung der Ursachen des Abbruchs war aus datenschutztechnischen und finanziellen Gründen nicht möglich. Darum konzentriert sich die folgende Darstellung auf Merkmale bestimmter Fragen und Seiten der Erhebung, bei denen auffallend viele Personen abbrechen. Es wird außerdem diskutiert, wie befragungsspezifische Faktoren der Erhebung einen Abbruch begünstigt haben könnten.

3.3.3 Fragebogenspezifische Faktoren

In Abbildung 3.3 sind die Abbrüche je Befragungsseite im Fragebogenverlauf abgebildet. Zehn Prozent der Abbrüche geschehen noch auf der Startseite. Bei diesen Befragten war das Interesse durch die Einladung groß genug, dass sie immerhin den Link zur Umfrage sowie ihr persönliches Kennwort eingegeben und die Umfrage aufgerufen haben. 1.535 (=n1) Befragte besuchten die Startseite, aber nur 1.489 (=n2) haben die Umfrage tatsächlich begonnen. Der Motivationserfolg, also der Anteil von Befragten, der über die Startseite hinausgeht, beträgt 97 % (n2/n1, vgl. auch Theobald, 2000, S. 71 ff.; El-Menouar, Blasius, 2005).

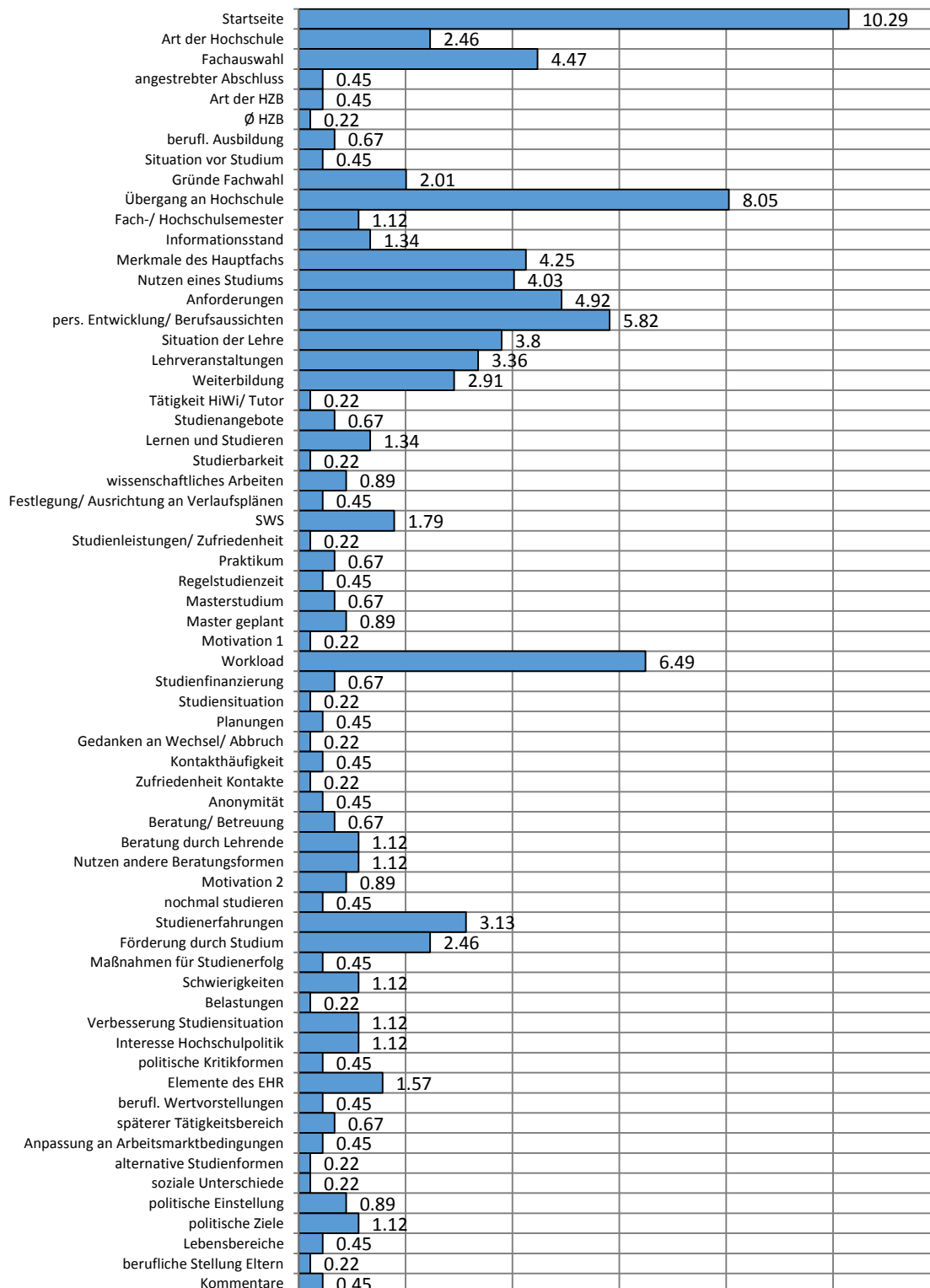
Die Hälfte der Abbrüche erfolgt bis zu Frage 17, also noch relativ früh im Fragebogen. Hierbei handelt es sich um Befragte, die sich die Umfrage erst einmal anschauen wollen und sich dann gegen eine ernsthafte Teilnahme entscheiden. Über die Gründe lässt sich bestenfalls spekulieren, aber mit einiger Wahrscheinlichkeit haben frühe Abbrecher wenig Interesse an den abgefragten Themen.

Einige Befragte stellen vielleicht auch schnell fest, dass eine korrekte Bearbeitung des Fragebogens mehr Zeit in Anspruch nimmt, als sie im Moment erübrigen können oder als sie generell bereit sind zu investieren.

Abbrüche bei zwei Fragen pro Seite

An drei Stellen im Web-Survey wurden in Analogie zum schriftlichen Fragebogen zwei Fragen auf einer Seite präsentiert (vgl. Abschnitt 3.2.3). Insbesondere an den ersten beiden Stellen (Frage 9: „Übergang an die Hochschule“ und Frage 16: „Persönliche/berufliche Entwicklung“) verursachte die Darstellung bzw. der Inhalt der Fragen viele Abbrüche: acht bzw. sechs Prozent über alle Gruppen hinweg (randomisiert und selbstselektiert). Die Abbrüche bei den Fragen zum Übergang an die Hochschule sind allerdings darauf zurückzuführen, dass hier nur ein Teil der Befragten Angaben machen konnte:

Abbildung 3.3
Abbruch in der Online-Befragung nach Fragebogenseite im Verlauf der Befragung
 (Angaben in Prozent)



Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

Die erste Frage auf der Seite bezog sich auf Angebote und Maßnahmen zum Übergang an die Hochschule. In der zweiten Frage sollten die Teilnehmer den Nutzen der Angebote bewerten, was streng genommen wiederum nur Personen konnten, die die Angebote genutzt hatten.

Dass aber nicht nur der Inhalt, sondern auch die Darstellungsform den Befragten Probleme bereitet, zeigen die Abbrüche bei Frage 16 und 19, bei denen ebenfalls zwei Fragen untereinander abgebildet wurden. Hier brachen zwischen sechs und drei Prozent der Teilnehmer ab. Durch die Darstellung mussten die Befragten auf der Bildschirmseite navigieren; zudem wirkt die Seite überfüllt, was die Abbruchbereitschaft der Befragten erhöht.

Abbrüche bei langen Itembatterien und viel Text

Ein ähnliches Problem betrifft Seiten, auf denen Fragen mit langen Itembatterien abgebildet wurden (zehn und mehr Items). Auch hier wurde aus Gründen der Vergleichbarkeit zur schriftlichen Befragung darauf verzichtet, die Itembatterien auf zwei oder mehr Seiten zu splitten. So wirken die Seiten, je nach Bildschirmauflösung bzw. verwendetem Gerät, schnell unübersichtlich und überfrachtet und damit demotivierend. Bei der Überarbeitung des Instruments ist auf eine ausgewogenere Darstellung der Itembatterien zu achten.

Aber auch bei weniger umfangreichen Itembatterien, z. B. bei acht Items, lassen sich viele Abbrüche registrieren. Der Grund hierfür liegt in den Items an sich, die sehr textlastig sind und daher ebenfalls viel Raum auf der Bildschirmseite einnehmen (z.B. Frage 17: „Situation der Lehre“ und Frage 18: „Lehrveranstaltungen“). Vor dem Hintergrund der „abschreckenden“ Wirkung der Textmenge, gerade in Online-Befragungen, sollten alternative und vor allem kürzere Formulierungen der Items geprüft werden. Problematisch bei diesem Vorgehen ist der Bruch mit der Zeitreihe.

Studentische Arbeitslast – Zusammenhang mit Motivationsseiten

Die Schätzung des studentischen Workloads scheint ebenfalls mit Schwierigkeiten verbunden zu sein. Die Befragten müssen nicht nur Überlegungen über ihre Studienaktivitäten anstellen, sie müssen die Dauer dieser Tätigkeiten außerdem möglichst genau schätzen und vor dem Absenden der Information nochmals prüfen, ob ihre Schätzung plausibel ist. Die Frage nach der studentischen Arbeitslast ist also mit einem hohen Aufwand verbunden. Interessant ist aber, dass der Abbruch vor allem auf die randomisierte Onlinegruppe mit Motivationsseiten (Gruppe 3) zurückzuführen ist (vgl. Abbildung 3.4).

Nicht nur brachen Teilnehmer direkt auf der ersten Motivationsseite ab (4 Abbrüche, 2,4 %), sie provozierte außerdem auf der nachfolgenden Seite weitere Abbrüche (15,9 % der Abbrüche in Gruppe 3). Bei der zweiten und dritten Motivationsseite zeigt sich hingegen kein Anstieg der Abbrüche.

Die Motivationsseiten könnten den Befragten auch als Ausstiegsseiten dienen, allerdings brechen auf der zweiten Motivationsseite ebenfalls wieder nur vier Befragte ab und auf der dritten Motivationsseite schließlich kein Befragter mehr. Auch kommt es bei den Folgefragen nur nach der ersten Motivationsseite zu einem vermehrten Abbruch. Im Gegenteil: Nach der zweiten Motivationsseite bricht in beiden Onlinegruppen (Gruppe 3 und 4) jeweils ein Befragter ab; die dritte Motivationsseite scheint sogar einen positiven Effekt zu haben, da weniger Personen in der Gruppe mit Motivation abrechen als in der Gruppe ohne Motivation (1 zu 3 Befragte).

Zudem nehmen die Abbrüche in Gruppe 3 im Befragungsverlauf ab; es werden weniger Abbrüche als in den beiden anderen Onlinegruppen verzeichnet. Das deutet darauf hin, dass die Teilnehmer der Gruppe mit Motivationsseiten vor allem in der fortgeschrittenen Befragung besser durchhalten als die übrigen Teilnehmer der Online-Befragung (vgl. Abbildung 3.4).

Zwischen der Startseite und der ersten Motivation (Frage 1 bis Frage 36 = 1. Abschnitt) erfolgten in der selbstselektierten Onlinegruppe knapp 72 % der Abbrüche, in der Gruppe mit Motivationsseiten waren es 68 % und in der Gruppe ohne Motivationsseiten 62 % (vgl. Tabelle 3.8). Der Abbruch zwischen 1. und 2. Motivationsseite (Frage 37 bis Frage 53 = 2. Abschnitt) liegt in Gruppe 3 deutlich höher als in den beiden anderen Gruppen: hier ereigneten sich 21 % der Abbrüche. Ab Frage 54 ändert

sich dann aber das Bild, denn nach etwas mehr als der Hälfte des Fragebogens gehen die Abbrüche in Gruppe 3 zurück. Zwischen der 2. und 3. Motivationsseite (3. Abschnitt) entspricht die Abbruchquote jener der Wahlgruppe (7,3 % zu 7,5 %); in Gruppe 4 liegt sie mit 15 % deutlich höher. Im letzten Abschnitt ist die Abbruchquote der Gruppe 3 schließlich am niedrigsten, während sie in der Wahlgruppe noch einmal deutlich ansteigt (14,9 %) und in Gruppe 4 nur leicht abnimmt (12,4 %).

Tabelle 3.8				
Abbruch nach Befragungsabschnitt in den Gruppen der Online-Befragung (WS 2015/16)				
(Angaben in Prozent)				
	1. Abschnitt	2. Abschnitt	3. Abschnitt	4. Abschnitt
Wahlgruppe	71,6	6,0	7,5	14,9
Gruppe 3*	68,3	21,3	7,3	3,0
Gruppe 4	61,6	10,7	15,3	12,4

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz
* Inklusive Abbrüche auf Motivationsseiten

Um zu überprüfen, ob sich die Anzahl der Abbrüche in den verschiedenen Fragebogenabschnitten zwischen den Gruppen mit und ohne Motivationsseiten signifikant voneinander unterscheiden, wird ein T-test berechnet (vgl. Tabelle 3.9). Nur für die letzte Phase zwischen der dritten Motivationsseite bis zum Ende der Befragung wird der Unterschied signifikant. Das bedeutet, dass es in der Gruppe mit Motivationsseiten im letzten Abschnitt zu weniger Abbrüchen kommt als in der Gruppe ohne Motivationsseiten. Daraus lässt sich ein später positiver Effekt der Motivationsseiten ableiten. Für künftige Befragungen ist jedoch darauf zu achten, dass die Folgefragen nach den Motivationsseiten nicht demotivierend wirken, also nicht zu anspruchsvolle Fragen gewählt werden.

Tabelle 3.9					
T-tests für Abbruch zwischen Gruppe 3 und 4 in den verschiedenen Abschnitten des Fragebogens (WS 2015/16)					
(T-Werte und Überschreitungswahrscheinlichkeiten p)					
	Gesamt	1. Abschnitt	2. Abschnitt	3. Abschnitt	4. Abschnitt
t-wert	-0,302	0,095	0,619	-1,532	-4,389
P	0,763	0,924	0,547	0,154	0,000

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
p > 0,05 = nicht signifikant, p <= 0,05 = *, p <= 0,01 = **, p <= 0,001 = ***

3.3.4 Abbruch und Antwortqualität

In Online-Befragungen, die häufiger als schriftliche Befragungen nur unvollständig bearbeitet werden, ist von besonderem Interesse, von welcher Qualität die Antworten der Abbrecher sind und ob sie gemeinsam mit den vollständigen Daten jener Befragten ausgewertet werden können, die die Umfrage vollständig bearbeitet haben. Hierzu werden die inhaltlichen Angaben von Abbrechern mit den Angaben der Personen, die die Befragung erfolgreich abgeschlossen haben, verglichen. Da ein Großteil der Abbrüche bereits sehr früh im Befragungsverlauf passiert, muss sich der Vergleich auf Fragen im ersten Fünftel des Fragebogens beschränken (vgl. Abschnitt 3.3.1: 50 % der Abbrüche geschehen bereits bis Frage 17). Alle Teilnehmer der Online-Befragung werden gemeinsam betrachtet und nur nach ihrem Status (abgebrochen vs. beendet) unterschieden.

Studiensicherheit nach Teilnehmerstatus

Die Teilnehmer sollten ihre Situation vor der Studienaufnahme beschreiben und angeben, wie sicher das Studium bereits feststand. Dafür standen ihnen vier Antwortmöglichkeiten zur Auswahl. Es bestehen geringfügige Unterschiede in der Verteilung nach Status der Befragten ($\chi^2(3) = 9,94, p = 0,019$). Für Studierende, die die Befragung abgebrochen haben, stand das Studium etwas seltener schon im Voraus fest und sie geben geringfügig häufiger an, dass sie ursprünglich nicht studieren wollten. Der Unterschied beträgt jedoch nur zwischen zwei und vier Prozentpunkten.

Tabelle 3.10		
Studiensicherheit nach Teilnehmerstatus (WS 2015/16)		
(Angaben in Prozent für die Kategorien)		
	Abgebrochen	Beendet
Studiensicherheit		
Ich wollte eigentlich nicht studieren	5	3
Ich war mir lange Zeit unsicher, ob ich studieren soll	11	15
Ich war mit ziemlich sicher, dass ich studieren wollte	36	30
Für mich stand von vornherein fest, dass ich studieren werde	48	52
Chi ² = 9,94, p = 0,019 **		
Cramers V = 0,08		
Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz		
p > 0,05 = nicht signifikant, p <= 0,05 = *, p <= 0,01 = **, p <= 0,001 = ***		

Werden neben den inhaltlichen Antworten auch die fehlenden Werte näher betrachtet, so zeigt sich zu diesem Zeitpunkt in der Befragung ebenfalls nur eine geringe Differenz bei der Antwortverweigerung: Von jenen Personen, die die Befragung später abbrechen, haben vier Personen keine inhaltliche Angabe gewählt (1 %), von jenen, die die Befragung abschließen, waren es drei (0,28 %).

Informationsstand

Bei dieser Frage wird der Informationsstand der Studierenden in verschiedenen Bereichen erhoben. Die Befragten haben die Möglichkeit, bei ihrer Antwort auf einer sechsstufigen Skala (viel zu wenig, etwas zu wenig, ausreichend, gut, sehr gut, dieser Bereich interessiert mich nicht) eine Abstufung vorzunehmen. Anders als bei der Sicherheit der Studienaufnahme ergeben sich für den Informationsstand keine signifikanten Unterschiede nach Teilnehmerstatus.

Tabelle 3.11**Informationsstand nach Teilnehmerstatus (WS 2015/16)**

(Skala von 1 = viel zu wenig bis 5 = sehr gut, 6 = dieser Bereich interessiert mich nicht. Angaben in Prozent für Kategorien 4 – 5 = gut)

Wie gut informiert fühlen Sie sich über folgende Bereiche?	Abgebrochen	Beendet	Chi	V
Studien- und Prüfungsordnung	45	48	3,001	0,047
Möglichkeiten der Studienberatung	37	40	1,642	0,035
Angebote zur Studienfinanzierung	24	26	1,417	0,032
Möglichkeiten für ein Studium im Ausland	33	32	5,124	0,061
Arbeitsmarktsituation in angestrebten Tätigkeitsfeld	36	35	0,469	0,018

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

p > 0,05 = nicht signifikant, p <= 0,05 = *, p <= 0,01 = **, p <= 0,001 = ***

Anforderungen

Die Studierenden sollen eine Einschätzung darüber liefern, ob die Anforderungen, die ihnen im Fachbereich gestellt werden, angemessen sind oder ob darauf zu viel bzw. zu wenig Wert gelegt wird. Insgesamt sollen zehn Bereiche bewertet werden (vgl. Tabelle 3.12). Die beiden Gruppen unterscheiden sich lediglich in ihrer Einschätzung hinsichtlich der Kritik an Lehrmeinungen: Teilnehmer, die die Befragung erfolgreich abgeschlossen haben, sind häufiger der Ansicht, der Fachbereich würde hierauf zu wenig Wert legen (Kategorien „viel zu wenig“ und „etwas zu wenig“ 51 % zu 48 %). Dementsprechend geben Teilnehmer, die später abbrechen, häufiger an, dass die Anforderungen in diesem Bereich überzogen wären (Kategorien „etwas zu viel“ und „viel zu viel“ 5 % zu 2 %). Die Unterschiede sind allerdings äußerst gering.

Tabelle 3.12**Anforderungen im Fachbereich nach Teilnehmerstatus (WS 2015/16)**

(Skala von 1 = viel zu wenig bis 5 = viel zu viel, 6 = kann ich nicht beurteilen. Angaben in Prozent für Kategorie 3 = gerade richtig)

Wie beurteilen Sie die Anforderungen, die an Studierende in Ihrem Fachbereich gestellt werden?	Abgebrochen	Beendet	Chi	V
Faktenwissen erwerben	46	47	0,627	0,052
zugrunde liegende Prinzipien verstehen	64	66	1,173	0,030
Interessenschwerpunkte entwickeln	38	39	9,887	0,087
Kritik an Lehrmeinungen üben	34	35	11,165*	0,093
viel und intensiv für das Studium arbeiten	39	40	4,268	0,512
regelmäßige Leistungsnachweise erbringen	55	54	1,512	0,034
an Diskussionen beteiligen	48	43	5,831	0,323
praktische Anwendung des Gelernten	40	37	10,338	0,089
sich mit ethischen Fragestellungen befassen	30	31	2,688	0,046
Forschungsmethoden anwenden können	32	36	6,475	0,071

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

p > 0,05 = nicht signifikant, p <= 0,05 = *, p <= 0,01 = **, p <= 0,001 = ***

Motive für die Studienfachwahl

Bei dieser Frage sollten die Studierenden bewerten, inwiefern verschiedene Motive die Wahl ihres derzeitigen Studienfachs beeinflusst haben. Insgesamt sollten sieben Items auf einer Skala von 0 = sehr unwichtig bis 6 = sehr wichtig bewertet werden. Bei drei der sieben Items zeigen sich signifikante Unterschiede zwischen den Gruppen. Demnach war das spezielle Fachinteresse für Personen, die die Umfrage erfolgreich beendet haben, noch etwas wichtiger, als für Teilnehmer, die die Befragung vorzeitig abgebrochen haben. Die Aussicht, durch das Studium später in eine Führungsposition

aufsteigen zu können sowie die Möglichkeit, durch das Studium die Berufstätigkeit weiter hinauszuschieben, ist hingegen für Personen bedeutender, die die Befragung abbrechen. Zwischen einzelnen Motiven der Fachwahl und dem Teilnehmerstatus besteht demnach ein nur sehr geringer, aber signifikanter Zusammenhang.

Tabelle 3.13			
Motive der Fachwahl nach Teilnehmerstatus (WS 2015/16)			
(Skala von 0 = sehr unwichtig bis 6 = sehr wichtig. Angaben = Mittelwerte)			
Fachwahlmotive	Abgebrochen	Beendet	t-Werte
spezielles Fachinteresse	4,99	5,13	-2,028*
eigene Begabung, Fähigkeiten	4,68	4,75	n.s.
Einkommenschancen im späteren Beruf	3,51	3,35	n.s.
fester Berufswunsch	3,14	3,24	n.s.
sicherer Arbeitsplatz	3,84	3,78	n.s.
Aussicht auf Führungsposition	2,73	2,44	2,510**
Ausweichlösung	1,02	0,68	3,404***

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
 p > 0,05 = nicht signifikant, p <= 0,05 = *, p <= 0,01 = **, p <= 0,001 = ***

Charakterisierung des Hauptfachs

Bei dieser Frage sollten die Studierenden angeben, durch welche Merkmale sie ihr Hauptstudienfach gekennzeichnet sehen. Insgesamt zehn Items sind auf einer Skala von 0 = überhaupt nicht bis 6 = sehr stark zu bewerten. Nur in zwei Merkmalen unterscheiden sich die beiden Gruppen in ihrer Bewertung signifikant: bei der „Konkurrenz zwischen Studierenden“ sowie bei dem Merkmal „überfüllte Lehrveranstaltungen“.

Tabelle 3.14			
Charakterisierung des Hauptfachs nach Teilnehmerstatus (WS 2015/16)			
(Skala von 0 = überhaupt nicht bis 6 = sehr stark. Angaben = Mittelwerte)			
Wie stark ist Ihr Hauptstudienfach aus Ihrer Sicht charakterisiert durch ...	Abgebrochen	Beendet	t-Werte
hohe Leistungsnormen	4,35	4,26	0,995
gut gegliederten Studienaufbau	3,78	3,74	0,360
Klarheit der Prüfungsanforderungen	3,70	3,67	0,255
Forschungsbezug der Lehre	3,52	3,57	-0,525
engen Praxisbezug	3,29	3,36	-0,597
gute Berufsvorbereitung	2,82	2,85	-0,324
Konkurrenz zwischen den Studierenden	2,30	2,04	2,222*
gute Beziehungen zwischen Studierenden und Lehrenden	3,88	3,97	-0,853
überfüllte Lehrveranstaltungen	2,32	2,07	1,943*
internationale Ausrichtung	2,88	2,69	1,626

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
 p > 0,05 = nicht signifikant, p <= 0,05 = *, p <= 0,01 = **, p <= 0,001 = ***

Die Differenz zwischen den beiden Gruppen ist allerdings in beiden Fällen sehr gering. Teilnehmer, die die Befragung abbrechen, sehen ihr Hauptfach etwas häufiger durch Konkurrenz (Kategorien 0-2 zusammen 57 % zu 63 %) und Überfüllung (Kategorien 0-2 zusammen 54 % zu 62 %) geprägt als Teilnehmer, die die Befragung abschließen. In der Tendenz scheinen Abbrecher etwas kritischer in ihrer Bewertung ihres Hauptfaches zu sein als Teilnehmer, die die Befragung abgeschlossen haben.

Wird gleichzeitig die Hochschulart kontrolliert, so bleibt beim Merkmal „Konkurrenz“ der Effekt des Teilnehmerstatus (abgebrochen vs. abgeschlossen) bestehen, im Fall der überfüllten Lehrveranstaltungen verschwindet der Effekt hingegen.

Es lassen sich durchaus inhaltliche Unterschiede nach dem Teilnehmerstatus feststellen. Dabei zeigen Personen, die die Befragung vorzeitig abbrechen, eine niedrigere Studiersicherheit als Teilnehmer, die die Umfrage bis zum Ende bearbeiten. Sie stufen die Anforderungen Kritik an Lehrmeinungen zu äußern häufiger als überzogen ein, betonen einzelne extrinsische Motive der Fachwahl stärker und sehen sich etwas öfter mit Konkurrenz und überfüllten Lehrveranstaltungen konfrontiert.

Die vereinzelt auftretenden Differenzen zwischen beiden Gruppen sind nur sehr gering ausgeprägt und nicht systematisch, lassen also kein eindeutiges Muster erkennen. Um diese Unterschiede besser analysieren zu können, sollte in künftigen Befragungen der Versuch unternommen werden, mehr über Abbrecher zu erfahren. Entweder, indem vor dem endgültigem Verlassen der Umfrage den Personen noch einzelne Fragen gestellt werden, oder als kurze Follow-Up Befragung. Solche Follow-Up Befragungen sind aber nur möglich, wenn bspw. die E-Mail-Adressen der Abbrecher bekannt sind und für den erneuten Kontakt genutzt werden können. Dem stehen unter Umständen datenschutztechnische Bedenken entgegen, weshalb verschiedene technische Maßnahmen geprüft und gegeneinander abgewogen werden müssen.

4. Analyse von Parادات

In Befragungen entstehen neben den interessierenden Umfragedaten auch immer Daten durch den Umfrageprozess an sich (Felderer, Birg, Kreuter, 2014). Das können im Falle von Online-Befragungen zum Beispiel timestamps, also Zeitmarken sein, die den Wechsel des Befragten von einer Seite zur nächsten und damit die Verweildauer pro Seite messen. Aber auch Tastenanschläge oder Mausbewegungen können parallel zur eigentlichen Befragung aufgenommen werden. Zentral ist, dass Parادات keine direkte Eingabe von den Befragten erfordern (Kaczmarek & Neubarth, 2007; Heerwegh, 2003; Bosnjak & Tuten, 2001), sondern automatisch mitaufgezeichnet werden. Es lassen sich drei Arten von Parادات unterscheiden (Callegaro et. al, 2015): Parادات zur Kontaktinformation, zum Gerätetyp und zur Fragebogennavigation. Die folgenden Abschnitte konzentrieren sich vor allem auf Informationen zur Fragebogennavigation (Bearbeitungsdauer) sowie den verwendeten Gerätetyp.

Der größte Unterschied zwischen Umfrage- und Parادات liegt in der Reaktivität: Während Umfragedaten durch das Instrument bzw. die Messung beeinflusst werden können (vgl. Abschnitt 3.2), sind sich Befragte über die Aufzeichnung der Parادات nicht bewusst. Es handelt sich um „Verhaltensspuren aus Handlungsprozessen“ (Kaczmarek & Neubarth, 2007, S.294). Die Analyse von Abbrüchen gibt Hinweise auf problematische Fragen, die Bearbeitungsdauer zeigt den tatsächlichen zeitlichen Aufwand für die Befragung und liefert Hinweise darauf, wie lange der Fragebogen maximal sein darf. Informationen über verwendete Geräte zeigen, welche Technik bevorzugt wird und für welche Plattformen das Instrument entwickelt/optimiert werden muss. Parادات bieten daher eine interessante Erweiterung zur Verbesserung des Befragungsprozesses (Parادات zur Coverage-Forschung, vgl. Felderer et. al, 2014) sowie insbesondere zur Überarbeitung des Befragungsinstruments. Allerdings ist sowohl die Erhebung als auch die Analyse von Parادات nicht ohne Probleme; insbesondere Fragen des Datenschutzes müssen berücksichtigt werden, denn je nach Art der erhobenen Informationen handelt es sich um personenbezogene Daten (IP-Adresse⁸). Parادات unterliegen dann den gleichen datenschutzrechtlichen Vorschriften wie andere personenbezogene Daten und müssen daher mit derselben Sorgfalt behandelt werden (vgl. Kapitel 9).

4.1 Bearbeitungsdauer

Die Bearbeitungsdauer kann wichtige Informationen über die Bearbeitungsqualität der Befragten liefern und gibt Aufschluss darüber, ob die in den Pretests ermittelte durchschnittliche Bearbeitungsdauer im Feld zutrifft oder ob die Befragten tatsächlich mehr bzw. weniger Zeit benötigen. Angaben über die Bearbeitungsdauer liegen nur für die Online-Befragung vor, da hier Zeitstempel verwendet wurden. In der schriftlichen Befragung wurde nach den Erfahrungen in Pretests und Rückmeldungen vergangener Befragungen als Richtwert 60 Minuten für die Bearbeitung angesetzt (die Teilnehmer im Pretest benötigten zwischen 50 und 120 Minuten).

Lange Fragebogen können Abbrüche provozieren und die Antwortqualität negativ beeinflussen („satisficing“). In der einschlägigen Literatur wird für Online-Befragungen daher häufig empfohlen, die Bearbeitungsdauer unter 20 Minuten zu halten (für eine knappe Zusammenfassung: Callegaro, Manfreda, Vehovar, 2015, S. 101 f.). Die maximale Bearbeitungsdauer hängt aber auch von der Motivation der Befragten ab (Incentives, möglicher Impact einer Teilnahme) oder deren Thematik (ist das Thema relevant für die Befragten, steigt auch die Motivation einer Teilnahme). Zudem handelt es sich bei der untersuchten Gruppe um Personen mit hohem Bildungsniveau und hohen kognitiven Kapazitäten, die unter Umständen auch eine längere Bearbeitungsdauer „aushalten“. Callegaro et. al

⁸ Die verwendete Befragungssoftware des Zofar-Teams des DZHW speichert nicht die IP-Adresse der Befragten. Es werden auch keine anderen personenbezogenen technischen Daten erhoben bzw. gespeichert.

verweisen auf den European Social Survey, welcher ebenfalls einen sehr umfangreichen Online-Fragebogen nutzt und Bearbeitungszeiten von bis zu einer Stunde verzeichnet, ohne dabei an Qualität einzubüßen (Callegaro et. al, 2015, S. 101; Ainsaar et al., 2013).

4.1.1 Bearbeitungsdauer in den Gruppen der Online-Befragung

Um für die online befragten Personen realistische Schätzungen der Bearbeitungsdauer zu erhalten, werden nur Personen mit verwertbaren Angaben (mindesten 50% der Angaben) berücksichtigt, um eine Verzerrung durch „Lurker“ zu vermeiden.

Ohne weitere Einschränkungen, bspw. nach vorzeitigem oder erfolgreichem Abschluss, benötigen Personen in der randomisierten Online-Befragung mit Motivationsseiten durchschnittlich 65 Minuten (Median: 53 Minuten), in der Gruppe ohne Motivation waren es durchschnittlich 62 Minuten (Median: 54 Minuten). Teilnehmer in der Wahlgruppe sind mit 61 Minuten (Median = 48 Minuten) geringfügig schneller. Damit liegt die durchschnittliche Bearbeitungsdauer der Web-Befragung nah an den bisher ermittelten Angaben für die schriftliche Erhebung.

Um eine noch präzisere Schätzung der realen Bearbeitungsdauer zu erhalten, sollten allerdings nur Personen betrachtet werden, die den Fragebogen am selben Tag bearbeitet und erfolgreich abgeschlossen haben, da eine Unterbrechung die Aufzeichnung der Bearbeitungsdauer verfälscht⁹ (vgl. Tabelle 4.1). Hier verringert sich die Bearbeitungsdauer in der Wahlgruppe (Gruppe 2) auf 55 Minuten (Median = 46 Minuten), in der randomisierten Onlinegruppe mit Motivationsseiten (Gruppe 3) auf 58 Minuten (Median = 46 Minuten) und in der Gruppe ohne Motivationsseiten auf 59 Minuten (Median = 51 Minuten).

Tabelle 4.1 Bearbeitungsdauer in den Gruppen der Online-Befragung (min. 50% der Angaben und am selben Tag abgeschlossen) (WS 2015/16) (Mittelwerte und Mediane)			
	Gruppe 2	Gruppe 3	Gruppe 4
Mittelwert	54,59	58,39	58,62
Median	46,12	48,73	50,69

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Ein Kruskal-Wallis-Test wurde durchgeführt um zu untersuchen, ob sich die drei Gruppen der Online-Befragung voneinander unterscheiden. Hier konnten keine signifikanten Unterschiede festgestellt werden: $\chi^2(2) = 2,267, p = 0,3220$.

⁹ Diese Entscheidung hat zwei Gründe:

1) Beim erstmaligen Einloggen in die Umfrage wird die Uhrzeit registriert. Unterbricht eine Person die Umfrage und loggt sich an einem anderen Tag wieder ein, wird die Uhrzeit nicht erneut gespeichert. Für die Berechnung der Bearbeitungsdauer wird die Verweildauer pro Fragebogenseite aufsummiert. Bei Personen, die die Umfrage unterbrochen haben, wird die Dauer der Abwesenheit und das Timeout auf der Ausstiegsseite (Timeout automatisch nach 30 Minuten) miteinbezogen, weshalb die Bearbeitungsdauer für alle Personen, die den Fragebogen nicht in einer Sitzung bearbeitet haben, verzerrt ist.

2) Selbst wenn die Dauer zwischen Erstkontakt und erneutem Login korrigiert wird, was nur annäherungsweise möglich ist, kann die Bearbeitungsdauer dieser Gruppe von jenen Personen abweichen, die die Befragung in einer Sitzung abgeschlossen haben, weil sie bspw. nach der Unterbrechung erholter sind oder sich in einer vorteilhafteren Befragungssituation (ruhig, keine Ablenkung, etc.) befinden.

Wird die Befragung vorzeitig abgebrochen¹⁰, beträgt die durchschnittliche Bearbeitungsdauer in der randomisierten Onlinegruppe mit Motivationsseiten 16 Minuten (Median = 11 Minuten) in der Wahlgruppe 17 Minuten (Median = 11 Minuten) und in der randomisierten Onlinegruppe ohne Motivationsseiten 24 Minuten (Median = 15 Minuten). Werden die Gruppen gemeinsam betrachtet, liegt die durchschnittliche Bearbeitungsdauer bei 20 Minuten (Median = 13 Minuten). Damit unterstützen die Ergebnisse gängige Vorgaben zur Fragebogenlänge in Web-Surveys, wonach ab einer Befragungsdauer von 20 Minuten das Abbruchrisiko ansteigt.

4.1.2 Item Non-Response und Bearbeitungsdauer

Aufschlussreich ist die Prüfung, wie die Bearbeitungsdauer mit der Anzahl fehlender Angaben zusammenhängt. Geht eine zügige Bearbeitung zu Lasten der Antwortqualität? Zu diesem Zweck wird der Item Non-Response zwischen Teilnehmern verglichen, die entweder mehr oder weniger als die durchschnittliche Bearbeitungsdauer für die Beantwortung der Fragen benötigt haben. Dabei werden nur Teilnehmer berücksichtigt, die die Befragung am selben Tag erfolgreich abgeschlossen haben¹¹. Es erfolgte keine Einschränkung nach Item Non-Response. Die durchschnittliche Bearbeitungsdauer liegt in der Wahlgruppe bei 54 Minuten (Median: 46 Minuten), in Gruppe 3 und 4 bei 58 bzw. 59 Minuten (Median: 49 und 50 Minuten).

Tabelle 4.2 Durchschnittlicher Item Non-Response in Abhängigkeit von der Bearbeitungsdauer in den Gruppen der Online-Befragung – Befragung am selben Tag erfolgreich abgeschlossen (WS 2015/16) (Häufigkeiten)				
	IN für $t < \bar{t}$	IN für $t \geq \bar{t}$	\bar{t} IN in Gruppe	Korrelation zwischen IN und Bearbeitungsdauer (ρ)
Wahlgruppe	14	8	12	- 0,1161 (n.s.)
Online randomisiert mit Motivationsseiten	15	8	13	- 0,0560 (n.s.)
Online randomisiert ohne Motivationsseiten	18	7	14	- 0,1963 (***)

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
 $p > 0,05$ = nicht signifikant, $p \leq 0,05$ = *, $p \leq 0,01$ = **, $p \leq 0,001$ = ***

Obwohl es in der Wahlgruppe deutliche Unterschiede im durchschnittlichen IN zwischen den Teilnehmern gibt, ist die Differenz nicht signifikant: Im Schnitt fehlen Personen der Wahlgruppe zwölf Angaben. Personen, die weniger als die mittlere Bearbeitungsdauer benötigt haben, weisen 14 fehlende Werte auf. Teilnehmer, die im Schnitt 54 Minuten oder länger benötigen, weisen hingegen nur acht fehlende Angaben auf. Die Korrelation nach Spearman weist auf einen negativen Zusammenhang hin, wonach mit steigender Bearbeitungsdauer die Anzahl fehlender Werte abnimmt. Der Zusammenhang ist allerdings nicht signifikant.

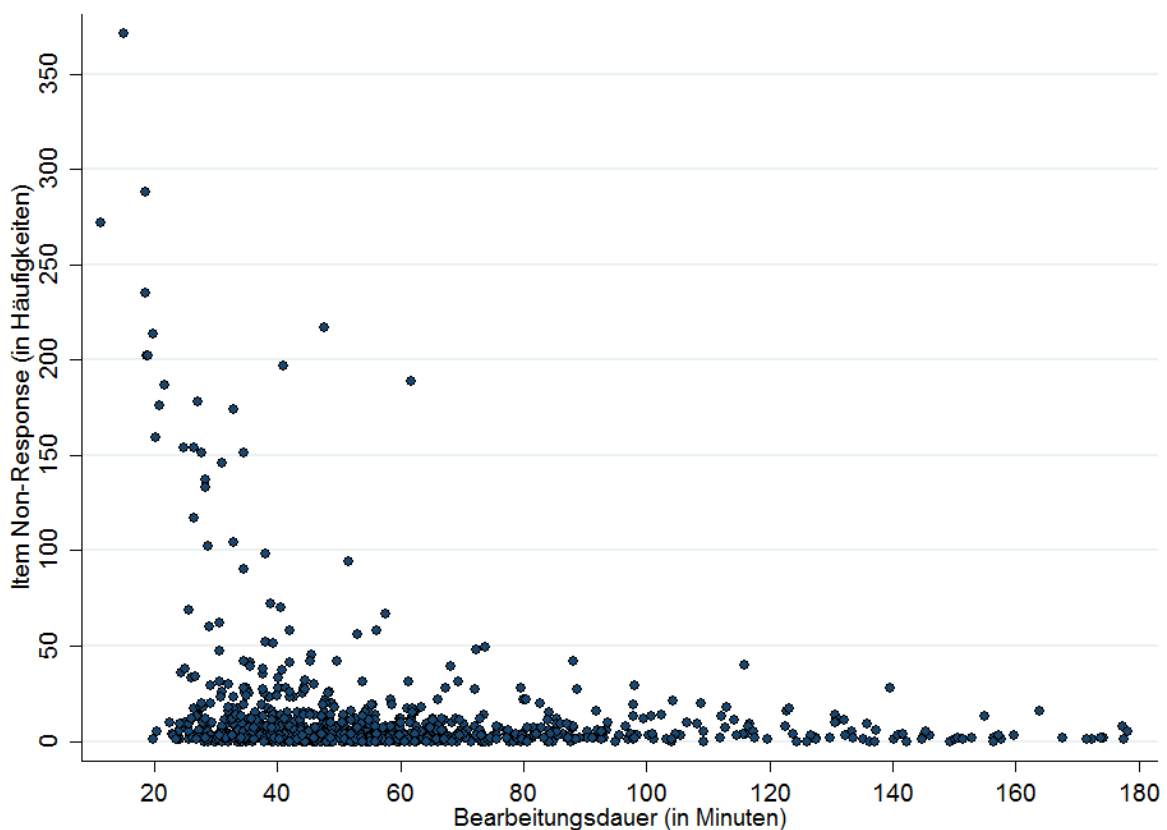
Das gleiche Ergebnis zeigt sich in Gruppe 3, die noch zusätzlich Motivationsseiten angezeigt bekam. Hier fehlen den Befragten durchschnittlich 13 Angaben. Personen, die den Fragebogen schneller als im Schnitt bearbeitet haben, fehlen 15 Angaben; 7 mehr als den restlichen Teilnehmern. Auch hier zeigt sich ein negativer Zusammenhang, der nicht signifikant ist.

¹⁰ In der Gruppe der Abbrecher wurde nicht nach dem Anteil verwertbarer Angaben kontrolliert.

¹¹ Abbrecher würden zum einen die Bearbeitungsdauer nach unten verzerren. Zum anderen interessiert nur die Antwortqualität der Personen, die die Umfrage regulär abgeschlossen haben und ob eine kurze Bearbeitungsdauer mit einem erhöhtem Item Non-Response zusammenhängt.

In Gruppe 4 wird der Zusammenhang erstmals signifikant ($-0,1963$, $p < 0,001$). Hier weisen Teilnehmer, die weniger Zeit benötigt haben als der Durchschnitt, deutlich mehr fehlende Werte auf als Befragte, die die durchschnittliche Dauer oder mehr Zeit benötigt haben (18 zu 7). Grundsätzlich lässt sich also festhalten, dass ein negativer Zusammenhang zwischen dem IN und der Bearbeitungsdauer besteht (vgl. Abbildung 4.1), schnelle Befragte also auch mehr fehlende Werte produzieren. Dieser Zusammenhang ist allerdings nur in Gruppe 4 signifikant.

Abbildung 4.1
Item Non-Response in Abhängigkeit von der Bearbeitungsdauer – nur Teilnehmer, die die Befragung am selben Tag abgeschlossen haben (WS 2015/16)
 (Angaben Item Non-Response = Häufigkeiten, Bearbeitungsdauer in Minuten)



Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Offensichtlich gibt es zwei Arten von Befragten: 1) Teilnehmer, die trotz der Länge des Fragebogens diesen gewissenhaft bis zu Ende ausfüllen, und 2) Teilnehmer, die durch die Länge des Instruments frustriert sind, nur unvollständige Angaben produzieren und häufiger Fragen/Items auslassen (Lurker). Mit der Kombination aus Item Non-Response und Bearbeitungsdauer lassen sich diese Teilnehmer identifizieren, markieren und gegebenenfalls von der weiteren Untersuchung ausschließen.

4.1.3 Dauer zwischen erstmaligem Zugriff und Abschluss/ Abbruch der Befragung

Auch hier beschränkt sich die Betrachtung wieder auf Personen, die die Befragung erfolgreich abgeschlossen haben. Generell bearbeiten die meisten Studierenden den Fragebogen nach dem ersten Zugriff noch am selben Tag: In der randomisierten Onlinegruppe sind es 84 % (Gruppe 4) bzw. 85 % (Gruppe 3) der Teilnehmer, in der Wahlgruppe 87 %. Weitere elf Prozent in beiden Gruppen beenden die Befragung innerhalb einer Woche nach ihrem ersten Login (vgl. Tabelle 4.3). Danach erfolgen nur

noch vereinzelte Zugriffe. Insbesondere in der randomisierten Onlinegruppe loggen sich Personen auch noch nach drei Wochen wieder in den Web-Survey ein (eine genauere Übersicht über die Dauer seit dem ersten Login ist Tabelle A1 im Anhang zu entnehmen). Über alle Gruppen der Online-Befragung hinweg machen 15 % von der Möglichkeit eines erneuten Logins Gebrauch.

Tabelle 4.3			
Dauer seit erstem Login – nur Personen, die die Befragung erfolgreich abgeschlossen haben (WS 2015/16)			
(Häufigkeiten und Prozente in Klammern)			
Dauer zwischen erstem Login und Abschluss	Wahlgruppe Online	Rand. Online mit Motivation	Rand. Online ohne Motivation
Am selben Tag	193 (87)	360 (85)	373 (84)
1. Woche (1 – 7 Tage)	24 (11)	46 (11)	48 (11)
2. Woche (8 – 14 Tage)	2 (1)	4 (1)	98 (2)
3. Woche (15 – 21 Tage)	0 (0)	5 (1)	4 (1)
4. – 12. Woche (22 – 90 Tage)	2 (1)	10 (2)	9 (2)

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Für Teilnehmer, die die Befragung abbrechen, gestaltet sich die Situation noch eindeutiger. Über alle Gruppen der Online-Befragung hinweg beenden 93 % die Umfrage noch am selben Tag und loggen sich nicht erneut ein. Das bedeutet, dass insgesamt sieben Prozent (28 Personen) noch einmal die Befragung aufrufen und erst danach abbrechen. Bei der Betrachtung der zuletzt angezeigten Seiten dieser Personen zeigt sich, dass sie an verschiedenen Stellen im Fragebogen abbrechen und zum Teil erstaunlich viele Fragen beantworten, bevor sie die Umfrage endgültig beenden. Knapp mehr als die Hälfte dieser Personen steigt erst ab Frage 51 aus, also erst nachdem sie bereits über die Hälfte des Fragebogens ausgefüllt haben. Der späteste Abbruch erfolgte bei Frage 76, also bei etwa 82 % des Fragebogens.

Das erneute Einloggen in den Web-Survey zeugt von einem starken Interesse an der Befragung und einer hohen Motivation der Befragten. Auch wenn insgesamt betrachtet nur ein kleiner Teil der Befragten diese Möglichkeit nutzt (15 % der Teilnehmer¹²), handelt es sich um eine wichtige, den Rücklauf fördernde Maßnahme.

4.2 Devices

Onlineumfragen stellen andere Anforderungen an das Design des Untersuchungsinstruments als schriftliche Befragungen. Der Onlinefragebogen des 13. Studierendensurveys ist noch nicht vollständig für die Web-Darstellung optimiert und wurde auch nicht für mobile Endgeräte programmiert. Das bedeutet, dass die Darstellung des Fragebogens nur für den Gebrauch an einem Desktop-PC ausgelegt ist. Zentral für die weitere Optimierung des Fragebogens und künftige Erhebungen ist daher die Information, wie viele Personen in der aktuellen Erhebung welches Gerät genutzt haben.

Hierzu wird der User Agent String (UAS) der Teilnehmer ausgelesen, welcher von der Software aufgezeichnet wird¹³. Die Informationen des UAS werden mittels des von Rossmann und Gummer entwi-

¹² Ohne Abbrecher. Mit Abbrecher: 13 %.

¹³ IP-Adressen werden nicht aufgezeichnet bzw. gespeichert.

ckelten Ado-File für die Statistik-Software Stata ausgelesen und in separaten Variablen für die verwendeten Geräte gespeichert¹⁴. In einem ersten Schritt wird für alle Teilnehmer der Online-Befragung (Wahlgruppe, Lang- und Kurzfragebogen, also Gruppe 2 – 9) die Gerätenutzung berichtet. In einem zweiten Schritt werden dann einzelne Merkmale der Datenqualität näher betrachtet und zwischen den Nutzern unterschiedlicher Geräte verglichen. In der folgenden Darstellung werden Device-Switcher, also Teilnehmer, die während der Befragung das Gerät gewechselt haben, zunächst nicht näher berücksichtigt, da sie eine eigene Untersuchungsgruppe bilden und ihre ausführliche Analyse den Rahmen dieses Berichts sprengen würde. In insgesamt 260 Fällen müsste geprüft werden, ob es sich um Device-Switcher handelt, oder um doppelte, aber fehlerhafte user agent strings.

4.2.1 Verwendete Geräte in der Online-Befragung

Im Einladungsschreiben wurden die Befragten darauf hingewiesen, dass der Fragebogen nicht für die Bearbeitung auf Smartphones oder Tablets ausgelegt ist, weshalb nur mit einer geringen Nutzung dieser Geräte zu rechnen ist. Wie die Ergebnisse in Tabelle 4.4 zeigen, ist diese Vermutung korrekt: nur 84 Personen (3 %) haben die Befragung mit einem Smartphone durchgeführt, 154 (5 %) mit einem Tablet. Teilnehmer in der Gruppe mit langem Fragebogen verwenden häufiger Tablets ($p \leq 0.000$) und Smartphones ($p \leq 0.000$) als Teilnehmer, die den Kurzfragebogen bearbeitet haben.

Tabelle 4.4				
Von den Teilnehmern verwendete Geräte (WS 2015/16)				
(Häufigkeiten und Prozente in Klammern)				
	Gruppe 2	Langer FB Gruppe 3, 4, 9*	Kurzer FB Gruppe 4-8	Insgesamt
PC	242 (93)	1.389 (88)	1.112 (97)	2743 (92)
Tablet	12 (5)	114 (7)	28 (2)	154 (5)
Smartphone	6 (2)	69 (4)	9 (1)	84 (3)
Insgesamt	260	1572	1149	2981

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
* Differenz zu 100% aufgrund von Rundungsfehlern

Der Vergleich zur Wahlgruppe ist aufgrund der geringen Zellbesetzung unsicher; hier zeigen sich keine signifikanten Unterschiede in der Nutzung der Geräte. Allerdings ist ein Großteil der Tablet- und Smartphone-Nutzer in Gruppe 9 zu finden: 82 Personen haben für die Bearbeitung des Fragebogens ein Tablet und 53 ein Smartphone gewählt. Das entspricht 53 % (82/154) bzw. 63 % (53/84) der jeweiligen Nutzergruppe. Gruppe 9 besteht nur aus einer einzigen Hochschule, an der die Studierenden die Einladung zur Studie per E-Mail erhielten. Vieles deutet auf Probleme innerhalb dieser Gruppe hin: Es haben bspw. überdurchschnittlich viele Teilnehmer die Befragung abgebrochen (63 % von jenen, die auf die Befragung zugegriffen haben). Trotz des Hinweises in der Einladung wählten hier viele Befragte ein anderes Gerät, was zumindest teilweise den erhöhten Item Non-Response und die vielen Abbrüche erklären könnte.

¹⁴ Roßmann, Joss, Tobias Gummer. 2016. *PARSEUAS: Stata Module to Extract Detailed Information From User Agent Strings (Version: 1.3): [Computer Software]*. Chestnut Hill, MA: Boston College. <http://econpapers.repec.org/software/bocbocode/S457937.htm>

Abbruch nach verwendeten Geräten

Studierende, die den Fragebogen an einem Desktop-PC ausgefüllt haben, schließen die Befragung deutlich häufiger erfolgreich ab (73 %). Weder das Tablet noch das Smartphone scheinen für die Bearbeitung geeignet zu sein, denn 60 % bzw. 75 % der Teilnehmer brechen hier ab. Immerhin: 21 Studierende haben die Umfrage mit dem Smartphone erfolgreich abgeschlossen; unter den Studierenden, die ein Tablet genutzt haben, sind es 61 Personen. Diese Gruppe ist zu klein um wirklich aussagekräftige Ergebnisse berechnen zu können. Allerdings liefert ein Blick auf verschiedene Merkmale der Datenqualität in den Gruppen erste Hinweise auf potentielle Unterschiede zur Bearbeitung an einem Desktop-PC. Die folgenden kurzen Analysen beschränken sich auf Befragte, die den langen Fragebogen bearbeitet haben.

Item Non-Response, Straightlining und Bearbeitungsdauer nach verwendetem Device

Item Non-Response: Aufgrund der sehr kleinen Fallzahlen in der Gruppe mit Smartphone und Tablet-Nutzung sollte die Stichprobe nicht noch weiter eingeschränkt werden, weshalb die in Abschnitt 3.2.1 verwendete restriktive Definition des Unit Non-Response nicht verwendet werden kann. Zudem wird bei dieser Analyse die Gruppe 9 mit einbezogen. Der IN ist daher sehr stark durch Abbrüche und Lurker verzerrt. Zudem ist die Anzahl der Personen in den zu vergleichenden Gruppen sehr unterschiedlich, weshalb die Aussagekraft der berichteten Tests eingeschränkt ist (was auch für die weiteren Analysen gilt).

Straightlining: Als Straightlining wird ein Antwortverhalten bezeichnet, bei dem der Befragte dieselbe Antwortkategorie für alle Items einer Frage wählt. Um zu untersuchen, ob es aufgrund von Darstellungsunterschieden zu straightlining kommt, werden mehrere Fragen betrachtet, die an verschiedenen Stellen im Fragebogen beantwortet werden sollten. Denn es ist denkbar, dass das straightlining eine Art Erschöpfungserscheinung ist, die erst im fortgeschrittenen Befragungsverlauf auftritt. Es werden die Fragen nach Gründen für die Fachwahl (Frage 8), nach dem Nutzen eines Studiums (Frage 14), nach den bisherigen Erfahrungen im Studium (Frage 56), und schließlich nach den Berufswerten betrachtet (Frage 68).

Bearbeitungsdauer: Weil die Befragung nicht auf mobile Geräte ausgelegt ist, weicht die Bearbeitungsdauer dieser Gruppe sehr wahrscheinlich von jener der PC-Nutzer ab. Die Richtung ist allerdings unklar: Teilnehmer, die mobile Geräte nutzen, können aufgrund der umständlicheren Handhabung entweder mehr Zeit benötigen; sie könnten aber auch dazu übergehen, die Befragung nicht ernsthaft, sondern nur flüchtig zu bearbeiten, und die Befragung so schneller beenden.

Für die Analyse intervallskalierter Maße (Item Non-Response, Bearbeitungsdauer) wird eine einfaktorielle Varianzanalyse bzw. ein Kruskal-Wallis-Test berechnet. Für das straightlining wird eine Variable berechnet, die den Wert 1 annimmt, wenn ein Befragter in allen Items einer Frage dieselbe Antwort gegeben hat und 0, wenn seine Antworten variieren. Anschließend werden die Zusammenhänge mit Chi-Quadrat Tests überprüft.

Tablet- und Smartphone-Nutzer haben im Vergleich zu PC-Nutzern einen deutlich erhöhten IN. Auch wenn hier alle Befragten enthalten sind, die die Befragung abgebrochen oder nicht ernsthaft bearbeitet haben, sind die Unterschiede deutlich. Smartphone-Nutzer produzieren durchschnittlich bis zu 73 % fehlende Angaben, während PC-Nutzer nur 30 % fehlende Werte aufweisen. Teilnehmer, die ein Tablet genutzt haben, liegen mit 62 % IN dazwischen. Auch die Varianzanalyse bestätigt einen signifikanten Unterschied zwischen den Gerätetypen, allerdings nur zwischen PC-Nutzern und mobilen Endgeräten, nicht zwischen Tablet- und Smartphone-Nutzern.

Tabelle 4.5				
Merkmale der Datenqualität nach verwendetem Device (WS 2015/16)¹⁾				
(IN: Angaben in Häufigkeiten, Prozente in Klammern; Straightlining: Angaben in Prozent; Bearbeitungsdauer: Angaben in Minuten)				
	PC	Tablet	Smart-phone	ANOVA/ Chi-Quadrat
Durchschnittliche Anzahl an IN ¹	112 (30)	232 (62)	276 (73)	F(2, 1.829) = 79,64 P<0.001, $\eta^2 = 0,080$
Straightlining				
Frage 8	0,40	1,89	0,95	Fisher's exact = 0,157
Frage 14	0,97	3,23	4,69	Fisher's exact = 1.000
Frage 56	3.15	9,09	5,97	Fisher's exact = 0,073
Frage 68	0,74	-	3,23	Fisher's exact = 0,136
Durchschnittliche Bearbeitungsdauer	47,9 min	38,2 min	25,6 min	F(2, 1.692) = 12,49, p<0.001, $\eta^2 = 0,028$
Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz				
1) Nur Gruppen 2, 3, 4 und 9 (langer Fragebogen)				

Eine Einschränkung auf Personen, die die Befragung erfolgreich abgeschlossen haben, reduziert die Anzahl gültiger Fälle noch einmal deutlich (PC 1.050, Smartphone 15, Tablet 38). Auch nach der Einschränkung bestehen deutliche Unterschiede im Item Non-Response zwischen den drei Gruppen ($F(2, 1.100) = 15.62, p = 0,000$): PC-Nutzer weisen nur noch durchschnittlich 12 fehlende Angaben auf, Tablet-Nutzer schon 24 und Smartphone-Nutzer 53.

Die Vergleiche zwischen den drei Gruppen zeigen signifikante Differenzen zwischen PC-Nutzern und Tablet- bzw. Smartphone-Nutzern. Außerdem unterscheiden sich auch die Nutzer mobiler Endgeräte signifikant voneinander in der Anzahl fehlender Werte (Tablet- versus Smartphone-Nutzer).

Das straightlining scheint etwas häufiger bei Nutzern von Tablets und Smartphones vorzukommen: Bei Frage 8, Frage 14, Frage 56 und Frage 68 zeigt sich das Antwortmuster öfter in der Gruppe der mobilen Endgeräte. Allerdings sind es nicht immer die Smartphone-Nutzer, die hauptsächlich auf dieses Antwortmuster zurückgreifen. Teilnehmer, die ein Tablet genutzt haben, weisen bei Frage 56 bspw. den höchsten Anteil an straightlining auf: neun Prozent haben hier immer dieselbe Antwortkategorie gewählt. In der Gruppe mit PC- bzw. Laptop-Nutzung liegt der Anteil mit drei Prozent deutlich niedriger.

Bei der Betrachtung ist allerdings zu beachten, dass die Smartphone- und Tablet-Nutzer die Befragung häufiger abbrechen, weshalb nur sehr wenige Befragte überhaupt auf straightlining hin untersucht werden können. Mit weiterem Fortschritt im Fragebogen stehen dann immer weniger Nutzer mobiler Endgeräte zur Verfügung, was die Aussagekraft des Vergleichs zusätzlich einschränkt. In der Tendenz scheint straightlining als Antwortmuster dennoch ein größeres Problem der Tablet- und Smartphone-Nutzer zu sein.

Die Angaben zur Bearbeitungsdauer stützen die Annahme, dass das unangepasste Design des Fragebogens zu einer hastigeren und weniger sorgfältigen Bearbeitung führt. Smartphone-Nutzer benötigen am wenigsten Zeit: Nach ungefähr 26 Minuten beenden sie die Umfrage bzw. brechen diese ab. PC- Nutzer benötigen fast doppelt so lange (48 Minuten) und Tablet-Nutzer liegen mit 38 Minuten durchschnittlicher Bearbeitungszeit dazwischen. Die anschließenden Vergleichstests zeigen signifikante Unterschiede zwischen PC-Nutzern und Personen, die mobile Endgeräte nutzen, aber nicht zwischen Tablet- und Smartphone-Nutzern.

Die Einschränkung auf Personen, die die Befragung erfolgreich abgeschlossen haben, ändert das Bild leicht: PC-Nutzer schließen die Umfrage nach durchschnittlich 61 Minuten ab, Tablet-Nutzer benötigen fast genauso viel Zeit (62 Minuten), und Smartphone-Nutzer schließen noch immer deutlich schneller die Befragung ab (55 Minuten). Die Unterschiede sind nicht mehr signifikant.

Auch wenn diese kurze Untersuchung der Gerätetypen noch wenig elaboriert ist, liefert sie doch erste wichtige Hinweise für zukünftige Befragungen. Trotz des Hinweises, dass die Befragung nicht für die Bearbeitung an mobilen Geräten ausgelegt ist, ruft ein kleiner Teil der Studierenden den Web-Survey dennoch über ein Tablet oder ein Smartphone auf. Bei einer Zielgruppe mit einer vergleichsweise hohen Medienkompetenz, ist eine Beschränkung auf Desktop-PCs oder Laptops zu eng und nicht länger zeitgemäß. Kommende Erhebungen müssen daher ein responsives Design aufweisen, um den unterschiedlichen Gerätetypen zu entsprechen und die user experience zu verbessern.

Zudem gibt es Hinweise darauf, dass die Nutzung mobiler Geräte mit neuen Problemen wie zum Beispiel einem hohen Item Non-Response und unerwünschte Antworttendenzen verbunden ist. Allerdings kann zum jetzigen Zeitpunkt nicht ausgeschlossen werden, dass es sich hierbei um Effekte des nicht-responsiven Designs handelt. Bei einer adäquaten Adaption des Befragungsinstruments auf andere Plattformen/Geräte können solche Probleme eventuell eingegrenzt oder komplett vermieden werden.

5. Strukturvergleiche zwischen Untergruppen

In diesem Kapitel sollen die verschiedenen Untergruppen (vgl. Kapitel 1-2) miteinander verglichen werden. Dabei werden zum einen Mode- und Selektionseffekte untersucht, zum anderen Strukturvergleiche und Kennwertprüfungen durchgeführt.

Der Mode-Effekt wird geprüft, indem Gruppen mit unterschiedlichen Erhebungsarten verglichen werden (Papier vs. Online). Der Selbstselektionseffekt wird geprüft, indem die randomisierten Gruppen (Papier bzw. Online) mit der jeweiligen selbstselegierten Wahlgruppe verglichen werden. In Kombination beider Modi können Interaktionseffekte geprüft werden.

Für weitere Untersuchungen ließen sich auch Vergleiche zwischen den Online-lang und Online-kurz Varianten durchführen sowie zu der Spezialgruppe, die auch online eingeladen wurde (Gruppe 9). Diese Vergleiche werden ausführlich erst in einem späteren Bericht durchgeführt werden, da hier vorab inhaltliche Analysen notwendig werden.

Die Art der Prüfungen zwischen den Designgruppen bezieht sich hauptsächlich auf zwei Arten von Vergleichen: zum einen der Vergleich der Stichprobenstruktur nach ausgewählten personenbezogenen Merkmalen mit Daten des statistischen Bundesamtes, zum anderen der Vergleich von Kennwerten aus Items des Studierendensurveys zwischen den Designgruppen.

Strukturvergleiche der resultierenden Designstichproben

Die Struktur der Stichproben wird über die Verteilung wichtiger soziodemographischer und indikativer Merkmale unter den Studierenden erfasst. Der Vergleich dieser Verteilungen und Kennwerte erlaubt Aussagen darüber, wie ähnlich oder unähnlich sich die Stichproben untereinander und zu der Grundgesamtheit sind.

Für die Vergleiche werden fünf der zehn Designgruppen verwendet, wobei die beiden Online-lang-Gruppen zusammengefasst werden. Die Gruppen der Kurzversionen sowie die Gruppe 9 werden in diese Analysen nicht einbezogen.

	Design	Gesamt	Uni	FH
Gruppe 0	Papier	4.959	3.885	1.074
Gruppe 1	Wahl Papier	661	508	153
Gruppe 2	Wahl online	273	201	70
Gruppe 3,4	Online lang	1160	847	320
Gesamt		7.053	5.441	1.617

Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

In den folgenden Tabellen werden verschiedene Kennwerte berechnet und angezeigt. Für die Verteilungen der nominalen und der Klassifikationsvariablen werden Chi-Quadrat-Tests auf Unabhängigkeit (χ^2) berechnet. Dazu werden die zugehörigen Freiheitsgrade (df) und die zugehörige Überschreitungswahrscheinlichkeit (p) angegeben, die besagt, wie wahrscheinlich die Nullhypothese der Gleichheit der Verteilungen ist. Dabei werden für die nominalskalierten Variablen nur drei Angaben gemacht: ** für signifikante Ergebnisse auf dem 1 %-Niveau, * für signifikante Ergebnisse auf den 5 %-Niveau, n.s. für nicht signifikante Ergebnisse. Für intervallskalierte Variablen werden die p-Werte angegeben. Zusätzlich werden für die signifikanten Ergebnisse noch die Effektstärken berechnet. Als Maß für die Größe der Effektstärke wird Cramers V angegeben, das zwischen 0 und 1 rangiert, und

bei dem ein Wert von 0.1 als klein, ein Wert von 0.3 als mittel und ein Wert von 0.5 als groß gilt (vgl. Cohen 1988).

Für intervallskalierte Items werden die Mittelwerte und t-Werte angegeben, sowie das Ergebnis der Varianzgleichheit und weitere Effektgrößen. Dabei ist Eta der Determinationskoeffizient der Varianzanalyse, der angibt, wieviel Variation das Modell aufklären kann.

5.1 Papier vs. Onlinegruppe (Mode-Effekt) bei Fremdselektion

Zuerst soll der Vergleich der beiden Erhebungsarten durchgeführt werden. Dabei wird die Struktur jener Studierenden, die nur einen Papierfragebogen zugeschickt bekamen, verglichen mit denen, die nur einen Onlinezugang erhielten. Um die Bedingungen möglichst konstant zu halten, werden nur die Gruppen mit der Langversion einbezogen. Es wird also der Mode-Effekt bei Fremdselektion betrachtet (vgl. Tabelle 5.2 und 5.3).

Auffällige Befunde ergeben sich für das Alter, das Geschlecht, die Hochschulart, die Abschlussart und den Stand der Berufstätigkeit vor dem Studium:

- Studierende der Onlinegruppe sind im Vergleich zur Papiergruppe im Schnitt etwas älter (25 zu 24,2 Jahre);
- In der Onlinegruppe ist der Frauenanteil geringer als in der Papiergruppe (56 % zu 62 %);
- Studierende der Onlinegruppe studieren häufiger an Fachhochschulen (27 % zu 22 %);
- Studierende der Onlinegruppe sind häufiger im Bachelor- (51 % zu 48 %) und seltener im Masterstudium (23 % zu 27 %);
- Studierende der Onlinegruppe waren häufiger vor dem Studium berufstätig (30 % zu 26 %);
- Gleichzeitig haben sie etwas häufiger eine Berufsausbildung vor dem Studium absolviert (25 % zu 22 %).

16 weitere geprüfte Merkmale zeigen keine Unterschiede zwischen den beiden Gruppen auf.

Werden die Effekte der gefundenen Unterschiede betrachtet, so stellt sich heraus, dass alle Effekte sehr gering sind. Kein Wert der Effektgröße (Cramers V) steigt über 0.1. Für das Alter ist ein Effekt von 0,29 (Eta) auszumachen, was bedeutet, dass nur 0,29 % der Datenvarianz durch das Modell (die Gruppenzugehörigkeit) erklärt werden kann. Auch der Blick auf die Größenordnung der Unterschiede verweist darauf, dass es sich um eher kleine Differenzen handelt, die nur für das Geschlecht mehr als fünf Prozentpunkte betragen.

Der Effekt des Altersunterschieds findet sich auch unabhängig von den Designgruppen zwischen den Hochschularten und zwischen den Geschlechtern. In Kombination mit dem Modus bleibt der Effekt erhalten, außer bei den Frauen an Universitäten. Diese unterscheiden sich im Alter nicht in den beiden Designgruppen. An den Fachhochschulen sind die Altersunterschiede zwischen den Designgruppen größer als an Universitäten.

Tabelle 5.2
Vergleich der vorgegebene Erhebungsarten, Papier vs. Online lang (Teil 1)
(Angaben in Prozent)

Merkmale	Papier (Grp. 0)	Online lang (Grp. 3,4)	df	Chi²	p¹⁾	Cramers V
Alter						
<= 21	27	25				
22-23	23	22				
24-25	21	19	5	26,3	**	0,068
26-27	14	12				
28-29	7	8				
=> 30	8	14				
Geschlecht						
männlich	38	44	1	10,7	**	0,043
weiblich	62	56				
Region						
alte Länder	71	71	1	0,31	n.s.	
neue Länder	29	29				
Elternschaft						
ohne Kind	94	95	1	0,104	n.s.	
mit Kind	6	5				
Bildungshintergrund						
nichtakad. Elternhaus	46	46	1	0,032	n.s.	
akademisches Elternhaus	54	54				
Hochschulart						
Universitäten	78	73	1	14,1	**	0,048
Fachhochschulen	22	27				
Studienform						
Vollzeit	75	76	2	0,46	n.s.	
Teilzeit	22	21				
Pro-Forma	3	3				
Fächergruppe						
Kulturwiss.	15	15	6	10,7	n.s.	
Sozialwiss.	21	22				
Rechtswiss.	5	3				
Wirt.wiss.	15	13				
Medizin	8	8				
Naturwiss.	17	20				
Ing.wiss.	19	19				
Abschlussart						
Bachelor	48	51	3	12,5	**	0,046
Master	27	23				
Staatsexamen	18	17				
sonstiges	7	9				
Zentralität						
Wichtigkeit von Studium/HS						
kaum	4	4	2	0,31	n.s.	
teilweise	37	38				
wichtig	59	58				

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

¹⁾ p, Überschreitungswahrscheinlichkeit:

** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant

Tabelle 5.3**Vergleich der vorgegebene Erhebungsarten, Papier vs. Online lang (Teil 2)**

(Angaben in Prozent)

Merkmale	Papier (Grp. 0)	Online lang (Grp. 3,4)	df	Chi²	P¹⁾	Cramers V
Berufsausbildung vor Studium						
Nein	78	75	1	3,88	*	0,025
Ja	22	25				
berufstätig vor Studium						
Nein	74	70	1	7,58	**	0,035
ja	26	30				
Verzug im Studium						
Nein	57	57	3	2,41	n.s.	
1 Semester	24	23				
2-3 Sem.	14	15				
4 und mehr S.	5	6				
Fachidentifikation						
nochmals gleiches Studium	76	75	2	0,54	n.s.	
anderes Fach	18	18				
kein Studium	6	7				
Berufswunsch steht fest						
nein, noch offen	34	33	2	3,76	n.s.	
ja, mit einiger Sicherheit	36	40				
ja, mit großer Sicherheit	30	27				
Fachsemester						
1-2 Semester	24	24	6	4,20	n.s.	
3-4	25	23				
5-6	21	23				
7-8	14	14				
9-10	8	8				
11-12	4	4				
13 u. mehr	4	4				
Mittelwertvergleiche						
	MW	MW	V-Sked. ²⁾	t-Wert	p-Wert	Eta ³⁾
Fachsemester	4,93	5,13		-1,46	0,07	0,04
Hochschulsemester	7,50	7,70	u	-1.20	0.11	0,03
Notenschnitt	2,15	2,26	u	-0.26	0.39	0,00
Abbruchneigung	0,45	0,42		0.59	0.72	0,01
Studienidentifikation	4,90	4,87	u	0.66	0.74	0,01
Wichtigkeit von Studium	4,58	4,87	u	0.59	0.72	0,01
Alter	24,20	24,98	u	3.17	0,001	0,29

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

¹⁾ ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant²⁾ Varianztest auf Homoskedastizität, u = ungleiche Varianzen³⁾ Eta: Determinationskoeffizient der Varianzanalyse, durch Modell erklärter Varianzanteil

5.2 Papier vs. Onlinegruppe (Mode-Effekt) bei Selbstselektion

In der Wahlgruppe können die Studierenden, die einen Papierfragebogen beantwortet haben, mit jenen verglichen werden, die an der Befragung über das Internet teilgenommen haben. Da alle Studierenden in dieser Gruppe beide Zugänge hatten, ist die Aufteilung der beiden Mode-Gruppen selbstgewählt. Der Mode-Vergleich bezieht sich hier also auf die Selbstselektion (vgl. Tabelle 5.4/5.5).

Auffällige Befunde hinsichtlich des Vergleichs der Stichprobenstruktur ergeben sich hier für das Geschlecht und die Fächergruppe.

- In der selbstgewählten Papiergruppe sind deutlich mehr Frauen (66 % zu 49 %).
- In der Papiergruppe sind mehr Studierende aus den Kultur- und Sozialwissenschaften, aber weniger aus den Wirtschafts-, Natur- und Ingenieurwissenschaften.

Alle anderen Vergleiche ergeben keine signifikanten Unterschiede.

Die gefundenen Effekte sind wiederum sehr gering. Die Effektmaße (Cramers V) liegen bei 0,15 bzw. 0,16, was einem eher kleinen Effekt entspricht. Betrachtet man die Größenordnungen der Unterschiede, so fällt auf, dass der Geschlechtereffekt in der Wahlgruppe größer ist als bei Fremdselektion. Der Effekt der Fächergruppen hängt mit dem Geschlechtseffekt zusammen, da die Kultur- und Sozialwissenschaften jene Fächergruppen sind, in denen die weiblichen Studierenden dominieren. Kontrolliert nach dem Geschlecht werden die Mode-Effekte deutlich kleiner und nur noch auf dem 5 %-Niveau signifikant. Gleichzeitig wird deutlich, dass der Fächergruppen-Effekt bei den Männern größer ist als bei den Frauen.

Da die Stichproben in der Wahlgruppe kleiner sind als bei den vordefinierten Mode-Gruppen, werden gleiche Realunterschiede teilweise nicht signifikant. Umgedreht bedeutet dies, dass gefundene Effekte bei den fremdselektierten Mode-Gruppen nur signifikant werden, weil die Stichproben größer sind.

5.2.1 Selektionseffekt bei Papierbefragung

In den Tabellen 5.6 und 5.7 sind die beiden Papiergruppen gegenübergestellt. Gruppe 0 bekam den Papierfragebogen zugesendet (Fremdselektion), während die Wahlgruppe die Papiervariante selbst gewählt hat. Es wird damit der Selektionseffekt bei der Papierbefragung geprüft.

Drei Vergleiche werden auf dem 5 %-Niveau signifikant:

- In der Wahlgruppe sind etwas mehr Frauen (66 % zu 62 %).
- In der Papiergruppe sind etwas mehr Studierende aus akademischen Elternhäuser (54 % zu 49 %).
- Studierende der Wahlgruppe sind etwas älter (24,6 zu 24,2).

Alle anderen Vergleiche ergeben keine signifikanten Unterschiede.

Die Effekte der Unterschiede sind alle sehr gering. Bei den nominalen Variablen liegt die Effektgröße (Cramers V) unter 0,1 und bei der intervallskalierten Variablen (Alter) kann die Gruppenzugehörigkeit weniger als 0,1 % aufklären (η^2).

Der Effekt beim Geschlecht und beim Alter wurde auch in den anderen Vergleichen gefunden. Neu ist der Effekt für den Bildungshintergrund bzw. die soziale Herkunft. In Abhängigkeit vom Geschlecht stellt sich heraus, dass der Effekt nur bei den Studentinnen auftritt, dort aber auf dem 1 %-Niveau signifikant wird. Allerdings bleibt die Effektgröße (Cramers V) mit 0,047 weiterhin sehr gering.

Tabelle 5.4**Vergleich der vorgegebene Erhebungsarten, Papier vs. Online in der Wahlgruppe**

(Angaben in Prozent)

Merkmale	Wahlgruppen	Papier (Grp. 1)	Online (Grp. 2)	df	Chi²	P¹⁾	Cramers V
Alter	<= 21	30	29	5	1,04	n.s.	
	22-23	21	22				
	24-25	20	22				
	26-27	11	12				
	28-29	7	6				
	=> 30	11	9				
Geschlecht	männlich	34	51	1	21,07	**	0,15
	weiblich	66	49				
Region	alte Länder	67	72	1	1,74	n.s.	
	neue Länder	33	28				
Elternschaft	ohne Kind	93	96	1	1,95	n.s.	
	mit Kind	7	4				
Bildungshintergrund nichtakad. Elternhaus		51	46	1	1,93	n.s.	
	akademisches Elternhaus	49	55				
Hochschulart	Universitäten	77	74	1	0,76	n.s.	
	Fachhochschulen	23	26				
Studienform	Vollzeit	74	76	2	3,48	n.s.	
	Teilzeit	24	22				
	Pro-Forma	2	2				
Fächergruppe	Kulturwiss.	19	10	6	21,97	**	0,16
	Sozialwiss.	24	17				
	Rechtswiss.	4	3				
	Wirt.wiss.	12	18				
	Medizin	7	7				
	Naturwiss.	17	22				
	Ing.wiss.	17	23				
Abschlussart	Bachelor	51	49	3	1,64	n.s.	
	Master	24	25				
	Staatsexamen	18	20				
	sonstiges	7	6				
Zentralität Wichtigkeit von Studium/HS	kaum	3	6	2	2,89	n.s.	
	teilweise	37	34				
	wichtig	60	60				

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

¹⁾ p, Überschreitungswahrscheinlichkeit:

** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant

Tabelle 5.5**Vergleich der vorgegebene Erhebungsarten, Papier vs. Online in der Wahlgruppe**

(Angaben in Prozent)

Merkmale	Wahlgruppe	Papier (Grp. 1)	Online (Grp. 2)	df	Chi²	P¹⁾	Cramers V
Berufsausbildung vor Studium							
	Nein	76	77	1	0,04	n.s.	
	Ja	24	23				
berufstätig vor Studium							
	Nein	70	73	1	0,89	n.s.	
	ja	30	27				
Verzug im Studium							
	Nein	61	59	3	2,44	n.s.	
	1 Semester	24	23				
	2-3 Sem.	11	13				
	4 und mehr S.	4	5				
Fachidentifikation							
nochmals gleiches Studium		79	76	2	0,88	n.s.	
anderes Fach		15	18				
kein Studium		6	6				
Berufswunsch steht fest							
nein, noch offen		34	35	2	2,03	n.s.	
ja, mit einiger Sicherheit		37	40				
ja, mit großer Sicherheit		29	25				
Fachsemester							
	1-2 Semester	26	26	6	2,77	n.s.	
	3-4	27	22				
	5-6	19	22				
	7-8	14	15				
	9-10	8	8				
	11-12	2	2				
	13 u. mehr	4	5				
Mittelwertvergleiche							
		MW	MW	V-Sked. ²⁾	t-Wert	p-Wert	Eta ³⁾
	Fachsemester	4,75	4,97		-0,80	0,21	0,07
	Hochschulsemester	7,26	7,35		-0,23	0,40	0,01
	Notenschnitt	2,15	2,23		-1,42	0,07	0,24
	Abbruchneigung	0,48	0,38		1,23	0,89	0,17
	Studienidentifikation	4,97	4,83	u	1,26	0,89	0,21
	Wichtigkeit von Studium	4,61	4,58	u	0,37	0,64	0,02
	Alter	24,65	24,10		1,17	0,87	0,16

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
¹⁾ ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant
²⁾ Varianztest auf Homoskedastizität, u = ungleiche Varianzen
³⁾ Eta: Determinationskoeffizient der Varianzanalyse, durch Modell erklärter Varianzanteil

Tabelle 5.6**Vergleich innerhalb der Erhebungsart Papier: Selbst vs. Fremdselektion (Gruppe 0 vs. Gruppe 1)**

(Angaben in Prozent)

Merkmale	Papier (Grp. 0)	Wahl-Papier (Grp. 1)	df	Chi²	p¹⁾	Cramers V
Alter						
<= 21	27	30				
22-23	23	21				
24-25	21	20	5	8,98	n.s.	
26-27	14	11				
28-29	7	7				
=> 30	8	11				
Geschlecht						
männlich	38	34	1	5,02	*	0,03
weiblich	62	66				
Region						
alte Länder	71	67	1	3,69	n.s.	
neue Länder	29	33				
Elternschaft						
ohne Kind	94	93	1	1,27	n.s.	
mit Kind	6	7				
Bildungshintergrund						
nichtakad. Elternhaus	46	51	1	6,07	*	0,03
akademisches Elternhaus	54	49				
Hochschulart						
Universitäten	78	77	1	0,75	n.s.	
Fachhochschulen	22	23				
Studienform						
Vollzeit	75	74	2	2,60	n.s.	
Teilzeit	22	24				
Pro-Forma	3	2				
Fächergruppe						
Kulturwiss.	15	19				
Sozialwiss.	21	24				
Rechtswiss.	5	4				
Wirt.wiss.	15	12	6	12,26	n.s.	
Medizin	8	7				
Naturwiss.	17	17				
Ing.wiss.	19	17				
Abschlussart						
Bachelor	48	51	3	2,77	n.s.	
Master	27	24				
Staatsexamen	18	18				
sonstiges	7	7				
Zentralität						
Wichtigkeit von Studium/HS						
kaum	4	3	2	1,32	n.s.	
teilweise	37	37				
wichtig	59	60				

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

¹⁾ p, Überschreitungswahrscheinlichkeit:

** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant

Tabelle 5.7**Vergleich innerhalb der Erhebungsart Papier: Selbst vs. Fremdselektion (Gruppe 0 vs. Gruppe 1)**

(Angaben in Prozent)

Merkmale	Papier (Grp. 0)	Wahl-Papier (Grp. 1)	df	Chi²	P¹⁾	Cramers V
Berufsausbildung vor Studium						
nein	78	76				
ja	22	24	1	1,42	n.s.	
berufstätig vor Studium						
nein	74	70				
ja	26	30	1	3,51	n.s.	
Verzug im Studium						
nein	57	61				
1 Semester	24	24	3	6,03	n.s.	
2-3 Sem.	14	11				
4 und mehr S.	5	4				
Fachidentifikation						
nochmals gleiches Studium	76	79				
anderes Fach	18	15	2	3,19	n.s.	
kein Studium	6	6				
Berufswunsch steht fest						
nein, noch offen	34	34				
ja, mit einiger Sicherheit	36	37	2	0,39	n.s.	
ja, mit großer Sicherheit	30	29				
Fachsemester						
1-2 Semester	24	26				
3-4	25	27				
5-6	21	19				
7-8	14	14	6	5,8	n.s.	
9-10	8	8				
11-12	4	2				
13 u. mehr	4	4				
Mittelwertvergleiche						
	MW	MW	V-Sked. ²⁾	t-Wert	p-Wert	Eta ³⁾
Fachsemester	4,93	4,75		1,11	0,86	0,02
Hochschulsemester	7,50	7,26	u	1,07	0,87	0,02
Notenschnitt	2,15	2,15		-0,08	0,47	0,00
Abbruchneigung	0,45	0,48		-0,76	0,22	0,01
Studienidentifikation	4,90	4,97		-1,32	0,09	0,03
Wichtigkeit von Studium	4,58	4,61		-0,67	0,25	0,01
Alter	24,2	24,65	u	-1,83	0,03	0,09

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

¹⁾ ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant²⁾ Varianztest auf Homoskedastizität, u = ungleiche Varianzen³⁾ Eta: Determinationskoeffizient der Varianzanalyse, durch Modell erklärter Varianzanteil

Tabelle 5.8
Vergleich innerhalb der Erhebungsart Online: Selbst- vs. Fremdselektion
 (Angaben in Prozent)

Merkmale	Wahl-online (Grp. 2)	Online lang (Grp. 3,4)	df	Chi²	P¹⁾	Cramers V
Alter						
<= 21	29	25				
22-23	22	22				
24-25	22	19	5	5,71	n.s.	
26-27	12	12				
28-29	6	8				
=> 30	9	14				
Geschlecht						
männlich	51	44	1	3,42	n.s.	
weiblich	49	56				
Region						
alte Länder	72	71	1	0,17	n.s.	
neue Länder	28	29				
Elternschaft						
ohne Kind	96	95	1	0,54	n.s.	
mit Kind	4	5				
Bildungshintergrund						
nichtakad. Elternhaus	46	46	1	0,03	n.s.	
akademisches Elternhaus	55	54				
Hochschulart						
Universitäten	74	73	1	0,10	n.s.	
Fachhochschulen	26	27				
Studienform						
Vollzeit	76	76				
Teilzeit	22	21	2	1,92	n.s.	
Pro-Forma	2	3				
Fächergruppe						
Kulturwiss.	10	15				
Sozialwiss.	17	22				
Rechtswiss.	3	3				
Wirt.wiss.	18	13	6	12,59	n.s.	
Medizin	7	8				
Naturwiss.	22	20				
Ing.wiss.	23	19				
Abschlussart						
Bachelor	49	51				
Master	25	23	3	4,19	n.s.	
Staatsexamen	20	17				
sonstiges	6	9				
Zentralität						
Wichtigkeit von Studium/HS						
kaum	6	4				
teilweise	34	38	2	1,03	n.s.	
wichtig	60	58				

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

¹⁾ p, Überschreitungswahrscheinlichkeit:

** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant

Tabelle 5.9

Vergleich innerhalb der Erhebungsart Online: Selbst- vs. Fremdselektion

(Angaben in Prozent)

Merkmale	Wahl-online (Grp. 2)	Online lang (Grp. 3,4)	df	Chi²	P¹⁾	Cramers V
Berufsausbildung vor Studium						
nein	77	75	1	0,18	n.s.	
ja	23	25				
berufstätig vor Studium						
nein	73	70	1	1,40	n.s.	
ja	27	30				
Verzug im Studium						
nein	59	57	3	0,69	n.s.	
1 Semester	23	23				
2-3 Sem.	13	15				
4 und mehr S.	5	6				
Fachidentifikation						
nochmals gleiches Studium	76	75	2	0,16	n.s.	
anderes Fach	18	18				
kein Studium	6	7				
Berufswunsch steht fest						
nein, noch offen	35	33	2	0,59	n.s.	
ja, mit einiger Sicherheit	40	40				
ja, mit großer Sicherheit	25	27				
Fachsemester						
1-2 Semester	26	24	6	2,31	n.s.	
3-4	22	23				
5-6	22	23				
7-8	15	14				
9-10	8	8				
11-12	2	4				
13 u. mehr	5	4				
Mittelwertvergleiche						
	MW	MW	V-Sked.²⁾	t-Wert	p-Wert	Eta³⁾
Fachsemester	4,97	5,13		-0,57	0,28	0,02
Hochschulsemester	7,35	7,70		-1,01	0,15	0,08
Notenschnitt	2,23	2,16		1,33	0,90	0,16
Abbruchneigung	0,38	0,42		-0,55	0,29	0,03
Studienidentifikation	4,83	4,87		-0,35	0,36	0,01
Wichtigkeit von Studium	4,58	4,55		0,26	0,60	0,01
Alter	24,10	24,98	u	-1,98	0,02	0,28

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

¹⁾ ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant

²⁾ Varianztest auf Homoskedastizität, u = ungleiche Varianzen

³⁾ Eta: Determinationskoeffizient der Varianzanalyse, durch Modell erklärter Varianzanteil

5.2.2 Selektionseffekt bei Online-Befragung

In den Tabellen 5.8 und 5.9 sind die Vergleiche des Selektionseffektes bei der Onlinevariante abgebildet. Dazu wurden die beiden Onlinegruppen mit langem Fragebogen (3 und 4) mit der Onlinevariante der Wahlgruppe (Gruppe 2) verglichen.

Nur ein Vergleich wird signifikant:

- In den fremdselegierten Onlinegruppen sind die Studierenden älter (25 zu 24,1).

Die Effektgröße ist mit 0,28 allerdings wiederum gering. Der Blick auf die Altersverteilung lässt erkennen, dass in der Gruppe mit Fremdselektion vor allem mehr Studierende sind, die bereits über 30 Jahre sind.

Die anderen Effekte sind nicht signifikant. Allerdings weisen einige Vergleiche prozentuale Unterschiede auf, die bei den vorhandenen Stichprobengrößen nicht signifikant werden. So sind z.B. bei den fremdselegierten Studierenden der Onlinegruppen mehr Studentinnen (56 % zu 49 %). Dieser Befund deckt sich mit den bisherigen und verweist darauf, dass Frauen generell häufiger an Befragungen teilnehmen und, wenn sie die Wahl haben, lieber einen Papierfragebogen nutzen.

Die bis zu fünf Prozentpunkte Unterschiede bei den Fächergruppen verweisen auf die unterschiedlichen Geschlechterverhältnisse.

Als **Fazit** wird deutlich:

- Studierende nehmen eher an Papierbefragungen teil als an Online-Befragungen;
- Studentinnen nehmen häufiger an Befragungen teil;
- Studentinnen nutzen lieber Papierbefragungen, vor allem, wenn sie die Wahl haben;
- männliche Studierende an Fachhochschulen nehmen häufiger an Online-Befragungen teil, bzw. seltener an Papierbefragungen;
- ältere Studierende nutzen lieber einen Onlinezugang.

Konsequenz

Ziel der Untersuchung an den verschiedenen Designgruppen ist zu bestätigen, dass die Gruppen untereinander vergleichbar sind und als Gesamtstichprobe zusammengeführt werden dürfen.

Anhand der Vergleiche und den daraus folgenden Befunden wird deutlich, dass die beiden Papiergruppen (randomisierte Papiergruppe 0 und Papier-Wahlgruppe 1) sowie die beiden Onlinegruppen (randomisierte Online-Gruppe 3 und 4 sowie die Wahlgruppe-Online 2) zusammengenommen werden können. Der Effekt zwischen Fremd- und Selbstselektion ist nur sehr gering. Etwas größer bleibt der Mode-Effekt zwischen Papier und Online, der sich vor allem auf das Alter, das Geschlecht, die Region, Abschlussart und Fächergruppe bezieht. Allerdings bleiben auch hier die Effektgrößen relativ gering, weshalb auch diese Samples, auf Grundlage der Strukturvergleiche der Stichproben, eher zusammengefasst werden können.

5.3 Multivariater Vergleich: Einfluss von soziodemographischen Merkmalen

Anhand von logistischen Regressionen soll überprüft werden, welche soziodemographischen Daten einen Einfluss auf den Mode- und Selektionseffekt haben. Im Vergleich zu den bisher durchgeführten Analysen wird also ein multivariater Ansatz gewählt, in dem alle ausgewählten Merkmale simultan bzw. schrittweise in ein Modell eingeführt werden (ausführliche Ergebnisse siehe Anhang Tabellen A2 bis A5).

Zusätzlich zu den soziodemographischen Merkmalen werden in Erweiterung des Modells spezifische Einstellungsmerkmale schrittweise in das Modell eingeführt, die als Indikatoren für Hochschul- und Computeraffinität dienen sollen. In den Tabellen werden die Ergebnisse der Regressionsanalyse dargestellt. Dabei werden die Odds Ratio-Werte und die „average marginal effects“ (AME) für die Prädiktoren angegeben. Im unteren Teil stehen die Effektstärken für die Modellgüte.

In der ersten Tabelle werden die Selektionseffekte zwischen den randomisierten und den selbstselektierten Gruppen überprüft, sowohl für die Papiergruppen, als auch für die Onlinegruppen. Als Referenzkategorie wird dabei jeweils die randomisierte Gruppe verwendet, sodass die Wahrscheinlichkeitsaussagen jeweils für die Wahlgruppen gelten (vgl. Tabelle 5.10).

Tabelle 5.10					
Selektionseffekt bei Papier- und Onlinegruppen (Selbstselektiert vs. randomisiert (ref.))					
(Angaben: Odds Ratio, AME (average marginal effect), N, Pseudo-R ² -Maße)					
Merkmale (Referenzkat.)	Kategorie	Papiergruppe		Onlinegruppe	
		Odds Ratio	AME	Odds Ratio	AME
Alter ¹⁾ (<17 Jahre)		1,006	0,0006	0,969	-0,0047
Geschlecht (Männer)	Frauen	1,079	0,0078	0,859	-0,0230
Region (neue Länder)	alte Länder	0,786*	-0,0249*	1,026	0,0039
Elternschaft (ohne)	mit Kinder	0,909	-0,0098	1,268	0,0361
Bildung Eltern (n.akad.)	akademisch	0,850	-0,0167	0,875	-0,0202
Hochschulart (FH)	Uni	0,859	-0,0156	1,033	0,0049
Studienform (Vollzeit)	Teilzeit	0,992	-0,0007	1,021	0,0032
Familienstand (ledig)	verheiratet	1,259	0,0239	1,482	0,0598
Studienabschluss (ohne)	mit	0,897	-0,0111	0,982	-0,0026
Fächergruppe (Ing.wiss)	Kultur	1,528*	0,0458*	0,571	-0,0759
	Sozial	1,367*	0,0324*	0,616	-0,0671
	Jur/Wiwi	0,908	-0,0084	1,318	0,0480
	Medizin	1,045	0,0041	0,912	-0,0142
	Natur	1,236	0,0211	1,148	0,0232
Zentralität Studium ²⁾ (wichtig)	unwichtig	0,844	-0,0167	1,170	0,0257
	teilweise wichtig	0,959	-0,0042	0,859	-0,0226
virtuell studieren ³⁾ (nein)	ja	1,020	0,0020	1,369	0,0477
Konstante		1,391**		0,542	
n		4550		840	
Pseudo R ²		0,0090		0,0242	
McFadden		0,009		0,024	
Cragg-Uhler/Nagelkerke		0,013		0,038	
McKelvey / Zavoina			0,019		0,047

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
 1) Alter kategorisiert: <17, 17-21, 22-23,24-25,26-27,28-29, 30>;
 * signifikant auf 5%-Niveau, ** signifikant auf 1%-Niveau
 2) Skala von 0=völlig unwichtig bis 6=sehr wichtig, Kategorien: 0-2=unwichtig, 3-4=teilweise, 5-6=wichtig (ref.)
 3) Skala von 1=sicher nicht bis 5=sicher ja; Kategorien: 1-2=nein (ref.), 3-5=ja

Das logistische Regressionsmodell macht Aussagen darüber, mit welcher Wahrscheinlichkeit Prädiktoren eine Kriteriumsvariable erklären bzw. vorhersagen können. Dabei wird die Wahrscheinlichkeit des Eintretens einer bestimmten Ausprägung des Kriteriums in Abhängigkeit von den Ausprägungen der Prädiktoren ermittelt. Letztere sind binär kodiert, deren Ausprägungen jeweils 0 oder 1 annehmen können. Als Linearkombination aller Prädiktoren wird eine logistische Wahrscheinlichkeitsfunktion berechnet (Maximum Likelihood-Methode). Zur Darstellung der Effekte werden die Odds-Ratio und die AME-Werte verwendet. Die Odds stellen das Verhältnis der Eintrittswahrscheinlichkeit zur Gegenwahrscheinlichkeit dar, die Odds Ratio (OR) ist der Quotient der Odds pro Ausprägung des Prädiktors. Eine OR von 1,5 bedeutet, dass die Odds für ein Ereignis um das 1,5 fache steigen, wenn die unabhängige Variable x_i um eine Einheit steigt.

Die AME geben den Durchschnittseffekt eines Prädiktors als Mittelwert der marginalen Effekte über alle Beobachtungen hinweg wieder. Er kann Werte zwischen 1 und -1 annehmen. Marginale Effekte machen Aussagen darüber, wie sich Veränderungen der Prädiktoren auf Veränderungen des Kriteriums auswirken, wenn alle anderen Variablen auf ihrem Mittelwert konstant gehalten werden. Die Odds Ratio werden nur zu Vergleichszwecken mit anderen Studien dargestellt, die intuitiveren Aussagen und die bessere Vergleichbarkeit bieten die AME (vgl. Best/Wolf 2012). Mood identifiziert in ihrem Artikel mehrere Probleme bei der Interpretation von Odds und ORs, denn die in Analysen identifizierten Effekte sind 1) durch unbeobachtete Heterogenität verzerrt, 2) ist ein Vergleich von Odds/ORs über mehrere Modelle mit verschiedenen unabhängigen Variablen problematisch, weil die unbeobachtete Heterogenität zwischen den Modellen variiert, und 3) gilt das auch für den Vergleich von Modellen mit identischen Prädiktorvariablen über verschiedene Gruppen oder die Zeit hinweg, denn auch hier kann die unbeobachtete Heterogenität über die Zeit oder zwischen den Gruppen variieren (Mood, 2010). Der durchschnittliche marginale Effekt wird hingegen nicht von unkontrollierter unbeobachteter Heterogenität verzerrt und erlaubt daher den Vergleich von Koeffizienten in hierarchisch aufgebauten Modellen. Average marginal effects sind darüber hinaus leichter zu interpretieren. Sie geben den durchschnittlichen Effekt einer unabhängigen Variable auf die Wahrscheinlichkeit wieder: wenn x_i um eine Einheit steigt (sinkt), steigt (sinkt) die Wahrscheinlichkeit des Ereignisses **durchschnittlich** um den Wert des AME (Best & Wolf, 2010, 2012). Ziel der logistischen Regression ist hier letztlich zu prüfen, ob die Stichproben zusammengefasst werden dürfen.

5.3.1 Effekte der Selbstselektion in der Papiergruppe

Tabelle 5.10 bildet die Ergebnisse des Selektionseffekts für die Papiergruppe und die Onlinegruppen ab. Die Odds Ratios geben an, um das wieviel-fache sich das erwartete Chancenverhältnis für das Ereignis erhöht, wenn die Variable um eine Ausprägung erhöht wird. Die AME geben den durchschnittlichen Effekt des Merkmals auf das Kriterium an, unter Konstanthaltung aller anderen Merkmale.

Im Beispiel des Merkmals Geschlecht in der Papiergruppe beträgt die Odds Ratio 1,079 und bedeutet, dass Frauen eine 1,079-fache höhere Chance haben, zur Wahlpapiergruppe zu gehören, im Vergleich zu Männern. Der AME-Wert von 0,0078 besagt, dass Frauen um 0,78 % häufiger der Wahlgruppe zugehören als Männer. Im Vergleich beider Werte wird ein Unterschied von Faktor 10 ersichtlich. Daran wird deutlich, dass es sich nicht um vergleichbare Wahrscheinlichkeiten handelt. Bei den Odds Ratio wird die Erhöhung der Chance, bezogen auf die Basiswahrscheinlichkeit, abgebildet; bei den AME der durchschnittliche Effekt auf das Kriterium. Nur letztere können als Zunahme bzw. Abnahme der Wahrscheinlichkeiten des Eintritts des Ereignisses (Wahlgruppe) interpretiert werden. Eine wichtige Einschränkung der AME ist, dass es sich eben nur um einen Durchschnittseffekt handelt: In logistischen Modellen hängt der Effekt einer Variablen von den Ausprägungen der übrigen Variablen im

Modell ab. Bei der Berechnung des AME werden die übrigen Variablen auf ihren Mittelwert festgesetzt, was einem Informationsverlust entspricht¹⁵.

Von den verschiedenen Modellen, die die schrittweise eingeführten Merkmale (Bedeutung des Studiums und Affinität zu einem virtuellen Studium) betrachten, wird nur das jeweils vollständige angezeigt, da die schrittweisen Erweiterungen keine nennenswerten Veränderungen bzw. Effekte erbrachten.

Signifikante Ergebnisse besagen, dass die Odds Ratio-Werte bedeutsam von 1,0 abweichen, bzw. die AME-Werte bedeutsam von Null abweichen, womit geprüft wird, ob die Wahrscheinlichkeit der beiden Kategorien ungleich sind. Solche Werte werden in den Tabellen mit einem Stern gekennzeichnet (5 %-Niveau, zwei Sterne = 1 %-Niveau). In der Papiergruppe werden drei Merkmale auf dem 5 %-Niveau signifikant, zum einen die alten Bundesländer, und zum anderen in den Fächergruppen die Kultur- und die Sozialwissenschaften. Studierende aus den alten Ländern haben eine geringere Chance in die Wahlgruppe zu fallen, Studierende der Kultur- und Sozialwissenschaften eine höhere. D.h. Studierende aus den neuen Ländern und Studierende der Kultur- und Sozialwissenschaften entscheiden sich jeweils häufiger in der Wahlgruppe für einen Papierfragebogen.

Die Pseudo R^2 Werte machen eine Aussage über die Güte des Modells. Sie basieren in ihren Berechnungen in der Spalte der Odds alle auf dem Quotienten der Likelihoodfunktionen des Nullmodells (ohne Regressoren) und dem vollständigen Modell (mit allen Regressoren) und geben an, wie gut das Modell die Daten erklären kann. Für die AME kann ein R^2 auf Basis der Varianzzerlegung berechnet werden, der die Güte des Modells besser darstellen kann. Die abgebildeten Werte liegen zwischen ein und zwei Prozent, womit das Modell die Datenvariation kaum erklären kann, die Merkmale also nicht geeignet sind, die Personen einer Gruppe zuzuweisen. Das bedeutet, dass die Personen in den beiden Papiergruppen sich in diesen Merkmalen nicht bedeutsam unterscheiden.

Trotz der signifikanten Effekte in den drei Merkmalen bleibt das Gesamtmodell unbefriedigend, da es zu wenig Aussagekraft besitzt. Die soziodemografischen Merkmale sowie die zusätzlichen Indikatoren für Hochschule und Computeraffinität können die Gruppenzugehörigkeit der Personen nicht erklären. Damit darf angenommen werden, dass die Stichproben der beiden Gruppen gleich sind.

Bei den Onlinegruppen wird kein Merkmal signifikant und die Effektstärken bleiben unter fünf Prozent, womit auch diese Stichproben als gleich betrachtet werden können.

5.3.2 Mode-Effekte

Tabelle 5.11 stellt die Ergebnisse der Mode-Effekte zwischen den Papier- und den Onlinegruppen dar, wobei die entsprechenden Wahlgruppen einbezogen wurden. Als Referenzkategorie wird die Papiergruppe verwendet. Zuerst werden die Mode-Effekte ohne und dann mit dem Selektionseffekt betrachtet.

Für den Mode-Effekt werden fünf Merkmale signifikant: Alter, Geschlecht, Familienstand und Studienabschluss sowie in den Fächergruppen die Naturwissenschaften. Bei Einbeziehung des Selektionseffekts wird dieser ebenfalls signifikant und weist im Vergleich aller Merkmale den stärksten Effekt auf. Dagegen hat der Familienstand den geringsten Effekt.

¹⁵ Der AME ließe sich auch für bestimmte Ausprägungen der unabhängigen Variablen berechnen.

Tabelle 5.11**Mode-Effekt zwischen Papier- und Onlinegruppen, ohne und mit Selektionseffekt**(Angaben: Odds Ratio, AME (average marginal effect), N, Pseudo-R²-Maße)

Merkmale (Referenzkat.)	Kategorie	Ohne Selektionseffekt		Mit Selektionseffekt	
		Odds Ratio	AME	Odds Ratio	AME
Selektionseffekt (rand.)	selbstselektiert	-	-	1,784**	0,0747**
Alter ¹⁾ (<17 Jahre)		1,044**	0,0056**	1,044**	0,0055**
Geschlecht (Männer)	Frauen	0,764**	-0,0349**	0,762**	-0,0349**
Region (neue Länder)	alte Länder	1,014	0,0019	1,028	0,0036
Elternschaft (ohne)	mit Kinder	0,742	-0,0386	0,741	-0,0386
Bildung Eltern (n.akad.)	akademisch	1,048	0,0061	1,059	0,0074
Hochschulart (FH)	Uni	0,835	-0,0234	0,841	-0,0222
Studienform (Vollzeit)	Teilzeit	0,864	-0,0188	0,865	-0,0186
Familienstand (ledig)	verheiratet	0,600*	-0,0662*	0,586*	-0,0688*
Studienabschluss (ohne)	mit	0,783**	-0,0317**	0,786**	-0,0310**
Fächergruppe (Ing.wiss)	Kultur	1,000	0,0000	0,980	-0,0023
	Sozial	1,064	0,0078	1,052	0,0063
	Jur/Wiwi	0,892	-0,0135	0,892	-0,0134
	Medizin	1,178	0,0214	1,181	0,0216
	Natur	1,482**	0,0554**	1,462**	0,0530**
Zentralität (wichtig)	unwichtig	1,165	0,0206	1,171	0,0211
	teilweise wichtig	1,041	0,0052	1,045	0,0057
virtuell studieren (nein)	ja	1,143	0,0173	1,133	0,0162
Konstante		0,032**		0,075**	
n		5390		5390	
Pseudo R ²		0,0142		0,0209	
McFadden		0,014		0,021	
Cragg-Uhler/Nagelkerke		0,021		0,031	
McKelvey / Zavoina			0,027		0,039

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
1) Alter kategorisiert: <17, 17-21, 22-23, 24-25, 26-27, 28-29, 30>;
* signifikant auf 5%-Niveau, ** signifikant auf 1%-Niveau

Ältere, männliche, ledige Studierende im Erststudium in den Naturwissenschaften finden sich somit häufiger als ihre jeweiligen Pendanten in der Onlinegruppe. Allerdings bleiben die Effektstärken mit zwei bis vier Prozent sehr gering. Auch hier weisen die Modelle nur einen sehr geringen Anteil erklärter Varianz auf (zwischen ein und drei Prozent). Somit kann trotz signifikanter Unterschiede die Aussage getroffen werden, dass die Stichproben vergleichbar sind.

Ziel auch der multivariaten Analyse war die Prüfung der Frage, ob die unterschiedlichen Stichproben des Untersuchungsdesigns für Aussagen zusammengefasst werden dürfen oder nicht. Dazu wurden die strukturellen Merkmale der Stichproben miteinander verglichen. Die Analysen weisen zwar bei einzelnen Merkmalen signifikante Unterschiede auf, jedoch sind die Effekte so gering, dass sie nicht ausreichen, die Stichproben generell als ungleich zu bezeichnen. Insofern ist der Nachweis geführt, dass die Stichproben zusammengefasst werden dürfen.

6. Strukturvergleiche mit Grundgesamtheit

Damit die Befunde als weitgehend repräsentativ für die Grundgesamtheit der Studierenden in Deutschland angenommen werden können, sollte die erhobene Stichprobe die Struktur der Grundgesamtheit widerspiegeln können. Dazu wird anhand wichtiger sozialer Merkmale die Struktur der Stichproben mit der der Grundgesamtheit verglichen, wie sie aus den Daten des Statistischen Bundesamtes hervorgehen.

Verteilung der Hochschularten

Das Statistische Bundesamt benennt 2016 insgesamt 426 Hochschulen in Deutschland. Davon werden 215 als Fachhochschulen (ohne Verwaltungsfachhochschulen) ausgewiesen und 107 als Universitäten. Des Weiteren werden sechs Pädagogische Hochschulen, 16 Theologische Hochschulen, 52 Kunsthochschulen und 30 Verwaltungsfachhochschulen gezählt.

Die Anzahl der Studierenden beläuft sich zum WS 2015/16 in Deutschland auf 2.757.799 Studierende insgesamt. Deutsche plus Bildungsinländer summieren sich auf 2.595.020 Studierende. Davon studieren 1.551.282 an Universitäten und 865.829 an Fachhochschulen. In Anteilen entspricht dies einer Quote von 59,8 % für die Universitäten und von 33,3 % für die Fachhochschulen. Die restlichen 6,9 % Studierende sind an den anderen Hochschularten eingeschrieben (Statistisches Bundesamt, 2016).

Die Anteile der Hochschularten im Studierendensurvey unterscheiden sich von denen der Grundgesamtheit. Von den befragten Studierenden gehören 77,2 % zu den Universitäten und 22,8 % zu den Fachhochschulen. Damit sind im Survey die Universitäten überrepräsentiert, die Fachhochschulen unterrepräsentiert.

Tabelle 6.1
Verteilung der Designgruppen nach Hochschularten (WS 2015/16)
(Angaben in Prozent und Differenzen)

	Anteile in %						Abweichungen von GG				
	GG	Ges.	Papier		Online		Ges.	Papier		Online	
			r	W	r	W		r	W		
Uni	59,8	77,2	78,3	76,9	74,2	73,2	+17,4	+18,5	+17,1	+14,4	+13,4
FH	33,3	22,8	21,7	23,1	25,8	26,8	-10,5	-11,6	-10,2	-7,5	-6,5

Quelle: Statistische Bundesamt 2016, Studierendensurvey 2016, AG Hochschulforschung, Universität Konstanz
GG: Grundgesamtheit; r: randomisierte Gruppen; W: Wahlgruppe

Die im Vergleich zur Grundgesamtheit verschiedenen Anteile von Studierenden an Universitäten und Fachhochschulen im Studierendensurvey haben zwei Gründe:

- 1) Die Tradition der Erhebung: Zu Beginn der Erhebungen Anfang der 80er Jahre lag das Verhältnis von Studierenden an Universitäten und Fachhochschulen bei 72 % und 17 %. Bis zur Jahrtausendwende hin haben sich diese Anteile auf 65 % und 23 % verändert (WR 2002). Erst im neuen Jahrtausend (ab 2004) sind die Anteile der Fachhochschulen über 30 % gestiegen (Destatis 2016).
- 2) Die theoretische Stichprobe des Studierendensurveys: Um die zeitliche Entwicklung möglichst valide einschätzen zu können, war es in allen Erhebungen ein Anliegen, möglichst immer die gleichen Hochschulen zu befragen. Gleichzeitig sollte an den Hochschulen eine auf die Hochschulgröße vergleichbare angepasste Zufallsauswahl gezogen werden. Dadurch bleiben die Anteile der befragten Studierenden an den beiden Hochschularten relativ konstant.

Seit Beginn der Erhebungen rangieren die Anteile für die Studierenden an Universitäten in der Stichprobe zwischen 76,7 % und 83,3 %. In der ersten Erhebung lag der Anteil sogar bei 86,2 %, allerdings waren damals nur drei Fachhochschulen im Sample vertreten. Für die Zeitreihe sind die vergleichbaren Anteile über alle Erhebungen hinweg sinnvoll und für die Gesamtaussagen sogar valider.

Für die Darstellung der Befunde ergibt sich durch die zur Grundgesamtheit unterschiedlichen Anteile nur dann ein Problem, wenn Aussagen für alle Studierenden insgesamt, bezogen auf die aktuelle Situation, gemacht werden sollen. Alle Aussagen, die nach der Hochschulart unterschieden werden, bleiben dagegen valide.

Hinsichtlich der Designgruppen ist festzuhalten, dass die Überrepräsentativität der Universitäten in den Papiergruppen etwas größer ist als bei den Onlinegruppen. Gleichzeitig zeigt sich eine tendenzielle Minderung der Überrepräsentativität bei der Wahlgruppe im Vergleich zu den randomisierten Gruppen. Für die Fachhochschulen heißt das, dass die Unterrepräsentativität in den Onlinegruppen geringer ist als in den Papiergruppen. Und ebenso ist ein tendenzieller Rückgang der Unterrepräsentation bei der Wahlgruppe im Vergleich zu den randomisierten Gruppen festzustellen.

Verteilung nach Geschlecht

In den Erhebungen des Studierendensurveys beteiligen sich traditionell mehr Frauen als Männer. An den Universitäten entsteht dadurch eine Überrepräsentation von knapp elf Prozent, an den Fachhochschulen von rund 17 %. Diese Verteilung hat sich allerdings erst im Laufe der Jahre ergeben. Zu Beginn der Erhebungen, Anfang der 80er Jahre, lag der Frauenanteil im Studierendensurvey bei 34 %. Seit Ende der 80er Jahre ist er stetig angestiegen und überschritt mit dem Jahrtausendwechsel erstmals die 50 %-Marke. Ein Anteil von über 60 % ist allerdings erst in der 13. Erhebung 2016 aufgetreten.

	Anteile in %						Abweichungen von GG				
	GG	Ges.	Papier		Online		Ges.	Papier		Online	
			r	W	r	W		r	W	r	W
Frauen											
Ges.	47,8	61,0	61,7	66,3	48,9	55,8	+12,2	+13,9	+18,5	+1,1	+8,0
Uni	50,8	61,5	62,2	65,5	50,0	57,3	+10,7	+11,4	+14,7	-0,8	+6,5
FH	42,0	59,2	60,2	68,6	43,6	52,0	+17,2	+18,2	+26,6	+1,6	+10,0

Quelle: Statistische Bundesamt 2016, Studierendensurvey 2016, AG Hochschulforschung, Universität Konstanz
GG: Grundgesamtheit; r: randomisierte Gruppen; W: Wahlgruppe

Besonders hoch ist der Frauenanteil bei den Papierbefragungen. Den höchsten Wert weisen die Studentinnen dort in der Wahlgruppe an Fachhochschulen auf. In den Onlinegruppen liegen die Anteile deutlich näher an der Grundgesamtheit, vor allem in der randomisierten Gruppe.

Frauen nehmen an solchen Befragungen deutlich häufiger teil als Männer, vor allem, wenn es sich um eine Papierbefragung handelt und insbesondere, wenn sie das Medium auswählen können. Für die Papierbefragungen bedeutet dies, dass bei Befunden zur Gesamtheit die Studentinnen durch ihre Überrepräsentativität stärker ins Gewicht fallen. Deshalb ist für alle Aussagen zu prüfen, ob Geschlechtsunterschiede existieren.

Verteilung nach Fächergruppen

An den Universitäten weichen die Anteile der Fächergruppen nicht sehr stark von der Grundgesamtheit ab. Die beiden größten Abweichungen fallen bei den Kultur- und Sozialwissenschaften auf, belaufen sich allerdings nur auf rund vier Prozent, wobei die Kulturwissenschaften im Vergleich zur Grundgesamtheit im Survey leicht unter-, die Sozialwissenschaften leicht überrepräsentiert sind.

In den Designgruppen steigt die Überrepräsentativität der Sozialwissenschaften in der Papier-Wahlgruppe dann nochmals leicht an, auf 6,3 %, dagegen fällt die Unterrepräsentativität der Kulturwissenschaften weg. In den Onlinegruppen verstärkt sich die Unterrepräsentativität der Kulturwissenschaften und erreicht ein Minus von 10,3 % in der randomisierten Gruppe. Dagegen erreichen die Natur- und Ingenieurwissenschaften in dieser Gruppe ein Plus von 6,3 % bzw. 5 %. Die Online-Wahlgruppe weist eine leichte Überrepräsentativität der Sozial- und der Naturwissenschaften auf.

Die Überrepräsentativität der Sozialwissenschaften hängt mit der häufigeren Beteiligung von Studentinnen zusammen, die öfter als Studenten in sozialwissenschaftlichen Fächern eingeschrieben sind. Die Überrepräsentativität der Natur- und Ingenieurwissenschaften in der Onlinegruppe ist analog mit der geringeren Teilnahme von Studentinnen zu erklären, die weniger stark in Fächern der Natur- und Ingenieurwissenschaften vertreten sind.

	GG	Ges.	Anteile in %				Abweichungen von GG				
			Papier		Online		Ges.	Papier		Online	
			r	W	r	W		r	W	r	W
Universitäten											
Kultur.	22,0	18,3	17,9	22,6	11,7	19,1	-3,7	-4,1	+0,6	-10,3	-1,9
Sozial.	14,7	19,1	19,1	21,0	14,9	18,6	+4,4	+4,4	+6,3	+0,2	+3,9
Recht.	7,0	5,2	5,5	4,6	4,3	4,3	-1,8	-1,5	-2,4	-2,7	-2,7
Wiwi.	13,1	10,6	11,2	8,9	13,3	8,1	-2,5	-1,9	-4,2	+0,2	-5,0
Mediz.	7,4	9,8	9,9	8,9	8,0	10,0	+2,4	+2,5	+1,5	+0,6	+2,6
Natur.	21,4	21,4	20,5	20,4	27,7	24,8	0	-0,9	-1,0	+6,3	+3,4
Ing.	12,6	13,5	13,7	12,1	17,6	13,0	+0,9	+1,1	-0,5	+5,0	+0,4
sonst.	1,8	2,2	2,2	1,6	2,7	2,2	+0,4	+0,4	-0,2	+0,9	+0,4
Fachhochschulen											
Kultur.	4,1	3,7	4,2	2,6	0	3,0	-0,4	+0,1	-1,6	-4,1	-1,1
Sozial.	13,1	23,3	22,6	28,9	16,9	24,1	+10,2	+9,5	+15,8	+3,8	+11,0
Recht.	1,7	1,4	1,5	2,6	1,5	0,3	-0,3	-0,2	+0,9	-0,2	-1,4
Wiwi.	32,4	24,1	25,5	20,4	24,6	20,7	-8,3	-6,9	-12,0	-7,8	-11,7
Mediz.	5,5	5,8	6,0	3,3	15,4	11,4	+0,3	+1,5	-2,2	+9,9	+5,9
Natur.	7,9	7,1	6,3	7,9	1,5	4,0	-0,8	-1,6	0	-6,4	-3,9
Ing.	32,5	30,6	30,9	29,6	32,3	29,8	-1,9	-1,6	-2,9	-0,2	-2,7
sonst.	2,6	4,1	3,1	4,6	7,7	6,7	+1,5	+0,5	+2,0	+5,1	+4,1

Quelle: Statistische Bundesamt 2016, Studierendensurvey 2016, AG Hochschulforschung, Universität Konstanz
GG: Grundgesamtheit; r: randomisierte Gruppen; W: Wahlgruppe

An den Fachhochschulen fällt eine deutliche Überrepräsentativität der Sozialwissenschaften auf. Dies hängt vor allem mit der theoretischen Stichprobe zusammen, da die Sozialwissenschaften eine Voraussetzung für die Einbeziehung der Fachhochschulen in das Hochschulsample waren. Gleichzeitig fällt hier auch die häufigere Beteiligung von Studentinnen ins Gewicht. Auf der Gegenseite ergibt sich dadurch eine Unterrepräsentation der Wirtschaftswissenschaften, da diese nur gleichzeitig mit den

Sozial- und Ingenieurwissenschaften einbezogen werden, real aber häufiger an Fachhochschulen zu finden sind. In den Onlinegruppen fällt auf, dass überverhältnismäßig viele Studierende aus medizinisch-gesundheitlichen Studiengängen teilgenommen haben. Dies hängt ebenfalls zum Teil mit der höheren Beteiligung der Frauen zusammen.

Verteilung nach Alter

In der zentralen Tendenz, hier als Mediane angegeben, weichen die Altersangaben der befragten Studierende nicht sehr stark von der Grundgesamtheit ab. Den größten Unterschied insgesamt weisen die Studierenden der Online-Wahlgruppe auf (Median: +0,4). Bei der Altersverteilung fällt auf, dass unter den befragten Studierenden weniger ältere Studierende (über 29 Jahre) zu finden sind.

Tabelle 6.4											
Verteilung nach Alter und Designgruppen (WS 2015/16)											
(Angaben in Prozent Mediane und Differenzen)											
	GG	Ges.	Anteile in %				Abweichungen von GG				
			Papier		Online		Ges.	Papier		Online	
			r	W	r	W		r	W	r	W
<= 21	27,3	27,3	27,3	30,2	29,2	25,1	0	0	+ 2,9	+ 1,9	- 2,2
22-23	19,2	22,2	22,4	21,1	21,7	21,7	+3,0	+ 3,2	+ 1,9	+ 2,5	+ 2,5
24-25	18,0	20,8	21,1	20,2	21,7	18,8	+2,8	+ 3,1	+ 2,2	+ 3,7	+ 0,8
26-27	12,7	13,5	13,9	10,9	12,3	12,6	+0,8	+ 1,2	- 1,8	- 0,4	- 0,1
28-29	7,7	6,9	6,7	7,0	6,1	8,0	-0,8	- 1,0	- 0,7	- 1,6	+ 0,3
=>30	15,0	9,4	8,6	10,5	9,0	13,8	-5,6	- 6,4	- 4,5	- 6,0	- 1,2
Median	23,4	23,5	23,5	23,4	23,4	23,8	+0,1	+ 0,1	0	0	+ 0,4
Universitäten											
<= 21	26,5	29,1	28,5	33,9	32,3	27,8	+2,6	+2,0	+7,4	+5,8	+1,3
22-23	18,1	22,0	22,2	20,6	21,9	22,1	+3,9	+4,1	+2,5	+3,8	+4,0
24-25	17,9	20,9	21,5	20,0	20,0	18,5	+3,0	+3,6	+2,1	+2,1	+0,6
26-27	13,1	13,6	14,3	10,5	12,3	11,8	+0,5	+1,2	-2,6	-0,8	-1,3
28-29	8,1	6,7	6,6	5,8	6,5	8,0	-1,4	-1,5	-2,3	-1,6	-0,1
=>30	16,4	7,7	6,9	9,1	7,1	11,9	-8,7	-9,5	-7,3	-9,3	-4,5
Median	23,6	23,4	23,4	23,1	23,2	23,5	-0,2	-0,2	-0,5	-0,4	-0,1
Fachhochschulen											
<= 21	26,0	21,2	22,2	18,3	21,8	18,3	-4,8	-3,8	-7,7	-4,2	-7,7
22-23	20,8	22,7	23,4	22,9	18,2	20,5	+1,9	+2,6	+2,1	-2,6	-0,3
24-25	18,4	20,2	19,9	20,9	27,3	19,6	+1,8	+1,5	+2,5	+8,9	+1,2
26-27	12,8	13,1	12,7	12,4	12,7	15,2	+0,3	-0,1	-0,4	-0,1	+2,4
28-29	7,7	7,5	7,0	11,1	5,5	8,0	-0,2	-0,7	+3,4	-2,2	+0,3
=>30	14,4	15,3	14,8	14,4	11,9	18,3	+0,9	+0,4	0	-2,5	+3,9
Median	23,3	24,0	23,9	24,3	24,3	24,6	+0,7	+0,6	+1,0	+1,0	+1,3

Quelle: Statistische Bundesamt 2016, Studierendensurvey 2016, AG Hochschulforschung, Universität Konstanz
GG: Grundgesamtheit; r: randomisierte Gruppen; W: Wahlgruppe

Getrennt nach der Hochschulart liegen die Mediane für das Alter an den Universitäten nahe bei der Grundgesamtheit, während an Fachhochschulen größere Abweichungen zu beobachten sind. So sind die Studierenden dort älter, es haben weniger junge, dafür mehr ältere Studierende teilgenommen.

Anhand der Verteilungen lässt sich erkennen, dass die befragten Studierenden an Universitäten tendenziell jünger als in der Grundgesamtheit sind, an Fachhochschulen dagegen erkennbar älter. Dadurch wird der Altersunterschied zwischen den Hochschularten noch etwas vergrößert.

Die auffällig geringeren Anteile an älteren Studierenden unter den befragten Studierenden an Universitäten korrespondieren mit deutlich höheren Anteile an jüngeren Studierenden. Anscheinend sind jüngere Studierende besser zu einer Teilnahme zu bewegen als ältere, möglicherweise, weil sie bislang seltener zu Befragungen eingeladen wurden. Die spezielle Gruppe der älteren Studierenden an Universitäten bleibt daher in diesen Befragungen unterrepräsentiert.

An den Fachhochschulen lässt sich dagegen eine umgedrehte Situation beobachten. Die jüngeren Studierenden nehmen seltener teil, die älteren dagegen häufiger. Die Gründe hierfür sind noch offen.

Verteilung nach Fach- und Hochschulsesemester

Hinsichtlich der Verteilung nach Semester sind unter den befragten Studierenden weniger Studienanfänger als in der Grundgesamtheit angegeben. Ebenso haben weniger Studierende in höheren Semestern teilgenommen als in der Grundgesamtheit aufgelistet sind.

Tabelle 6.5											
Verteilung nach Semester und Designgruppen (WS 2015/16)											
(Angaben in Prozent und Differenzen)											
	Anteile in %						Abweichungen von GG				
	GG	Ges.	Papier		Online		Ges.	Papier		Online	
			r	W	r	W		r	W	r	W
Fachsemester											
1-2	30,1	24,5	24,4	26,3	26,6	23,6	-5,6	-5,7	-3,8	-3,5	-6,5
3-4	23,8	24,6	24,9	26,6	21,7	22,9	0,8	1,1	2,8	-2,1	-0,9
5-6	19,3	21,4	21,3	18,9	21,7	23,0	2,1	2	-0,4	2,4	3,7
7-8	12,1	14,4	14,4	13,7	14,8	14,5	2,3	2,3	1,6	2,7	2,4
9-10	6,2	7,8	7,7	8,1	8,4	7,9	1,6	1,5	1,9	2,2	1,7
11-12	3,3	3,3	3,5	2,3	2,3	3,6	0	0,2	-1	-1	0,3
13-14	1,9	2,2	2,1	3,1	1,9	1,9	0,3	0,2	1,2	0	0
15+	3,3	1,8	1,7	1,1	2,7	2,6	-1,5	-1,6	-2,2	-0,6	-0,7
Hochschulsesemester											
1-2	18,2	13,3	13,4	14,4	13,5	11,9	-4,9	-4,8	-3,8	-4,7	-6,3
3-4	15,5	14,1	13,6	16,9	14,3	15,1	-1,4	-1,9	1,4	-1,2	-0,4
5-6	14,3	14,8	14,7	14,7	16,6	15,3	0,5	0,4	0,4	2,3	1
7-8	12,8	17,0	17,0	17,2	16,6	17,0	4,2	4,2	4,4	3,8	4,2
9-10	11,5	16,3	16,7	15,3	18,1	14,8	4,8	5,2	3,8	6,6	3,3
11-12	8,2	9,8	10,3	8,4	7,7	9,1	1,6	2,1	0,2	-0,5	0,9
13-14	5,9	6,6	6,7	5,7	5,0	7,3	0,7	0,8	-0,2	-0,9	1,4
15+	13,5	7,9	7,6	7,4	8,1	9,5	-5,6	-5,9	-6,1	-5,4	-4

Quelle: Statistische Bundesamt 2016, Studierendensurvey 2016, AG Hochschulforschung, Universität Konstanz
GG: Grundgesamtheit; r: randomisierte Gruppen; W: Wahlgruppe

In den Designgruppen setzt sich dieser Befund fort, es gibt unter den Befragten weniger Studienanfänger und weniger Studierende, die sich bereits im 15. Semester oder darüber hinaus befinden (vor allem hinsichtlich der Hochschulsesemester). In den Studienjahren dazwischen sind die Abweichungen für die Fachsemester nicht auffällig, während für die Hochschulsesemester vermehrt Studierende beteiligt waren, die sich noch in der regulären Studienendphase befinden (7-10 Hochschulsesemester).

Tabelle 6.6
Verteilung nach Semester, Hochschulart und Designgruppen (WS 2015/16)
 (Angaben in Prozent und Differenzen)

	GG	Ges.	Anteile in %				Abweichungen von GG				
			Papier		Online		Ges.	Papier		Online	
			r	W	r	W		r	W	r	W
Universitäten											
Fachsemester											
1-2	30,3	24,3	24,6	24,8	27,6	22,0	-6,0	-5,7	-5,5	-2,7	-8,3
3-4	23,5	24,3	24,8	27,8	18,8	21,1	0,8	1,3	4,3	-4,7	-2,4
5-6	18,9	21,1	20,7	18,8	22,9	24,1	2,2	1,8	-0,1	4,0	5,2
7-8	11,3	13,1	12,9	12,3	12,0	14,4	1,8	1,6	1,0	0,7	3,1
9-10	6,3	8,4	8,4	8,3	9,9	8,5	2,1	2,1	2,0	3,6	2,2
11-12	3,6	3,9	4,1	3,0	2,6	4,2	0,3	0,5	-0,6	-1,0	0,6
13-14	2,2	2,7	2,6	3,6	2,6	2,6	0,5	0,4	1,4	0,4	0,4
15+	3,9	2,2	2,0	1,4	3,6	3,2	-1,7	-1,9	-2,5	-0,3	-0,7
Hochschulsemester											
1-2	16,7	12,8	13,2	12,7	14,2	10,7	-3,9	-3,5	-4,0	-2,5	-6,0
3-4	14,4	13,0	12,5	18,1	12,6	12,1	-1,4	-1,9	3,7	-1,8	-2,3
5-6	13,1	13,9	13,4	14,5	15,8	15,9	0,8	0,3	1,4	2,7	2,8
7-8	12,4	15,6	15,3	15,3	14,2	17,2	3,2	2,9	2,9	1,8	4,8
9-10	12,6	17,0	17,5	15,7	18,9	15,2	4,4	4,9	3,1	6,3	2,6
11-12	9,6	10,9	11,6	9,0	9,5	9,5	1,3	2,0	-0,6	-0,1	-0,1
13-14	6,9	7,7	7,8	6,4	4,2	8,7	0,8	0,9	-0,5	-2,7	1,8
15+	14,3	9,0	8,7	8,2	10,5	10,6	-5,3	-5,6	-6,1	-3,8	-3,7
Fachhochschulen											
Fachsemester											
1-2	30,5	25,2	23,5	31,1	24,6	28,3	-5,3	-7,0	0,6	-5,9	-2,2
3-4	25,4	25,8	25,5	22,5	29,0	27,9	0,4	0,1	-2,9	3,6	2,5
5-6	21,0	22,3	23,6	19,2	18,8	20,0	1,3	2,6	-1,8	-2,2	-1,0
7-8	13,3	18,8	19,7	18,5	21,7	14,8	5,5	6,4	5,2	8,4	1,5
9-10	5,3	5,6	5,3	7,3	4,3	6,2	0,3	0	2,0	-1,0	0,9
11-12	2,0	1,2	1,2	0	1,4	1,7	-0,8	-0,8	-2,0	-0,6	-0,3
13-14	1,0	0,4	0,5	1,3	0	0	-0,6	-0,5	0,3	-1,0	-1,0
15+	1,4	0,6	0,7	0	0	1,0	-0,8	-0,7	-1,4	-1,4	-0,4
Hochschulsemester											
1-2	21,3	14,9	14,3	20,0	11,9	15,0	-6,4	-7,0	-1,3	-9,4	-6,3
3-4	19,6	17,9	17,2	12,7	19,4	23,1	-1,7	-2,4	-6,9	-0,2	3,5
5-6	18,1	17,8	19,4	15,3	17,9	13,3	-0,3	1,3	-2,8	-0,2	-4,8
7-8	14,5	22,0	23,2	23,3	22,4	16,8	7,5	8,7	8,8	7,9	2,3
9-10	10,4	13,9	13,8	14,0	16,4	14,0	3,5	3,4	3,6	6,0	3,6
11-12	6,2	6,1	5,7	6,7	3,0	7,7	-0,1	-0,5	0,5	-3,2	1,5
13-14	3,7	3,2	2,8	3,3	7,5	3,5	-0,5	-0,9	-0,4	3,8	-0,2
15+	6,2	4,1	3,6	4,7	1,5	6,6	-2,1	-2,6	-1,5	-4,7	0,4

Quelle: Statistische Bundesamt 2016, Studierendensurvey 2016, AG Hochschulforschung, Universität Konstanz
 GG: Grundgesamtheit; r: randomisierte Gruppen; W: Wahlgruppe

Die erkennbaren Abweichungen von der Grundgesamtheit bei den Studierenden in höheren Hochschulsemestern sind an Universitäten deutlich größer als an Fachhochschulen. Gleichzeitig sind die Abweichungen von der Grundgesamtheit bei den Studienanfängern an Universitäten hinsichtlich der Fachsemester größer als hinsichtlich der Hochschulsemester, während es an den Fachhochschulen tendenziell umgekehrt ist. An den Fachhochschulen fällt auf, dass die beiden randomisierten Gruppen insgesamt geringere Abweichungen aufweisen als die beiden Wahlgruppen.

Im Vergleich mit den Befunden hinsichtlich des Alters lässt sich nun konkretisieren, dass an Universitäten zwar jüngere Studierende häufiger teilnehmen, diese jedoch nicht automatisch zu den Studienanfängern gehören, die ihrerseits seltener an den Befragungen teilnehmen. Das bedeutet, dass sich jüngere Studierende häufiger bereits im zweiten oder dritten Studienjahr befinden.

An den Fachhochschulen nehmen jüngere Studierende seltener teil, und gleichzeitig sind auch weniger Studienanfänger im Sample. Hier zeigt sich eine Korrespondenz zwischen Alter und Semestern.

Allerdings sind die Zusammenhänge zwischen Alter und Semester nicht sehr deutlich. An Universitäten beläuft sich die Korrelation zwischen Alter und Fachsemester auf $R=0.33$, an Fachhochschulen liegt sie bei 0.20 . Bessere Werte erbringt die Korrelation zwischen Alter und Hochschulsemester: An Universitäten beträgt sie 0.67 und an Fachhochschulen 0.41 .

Die schwache Korrelation mit den Fachsemestern geht zu einem großen Teil auf die Masterstudierenden zurück, die bei Studienbeginn ihre Fachsemester neu zu zählen beginnen. Die schwächeren Korrelationen an den Fachhochschulen gehen auf die größeren Anteile an älteren Studierenden zurück, die nicht gleich nach dem Schulabschluss ein Studium aufnehmen.

Fazit

Die Strukturvergleiche der Stichproben mit der Grundgesamtheit lässt folgende Folgerungen zu:

Bestimmte Unterschiede bzw. Abweichungen der Stichprobenstruktur von der Grundgesamtheit stehen mit der Vorauswahl der theoretischen Stichprobe in Zusammenhang und stellen damit keine Fehler dar, sondern sind als methodisch gewollte bzw. akzeptierte Unterschiede zugunsten der Homogenität der Hochschulen im Sample anzusehen. Dies betrifft das Verhältnis der Hochschularten, der Fächergruppen und teilweise auch des Geschlechts.

Diesen Abweichungen wurde in der Vergangenheit dadurch gerecht, dass die Kennwerte der Studiensituation und der studentischen Orientierungen stets im Vergleich dieser Gruppenmerkmale analysiert wurden, sodass die Über- bzw. Unterrepräsentationen keine Relevanz hatten.

Die Abweichungen von der Grundgesamtheit hinsichtlich Alter und Semester verweisen darauf, dass vorrangig Studierende teilnehmen, die bereits Erfahrungen im Studium gesammelt haben, sich jedoch immer noch in der Regelstudienzeit befinden. Es sind dies somit die Studierenden, die zum einen ausreichend Erfahrung besitzen, um Aussagen machen können, und zum anderen keine Sonderstellungen einnehmen (wie z.B. Langzeitstudierende). Somit sind diese Studierenden sogar besonders geeignet für Aussagen über die konstitutionelle Grundgesamtheit.

Da die Größenordnungen der Abweichungen alle unter zehn Prozentpunkten, zumeist sogar unter fünf Prozentpunkten bleiben, und gleichzeitig die Effektstärken der Vergleiche vernachlässigbare Größen aufweisen, besteht kein Grund, die Repräsentativität der Stichprobenstrukturen anzuzweifeln.

6.1 Vergleich der Ergebnisse mit der 20. Sozialerhebung

Für die 20. Sozialerhebung wurde ebenfalls ein Stichprobendesign getestet, das Selektions- und Mode-Effekte über verschiedene Stichproben untersuchte. Dabei wurden die Stichproben in ihren strukturellen Merkmalen auf Ungleichheit getestet. Von den dafür verwendeten soziodemografischen Merkmalen waren die meisten unauffällig (vgl. Poskowsky u.a. 2014).

Signifikante Unterschiede zeigten sich in den univariaten Vergleichen bei:

- den randomisierten Befragungsgruppen (fremdselektiert Papier-Online): Geschlecht, Region, Hochschulart, Zentralität des Studiums.

- den selbstselektierten Befragungsgruppen (Wahlgruppe: Papier-Online): Alter, Geschlecht, Fächergruppen.
- dem modespezifischen Vergleich in den Befragungsgruppen (Papier - Wahl-Papier und Online - Wahl-Online):
Papier: Alter, Geschlecht, Studienform,
Online: Geschlecht, Region, Fächergruppe.

In den multivariaten Vergleichen bleiben als signifikante Unterschiede erhalten:

- Selbstselektion bei papierbasierten Befragungen: Geschlecht,
- Selbstselektion bei webbasierten Befragungen: Geschlecht, Region, computerbezogene Selbstwirksamkeit.

Im Vergleich zu den Befunden des Studierendensurveys sind folgende Aussagen zu machen:

In den univariaten Fällen:

Befunde in den randomisierten Befragungsgruppen:

- Die Verteilung des Alters wird im Survey signifikant, ebenso in der Sozialerhebung. Die Verteilungen der Altersklassen unterscheiden sich etwas, im Survey sind in der Papiergruppe mehr ältere Studierende (bereits ab 24 Jahre).
- Das Geschlecht wird in beiden Untersuchungen signifikant und die Verteilung weist in die gleiche Richtung.
- Die Region wird in der Sozialerhebung signifikant, weist jedoch im Survey keine Unterschiede auf.
- Die Hochschulart wird im Survey signifikant, in der Sozialerhebung nicht; die Verteilungen weisen in unterschiedliche Richtungen.
- Die Abschlussart wird im Survey signifikant, wurde in der Sozialerhebung aber nicht erhoben.
- Der Bildungshintergrund wird in der Sozialerhebung signifikant, im Survey nicht.
- Die Fächergruppen werden in der Sozialerhebung signifikant, im Survey nicht.

Befunde in den selbstselektierten Gruppen:

- Die Altersverteilung unterscheidet sich in der Sozialerhebung signifikant zwischen den Gruppen, im Survey nicht. Im Survey sind in der Papiergruppe mehr ganz junge und mehr ältere Studierende als in der Sozialerhebung.
- Das Geschlecht wird in beiden Erhebungen signifikant und die Verteilungsunterschiede weisen in die gleiche Richtung.
- Die Fächergruppen werden in beiden Erhebungen signifikant und die Unterschiede weisen alle in die gleiche Richtung.

Befunde bei modespezifischen Vergleichen: Papiergruppe

- Der Effekt des Geschlechts wird in beiden Erhebungen signifikant, und die Unterschiede weisen in die gleiche Richtung.
- Der Bildungshintergrund der Eltern wird nur im Survey signifikant, weist aber in der Sozialerhebung in die gleiche Richtung.
- Das Alter wird in der Sozialerhebung signifikant, nicht aber im Survey.

Unterschiede bei modespezifischen Vergleichen: Onlinegruppe

- Das Geschlecht wird in der Sozialerhebung signifikant, die Verteilung weist im Survey in die gleiche Richtung.
- Die Region wird in der Sozialerhebung signifikant, nicht im Survey. Die Verteilung weist tendenziell in die andere Richtung.
- Die Fächergruppen werden in der Sozialerhebung signifikant, nicht im Survey. Die Verteilung weist überwiegend in die gleiche Richtung.

Ein Vergleich der Effektstärken ist für die univariaten Fälle nicht möglich, da sie in der Sozialerhebung nicht mit angegeben sind.

Im multivariaten Fall:

Die Selektionseffekte in den Papiergruppen unterscheiden sich zwischen den Erhebungen:

- Das Geschlecht erhält im Survey keinen signifikanten Effekt, weist aber in die gleiche Richtung wie in der Sozialerhebung.
- Die Region wird im Survey signifikant, weist jedoch in die andere Richtung als in der Sozialerhebung.
- Die Kultur- und Sozialwissenschaften werden im Survey signifikant, weisen aber in die gleiche Richtung wie in der Sozialerhebung.

Die Selektionseffekte in den Onlinegruppen unterscheiden sich ebenfalls:

- Das Geschlecht wird in der Sozialerhebung signifikant und weist in die gleiche Richtung wie beim Survey.
- Die Region wird in der Sozialerhebung signifikant, weist aber in die gleiche Richtung.
- Die Computeraffinität wird in der Sozialerhebung signifikant, weist aber in die gleiche Richtung wie im Survey, wobei jedoch unterschiedliche Fragen zur Operationalisierung verwendet wurden.
- Computerbezogene Nutzungsmotive, die in der Sozialerhebung signifikant werden, existieren im Survey nicht.

Die Effektstärken der multivariaten Modelle sind in beiden Erhebungen sehr gering. In der Sozialerhebung erreichen sie maximal drei, im Survey maximal fünf Prozent.

Vergleiche zwischen den Erhebungen zur Grundgesamtheit

Zu Beginn des Kapitels wurden die Abweichungen ausgewählter Merkmale zur Grundgesamtheit dargestellt. Solche Abweichungen wurden auch von der Sozialerhebung dokumentiert. Im Vergleich beider Untersuchungen ergeben sich folgende Ergebnisse:

- Die Abweichungen hinsichtlich Hochschulart gehen in die gleiche Richtung, sind im Survey für die Papiergruppen allerdings größer.
- Die Abweichungen beim Geschlecht gehen in beiden Untersuchungen für die Papiergruppen in die gleiche Richtung, bei den Onlinegruppen unterscheiden sie sich. Im Survey sind die Frauen auch in der Wahl-Onlinegruppe überrepräsentiert.
- Die Abweichungen in den Fächergruppen weisen bis auf kleinere Abweichungen in die jeweils gleichen Richtungen.
- Die Abweichungen in den Altersklassen weisen zumeist in die gleiche Richtung. In der Sozialerhebung sind die Abweichungen in den einzelnen Klassen größer.

Fazit

Der Vergleich der Befunde mit der Sozialerhebung zur strukturellen Gleichheit der Stichproben stellt heraus, dass die meisten gefundenen signifikanten Unterschiede ähnliche Ergebnisse aufzeigen wie in den Analysen zum Survey. Der größte Effekt ist in beiden Erhebungen der Unterschied zwischen Männern und Frauen, wobei letztere deutlich häufiger Papierfragebogen bevorzugen. Im Vergleich beider Erhebungen ist die Überrepräsentativität der Frauen im Survey durchgängiger vorhanden.

Unterschiede zwischen den Studien fallen nur in drei Bereichen auf: in der Altersverteilung, der Hochschulart und der Region. Dabei treten die Unterschiede im Alter eher generell auf, während die Differenzen bezüglich Hochschulart und Region nur in einzelnen Vergleichen erscheinen. In der Sozialerhebung sind in den Onlinegruppen im Vergleich zum Survey mehr Studierende der neuen Länder, aber weniger Studierende von Fachhochschulen. Bei diesen Unterschieden muss jedoch die theoretische Stichprobe des Studierendensurveys in Betracht gezogen werden, die nur Fachhochschulen eines bestimmten Fächerprofils einbezieht, von denen nur drei Fachhochschulen in den neuen Ländern liegen. Hinsichtlich der Altersverteilung liegt der Survey näher an der Grundgesamtheit. Insgesamt sind die Unterschiede jedoch eher gering.

Damit kann geschlossen werden, dass nicht nur die unterschiedlichen Designstichproben untereinander vergleichbar sind und somit als Gesamtstichprobe analysiert werden dürfen, ohne dass damit die Zeitreihe gefährdet würde, sondern es lässt sich auch folgern, dass die Stichproben zwischen den beiden Erhebungen, Studierendensurvey und Sozialerhebung, vergleichbare Strukturen aufweisen, die Befunde somit vergleichbar werden.

7. Vergleich der Designgruppen an inhaltlichen Indikatoren

Eine wichtige Voraussetzung für die gemeinsame Nutzung der unterschiedlichen Designgruppen war die Strukturgleichheit der Designstichproben. Für weitere Analysen zur Prüfung der (Un-)Gleichheit der Untergruppen sollen nun inhaltliche Indikatoren verwendet werden. Dabei werden solche Merkmale ausgewählt, die:

- 1) bedeutsame stabile Unterschiede nach wichtigen Strukturmerkmalen wie z.B. der Hochschulart aufweisen;
- 2) Fragen, die lange Itembatterien beinhalten;
- 3) Fragen, deren Darstellung online abweicht von der Darstellung im Papierfragebogen.

Damit sollen zum einen differenziertere Analysen geprüft werden und zum anderen jene Darstellungen speziell kontrolliert werden, die in der Online-Darstellung von der Papierform abweichen. Denn Mode-Effekte sollten sich gerade an solchen Designunterschieden nachweisen lassen.

Für nominale Items werden Chi-Quadrat-Tests durchgeführt, sowie als Effektgröße Cramers V berechnet. Für intervallskalierte Items werden Varianzanalysen durchgeführt und als Effektstärkenmaß η^2 angegeben. Dabei werden aus den Designgruppen zwei Faktoren kreiert, die zum einen den Mode-Effekt und zum anderen den Selektionseffekt abbilden. Der Vorteil der zweifaktoriellen Varianzanalyse ist dabei die Möglichkeit der Darstellung des Interaktionseffektes zwischen beiden Faktoren. Bei Analysen, die signifikante Haupt- oder Interaktionseffekte ergeben, werden zusätzlich jene Items kontrolliert, die sich in den Strukturvergleichen als unterschiedlich herausgestellt haben. Dies sind vor allem das Geschlecht und das Alter.

Ziel der Analysen ist der Nachweis, dass sich die Merkmale der Studiensituation nicht zwischen den Designgruppen unterscheiden, sodass sich die Befragung der Studierenden zur Studiensituation und den studentischen Orientierungen ebenso über eine Online-Erhebung durchführen lässt wie über einen Papierfragebogen, ohne dass die Zeitreihen dadurch gefährdet werden. Dabei geht es allerdings nicht darum, dass gar keine Unterschiede auftreten – das wird nach der wissenschaftlichen Logik bereits zufällig aufgrund der großen Anzahl der Analysen passieren – sondern, dass Unterschiede nicht häufig und wenn dann nicht bedeutsam oder systematisch auftreten. Daher werden jeweils auch die Effektstärken und die relativen Differenzen der Unterschiede betrachtet und interpretiert. Und es werden weitergehende Analysen angeschlossen, um zu klären, worauf Unterschiede zurückgehen. Diese Klärung bestehender Unterschiede ist besonders wichtig, um einerseits Mode- und Selektionseffekte und andererseits Folgerungen für die Studiensituation korrekt einordnen zu können.

7.1 Prüfung nominaler bzw. ordinaler Merkmale

Im Studierendensurvey kommen unterschiedliche Fragearten zum Einsatz. Ein Teil der Fragen hat feste Antwortvorgaben, die ein nominales oder ordinales Skalenniveau besitzen. Aus dieser Gruppe an Fragearten werden vier Fragen ausgewählt und nach den Designgruppen überprüft.

Studiensicherheit

Die Sicherheit, mit der ein Studium aufgenommen wurde, ist ein Merkmal, das in der Zeit sehr stabil ist und deutliche Unterschiede zwischen Universitäten und Fachhochschulen aufweist. Im Studierendensurvey wird dieses Merkmal nominal bzw. ordinal anhand von vier Antwortvorgaben erhoben, die unterscheiden, wie sicher sich die Studierenden vor dem Studium waren, dass sie studieren wollen.

Die bisherigen Befunde (bis zur 12. Erhebung im WS 2012/13) zeigen, dass es an Universitäten für 58 % der Studierenden von vorneherein feststand, dass sie studieren wollen, und weitere 29 % waren sich ziemlich sicher. An Fachhochschulen ist die Studiersicherheit deutlich geringer, nur für etwa jeden Dritten stand von vorneherein fest zu studieren, und nochmal so viele waren sich ziemlich sicher. Damit haben an Universitäten 13 %, an Fachhochschulen 31 % ein Studium aufgenommen, obwohl sie ursprünglich lange unsicher waren bzw. gar nicht studieren wollten.

Die Befunde der 13. Erhebung im WS 2015/16 unterstützen die bisherigen Befunde. In allen Designgruppen ist an Universitäten bei rund drei Fünftel der Studierenden eine hohe Studiersicherheit nachzuweisen, an Fachhochschulen bei etwa einem Drittel. Die größten Unterschiede sind zwischen den beiden Onlinegruppen festzustellen. In der Wahlgruppe waren mehr Studierende von vorneherein sicher, studieren zu wollen. Auf Ebene der Designgruppen wird dieser Unterschied aber nicht signifikant. Die Verteilungen zwischen den Designgruppen erweist sich für die Gesamtstichprobe nach Chi-Quadrat-Testung als gleich.

Tabelle 7.1										
Studiersicherheit nach Designgruppen (WS 2015/16)										
(Angaben in Prozent, Chi-Quadrat, Cramers V)										
Studiersicherheit	Papier		Papier-Wahl		Online-Wahl		Online		Gesamt	
wollte eigentlich nicht stud.	3		3		3		4		3	
war mir lange Zeit unsicher	14		14		10		14		14	
war ziemlich sicher	31		32		30		32		31	
stand von vorneherein fest	52		51		57		50		52	
Chi ²	10,16 n.s.									
Nach Hochschulart	Papier		Papier-Wahl		Online-Wahl		Online		Gesamt	
	Uni	FH	Uni	FH	Uni	FH	Uni	FH	Uni	FH
wollte eigentlich nicht stud.	2	6	3	6	1	7	1	10	2	7
war mir lange Zeit unsicher	11	25	11	21	9	13	12	23	11	24
war ziemlich sicher	29	37	28	44	27	39	32	33	29	37
stand von vorneherein fest	58	32	58	29	63	41	55	34	58	32
	Chi ² =420 V=0,24									
Universitäten	Chi ² =10.4 n.s.									
Fachhochschulen	Chi ² =19,6 * (Cramers V = 0.06)									
	Papier mit Online: Chi ² = 10,5 * (V=0,08)									
Vergleiche der HS-art	** für alle Designgruppen									
Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 7										
1) ** = auf 1 %-Niveau signifikant, * = auf 5 %-Niveau signifikant; n.s. = nicht signifikant										

Die höhere Studiersicherheit der Online-Wahlgruppe ist sowohl bei den Universitäten als auch bei den Fachhochschulen festzustellen. Innerhalb der Designgruppen werden alle Vergleiche der Hochschulart hoch signifikant. Hinsichtlich der Verteilung sind für die Universitäten keine Unterschiede zwischen den Designgruppen auszumachen, für die Fachhochschulen wird ein Unterschied auf dem Fünf-Prozent-Niveau erkennbar. Der Einzelvergleich stellt heraus, dass nur der Vergleich zwischen der Papiergruppe und der Onlinegruppe signifikant wird. Allerdings bleibt die Effektgröße sehr gering (V=0,08). Das signifikante Ergebnis spiegelt damit den absolut geringsten Unterschied wider und

nicht den größten Unterschied zwischen den beiden Wahlgruppen. Dies ist der Stichprobengröße geschuldet, was darauf verweist, wie sensibel Signifikanztests hinsichtlich der Stichprobengrößen sind.

Bei weitergehender Prüfung nach dem Geschlecht stellt sich heraus, dass der gefundene signifikante Unterschied nur bei den Männern auftritt ($\chi^2 = 8,2$, $V=0,124$) und dort nur bei den älteren Studenten über 26 Jahren ($\chi^2=9,4$, $V=.249$).

Stand der Berufswahl

Bei der Frage, ob die Studierenden schon wissen, welchen Beruf sie später ergreifen möchten, lassen sich zeitlich stabile Differenzen zwischen den Hochschularten feststellen. Studierende an Universitäten zeigen sich bezüglich ihrer Berufswahl noch häufiger unentschlossen.

Tabelle 7.2										
Stand der Berufswahl nach Designgruppen (WS 2015/16)										
(Angaben in Prozent, Chi-Quadrat, Cramers V)										
Wissen Sie schon, welchen Beruf Sie ergreifen wollen?	Papier		Papier-Wahl		Online-Wahl		Online		Gesamt	
nein, ist noch offen	34		34		35		33		34	
ja, mit einiger Sicherheit	36		37		40		39		37	
ja, mit großer Sicherheit	30		30		25		28		29	
Chi ²										3,41 n.s.
nach Hochschulart	Papier		Papier-Wahl		Online-Wahl		Online		Gesamt	
	Uni	FH	Uni	FH	Uni	FH	Uni	FH	Uni	FH
nein, ist noch offen	36	27	35	30	35	33	36	25	36	27
ja, mit einiger Sicherheit	34	44	35	41	39	45	36	46	34	44
ja, mit großer Sicherheit	30	29	30	29	26	22	28	29	30	29
										Chi ² = 54,9 V=0,09
Universitäten	n.s.									
Fachhochschulen	n.s.									
nach Mode-Gruppen	Papier (0,1): $\chi^2=45,4$, $V=0,09$; Online (2-4): 8,7, $V=0,09$									
Vergleiche der HS-art	Innerhalb der Designgruppen weist nur die Papiergruppe Effekte auf									
Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 66										
1) ** = auf 1 %-Niveau signifikant, * = auf 5 %-Niveau signifikant; n.s. = nicht signifikant										

Für die Gesamtstichprobe sind keine Effekte nach den Designgruppen festzustellen, d.h. die Verteilungen sind gleich. Dies ist auch bei der Kontrolle der Hochschulart zu beobachten.

Bei Unterscheidung nach der Hochschulart zeigen sich die Differenzen zwischen Universitäten und Fachhochschulen in allen Designgruppen. Allerdings erbringt nur die Papiergruppe einen signifikanten Unterschied. Werden die Mode-Gruppen zusammengenommen, dann werden die Effekte zwischen den Hochschularten wieder signifikant.

Die bisher gefundenen Effekte (häufigere Offenheit an Universitäten) sind auch in den Designgruppen zu finden.

Studienstrategien

Im Studierendensurvey sollen die Befragten für verschiedene Studiererfahrungen angeben, ob sie hinsichtlich der persönlichen Entwicklung und hinsichtlich der Berufsaussichten einen Nutzen erbringen. Diese Strategien sind im Papierfragebogen nebeneinander dargestellt, sodass die Befragten für jedes Item den Nutzen für beide Entwicklungen auf einer Zeile angeben können. In der Onlineversion war diese Darstellung nicht möglich, sodass die beiden Strategieeinschätzungen untereinander abgebildet werden mussten. Insofern besteht bei dieser Frage ein Darstellungsunterschied zwischen den beiden Modi.

Tabelle 7.3

Studienstrategien nach Designgruppen (WS 2015/16)

(Kategorien von 1 = eher nachteilig, 2 = wenig nützlich, 3 = nützlich, 4 = sehr nützlich, 5 = kann ich nicht beurteilen; Angaben in Prozent für Ausprägung „sehr nützlich“)

Strategien	Papier	Papier- Wahl	Online- Wahl	Online	p	Chi ²	V
Nutzen für persönliche Entwicklung							
an Forschungsprojekt teilnehmen	35	35	37	35	n.s.		
zeitweise im Ausland studieren	52	53	45	45	**	45,7	0,05
Arbeiterfahrungen außerhalb der HS	57	61	56	54	n.s.		
sich hochschulpolitisch engagieren	12	15	8	8	**	41,1	0,04
schneller Studienabschluss	20	21	22	23	**	30,7	0,04
möglichst gute Examensnote	27	28	25	28	n.s.		
einen Master anschließen	38	40	33	37	n.s.		
Nutzen für Berufsaussichten							
an Forschungsprojekt teilnehmen	40	40	40	44	**	30,1	0,04
zeitweise im Ausland studieren	44	43	45	51	**	33,0	0,04
Arbeiterfahrungen außerhalb der HS	61	63	63	67	**	29,1	0,04
sich hochschulpolitisch engagieren	8	10	12	15	**	99,4	0,07
schneller Studienabschluss	33	33	35	39	**	51,8	0,05
möglichst gute Examensnote	58	58	57	58	n.s.		
einen Master anschließen	54	53	47	52	*	22,6	0,03

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 16
¹⁾ ** = auf 1 %-Niveau signifikant, * = auf 5 %-Niveau signifikant; n.s. = nicht signifikant

Die Verteilungstests zwischen den Designgruppen ergeben für die erste Fragebatterie (Nutzen für persönliche Entwicklung) drei signifikante Ergebnisse. Die Antworten zu den Items Auslandsstudium, hochschulpolitisches Engagement und rascher Studienabschluss unterscheiden sich zwischen den Gruppen. Dabei halten die Studierenden der Papiergruppen das Auslandsstudium und das hochschulpolitische Engagement für nützlicher als die Onlinegruppen, während der rasche Abschluss den Onlinegruppen mehr Nutzen erbringt. Die Effektstärken sind allerdings alle sehr gering. Die absoluten Differenzen bleiben unter zehn Prozentpunkten, bei raschem Abschluss sogar unter fünf Prozentpunkten.

In der zweiten Fragebatterie (Nutzen für Berufschancen) werden die Verteilungstests für alle Items signifikant, außer beim möglichst guten Abschluss. Dabei fällt auf, dass die Studierenden der Onlinegruppen die Strategien für nützlicher halten, außer bei dem Item „Anschluss eines Masters“. Innerhalb der Onlinegruppen fällt auf, dass die randomisierten Onlinegruppen noch häufiger einen Nutzen sehen als die Studierenden der Wahl-Onlinegruppe. Die Effektstärken sind allerdings wiederum alle sehr gering. Die absoluten Differenzen bleiben alle unter zehn Prozentpunkten.

Unterteilt nach der Hochschulart fallen die Verteilungseffekte an den Fachhochschulen für die erste Fragebatterie weg. An Universitäten bleiben die Effekte für das Auslandsstudium und das hochschulpolitische Engagement erhalten. Der Effekt für den raschen Abschluss verringert sich auf das 5 %-Niveau, und es kommt ein zusätzlicher Effekt auf dem 5 %-Niveau für das Item Arbeitserfahrung außerhalb der Hochschule hinzu. Für die zweite Fragebatterie bleiben die Effekte an Universitäten erhalten, allerdings verringern sich drei Effekte auf das 5 %-Niveau (Auslandsstudium, Arbeitserfahrungen, Master). An Fachhochschulen bleibt nur der Effekt für das hochschulpolitische Engagement bestehen.

Unterschieden nach dem Geschlecht fällt für die zweite Fragebatterie der Effekt für das Masterstudium sowohl bei Männern wie Frauen weg. Alle anderen Effekte bleiben erhalten.

Wird nur der Mode-Effekt betrachtet, also beide Papiergruppen und beide Onlinegruppen zusammengefasst, dann treten in der ersten Fragebatterie vier Verteilungseffekte auf. Neben dem Auslandsstudium, der Hochschulpolitik und dem raschen Abschluss werden nun auch die Verteilungsunterschiede bei der Güte der Examensnote signifikant, wenn auch nur auf dem 5 %-Niveau. Für die zweite Fragebatterie bleiben alle Effekte vorhanden.

Werden die vier ordinalen Ausprägungen verwendet, dann reduziert eine parameterfreie Varianzanalyse (Kruskall-Wallis) die Effekte auf vier Items: die Forschungsteilnahme und das hochschulpolitische Engagement für die Berufsaussichten sowie das Engagement und den raschen Abschluss für die persönliche Entwicklung.

Eine zweifaktorielle Varianzanalyse erbringt sieben Haupteffekte für den Modus, aber keine für den Selektionseffekt oder die Interaktion. Dabei fehlten die Effekte für die Forschung, den Master und den raschen Abschluss für die Berufsaussichten, dafür kommt der Effekt für das gute Examen in persönlicher Hinsicht hinzu. Da hier aber die Voraussetzungen nicht erfüllt sind, wird die Analyse nicht weiter interpretiert, sondern nur zum Vergleich herangezogen.

Bei sämtlichen Analysen verweisen die Effektstärken auf sehr geringe Größenordnungen, sodass die Unterschiede zwischen den Designgruppen vernachlässigbar sind. Die Zunahme der Unterschiede in der zweiten Fragebatterie deutet allerdings auf einen anderen Effekt hin, nämlich auf Mode-Effekte aufgrund von Darstellungsunterschieden zwischen Papier- und Onlineversion. Da in der Onlineversion die beiden Teilfragen untereinander stehen, wiederholen sich die Fragen und Items bis auf einen Teil der Fragestellung. Dies kann zu Irritationen führen, da nicht gleich zu erkennen sein könnte, warum die Fragen sich wiederholen. In der Papierversion besteht dieses Problem nicht, da die Teilfragen nebeneinander stehen.

Möglichkeiten der Weiterbildung

In dieser Frage werden fünf Möglichkeiten der Weiterbildung nachgefragt, zum einen wie die Studierenden die Angebote bisher genutzt haben, und zum anderen wie sie diese zukünftig nutzen wollen. Dabei ist auch in diesem Beispiel die Darstellung zwischen den Modi unterschiedlich. In der Papiervariante konnten beide Fragen nebeneinander dargestellt werden, in der Onlineversion stehen die beiden Fragen untereinander.

In der ersten Fragebatterie (bisher genutzt) werden die Verteilungsunterschiede zwischen den Designgruppen nur bei einem Item signifikant, und zwar bei der Frage nach den Fremdsprachenkursen. Dabei haben Studierende aus den Papiergruppen diese Form der Weiterbildung häufiger genutzt. In der zweiten Fragebatterie (zukünftig nutzen) werden drei Effekte signifikant: Neben den Fremdsprachenkursen werden nun auch die Vorlesungen anderer Fächer sowie die öffentlichen Vorträge signifikant unterschiedlich. Dabei ist nur für die Fremdsprachen ein deutlicher Mode-Effekt zu beobachten.

Unterschieden nach dem Geschlecht fällt in der ersten Fragebatterie bei den Männern der Unterschied in den Fremdsprachen weg. Sie weisen damit keine Effekte mehr auf. Bei den Frauen werden auf dem 5 %-Niveau neben den Fremdsprachen auch die Berufsveranstaltungen signifikant. Bei der zweiten Fragebatterie bleiben die gefundenen Effekte bei den Männern erhalten. Bei den Frauen wird nur der Unterschied bei den Fremdsprachen signifikant.

Unterschieden nach der Hochschulart bleiben in beiden Fragebatterien an Universitäten die Effekte erhalten, während an Fachhochschulen keine Effekte auftreten.

Tabelle 7.4
Weiterbildungsmöglichkeiten nach Designgruppen (WS 2015/16)
(Kategorien für bisher genutzt: von 1 = nein, 2 = ja, ab und zu, 3 = ja, häufiger; Skala für zukünftig nutzen: von 1 = nein, 2 = ja, vielleicht, 3 = ja, auf jeden Fall; Angaben in Prozent für Ausprägung „ja, häufiger“ bzw. „ja, auf jeden Fall“)

Strategien	Papier-		Online-		p	CHI	V
	Papier	Wahl	Wahl	Online			
bisher genutzt							
Vorlesungen anderer Fächer	9	8	8	6	n.s.		
öffentliche Vorträge	9	11	6	7	n.s.		
Fremdsprachenkurse	21	21	16	16	**	20,5	0,04
Kurse zu Schlüsselqualifikationen	13	11	10	10	n.s.		
Veranstaltungen für Berufsübergang	4	5	2	3	n.s.		
zukünftig nutzen							
Vorlesungen anderer Fächer	17	20	16	19	**	20,7	0,04
öffentliche Vorträge	20	24	16	25	**	34,5	0,05
Fremdsprachenkurse	31	33	23	23	**	59,2	0,07
Kurse zu Schlüsselqualifikationen	22	25	21	22	n.s.		
Veranstaltungen für Berufsübergang	20	18	21	19	n.s.		

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 19
¹⁾ ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant

Eine Varianzanalyse erbringt im parameterfreien Falle (Kruskal-Wallis) nur die drei bekannten Effekte in der zweiten Fragebatterie, da die bisherigen Fremdsprachen wegfallen; im parametrischen Falle kommt ein zusätzlicher Mode-Effekt für das bisherige Kompetenztraining hinzu.

Werden nur die Mode-Effekte untersucht, dann kommen in der ersten Fragebatterie zu den signifikanten Unterschieden bei den Fremdsprachen vier weitere Effekte auf dem 5 %-Niveau hinzu, nur die öffentlichen Vorträge weisen dann keine Effekte auf. Für die zweite Fragebatterie bleiben die drei gefundenen Effekte erhalten. Alle gefundenen Effekte weisen nur sehr geringe Effektstärken auf, und die absoluten Differenzen erreichen nur in einem Falle zehn Prozentpunkte. Damit können die Unterschiede vernachlässigt werden. Die Zunahme der Mode-Effekte in der zweiten Fragebatterie verweisen auch in diesem Beispiel darauf, dass es sich um Darstellungsunterschiede handeln könnte.

7.2 Prüfung intervallskalierter Merkmale der Studiensituation

Die folgenden Übersichten stellen die Ergebnisse der varianzanalytischen Prüfungen für Itembatterien aus fünf ausgewählten Fragen des Studierendensurveys dar. Sie alle umfassen siebenstufige Skalen und sind intervallskaliert. Für die Vergleiche wurden aus den Designgruppen zwei Faktoren gebildet. Der Modefaktor trennt die Papiergruppen von den Onlinegruppen, und der Selektionsfaktor trennt die randomisierten Gruppen von den selbstselegierten Wahlgruppen. Somit können über eine zweifaktorielle Varianzanalyse Haupteffekte (HE) und Interaktionseffekte (IA) betrachtet werden.

Damit die Varianzanalyse durchgeführt werden darf, werden vorher die Voraussetzungen geprüft und angegeben. Falls bestimmte Voraussetzungen nicht erfüllt sind, werden alternative Vorgehensweisen (z.B. parameterfreie Verfahren) durchgeführt.

Charakterisierung des Hauptstudienfaches

Bei dieser Frage werden die Studierenden anhand von zehn Items gebeten anzugeben, wie stark ihr Hauptfach durch verschiedene Merkmale gekennzeichnet ist. Die Merkmale beziehen sich dabei auf Leistungs- und Gliederungsaspekte, auf Bezüge und das soziale Klima.

Kennzeichen Fach	Papier (0,1)		Online (2-4)		p-Werte ¹⁾	Eta
	Random. Papier (0)	selbst Wahl (1)	selbst Wahl (2)	Random. Lang (3,4)		
hohe Leistungsansprüche	4,33	4,32	4,36	4,26	n.s.	
gute Studiengliederung	3,72	3,77	3,72	3,75	n.s.	
klare Prüfungsanforderungen	3,67	3,76	3,71	3,67	n.s.	
Forschungsbezug der Lehre	3,58	3,61	3,52	3,57	n.s.	
enger Praxisbezug	3,27	3,36	3,28	3,36	n.s.	
gute Berufsvorbereitung	2,81	2,80	2,85	2,85	n.s.	
Konkurrenz unter Studierenden	2,25	2,15	2,15	2,08	**HE Mode	0,11
gute Beziehungen zu Lehrenden	3,86	3,99	3,91	3,96	n.s.	
überfüllte Lehrveranstaltungen	2,29	2,29	2,22	2,10	** HE Mode	0,12
internationale Beziehungen	2,66	2,64	2,82	2,71	n.s.	
Bartlett-Tests n.s.						
Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 13						
¹⁾ ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant;						
HE = Haupteffekt, IE= Interaktionseffekt						

Bei den zehn nachgefragten Merkmalen ergeben sich bei acht Items keine Unterschiede zwischen den Designgruppen. Nur bei zwei Items fallen Effekte auf, die beide einen Haupteffekt für den Modus Papier oder Online abbilden. Der zweite Haupteffekt für die Selektion sowie die Interaktionen bleibt nicht signifikant.

Beide Items mit signifikanten Haupteffekten gehören zum sozialen Klima. Zum einen die Kennzeichnung des Faches durch Konkurrenz unter den Studierenden, und zum anderen durch die Überfüllung der Lehrveranstaltungen. Beide Merkmale werden von Studierenden der Papiergruppen als ein etwas stärkeres Kennzeichen des Faches betrachtet als von Studierenden der Onlinegruppen. Die Effektstärken fallen allerdings eher gering aus (Eta=0,11 bzw. 0,12). Nur etwa 0,1 % der Datenvariationen können über das Modell, hier die Unterteilung nach dem Modus, erklärt werden.

Unterschieden nach dem Geschlecht bleiben der Haupteffekt für die Konkurrenz nur bei den Studentinnen und der Haupteffekt für die Überfüllung nur bei den männlichen Studierenden erhalten (Eta=0,29 bzw. 0,13). Das Alter als Kovariate eingeführt weist keine Effekte auf.

Werden die tatsächlichen Mittelwertunterschiede betrachtet, dann erreichen diese maximal eine Differenz von 0,17 bzw. 0,19 Skalenpunkten, was nur einem Unterschied von 2,4 % bzw. 2,7 % der Skalenbreite entspricht.

Gründe für die Wahl des Hauptfaches

Die Motive der Fachwahl werden im Studierendensurvey mit sieben Items erhoben. Die Merkmale unterscheiden dabei intrinsische und extrinsische Motive sowie den Berufswunsch und die Auswech-lösung.

Bei zwei dieser Items sind Effekte festzustellen. Die eigene Begabung weist einen Haupteffekt für den Modus auf, während bei den Einkommenschancen ein Interaktionseffekt zu beobachten ist. Die an-deren fünf Items sind zwischen den Designgruppen gleich.

Tabelle 7.6						
Motive der Fachwahl nach Effekt: Mode – Selektion (WS 2015/16)						
(Mittelwerte)						
Kennzeichen Fach	Papier (0,1)		Online (2-4)		p-Werte ¹⁾ für HE u. IE	Eta
	Random. Papier (0)	selbst Wahl (1)	selbst Wahl (2)	Random. Lang (3,4)		
spezielles Fachinteresse ²⁾	5,09	5,18	5,09	5,10	n.s.	
eigene Begabung	4,66	4,68	4,75	4,73	* HE Mode	0,07
Einkommenschancen	3,41	3,31	3,54	3,35	* IA	0,07
fester Berufswunsch	3,19	3,28	3,19	3,22	n.s.	
Chance auf sicheren Arbeitsplatz	3,76	3,64	3,86	3,78	n.s.	
Chance auf Führungsposition	2,54	2,59	2,66	2,48	n.s.	
Auswech-lösung NC	0,74	0,63	0,70	0,78	n.s.	
Bartlett-Tests n.s. (bei 6 Items)						
Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 8						
1) ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant;						
HE = Haupteffekt, IE= Interaktionseffekt						
2) ungleiche Varianzen, Varianzhomogenität ist nicht erfüllt						

Beim dem Item spezielles Fachinteresse wird der Test auf Varianzhomogenität signifikant, d.h. die Voraussetzung der Varianzanalyse ist verletzt. Im Einzelvergleich unterschieden sich die Varianzen der Papier-Wahlgruppe von den anderen Designgruppen sowie die Varianzen der beiden randomi-sierten Gruppen. Anhand von t-tests für ungleiche Varianzen ergibt nur der Vergleich zwischen den beiden Papiergruppen ein signifikantes Ergebnis.

Der Haupteffekt für den Modus bei der eigenen Begabung geht auf den Unterschied im Einzelver-gleich zwischen den beiden randomisierten Gruppen zurück. Der Interaktionseffekt beim Item Ein-kommenschancen tritt auf Ebene der Einzelvergleiche zwischen den beiden Wahlgruppen sowie zwi-schen den beiden Onlinegruppen auf. Die Effektstärken liegen bei beiden Effekten bei Eta=0,07. Die tatsächlichen Mittelwertunterschiede bleiben allerdings unter 5 % der Skalenbreite.

Im einfaktoriellen Design, indem nur die vier Designgruppen gegeneinander getestet werden, sind keine Effekte signifikant. Im zweifaktoriellen Design verschwinden die Effekte bei Unterscheidung nach dem Geschlecht. Nur noch der Haupteffekt für die eigene Begabung bleibt bei den männlichen Studierenden erhalten (Eta=0,65).

Studienqualität

Im Survey werden sechs Grundelemente der Studienqualität als Indikatoren für dahinterstehende Dimensionen direkt erfragt. In der Zeitreihe weisen sie einen stetigen positiven Trend auf, und sie weisen deutliche Unterschiede zwischen den Hochschularten auf.

Tabelle 7.7						
Grundelemente der Studienqualität nach Effekt: Mode – Selektion (WS 2015/16)						
(Mittelwerte)						
Merkmale	Papier (0,1)		Online (2-4)		p-Werte ¹⁾ HE u. IE	eta
	Random. Papier (0)	selbst Wahl (1)	selbst Wahl (2)	Random. Lang (3,4)		
Bilanz Studienqualität						
Aufbau	4,80	4,87	4,81	4,83	n.s.	
Inhalt	5,24	5,28	5,16	5,26	n.s.	
Durchführung	4,78	4,90	4,67	4,87	** IE	0,13
Beratung	4,61	4,73	4,54	4,69	* IE	0,08
Ausstattung	4,74	4,73	4,77	4,74	n.s.	
Ertrag gesamt	5,18	5,25	5,12	5,17	n.s.	
Bartlett-Tests n.s						
Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 56						
¹⁾ ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant;						
HE = Haupteffekt, IE= Interaktionseffekt						

Bei zwei Grundelementen treten Interaktionseffekte, aber keine Haupteffekte auf. Beim Item Durchführung werden alle Vergleiche signifikant außer Papier zu Wahl-Online und Wahl-Papier zu Online. Bei der Beratung werden zwei Einzelvergleiche signifikant: die beiden Papiergruppen untereinander sowie die beiden Wahlgruppen untereinander.

Wird das Geschlecht als dritter Faktor eingeführt, dann ergibt sich bei allen Items außer bei der Betreuung ein deutlicher Haupteffekt für das Geschlecht. Die beiden Interaktionseffekte bleiben bestehen, hinzu kommt ein Interaktionseffekt 1. Art zwischen Mode und Geschlecht.

Getrennt nach Geschlecht ist bei den Männern nur ein Effekt auszumachen, ein Haupteffekt für den Modus bei dem Item Ertrag (Eta=0,18). Bei den Studentinnen bleiben die beiden Interaktionseffekte für Durchführung und Betreuung bestehen, d.h. sie gehen also auf die Frauen zurück. Die größte Skalendifferenz erreicht dabei 0,4 Skaleneinheiten, was etwa sechs Prozentpunkten Unterschied entspricht. Die Effektstärken liegen bei Eta 0,27 bzw. 0,13.

Erfahrene Belastungen im Studium

Ein Fragebereich beschäftigt sich mit möglichen Belastungsfaktoren. Dabei werden die Studierenden danach gefragt, wie stark sie sich durch Leistungs-, soziale und externe Aspekte belastet fühlen.

Von den elf nachgefragten Items sind sechs ohne Befund. Die restlichen fünf Items weisen einen Mode-Haupteffekt auf, wobei zwei Effekte auf dem 1 %-Niveau signifikant werden. Der zweite Haupteffekt für die Selbstselektion sowie die Interaktionen bleibt unauffällig.

Die Anonymität im Studium, die unsicheren Berufsaussichten, der Druck der Prüfungstermine und die Erwerbsarbeit neben dem Studium belastet Studierende aus der Papiergruppe stärker als aus der Onlinegruppe. Das Fehlen einer festen Partnerbeziehung ist für Studierende der Papiergruppe dagegen weniger belastend. Die Effektstärken liegen zwischen Eta=0,06 (unsichere Berufsaussichten) und 0,2 (Anonymität). Das Item „fehlende Partnerbeziehung“ weist ungleiche Varianzen zwischen den Gruppen auf, weshalb hier eine Alternative berechnet werden muss. Die parameterfreie Alternative (Kruskal-Wallis-Test) erbringt ebenfalls ein signifikantes Ergebnis.

Tabelle 7.8**Belastungen im Studium nach Effekte Mode – Selektion (WS 2015/16)**

(Mittelwerte)

Kennzeichen Fach	Papier (0,1)		Online (2-4)		p-Werte ¹⁾ für HE u. IE	eta
	Random. Papier (0)	selbst Wahl (1)	selbst Wahl (2)	Random. Lang (3,4)		
Leistungsanforderungen	3,60	3,56	3,62	3,50	n.s.	
Orientierungsprobleme	2,40	2,33	3,32	2,30	n.s.	
Anonymität	1,80	1,82	1,63	1,59	** HE Mode	0,2
bevorstehende Prüfungen	3,62	3,59	3,67	3,55	n.s.	
finanzielle Lage	2,57	2,56	2,38	2,55	n.s.	
persönliche Probleme	2,26	2,26	2,24	2,35	n.s.	
Fehlen fester Partnerbeziehung ²⁾	1,33	1,34	1,71	1,51	** HE Mode	0,18
unsichere Berufsaussichten	2,07	2,07	2,02	1,92	* HE Mode	0,06
Druck durch Prüfungstermine	3,57	3,53	3,55	3,40	* HE Mode	0,07
große Zahl Studierender	1,49	1,47	1,59	1,34	n.s.	
Erwerbsarbeit neben Studium	2,26	2,21	1,83	2,13	** HE Mode	0,12
Bartlett-Tests n.s. (-1)						
Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 60						
1) ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant; HE = Haupteffekt, IE= Interaktionseffekt						
2) ungleiche Varianzen						

Bei Einführung des Geschlechts als dritten Faktor bleiben die Haupteffekte erhalten. Allerdings wird bei vier Items der Geschlechtsfaktor signifikant und deutlich größer als der jeweilige Mode-Effekt (die Effektstärken für den Haupteffekt Geschlecht weisen Werte von 1,1 bis 2,6 auf. Nur bei der Anonymität ist kein solches Ergebnis festzustellen.

Wird nach Geschlecht getrennt, fällt der Haupteffekt für die unsicheren Berufsaussichten weg. Der Haupteffekt für Anonymität und das Fehlen einer Partnerbeziehung tritt nur bei den Männern auf. Für die Männer wird ein Interaktionseffekt bei der Erwerbsarbeit signifikant ($\eta=0,19$). Der Mode-Effekt für den Druck der Prüfungstermine und die Erwerbsarbeit neben dem Studium wird nur bei den Frauen signifikant (jeweils $\eta=0,1$).

Die tatsächlichen Mittelwertunterschiede bewegen sich zwischen 0,15 und 0,43 Skalenpunkten, was einer prozentualen Differenz von 2,1 bzw. 6,1 Prozentpunkten entspricht.

Studienerrträge

Ein wichtiger Output des Studiums sind die erfahrenen Förderungen in Kompetenzen, Fähigkeiten und Fertigkeiten. Sie werden im Studierendensurvey anhand von 14 Items abgefragt. Dabei wird nach fachlichen, überfachlichen und allgemeinen Fähigkeiten unterschieden.

Von den 14 nachgefragten Items erweisen sich elf Items als unauffällig, bzw. als zwischen den Designgruppen gleich. Bei zwei Items sind Haupteffekte für die Selektion zu beobachten (bei einem auf dem 1 %-Niveau) und bei einem Item wird ein Mode-Effekt signifikant.

Die Förderungen der Allgemeinbildung und der persönlichen Entwicklung werden von Studierenden aus der Wahlgruppe als etwas ertragreicher bewertet als aus den randomisierten Gruppen. Die Auto-

nomie wird in der Papiergruppe als besser gefördert erlebt als in der Onlinegruppe. Der Interaktionseffekt für das soziale Verantwortungsbewusstsein liegt zwischen der Wahl-Papiergruppe und der Papier- bzw. der Onlinegruppe. Die Effektstärken erreichen höchstens $\eta^2=0,19$.

Tabelle 7.9
Studienerrträge nach Effekte Mode – Selektion (WS 2015/16)
(Mittelwerte)

Kennzeichen Fach	Papier (0,1)		Online (2-4)		p-Werte ¹⁾ für HE u. IE	eta
	Random. Papier (0)	selbst Wahl (1)	selbst Wahl (2)	Random. Lang (3,4)		
selbständiges Weiterlernen	3.77	3.82	3.80	3.71	n.s.	
fachliche Kenntnisse	4.43	4.47	4.43	4.44	n.s.	
praktische Fähigkeiten	2.96	3.05	2.92	3.04	n.s.	
fachübergreifendes Wissen	3.17	3.24	3.13	3.30	n.s.	
Teamfähigkeit	3.47	3.45	3.52	3.53	n.s.	
Planungsfähigkeit	3.72	3.74	3.80	3.64	n.s.	
Allgemeinbildung	3.08	3.20	3.18	3.03	* HE Selekt	0,07
Autonomie	4.32	4.40	4.32	4.12	** HE Mode, *HE Selekt	0,19 0,06
Problemlösefähigkeit	4.00	4.00	4.05	3.99	n.s.	
Kritikfähigkeit	3.65	3.67	3.53	3.59	n.s.	
persönliche Entwicklung	3.80	3.94	3.83	3.73	* HE Selekt	0,11
soziales Verantwortungsbewusstsein	3.17	3.35	3.01	3.13	*IE	0,06
Kenntnisse wiss. Methoden	3.69	3.70	3.61	3.68	n.s.	
Berufsvorbereitung	2.52	2.51	2.63	2.51	n.s.	
Bartlett-Tests n.s.						

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz, Frage 57
¹⁾ ** = auf 1%-Niveau signifikant, * = auf 5%-Niveau signifikant; n.s. = nicht signifikant;
HE = Haupteffekt, IE= Interaktionseffekt

Bei Einführung des Geschlechts als drittem Faktor bleibt der Selektionseffekt für die Allgemeinbildung erhalten. Bei dem Item Autonomie werden alle drei Haupteffekte signifikant, wobei das Geschlecht den größten Effekt erbringt ($\eta^2=0,54$). Bei der persönlichen Entwicklung bleibt der Selektionseffekt erhalten, aber der Geschlechtseffekt kommt hinzu, mit einer größeren Effektstärke ($\eta^2=0,29$). Der Interaktionseffekt für das soziale Verantwortungsbewusstsein verschwindet, während der Haupteffekt Geschlecht signifikant wird ($\eta^2=2,35$).

Getrennt nach dem Geschlecht weisen die Männer keine Effekte für die Items „Allgemeinbildung“, „persönliche Entwicklung“ und „soziales Verantwortungsbewusstsein“ mehr auf. Der Mode-Effekt für Autonomie bleibt erhalten ($\eta^2=0,32$), aber der Selektionseffekt wird nicht mehr signifikant. Bei den Frauen wird kein Effekt zwischen den Designgruppen mehr signifikant.

Wird das Merkmal Alter als Kovariate in das zweifaktorielle Design eingeführt, ergibt sich für die „Allgemeinbildung“ ein Effekt der Kovariate, während das bereinigte Design nicht mehr signifikant wird. Bei dem Item „Autonomie“ hat die Kovariate zwar einen signifikanten Einfluss, doch bleiben beide Haupteffekte erhalten. Bei den Items „persönliche Entwicklung“ und „soziales Verantwortungsbewusstsein“ hat die Kovariate keinen Einfluss, die Effekte bleiben erhalten.

Mode-Vergleiche an weiteren Merkmalen der Studiensituation

An einer weiteren Auswahl von Items der Studiensituation soll nur noch der Mode-Effekt überprüft werden. Dabei werden 47 Items unterschiedlicher Fragen und Module ausgewählt und über die beiden Mode-Gruppen verglichen. Bei den Fragen handelt es sich um die Evaluation der Lehrsituation, Angaben zum eigenen Studierverhalten, soziales Klima, Wünsche zur Verbesserung der Studiensituation sowie Wichtigkeit von Aspekten im späteren Beruf. Im Anschluss werden die Analysen getrennt für das Geschlecht durchgeführt (vgl. Tabelle 7.10).

Bei diesen Vergleichen ergeben sich einige signifikante Unterschiede. Bei der Gesamtstichprobe werden 30 Vergleiche signifikant, 17 Vergleiche sind unauffällig. Der größte Effekt ($\eta=0,3$) findet sich für das Item „Gefühl, dass nur Leistung zählt“.

Werden die Analysen an der Teilstichprobe der Männer durchgeführt, dann stellen sich noch 14 Vergleiche als signifikant heraus, während 33 Vergleiche unauffällig bleiben. Größter Effekt bei den Männern ($\eta=0,5$) tritt bei dem Item „Wunsch nach Lehrveranstaltungen in kleinerem Kreis“ auf. Bei der Teilstichprobe der Frauen werden 23 Vergleiche signifikant und 24 bleiben unauffällig. Der größte Effekt ($\eta=0,24$) tritt hier bei dem Item „selbständige Entscheidungen treffen“ auf.

Der Blick auf die absoluten bzw. relativen Differenzen der Mittelwertunterschiede stellt heraus, dass die größten Unterschiede bei der Gesamtstichprobe sich auf 0,28 Skalenpunkten belaufen, was einer Differenz von vier Prozentpunkten entspricht. In den beiden Teilstichproben bleiben die Differenzen der Mittelwerte ebenfalls unter 5 Prozentpunkten. Nur bei den Männern erreicht die größte Differenz mit 0,35 Skalenpunkten exakt die fünf Prozentmarke.

Tabelle 7.10
Merkmale der Studiensituation nach Mode-Designgruppen (WS 2015/16)
(Mittelwerte)

	Mittelwerte		p (t-wert)		
	Papier (0,1)	Online (2-4)	ges.	nur männlich	nur weiblich
Terminausfälle	1,34	1,18	**	**	*
zeitgleiche Veranstaltungen	1,86	1,71	**	*	*
gute Vorbereitung der Lehrenden	4,40	4,43	n.s.	n.s.	n.s.
Lehrende erläutern Ergebnisse	2,53	2,71	n.s.	n.s.	n.s.
Forschungsfragen werden behandelt	3,00	2,98	n.s.	n.s.	n.s.
Lehrende helfen beim wiss. Arbeiten	3,04	2,95	n.s.	n.s.	n.s.
Lehrende vermitteln angekündigten Stoff	4,13	4,08	*	n.s.	n.s.
Lehrende bringen Praxisbeispiele	4,02	3,93	n.s.	*	n.s.
Sorge, Studium zu schaffen	2,37	2,26	*	n.s.	*
intensiv und viel für Studium arbeiten	3,94	3,90	n.s.	n.s.	n.s.
Ich kann konzentriert lernen	4,04	3,95	*	n.s.	n.s.
gutes Examen ist wichtig	4,48	4,32	**	n.s.	**
Faktenlernen fällt leicht	3,86	3,78	*	n.s.	**
in Prüfungen bin ich vergesslich	2,41	2,30	*	n.s.	*
habe meistens Angst vor Prüfungen	2,87	2,75	*	n.s.	n.s.
kann Lernstoff gut organisieren	3,76	3,67	*	n.s.	n.s.
möglichst rascher Abschluss ist wichtig	3,17	3,14	n.s.	n.s.	n.s.
mit Note zufrieden	3,68	3,78	n.s.	n.s.	n.s.
mit Kontakten zu Studierenden zufrieden	5,44	5,37	n.s.	n.s.	n.s.
mit Kontakten zu wiss. Mitarbeitern zufrieden	4,68	4,73	n.s.	n.s.	n.s.
mit Kontakten zu Professoren zufrieden	4,54	4,60	n.s.	n.s.	n.s.
habe genügend Ansprechpartner	3,75	3,79	n.s.	n.s.	n.s.
Gefühl, dass nur Leistung zählt	3,79	3,54	**	**	**
Abwesenheit würde niemandem auffallen	2,83	2,70	*	n.s.	**
Verringerung der Prüfungsanforderungen	2,63	2,46	**	n.s.	**
Konzentration der Studieninhalte	2,78	2,60	**	n.s.	**
stärkerer Praxisbezug	3,68	3,45	**	**	n.s.
stärkerer Forschungsbezug	2,60	2,46	**	*	*
Lehrveranstaltungen in kleinerem Kreis	3,20	2,95	**	**	*
intensivere Betreuung	3,21	3,00	**	**	*
Unterstützung bei Studienfinanzierung	3,07	2,82	**	n.s.	**
bessere Arbeitsmarktchancen	2,89	2,61	**	*	*
Einrichtung von Brückenkursen	2,44	2,20	**	*	**
Betreuungsangebote für Kinder	1,86	1,67	**	*	n.s.
gerne Student/in	4,91	4,86	n.s.	n.s.	n.s.
sicherer Arbeitsplatz	4,82	4,74	*	n.s.	*
eigene Ideen entwickeln	4,57	4,49	*	n.s.	*
hohes Einkommen	3,90	3,81	*	n.s.	*
selbständiges Entscheiden	4,62	4,47	**	*	**
immer neue Aufgaben	4,83	4,71	**	*	n.s.
wissenschaftliche Tätigkeit	2,97	3,02	n.s.	n.s.	n.s.
anderen helfen	4,13	4,05	n.s.	n.s.	n.s.
Aufstiegsmöglichkeiten	3,75	3,60	**	**	*
Unbekanntes erforschen	3,33	3,34	n.s.	n.s.	n.s.
Nützlich für die Allgemeinheit	4,26	4,15	*	n.s.	*
flexible Arbeitszeiten	4,27	4,28	n.s.	n.s.	n.s.
Vereinbarkeit von Familie und Beruf	5,29	5,20	**	n.s.	**

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Fazit

Die meisten Analysen ergeben keine Unterschiede zwischen den Designgruppen. Dort wo Differenzen auftreten, sind sie in ihren Effektgrößen sehr gering. Ebenso spiegeln die absoluten Unterschiede nur Verteilungs- bzw. Skalendifferenzen von weniger als zehn Prozentpunkten wieder, meist sogar weniger als fünf Prozentpunkte. Dabei sind einige Unterschiede mit dem Geschlecht oder dem Alter konfundiert. Auftretende Unterschiede sind überwiegend Mode-Effekte, Selektions- oder Interaktionseffekte sind eher selten.

Insgesamt lässt sich schlussfolgern, dass die Unterschiede gering sind und kein bedeutsamer Grund gegen die Zusammenlegung der Mode-Gruppen spricht.

Auffällig ist die Zunahme der Mode-Effekte bei Darstellung von Teilfragen, die in der Onlineversion untereinander stehen, während sie in der Papierversion nebeneinander stehen können. In der zweiten Teilfrage steigen die Mode-Effekte an, was auf eine Irritation hindeutet, da die Studierenden die gleichen Items wieder vorgelegt bekommen, wenn auch mit einer etwas anderen Frageausrichtung.

Bei den intervallskalierten Items fällt auf, dass die Einführung eines dritten Faktors Geschlecht oft höhere Haupteffekte herausstellen kann als für den Mode- oder den Selektionseffekt aufzufinden sind. Bestätigt wird dies durch die gleichen Analysen, bei denen das Geschlecht kontrolliert wurde. Damit sind die Unterschiede zwischen den Designgruppen geringer als zwischen den Geschlechtern.

8. Effekte von Designparameter

In der 13. Erhebung wurde neben der Papierbefragung eine Online-Erhebung durchgeführt. Das Hauptziel war dabei zu prüfen, ob der Studierendensurvey auch online durchgeführt werden kann, ohne dass die Befunde sich zu sehr von der Papierbefragung unterscheiden. Dabei hat es sich bei der Online-Befragung gleichzeitig angeboten, bestimmte zusätzliche Möglichkeiten der Kommunikation und der Onlinefähigkeit auszutesten. Zu diesem Zweck wurden verschiedene Maßnahmen erstellt und in die Befragung eingeführt, mit dem Ziel, in den Analysen zu prüfen, ob und inwieweit diese Maßnahmen Effekte verursachen.

Die einzelnen Maßnahmen waren dabei:

- 1) drei Motivationsseiten bei der Hälfte der Online-Befragungen,
- 2) verkürzter Online-Fragebogen,
- 3) unterschiedliche Module mit Zusatzfragen am Ende des Kurzfragebogens,
- 4) Versendung einer zweiten Erinnerung per E-Mail bei einem Teil der Samplehochschulen,
- 5) Ankündigung der Befragung auf der Website der Hochschule bei einem Teil der Samplehochschulen,
- 6) Ankündigung der Befragung auf der Facebook-Seite bei einem Teil der Samplehochschulen.

In diesem Bericht kann auf den Kurzfragebogen und die zusätzlichen Module aus Platzgründen nicht eingegangen werden. Daher werden im Folgenden nur die Effekte der Motivationsseiten, der Erinnerungen und der Ankündigungen untersucht.

8.1 Motivationsseiten

Die Designgruppen 3 und 4 erhielten den zur Papierversion „identischen“ Online-Fragebogen. Sie stellen die randomisierte Onlinegruppe dar. In Gruppe 3 wurden jedoch im Unterschied zu Gruppe 4 drei Motivationsseiten in den Fragebogen eingespielt. Diese Seiten sollten den Studierenden einerseits Hinweise geben, wie weit sie gekommen sind und welche Bereiche noch folgen, und zum anderen die Befragten auffordern, bei der Befragung durchzuhalten. Die Seiten wurden nach den Fragennummern 36, 53, 62 bzw. nach Onlinebildschirmseiten 31, 48, 58 eingeführt.

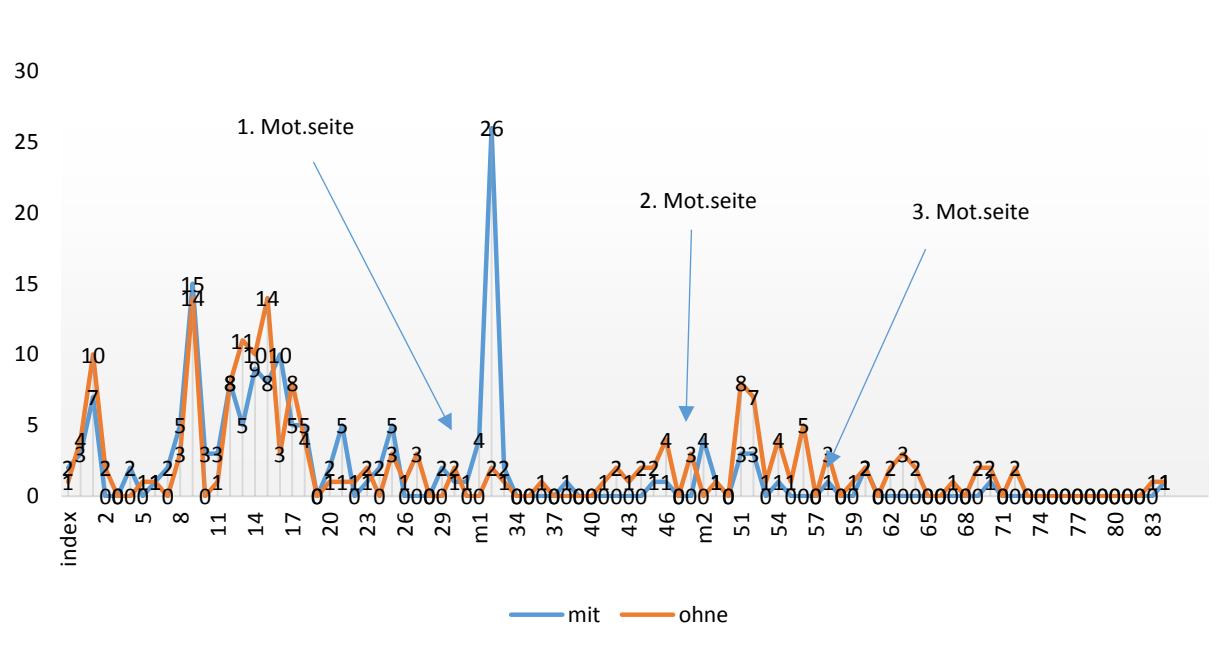
Insgesamt wurden in der randomisierten Onlinegruppe der Langversion 12.640 Studierende angeschrieben. Diese wurden beim Aufruf des Fragebogens zufällig auf die beiden Untergruppen (mit und ohne Motivationsseiten) aufgeteilt.

- in Gruppe 3 (mit Motivationsseiten) haben 589 Studierende (9,2 %) geantwortet,
- in Gruppe 4 (ohne Motivationsseiten) haben 620 Studierende (9,8 %) geantwortet.

Von diesen teilnehmenden Studierenden sind 425 in Gruppe 3 und 443 in Gruppe 4 bis zur letzten Fragebogenseite gelangt (6,8 % bzw. 6,9 %). D.h. 164 bzw. 177 Studierende haben vor dem Ende abgebrochen. Davon haben 112 bzw. 109 die Befragung abgebrochen, bevor die erste Motivationsseite erschienen ist. 40 bzw. 36 Studierende haben bereits bis zur zehnten Bildschirmseite abgebrochen, weitere 55 bzw. 60 Studierende bei den folgenden zehn Seiten.

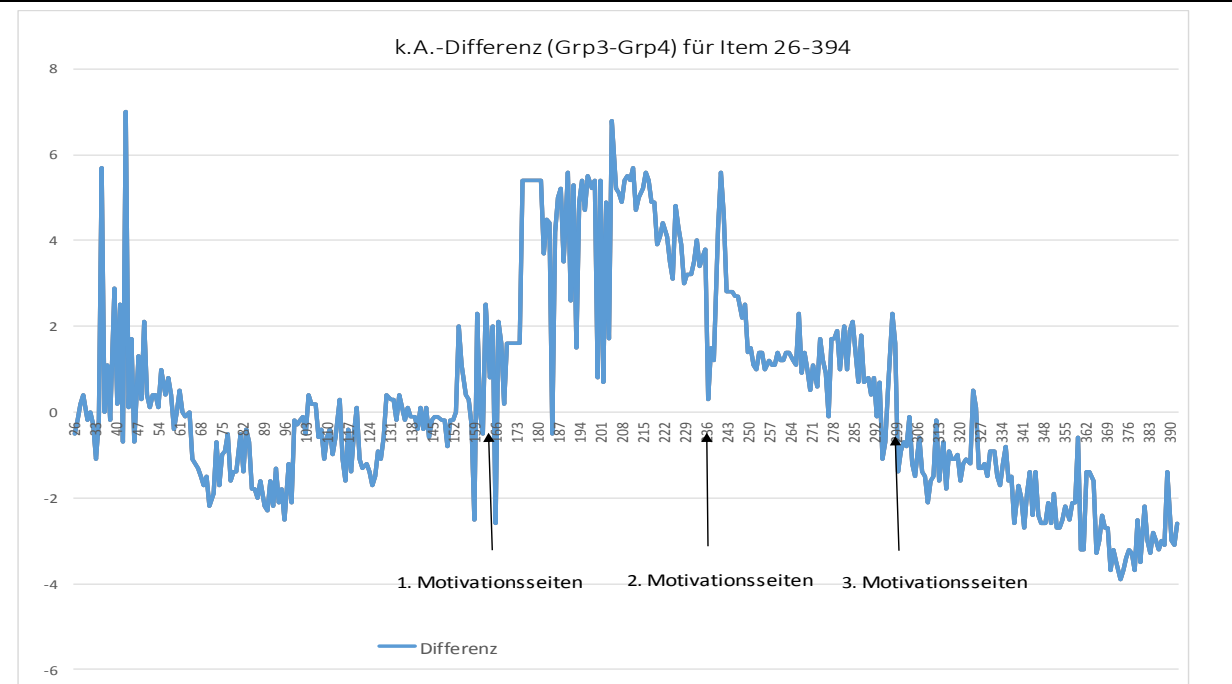
Der mit Abstand häufigste Abbruch lag in der Gruppe 3 (mit Motivationsseiten) direkt nach der ersten Motivationsseite. In der Gruppe 4 (ohne Motivationsseiten) waren häufig Abbrüche bei den Onlineseiten 9, 13-15 zu finden und nochmals bei Seite 51 und 52 (vgl. Abbildung 8.1).

Abbildung 8.1
Befragungsabbrüche in den Onlinegruppen mit und ohne Motivationsseiten pro Bildschirmseite des Onlinefragebogens (WS 2015/16)
 (Angaben in absolut)



Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Abbildung 8.2
Differenz des Item Non-Responses zwischen Onlinegruppe mit und ohne Motivationsseiten im Onlinefragebogen (WS 2015/16)
 (Angaben in absolut)



Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Wichtig für die Analysen ist die kontinuierliche Bearbeitung des Fragebogens. Nicht nur der Abbruch von teilnehmenden Studierenden führt zu Datenverlusten, sondern auch die Nonresponse-Anteile in

den einzelnen Items. Bis zur ersten Hälfte des Fragebogens steigt der Item Non-Response auf etwa 20 % an. Bis zum Ende des Fragebogens steigt dieser Anteil weiter auf bis zu ein Drittel.

Werden die beiden Onlinegruppen (mit und ohne Motivationsseiten) verglichen, dann fällt auf, dass zu Beginn des Fragebogens die Gruppe mit Motivationsseiten zuerst höhere Item Non-Response Raten (INR) aufweist, dann gleiche Anteile wie die Gruppe ohne Motivationsseiten hat und zur Mitte des Fragebogens hin wieder höhere Raten zeigt. In der zweiten Hälfte des Fragebogens gehen die Differenzen der INR zwischen beiden Gruppen zuerst stetig gegen Null und fallen dann in Gruppe 3 konstant und deutlich unter die der Gruppe 4 (vgl. Abbildung 8.2).

Ein t-test auf Unterschiedlichkeit, getrennt nach den Phasen der Einführung der Motivationsseiten, erbringt für die 4. Phase, nach Einführung der 3. Motivationsseite, einen signifikanten Effekt, der besagt, dass der Anteil des Non-Responses in Gruppe 4 höher ist als in Gruppe 3. In den drei vorangegangenen Phasen sind keine signifikanten Unterschiede festzustellen, wenn der t-test für unverbundene Stichproben verwendet wird. Bei verbundenen Stichproben werden auch die Unterschiede in der ersten (vor Einführung einer Motivationsseite) und zweiten Phase (nach Einführung der 1. Motivationsseite) signifikant. In der ersten Phase sind die Non-Responses dabei in Gruppe 4 etwas höher, in der zweiten Phase dagegen in Gruppe 3. Für die Überlegung, welcher t-test besser auf die Daten passt, lässt sich argumentieren, dass bei Differenzierung der Non-Responderaten über zwei Gruppen einerseits die Personen in beiden Gruppen unterschiedlich sind, weshalb es sich um unverbundene Stichproben handelt und andererseits die Items (als Untersuchungseinheiten) jeweils die gleichen sind, weshalb es sich um verbundene Stichproben handelt.

	Gesamt	1. Phase	2. Phase	3. Phase	4. Phase
Unverbunden					
t-wert	0,354	-0,164	1,28	0,51	-9,01
P	0,723	0,869	0,20	0,60	0,000
Verbunden					
t-wert	3,07	-2,72	19,32	0,87	-21,85
P	0,002	0,007	0,000	0,38	0,000

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Die Interpretation der Befunde verweist auf einige wichtige Erkenntnisse:

- Der geringe Rücklauf bestätigt die zunehmende Verweigerungshaltung der Befragten an solchen Untersuchungen.
- Die erhöhten Abbrüche im ersten Viertel des Fragebogens deuten darauf hin, dass die Befragten eher schnell die Motivation verlieren, wenn die Befragung länger als 20 Bildschirmseiten dauert.
- Der erhöhte Abbruch in der Gruppe mit Motivationsseiten direkt nach Einführung der ersten Motivationsseiten kann der nachfolgenden Frage geschuldet sein. Dort wird das Zeitbudget erfragt, bei dem die Studierenden zu unterschiedlichen Aktivitäten ihre wöchentliche Arbeitszeit angeben sollen. Für diesen Themenbereich hat sich die Motivationsseite als kontraproduktiv erwiesen.

- Die höhere Non-Responserate in Gruppe 4 (ohne) im Vergleich zu Gruppe 3 (mit) im letzten Viertel des Fragebogens bestätigen einen positiven Effekt der Einführung von Motivationsseiten, der sich allerdings erst spät im Fragebogen zu erkennen gibt.
- Die höhere Non-Responserate in Phase zwei (bei Betrachtung als abhängige Stichprobe) deutet auf einen nachteiligen Effekt der Einführung der ersten Motivationsseite hin.

Als Fazit zur Nutzung von Motivationsseiten lässt sich daher argumentieren, dass eine frühe Einführung von solchen unterstützenden Maßnahmen sich eher nachteilig auswirken kann, während zu späteren Zeiten der Befragung die Einführung solcher Motivationsseiten positive Effekte aufzeigen kann. Die Abfrage von für die Studierenden sehr aufwendigen Fragen sollte nicht direkt nach einer Motivationsseite erfolgen, da hier deutlich negative Auswirkungen aufzufinden sind.

Diese Befunde stellen allerdings nur erste Annäherungen dar, für konkretere Aussagen müssten weitere Tests erfolgen. So bleibt unklar, wie das Design der Motivationsseiten sich auswirkt, und welche Anzahl von solchen Seiten in welchen Abständen sich effizient auswirken kann.

Als Folgerung für zukünftige Online-Befragungen lässt sich nicht belegen, dass auf solche Maßnahmen verzichtet werden könnte. Eher sprechen die Befunde dafür, diese einzusetzen.

Prüfung der Datenqualität

Die Analyse des Non-Responses stellt einen Weg dar, die Bereitschaft der Studierenden zur Beantwortung der Fragen zu begutachten. Ein anderer Weg ist die Prüfung der Qualität der Daten, womit der Frage nachgegangen wird, ob zu späteren Zeiten die Motivation zu überlegten Antworten nachlässt.

Zur Analyse der Qualität der Antworten im Fragebogenverlauf bestehen zwei Möglichkeiten:

- 1) Vergleich der Kennwerte an valider Vergleichsstichprobe,
- 2) Interkorrelationen zwischen Items.

Zu 1: Vergleich der Kennwerte an valider Vergleichsstichprobe

Wird angenommen, dass die Papierversion validere Daten enthält, kann geprüft werden, ob die beiden Onlineversionen deutlich von der Papierversion abweichen. Dazu wird aus jeder Phase des Fragebogens eine Itematterie verwendet und anhand der Items geprüft, ob sie sich zwischen den Designgruppen (Papier-Online) unterscheiden. Diese Unterschiede sollen als Maße zwischen den Gruppen 3 und 4 fungieren.

Dazu werden insgesamt 22 Items analysiert, die aus den Fragen 23 (Aussagen über Lernen und Studieren; Items 5-9), 50 (Wichtigkeit von Beratungsbereichen, Items 1-6), 58 (Maßnahmen zur Erhöhung des Studienerfolgs, Items 1-6) und 76 (Meinungen zum politischen System, Items 1-5) stammen. Jede Frage liegt somit in einer anderen Motivationsphase (vgl. Tabelle 8.2).

Insgesamt lässt sich festhalten, dass zwischen der Papiergruppe und Gruppe 3 ebenso wie zwischen der Papiergruppe und Gruppe 4 in einzelnen Items signifikante Unterschiede auftauchen. Dabei handelt es sich allerdings um unterschiedliche Items, wobei die Effektstärken in der Gruppe 3 in allen fünf Fällen jeweils unter $\eta^2 = 0,1$ liegen, in Gruppe 4 bei den beiden Fällen $0,10$ bzw. $0,11$ betragen.

Tabelle 8.2											
Mittelwertvergleiche der Onlinegruppen (mit und ohne Motivationsseiten) mit Papiergruppe											
(Mittelwerte, Test auf Varianzgleichheit, T-werte, 1-p)											
	Papier	Gruppe 3 (mit Motivationsseiten)					Gruppe 4 (ohne Motivationsseiten)				
	MW	MW	Var=	T-Wert	1-p	Sign.	MW	Var=	T-Wert	1-P	Sign.
137	3.86	3.79	g	1.34	.911	*	3.85	g	.35	.635	n.s.
138	2.41	2.30	g	1.22	.889	n.s.	2.28	g	1.44	.925	n.s.
139	2.87	2.73	g	1.51	.935	n.s.	2.77	g	1.13	.871	n.s.
140	3.76	3.66	g	1.57	.942	n.s.	3.70	g	.98	.837	n.s.
141	3.17	3.14	g	.37	.644	n.s.	3.12	g	.60	.726	n.s.
223	3.91	3.76	u	1.57	.941	n.s.	3.90	g	.09	.534	n.s.
224	3.77	3.66	u	1.27	.899	n.s.	3.83	u	-.72	.235	n.s.
225	4.33	4.18	u	1.89	.971	*	4.30	g	.46	.676	n.s.
226	4.39	4.35	u	.51	.695	n.s.	4.39	u	-.12	.451	n.s.
227	4.02	3.99	u	.47	.681	n.s.	4.15	g	-1.73	.042	n.s.
228	3.51	3.36	u	1.73	.957	*	3.61	g	-1.25	.106	n.s.
264	4.67	4.52	u	1.72	.957	*	4.56	u	1.39	.917	n.s.
265	2.94	2.74	u	1.90	.971	*	2.89	u	.50	.691	n.s.
266	3.89	3.87	u	.27	.606	n.s.	3.77	u	1.47	.929	n.s.
267	3.60	3.61	u	-.14	.446	n.s.	3.56	u	.39	.651	n.s.
268	3.67	3.61	u	.74	.771	n.s.	3.59	g	1.01	.843	n.s.
269	3.19	3.18	u	.11	.544	n.s.	3.06	u	1.39	.917	n.s.
367	3.27	3.24	u	.40	.655	n.s.	3.14	u	1.37	.914	n.s.
368	3.72	3.60	u	1.20	.885	n.s.	3.50	u	2.26	.988	*
369	1.71	1.72	u	-.12	.454	n.s.	1.74	u	-.42	.339	n.s.
370	3.35	3.27	u	.85	.801	n.s.	3.33	u	.17	.569	n.s.
371	6.18	6.16	u	.31	.622	n.s.	6.03	u	2.17	.985	*

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
 MW= Mittelwert, Var= g=gleiche Varianzen, u=ungleiche Varianzen, 1-p= Überschreitungswahrscheinlichkeit,
 * = signifikant auf 5%-Niveau, n.s = nicht signifikant,
 137-141: Items 5-9 aus Frage 23; 223-228: Items 1-6 aus Frage 50; 264-269: Items 1-6 aus Frage 58;
 367-371: Items 1-5 aus Frage 76

Im Vergleich beider Onlinegruppen stellt sich heraus, dass in der dritten und vierten Phase die Abweichungen in Gruppe 4 von der Papiergruppe zunehmend größer werden als bei Gruppe 3, während in der zweiten Phase (nach Einführung der 1. Motivationsseite) Gruppe 3 stärker als Gruppe 4 abweicht. Allerdings sind die Effekte gering und nicht signifikant. Die Befunde belegen damit keine eindeutig schwächere Datenqualität in Gruppe 4.

Zu 2: Interkorrelationen zwischen Items

Um die Interkorrelationen zu analysieren, werden für die ausgewählten 22 Items die Korrelationen zwischen den Items pro Frage gemessen. Sollte die Motivation zur Beantwortung nachlassen, wäre mit vermehrt gleichen Antworten zu rechnen, was die Interkorrelationen erhöhen sollte. Hat die Einführung der Motivationsseiten einen Einfluss, sollten diese Interkorrelationen in Gruppe 3 geringer ausfallen (vgl. Tabelle 8.3).

Zwar weist Gruppe 4 im Vergleich zu Gruppe 3 häufiger höhere Interkorrelationen auf, doch werden die Analysen dazu allesamt nicht signifikant, sodass die Annahme verworfen wird. Auch in der Gruppe ohne Motivationsseiten lässt die Datenqualität damit nicht eindeutig nach.

Tabelle 8.3										
Interkorrelationen zwischen ausgewählten Items pro Motivationsphase										
(P-M-Korrelationen R)										
Gruppe 3 (mit Motivationsseiten)						Gruppe 4 (ohne Motivationsseiten)				
Item	137	138	139	140		137	138	139	140	
138	-.23					-.29				
139	-.13	.69				-.25	.61			
140	.32	-.13	-.04			.32	-.25	-.17		
141	.02	-.03	.02	.20		.02	-.05	-.07	.22	
Item	223	224	225	226	227	223	224	225	226	227
224	.56					.61				
225	.39	.43				.39	.40			
226	.22	.25	.32			.23	.33	.39		
227	.14	.20	.29	.37		.25	.29	.31	.48	
228	.33	.39	.37	.35	.38	.32	.33	.40	.35	.37
Item	264	265	266	267	268	264	265	266	267	268
265	.11					.15				
266	.04	.22				.19	.21			
267	.04	.07	.15			.15	.13	.25		
268	.18	.14	.15	.10		.22	.15	.09	.12	
269	.05	.19	.16	.07	.23	.08	.14	.13	.15	.30
Item	367	368	369	370		367	368	369	370	
368	.26					.35				
369	.07	.01				.18	.07			
370	.32	.19	-.00			.27	.23	.01		
371	-.17	-.19	-.14	-.09		-.15	-.17	-.18	-.17	

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz
137-141: Items 5-9 aus Frage 23; 223-228: Items 1-6 aus Frage 50; 264-269: Items 1-6 aus Frage 58;
367-371: Items 1-5 aus Frage 76

Als Fazit lässt sich anmerken: Es gibt zwar Hinweise darauf, dass in Gruppe 4 (ohne Motivationsseiten) mit fortschreitender Fragebogenlänge die Datenqualität tendenziell nachlässt, doch sind die Befunde dazu nicht eindeutig. Dennoch bieten die niedrigeren Interkorrelationen in Gruppe 3 Hinweise, dass die Einbringung von Motivationsseiten sich als nützlich erweist. Für detailliertere Aussagen müssen weitere Untersuchungen angestellt werden.

8.2 Ankündigung und Erinnerung

Die Samplehochschulen wurden zum einen angefragt, ob sie bereit wären, die Befragung in ihren sozialen Netzwerken bzw. auf ihrer Website anzukündigen und zum anderen, ob sie es ermöglichen können, eine zweite Erinnerung per E-Mail zu versenden.

Von den 28 Samplehochschulen sahen sich zehn Hochschulen nicht in der Lage, weitere Online-Unterstützungen durchzuführen. 15 Hochschulen haben sich bereit erklärt, eine 2. Online-Erinnerung zu versenden und 12 Hochschulen haben die Befragung angekündigt. Dabei haben sieben Hochschulen die Ankündigung auf ihre Website gestellt und sieben Hochschulen auf ihre Facebook-Seite.

Nur eine Hochschule hat alle drei Maßnahmen unterstützt, und eine Hochschule hat die Ankündigung in beide Onlineportale gestellt. Acht Hochschulen haben sowohl eine Online-Erinnerung verschickt als auch eine Ankündigung erstellt. Die übrigen acht Hochschulen haben nur eine der drei Maßnahmen durchgeführt (vgl. Tabelle 8.4).

Tabelle 8.4					
Zusätzliche Designparameter bei Samplehochschulen (WS 2015/16)					
(Angaben in absolut)					
	2. online Erinnerung		Ankündigung		
	ja	nein	ohne	Website	Facebook
Universitäten	7	9	9	4	3
Fachhochschulen	8	4	7	3	4
Gesamt	15	13	16	7	7

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Ankündigung über neue Medien

Die Ankündigung der Befragung auf den Webseiten der Hochschule bzw. auf den Facebook-Seiten sollte dazu dienen, die Studierenden auf die mögliche Einladung einzustimmen und so die Teilnahmebereitschaft zu erhöhen.

Die Hälfte der Sample-Hochschulen hat eine Ankündigung in den neuen Medien vorgenommen. Um zu prüfen, ob diese Maßnahmen die Teilnahme erhöhen konnte, können die Teilnahmezahlen der eingeladenen Studierenden verglichen werden. Dazu wird der bereinigte Rücklauf pro Hochschule verwendet, der sich aus der Teilnehmezahl durch die Anzahl der Einladungen, abzüglich der nichtzustellbaren Einladungen, ergibt (vgl. Tabelle 8.5).

Werden die bereinigten Rückläufe pro Hochschule nach den Designparametern unterschieden, dann stellt sich heraus, dass die Rückläufe bei den Hochschulen, die keine Ankündigung durchgeführt haben, sogar höher waren als bei den Hochschulen, die eine Ankündigung zur Studie geschaltet haben. Ein t-test erbringt allerdings nur für die Papiergruppe ein signifikantes Ergebnis ($t=-2,07$, $p=0,048$). Die parameterfreie Alternative (Mann-Whitney-U: $z=-2,03$, $p=0,04$) bestätigt den Effekt in der Papiergruppe. Bei der Unterteilung nach Universitäten und Fachhochschulen treten beim t-test keine signifikanten Ergebnisse auf. In der parameterfreien Alternative werden der Gesamtrücklauf (Papier und Online zusammen, $z=-2,11$, $p=0,03$) sowie der Rücklauf für die Papiergruppe ($z=-2,02$, $p=0,04$) bei den Fachhochschulen zusätzlich signifikant.

Für den Item Non-Response können nur eingeschränkt vergleichbare Analysen durchgeführt werden, da die Frage nach dem Hochschulort erst auf der letzten Seite des Fragebogens gestellt wurde und damit die vorzeitigen Abbrecher nicht mehr zugeordnet werden können. Ein Vergleich jener Studierenden, die bis zum Schluss den Fragebogen beantwortet haben, erbringt in den beiden Onlinegruppen 3 und 4 jeweils signifikante Unterschiede ($t=-4,62$, $p=0,000$ bzw. $t=-2,72$, $p=0,008$). In Gruppe 3 sind die Item Non-Responses im hinteren Teil des Fragebogens in der Gruppe mit Ankündigung größer, in Gruppe 4 kleiner als in der jeweiligen Gruppe ohne Ankündigung. Für den Gesamtrücklauf über alle Designgruppen hinweg treten keine signifikanten Unterschiede auf.

Eine parameterfreie Alternative erbringt zusätzlich signifikante Unterschiede in der Papiergruppe und der Papier-Wahlgruppe. Die Online-Wahlgruppe sowie die Gesamtstichprobe über alle Designgruppen weisen keine signifikanten Unterschiede auf. Für die beiden Onlinegruppen ergeben sich die gleichen Befunde wie beim t-test. Dabei ist in der Papiergruppe der Item Non-Response bei der Gruppe mit Ankündigung kleiner, in der Wahl-Papiergruppe größer als in der Gruppe ohne Ankündigung.

Tabelle 8.5
Bereinigter prozentualer Rücklauf nach Designgruppen und Maßnahmen (WS 2015/16)
 (Angaben in Prozent)

	Papier			Online			Wahl gesamt	Gesamt
	Papier o. Wahl	Wahl	gesamt	ohne Wahl	Wahl	gesamt		
2.E + Web + FB								
FH-1	12,8	10,9	12,5	7,2	5,1	5,7	16,0	9,9
2.E + Web								
Uni-1	11,7	5,8	10,7	6,6	1,4	5,2	7,2	8,3
Uni-2	19,0	14,8	18,3	6,6	3,1	5,3	17,9	12,7
Uni-3	-	-	-	11,3	-	11,3		11,3
FH-2	13,2	8,1	12,3	7,4	3,7	5,9	11,8	9,7
								10,5
Web + FB								
FH-3	9,1	7,1	8,8	5,6	3,2	4,6	10,3	7,1
2.E + FB								
Uni-4	13,9	11,4	13,5	8,8	2,8	7,0	14,2	10,8
Uni-5	12,8	8,8	12,1	4,5	2,8	3,6	11,6	8,5
FH-4	14,5	19,2	15,2	8,7	6,4	7,3	25,6	12,0
FH-5	13,6	13,9	13,6	5,6	2,8	4,6	16,7	9,7
								10,2
nur Web								
Uni-6	17,0	7,6	15,4	9,0	2,1	7,2	9,7	11,9
nur FB								
Uni-7	19,3	17,6	19,0	9,2	4,5	7,4	22,2	14,1
nur 2.E								
Uni-8	14,3	9,2	13,4	6,8	2,6	5,4	11,8	10,0
Uni-9	18,5	11,3	17,3	11,3	6,0	9,1	17,2	14,1
FH-6	13,2	5,9	12,0	8,5	4,2	6,8	10,2	10,0
FH-7	14,2	9,0	13,4	6,4	2,0	5,3	11,0	9,8
FH-8	17,3	13,2	16,7	11,2	0,9	9,2	14,2	13,3
FH-9	15,7	15,2	15,7	8,9	4,8	7,3	20,0	12,3
								11,6
Maßnahmen ges.								10,9
keine								
Uni-10	16,6	6,7	15,0	5,7	2,5	4,5	9,2	10,5
Uni-11	20,4	11,6	18,9	11,7	4,4	9,1	16,0	15,1
Uni-12	13,3	9,6	12,7	4,7	1,3	3,7	10,9	8,7
Uni-13	18,1	9,6	16,8	11,9	5,3	9,5	14,9	13,9
Uni-14	17,3	0,2	16,1	11,1	3,5	8,9	13,7	13,1
Uni-15	13,8	12,1	13,5	5,8	1,4	4,6	13,5	9,6
Uni-16	21,1	12,7	19,6	10,7	5,3	8,4	18,0	15,1
FH-10	22,7	13,6	21,3	13,2	4,5	10,8	18,2	16,9
FH-11	16,1	10,3	15,2	10,6	2,8	8,7	13,1	12,5
FH-12	14,3	11,9	13,9	7,2	4,8	5,8	16,7	10,6
								12,6
Universitäten	16,5	10,7	15,6	9,7	4,2	7,9	14,9	12,3
Fachhochschulen	14,7	11,7	14,2	9,8	5,1	8,0	16,4	11,8
Gesamt	16,1	10,8	15,2	9,7	4,5	8,0	15,3	12,2

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Ein Test auf Interaktionseffekte zwischen Ankündigung und Motivationsseiten erbringt ebenfalls keine signifikanten Ergebnisse.

Als Fazit lässt sich festhalten, dass die Ankündigungen nicht den erhofften Effekt erzielen konnten. Die Teilnahmebereitschaft hat sich dadurch nicht erhöht und auch die Antwortbereitschaft scheint dadurch nicht verbessert worden zu sein.

Ein weiteres Problem bei der Analyse ist der Umstand, dass nicht überprüft werden kann, welche Befragten die Ankündigung gelesen haben und welche nicht. Es kann sehr gut sein, dass ein Großteil der Befragten den Hinweis auf die Befragung nicht gesehen hat, was erklären könnte, wieso es zu keinem positiven Effekt auf die Rücklaufquote kam.

Zweite Online-Erinnerung

Von den beteiligten 28 Sample-Hochschulen haben sieben Universitäten und acht Fachhochschulen eine zweite Erinnerung per E-Mail versendet, während neun Universitäten und vier Fachhochschulen kein zweites Mal erinnern wollten.

Um zu prüfen, ob diese Maßnahme einen rücklaufsteigernden Erfolg hatte, können sowohl die endgültigen Rückläufe der Sample-Hochschulen betrachtet werden, als auch die Tages- bzw. Wochen-spezifischen Rücklaufprofile.

Gesamtrücklauf

Der endgültige Rücklauf der Fragebogen wird nach Bereinigung der unzustellbaren Sendungen pro Hochschule berechnet. Diese Daten werden zwischen den Hochschulen mit und ohne 2. Erinnerung verglichen. Die t-tests ergeben allerdings keine signifikanten Ergebnisse, auch nicht bei Unterscheidung nach der Hochschulart. Auch die parameterfreie Alternative, aufgrund der kleinen und unterschiedlichen Stichproben, ergeben keine signifikanten Unterschiede zwischen den Gruppen.

Rücklaufprofil der Samplehochschulen

Die Rückläufe der Papierfragebogen werden in der Arbeitsgruppe protokolliert. Insofern besteht die Möglichkeit, für jede Hochschule ein Profil der Tages- bzw. Wochenrückläufe zu erstellen. Unter der Voraussetzung, dass die Hochschulen sich an den Ablauf gehalten haben und der Annahme, dass Erinnerungen den Rücklauf steigern, sollten sich Effekte in den Rücklaufprofilen zeigen.

In Abbildung 8.3 sind die Profile der Rückläufe der Papierfragebogen über die Zeit (in Wochen) abgebildet. Dabei wurden Universitäten und Fachhochschulen getrennt danach unterschieden, ob sie eine zweite Erinnerung durchgeführt haben oder nicht. Die Daten sind dabei auf Wochen nach der jeweiligen Versendung angepasst, haben aber u.U. unterschiedliche Startzeiten. Daher werden keine Datumsangaben gemacht, sondern Wochen nach Versendung.

An der Graphik lässt sich ablesen, dass die Rückläufe ab der 5. Woche deutlich zurückgehen, und ab der 10. Woche nur noch vereinzelte Nachzügler eintreffen. Der Hauptanteil des Rücklaufes findet also in den ersten fünf Wochen statt.

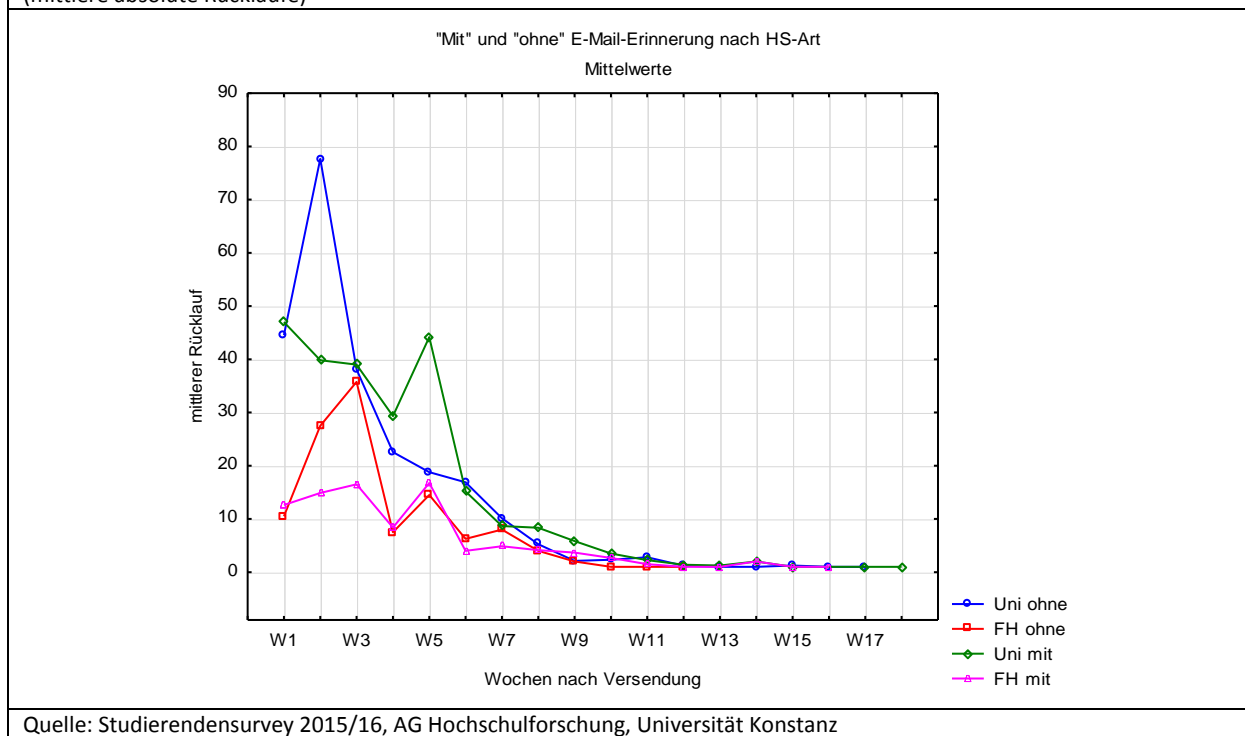
Bei den Universitäten ohne 2. Erinnerung ist der größte Rücklauf in der zweiten Woche nach Versendung zu beobachten, danach fällt die Kurve steil ab. Bei den Universitäten mit 2. Erinnerung ist ein Ausschlag in der fünften Woche zu beobachten, danach fällt auch dieser Rücklauf deutlich ab. An den Fachhochschulen steigen die Rückläufe zur dritten Woche hin an, vor allen in der Gruppe ohne 2. Erinnerung. In der fünften Woche ist ein weiterer Ausschlag zu beobachten, der bei der Gruppe mit 2. Erinnerung vergleichbar groß ist wie in der dritten Woche.

In drei Gruppen fallen also Ausschläge in der fünften Woche nach Versendung auf. Dieser Zeitraum fällt auf Mitte Januar bis Anfang Februar, je nachdem, wann die Hochschulen mit ihrer Versendung begonnen haben. Für die Hochschulen, die Anfang Dezember ins Feld gingen, liegt der Jahreswechsel

dazwischen, sodass für einen erhöhten Rücklauf eine Aufsummierung der angelaufenen Rücksendungen nicht ganz ausgeschlossen werden kann.

Dennoch stellt die fünfte Woche nach Versendung in etwa jenen Zeitpunkt dar, in dem die zweite Erinnerung ins Feld ging. Die Ausschläge der Gruppen, die eine solche Erinnerung bekamen, deuten somit auf einen Effekt hin. Bei den Universitäten ist dieser Effekt signifikant unterschiedlich zwischen den Gruppen mit und ohne 2. Erinnerung (MW-U-Test: $Z=2,125$, $p=0,03$), an Fachhochschulen bleibt er nicht signifikant. Dort wird der Unterschied in der dritten Woche signifikant ($Z=-2,12$, $p=0,03$), wobei der Wochenrücklauf hier in der Gruppe ohne 2. Erinnerung höher war.

Abbildung 8.3
Profil der Wochenrückläufe mit und ohne 2. Erinnerung an Universitäten und Fachhochschulen (WS 2015/16)
 (mittlere absolute Rückläufe)



Allerdings sind die Daten zur Feldphase, trotz klarer Instruktionen für den Ablauf, nicht eindeutig nachvollziehbar. Ziel war es, dass die Unterlagen (Papierfragebogen wie auch Einladung zur Online-Befragung) zu einem definierten Zeitpunkt versendet werden. Dann sollte drei Wochen später die erste Erinnerung verschickt werden und nochmals drei Wochen später die zweite Erinnerung per E-Mail. Unklar bleiben jedoch drei Aspekte seitens der Hochschulen:

- Es ist nicht bekannt, wie lange sich die Versendung der Unterlagen an den einzelnen Hochschulen hingezogen hat;
- Es ist nicht eindeutig festzustellen, wann genau die erste Erinnerung per Post versendet wurde;
- Es ist nicht eindeutig festzustellen, wann genau die zweite Erinnerung per Mail versendet wurde.

Hinzu kommt eine Ungenauigkeit aufgrund des speziellen Zeitraums der Erhebung. Da die Versendung Anfang Dezember begann, fällt der Jahreswechsel in die Rücklaufphase. In diesem Zeitraum konnten eine Woche lang keine Arbeiten erledigt werden, da manche Hochschulen geschlossen wa-

ren. Allerdings wird davon ausgegangen, dass die meisten Studierenden in dieser Phase an ihre Heimatorte zurückfahren und keine Arbeiten für das Studium durchführen. Dennoch kann der Anstieg in der fünften Woche auf eine Aufsummierung von den zwei Wochen des Jahreswechsels zurückgehen.

Diese Unklarheiten machen eine exakte Berechnung der Rückläufe schwierig und ungenau. Daher bleibt nur eine grobe Schätzung der Vorgaben übrig. Als Folgerung für zukünftige Untersuchungen bleibt festzuhalten, dass diese Angaben besser kontrolliert werden müssen.

Als vorläufiges Fazit lässt sich dennoch festhalten, dass eine 2. Erinnerung einen Effekt aufweist, so dass sich deren Einsatz lohnt.

9. Herausforderungen bei Online-Befragungen

Der 13. Studierendensurvey ist in vielerlei Hinsicht als Testlauf zu verstehen: Es sollte überprüft werden, ob die Umstellung auf eine Online-Befragung machbar ist und was die Umstellung für den Rücklauf, die Zusammensetzung der Stichprobe und die Qualität der Daten bedeutet. Auch wenn die inhaltlichen Analysen optimistisch stimmen, so sind doch vor allem die geringe Rücklaufquote, die vielen Abbrüche sowie der Item Non-Response Probleme, die in künftigen Erhebungen angegangen werden müssen. Abgesehen davon sind Online-Befragungen noch mit weiteren Herausforderungen verbunden, auf die im Folgenden näher eingegangen werden soll. Zu diesem Zeitpunkt kann noch keine finale Strategie für künftige Erhebungen präsentiert werden. Vielmehr geht es darum Themen zu identifizieren, die sich in der aktuellen Erhebung als problematisch herausgestellt haben, und Handlungsalternativen zu entwickeln. Zunächst werden Überlegungen zur Erhöhung der Teilnahmebereitschaft zusammengefasst. Im Anschluss werden die wichtigsten Aspekte zum Datenschutz diskutiert, da diesem Thema bisher noch zu wenig Rechnung getragen wurde. Daran schließt sich eine kurze Diskussion über das Thema responsive design und über technische Aspekte von Web-Surveys an. Kapitel 10 markiert mit einem Fazit den Abschluss des Berichts und gibt einen Ausblick auf die nächste Erhebung des Studierendensurveys.

9.1 Teilnahmebereitschaft

Die sinkende Bereitschaft der Studierenden, an Befragungen teilzunehmen, ist eines der Hauptprobleme der Studierendenforschung. In den letzten zehn Jahren ist diese Bereitschaft systematisch gesunken. Dies gilt für Befragungen über einen Papierfragebogen ebenso wie über einen Onlinezugang. Dabei handelt es sich nicht um ein regionales oder nationales Problem, sondern um ein fortschreitendes globales Phänomen (Ramm 2014). Ein Grund für diesen Rückgang ist sicherlich in der erhöhten Befragungsdichte von vielen unterschiedlichen Institutionen zu sehen, sodass die Studierenden, zusätzlich zu den Befragungen der eigenen Hochschule, schlichtweg überfordert und demotiviert sind.

Es sei an dieser Stelle herausgehoben, dass die viel beschworene „Überfischung“ der Studierenden als Untersuchungsobjekt nicht auf die beiden großen Studierendenbefragungen zurückgehen, die Sozialerhebung und den Studierendensurvey, da beide seit Bestehen gerade diese Überforderung durch einen 2-3 jährigen Erhebungsrhythmus bewusst beachtet haben.

Die Befunde des 13. Studierendensurveys verweisen darauf, dass die Online-Befragung einen deutlich geringeren Rücklauf erbringt als die Papierbefragung. Insofern ist dieses Problem für Online-Befragungen noch ausgeprägter vorhanden.

Für die Zukunft der Studierendenforschung ist die Teilnahmebereitschaft der Studierenden das bedeutendste Problem, dem durch verschiedene Maßnahmen entgegengewirkt werden muss. Allerdings sind die Studierenden nur die eine Seite des Problems, denn der Weg zu den Studierenden führt über die Hochschulen. Sie müssen zuerst zustimmen, dass eine Befragung bei ihren Studierenden durchgeführt werden darf. Erst dann kommt in einem zweiten Schritt der Zugang zu den Studierenden.

Insofern müssen also ebenso Maßnahmen entwickelt werden, um die Hochschulen zur Teilnahme zu bewegen, wie es auch für die Studierenden notwendig ist. Für die Hochschulen ist eine Online-Befragung insgesamt weniger aufwendig als eine Papierbefragung. Daher ist die Umstellung auf Online bereits eine Verringerung des Aufwandes für die Hochschulen. Gleichzeitig müssen sie überzeugt werden, dass die Befunde der Erhebungen auch für ihre Hochschule sinnvolle Daten und Folgerungen erlauben. Dazu müssen Überlegungen angestellt werden, welche Materialien oder Dienstleistungen den Hochschulen zur Verfügung gestellt werden können. Bisher erhalten die am Studierendensurvey beteiligten Hochschulen Indikatorenbände mit ihren Daten zu verschiedenen zentralen Fragen. Mit

Blick auf zukünftige Entwicklungen (Ausweitung des Samples) wäre auch ein Datenzugang für die Hochschulen eine interessante Lösung.

Hinzu kommt eine Absprache wichtiger Instanzen (z.B. HRK, DSW), die bestimmte Untersuchungen unterstützen könnten. Sowohl der Studierenden-Survey als auch die Sozialerhebung des Deutschen Studentenwerks sind aufgrund ihrer Auslegung hervorzuheben, da sie im Vergleich zu anderen Studierendenbefragungen eine gesamtdeutsche Betrachtung der Studiensituation erlauben und darüber hinaus weitestgehend repräsentative Aussagen ermöglichen.

Maßnahmen für die Studierenden müssen demgegenüber jedoch weiterreichen. Hier gilt es vorrangig die Motivation der Studierenden zu steigern, sich an solchen Befragungen zu beteiligen. Dazu können verschiedene Maßnahmen herangezogen werden:

- ausreichende Informationen über Sinn und Zweck der Befragung,
- deutliche Ankündigung,
- Verdeutlichung des eigenen Nutzens,
- ansprechendes, responsives Design (Umfrage darf nicht an wenige Gerätetypen gebunden sein),
- kurze Dauer der Befragung,
- gezieltes Erinnern (nur Personen, die noch nicht geantwortet haben),
- Incentives,
- komfortable Möglichkeit des Wiedereinstiegs,
- Unterstützung durch Institutionen und Stakeholder.

Online-Befragungen haben den Vorteil, dass das Medium erlaubt, auch sehr große Stichproben einzuladen, wodurch ausreichend große Stichproben gewährleistet werden können. Das Problem ist dann die Klärung der Repräsentativität. Des Weiteren ist das Medium geeignet, um Informationen zur Verfügung zu stellen und anspruchsvolle Designs zu erzeugen. Auch einen komfortablen Wiedereinstieg kann die Technik gewährleisten.

Durch die Vergabe von Zugangscodes (Tokens) können gezielte Erinnerungen geschaltet werden, da nur diejenigen Studierenden erinnert werden, die an der Befragung noch nicht teilgenommen haben, bzw. diese noch nicht abgeschlossen haben.

Ein weiterer Vorteil der Onlinenutzung sind Paradata. Wie bereits dargestellt erlauben sie Aussagen über die Navigation der Teilnehmer im Fragebogen, über deren Antwortverhalten sowie die benötigte Dauer pro Seite. So kann das Abbruchverhalten, die Dauer der Beantwortung, der Wiedereinstieg, etc. analysiert werden, eigenständig und in Kombination mit Daten des Surveys. Die Erhebung solcher Daten birgt also neue Analysepotentiale, weshalb die Nutzung dieser Daten von hohem Interesse ist. Allerdings stehen diesen Interessen u.U. Datenschutzbestimmungen entgegen.

9.2 Datenschutzbestimmungen

Das Thema Datenschutz hat in den letzten Jahren in der Öffentlichkeit enorm an Bedeutung gewonnen, und auch wenn die Forschung gewisse Privilegien genießt und nicht willkürlichen Einschränkungen unterworfen werden kann, so müssen sich Wissenschaftler doch eingehend mit den datenschutzrechtlichen Bestimmungen vertraut machen. Insbesondere wenn die Forschung die Erhebung personenbezogener Daten miteinschließt, muss die Wissenschaftsfreiheit gegen individuelle Persönlichkeitsrechte abgewogen und die Rechte des Individuums auf informationelle Selbstbestimmung gewahrt werden. Dass Datenschutz in der sozialwissenschaftlichen Forschung eine große Rolle spielt, zeigt sich in der festen Verankerung von Handlungsempfehlungen zum Thema in den Richtlinien des

Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute e.V., der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V., dem Berufsverband Deutscher Markt- und Sozialforscher e.V. sowie der Deutschen Gesellschaft für Online-Forschung e.V.. Des Weiteren beschreibt die Deutsche Gesellschaft für Soziologie in ihrem Ethik-Kodex (DGS, 2017) den korrekten Umgang mit Probanden, und auch GESIS, das Leibniz-Institut für Sozialwissenschaften, steuert ein Paper mit Hinweisen zum Datenschutz bei (Kinder-Kurlanda & Watteler, 2015). Im Folgenden sollen daher die wichtigsten Aspekte zum Thema personenbezogene Daten, Datenschutz und Anonymität in aller Kürze dargestellt werden.

9.2.1 Personenbezogene Daten

Personenbezogene Daten sind „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer natürlichen Person (Betroffener)“ (§3 Abs. 1 BDSG). Vor jeder Untersuchung ist daher zu prüfen, ob ein Personenbezug besteht, entweder durch das Datum selbst (Name, Matrikelnummer, in Online-Befragungen auch IP-Adresse) oder durch eine Kombination von Daten (bspw. durch die Kombination von Studienfach, Geschlecht und Fachsemester). Grundsätzlich sollte jeder Personenbezug vermieden werden, soweit die Forschungsfrage dies zulässt. Die Entscheidung für oder gegen die Erhebung personenbezogener Daten sollte sich dabei immer am Grundsatz der Datenminimierung orientieren: Es sollen „so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen“ (§3a BDSG). Vor jeder Untersuchung ist daher zunächst zu klären, ob ein Personenbezug vorliegt oder dieser nachträglich hergestellt werden kann.

Ein weiteres Hindernis stellen besonders sensible Informationen dar. Dazu zählen nach §3 Abs.9 BDSG Angaben über die ethnische Herkunft, politische, religiöse oder philosophische Einstellungen, die Gewerkschaftszugehörigkeit, Angaben zur Gesundheit oder zu sexuellen Neigungen. Für den Studierendensurvey sind hier insbesondere die Informationen zu politischen Einstellungen hervorzuheben, welche einen zentralen Themenbereich des Instruments ausmachen. Werden Daten dieser Art (automatisch) verarbeitet und ist ein Personenbezug gegeben, so ist eine *Vorabkontrolle* notwendig: Diese prüft die Rechtmäßigkeit des geplanten Verfahrens und ob alle gesetzlichen Vorschriften eingehalten werden. Eine Vorabkontrolle ist nicht nötig, wenn eine gesetzliche Mitwirkungspflicht der Personen besteht, sie bereits eine Einwilligung erteilt haben, oder wenn die Datenerhebung bzw. Datenverwendung vertraglich festgehalten wurde (§4d BDSG).

9.2.2 Einwilligungsprinzip – Informed Consent

Prinzipiell dürfen personenbezogene Daten nicht ohne Einwilligung der Befragten verarbeitet werden (§4a BDSG). Die Teilnehmer müssen ausreichend über das Forschungsvorhaben und die durchführende Institution aufgeklärt werden. Außerdem müssen sie auf die Freiwilligkeit ihrer Teilnahme und ihr Widerrufsrecht hingewiesen werden. Wird in der Einwilligung nicht auch explizit die Weitergabe der Daten in anonymisierter Form erlaubt, können nur die Primärforscher die Daten verarbeiten (Kinder-Kurlanda & Watteler, 2015).

9.2.3 Technische und organisatorische Maßnahmen

Werden personenbezogene Daten erhoben und verarbeitet, sind technische sowie organisatorische Maßnahmen zu treffen (vgl. §9 BDSG). Dazu zählen verschiedene Zugangs- bzw. Zugriffskontrollen; die Einzelheiten können in der Anlage zu §9 des BDSG nachgelesen werden¹⁶. Werden die Daten nicht von der Forschungsgruppe selbst, sondern z.B. von einem externen Dienstleister erhoben, muss

¹⁶ <https://dejure.org/gesetze/BDSG/9.html> (Letzter Zugriff: 19.07.2017)

dieser die zum Schutz ergriffenen Maßnahmen auflisten. Das DZHW, welches für die technische Umsetzung der Online-Befragung des 13. Studierendensurveys verantwortlich war, hat zu diesem Zweck eine ausführliche Dokumentation der allgemeinen technischen und organisatorischen Schutzmaßnahmen nach §9 des BDSG vorgelegt.

9.2.4 Anonymisierung und Pseudonymisierung

Des Weiteren sind die Daten so früh wie möglich zu anonymisieren/ pseudonymisieren. Das bedeutet, dass die Daten so zu verändern sind, dass sie „nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbar natürlichen Person zugeordnet werden“ können (vgl. §3 Abs. 6 BDSG). Dabei reicht es nicht aus, dass der Bezug rein theoretisch herstellbar ist, vielmehr muss das Unterfangen auch praktisch umsetzbar sein (Pflüger & Dobel, 2014). Für die Anonymisierung kommen mehrere Strategien mit unterschiedlicher Reichweite infrage. Bei der absoluten Anonymisierung werden die „Angaben durch Aggregation bzw. Verfremdung oder Schwärzen bzw. Löschen von Daten so verändert, dass eine Zuordnung [...] nach aktuellen technischen Möglichkeiten unmöglich wird“ (RatSWD, Handreichung Datenschutz, 2017). Bei der faktischen Anonymisierung soll durch eine Informationsreduktion bzw. Informationsveränderung eine Zuordnung zu Einzelpersonen eingeschränkt werden, ohne den analytische Gehalt der Daten einzuschränken. Die formale Anonymisierung schließlich sieht das Löschen des betreffenden Datums vor (Name, Matrikelnummer, IP-Adresse).

9.2.5 Situation und Vorgehen im Studierendensurvey

Im Falle des Studierendensurveys werden Informationen erhoben, die durch eine Kombination Einzelpersonen identifizieren könnten. Dies wäre aber nur den teilnehmenden Hochschulen möglich, denn der AG Hochschulforschung liegen zu keinem Zeitpunkt Adressen- oder Namenslisten der Studierenden vor. Die Befragungsunterlagen wurden verpackt an die Hochschulen geliefert, die diese dann mit den Adressen der zufällig ausgewählten Studierenden etikettierten.

Um einen Personenbezug herstellen zu können, müssten die Hochschulen die Adresslisten noch nach Abschluss der Feldphase speichern, und anschließend mit den Umfragedaten verknüpfen: Entweder über ein Matching der Adressdaten mit den für die Online-Befragung verschickten Zugangscodes, oder für Teilnehmer der schriftlichen Befragung über eine Kombination mehrerer Einzelmerkmale. Um ein solches Vorgehen im Vorhinein zu unterbinden, werden mehrere Maßnahmen von der AG durchgeführt:

- 1) Bevor die Hochschulen einen Datensatz mit den Umfragedaten ihrer Studierenden erhalten, müssen sie versichern, dass sie die Adressdatei gelöscht haben. Eine nachträgliche Rekonstruktion der Stichprobe sollte sich schwierig gestalten und einen unverhältnismäßig großen Aufwand darstellen.
- 2) Die beteiligten Hochschulen erhalten nur einen anonymisierten Datensatz: Die Angaben zum Alter und der Anzahl der Fach- und Hochschulsemeister werden kategorisiert, die Information über die Anzahl der Kinder wird dichotomisiert. Darüber hinaus werden die Zugangscodes der Studierenden für die Online-Befragung aus dem Datensatz gelöscht, ebenso wie die Angaben zu Staatsangehörigkeit und dem Geburtsort sowie zu chronischen Erkrankungen.
- 3) Nichtbeteiligte Dritte, die die Datensätze des Studierendensurveys für wissenschaftliche Arbeiten verwenden wollen, erhalten die Daten ohne die Einzelhochschulkennung.

Bei der Anonymisierung der Daten wird darauf geachtet, dass auch nach Kombination zentraler Merkmale die Zellbesetzung nie geringer ist als drei Fälle. In Zukunft sollen die Informationen zum Datenschutz weiterentwickelt und leichter zugänglich gemacht werden, um die Transparenz für Teilnehmer bzw. alle Interessierten zu erhöhen. Vor allem die Möglichkeit einer gezielten Erinnerung durch den Kontakt über individuelle E-Mail-Adressen muss unter datenschutzrechtlichen Aspekten geprüft werden. Denn hier wäre zumindest während der Feldphase prinzipiell ein Personenbezug gegeben. In diesem Fall wären alle datenschutztechnischen Maßnahmen (Einwilligung, Verfahrensverzeichnis, Vorabkontrolle) durchzuführen.

In Hinblick auf die neuen bzw. erweiterten Analysepotenziale durch die Erhebung und Verarbeitung von Paradata ist in zukünftigen Erhebungen auf eine offene Kommunikation bezüglich der damit verbundenen Probleme im Datenschutz zu achten. Wissenschaftler/innen müssen sich dazu verpflichtet fühlen, die Teilnehmer ihrer Untersuchungen offen und umfassend über diese Themen zu informieren, um ihnen zu vermitteln, dass ihre Daten sicher sind und nicht zweckentfremdet werden.

9.3 Responsive Design

Eine der größten Veränderungen im Bereich der Online-Forschung ist die zunehmende Nutzung mobiler Endgeräte, insbesondere von Smartphones und Tablets, zur Bearbeitung von Web-Surveys. Wo sich Forscher(innen) früher bei der Konzeption und Programmierung ihres Online-Fragebogens an einer überschaubaren Anzahl an Browsertypen, Betriebssystemen und Geräten orientieren konnten, entstehen immer neue technische Möglichkeiten und Kombinationen, die berücksichtigt werden müssen. Nach einem Bericht des Statistischen Bundesamtes zur Ausstattung privater Haushalte mit ausgewählten Gebrauchsgütern (Statistisches Bundesamt Fachserie 15, Reihe 2, 2016) besitzen 95 % der Haushalte in Deutschland ein Smartphone, etwas weniger (88,6 %) einen Desktop-PC, 68,5 % haben einen Laptop und immerhin 37,6 % ein Tablet. In Haushalten der Altersgruppe 18-25 Jahre liegen diese Anteile noch einmal deutlich höher: Smartphone 100 %, Desktop-PC 98,3 %, Laptop 97 %, Tablet 45,1 %.

Über einen Internetzugang verfügen allerdings auch weitere mobile und stationäre Geräte, wie bspw. Smart-TVs, Spielekonsolen oder E-Reader. Potentielle Teilnehmer könnten auch über diese Geräte auf das Internet zugreifen. Mit der ständigen Weiterentwicklung einzelner Geräte und ihrer Vernetzung mit dem Internet steigt diese Wahrscheinlichkeit an. Web-Surveys können somit nicht länger nur für einen Gerätetyp konzipiert werden, sondern müssen theoretisch flexibel auf alle verwendeten Geräte reagieren. Mit responsive design ist genau das gemeint: Ausgehend von den Eigenschaften des verwendeten Geräts wird ein entsprechend optimiertes Design angezeigt.

Damit müssen sowohl die Darstellungsgröße des Bildschirms, das Betriebssystem, die Browser-Software als auch die Eingabemöglichkeiten für die Antworten in Betracht gezogen werden. Ein Computerbildschirm als Arbeitsstation oder Laptop hat andere Darstellungs- und Eingabemöglichkeiten als ein Smartphone. Die Anzeige einer kompletten Fragebogenseite mit mehreren Items kann am Computer problemlos sein, während sie am Smartphone zu klein wird oder zu umständlich, weil der Bildschirm in zwei Richtungen verschoben werden muss. Die Antwortvorgaben können am Computer mit Tastatur oder Maus eingegeben werden, während am Touchscreen eines Smartphone meist der Finger verwendet wird.

Mario Callegaro (2010) schlägt verschiedene Lösungen mit unterschiedlicher Reichweite für die Herausforderung durch mobile Geräte vor:

- a. *Nichts verändern*: Es werden keine Änderungen am Instrument gemäß des verwendeten Geräts vorgenommen. Stattdessen wird der Gerätetyp erfasst (user agent string), Befragte mit

mobilen Endgeräten markiert und ggf. von der Analyse ausgeschlossen bzw. auf Unterschiede zum Rest der Befragten überprüft.

- b. *Nutzer mobiler Endgeräte abfangen/ blocken*: Die Teilnehmer werden bereits im Vorfeld (Einladung, Startseite) darauf hingewiesen, dass die Online-Befragung nicht für die Bearbeitung auf mobilen Endgeräten ausgelegt ist. Dieser Ansatz wurde im 13. Studierendensurvey verfolgt und von einem Großteil der Befragten auch beherzigt (92 % nutzen einen Desktop-PC oder Laptop zur Bearbeitung, 8 % ein Tablet/ Smartphone). Dennoch könnten Teilnehmer diese Hinweise übersehen oder gezielt ignorieren. Eine Alternative, die hier noch einen Schritt weiter geht, ist das gezielte Blockieren von Befragten, die versuchen, mit einem Tablet/ Smartphone auf die Umfrage zuzugreifen, indem auf einer Landingpage der user agent string ausgelesen und den Teilnehmern eine entsprechende Nachricht angezeigt wird¹⁷. Die Gefahr besteht in einem erhöhten Abbruch dieser „abgefangenen“ Befragten ohne erneutem Login an einem geeigneten Gerät.
- c. *Die Umfrage für alle gängigen Gerätetypen anpassen*: Der Fragebogen wird für die üblichen Gerätetypen angepasst, alle anderen werden wie in b) weitergeleitet. Dieses Vorgehen erfordert weitreichende Anpassungen im Layout und der Formatierung der Befragung, der Anzahl und Gestaltung der Antwortkategorien, der maximalen Anzahl an Items pro Seite sowie an Größe und Qualität von Multimediainhalten (Bilder, Videos, Ton).
- d. *Anpassung der Umfrage an alle Geräte*: Der Fragebogen wird für die Anzeige an allen Geräten optimiert.

Im Gegensatz zu gängigen Interpretationen von responsive (Web)design gehen die beiden letzten Strategien weiter, indem nicht einfach der bestehende Fragebogen neu skaliert oder Inhalte einer Seite anders verteilt werden, sondern der Fragebogen überall dort verändert/ angepasst wird, wo es die Darstellungsmöglichkeiten des verwendeten Geräts nötig machen. Wie weitreichend diese Anpassungen ausfallen (müssen), ist noch nicht absehbar. Sollte sich herausstellen, dass für die Nutzung via Smartphone/ Tablet etc. eine gänzlich andere Darstellungsform verwendet werden muss, gilt es Zugangsmöglichkeiten zu programmieren, die bei Verwendung mobiler Geräte automatisch die passende Darstellung zuordnen.

Da es sich bei Studierenden um eine mobile und technisch versierte Personengruppe handelt („digital natives“ (Prensky, 2001)), ist eine Berücksichtigung mobiler Geräte und damit eine umfassende Anpassung des Instruments an diese Geräte notwendig. Damit kann das Design zwischen den unterschiedlichen Gerätetypen variieren. Diese Unterschiede in der Darstellung und der Eingabe müssen im Vorfeld getestet werden, sodass auch bei unterschiedlicher Geräte-Nutzung gewährleistet werden kann, dass keine Behinderungen durch die Technik auftreten.

9.3.1 Device Effekte

Ein weiterer Punkt, der im Zusammenhang mit dem responsive design berücksichtigt werden muss, sind Effekte, die durch das verwendete Gerät entstehen können. In Kapitel vier wurde bereits darauf hingewiesen, dass Nutzer mobiler Endgeräte mehr Item Non-Response produzieren, eine Tendenz zu unerwünschtem Antwortverhalten besteht (straightlining) und außerdem die Bearbeitungsdauer der Smartphone-Nutzer auf eine schnelle, weniger ernsthafte Bearbeitung schließen lässt. Noch steht

¹⁷ Beispieltext: „Sie versuchen mit einem mobilen Endgerät auf die Befragung zuzugreifen. Diese Online-Befragung ist nicht für die Bearbeitung an Tablets, Smartphones oder anderen mobilen Geräten ausgelegt. Bitte wechseln Sie für die Bearbeitung an einen PC oder Laptop.“

nicht fest, ob diese Effekte durch die suboptimale Darstellung auf Tablets und Smartphones zurückzuführen ist, oder ob es sich tatsächlich um Device Effekte handelt. Fest steht jedoch, dass bei einer Anpassung des Instruments an mobile Endgeräte solche Effekte kontrolliert werden müssen, denn auch in der einschlägigen Literatur finden sich Unterschiede zwischen Nutzern verschiedener Gerätetypen.

Struminskaya, Weyandt und Bosnjak (2015) konnten in ihrer Analyse eine schlechtere Datenqualität für Befragte, die mobile Endgeräte nutzten, feststellen. Insbesondere Smartphone Nutzer weisen einen erhöhten Item Non-Response auf und zeigen häufiger straightlining als Antworttendenz. Zudem verfassten sowohl Tablet- als auch Smartphone-Nutzer kürzere Antworten in offenen Fragen als PC-Nutzer. Ein Effekt, den auch Mavletova (2013) in ihrer Studie feststellt. Allerdings gibt es auch gegenteilige Ergebnisse zur Datenqualität. Lugtig und Toepoel (2016) untersuchten in ihrer Studie den Umfang verschiedener Messfehler über mehrere Zeitpunkte hinweg. Die Unterschiede, die sie zwischen PC-Nutzern und Nutzern mobiler Endgeräte zunächst fanden, verschwanden, nachdem für den Wechsel zwischen Geräten kontrolliert wurde. Demnach produzieren einige Befragte unabhängig vom verwendeten Gerät mehr oder weniger Messfehler; es konnte also kein Mode-Effekt nachgewiesen werden.

Darüber hinaus benötigen Handy-Nutzer¹⁸ laut der Untersuchung von Mavletova (2013) länger als PC-Nutzer für eine Online-Befragung desselben Umfangs: Handy-Nutzer benötigen dreimal so lange wie Personen, die die Umfrage an einem PC bearbeiteten. Die Completion Rates waren deutlich niedriger in der Gruppe der Handy-Nutzer, die Breakoff Rates lagen dafür höher. Zu den gleichen Ergebnissen kommen auch de Bruijne und Wijnant (2013): Tablet und Smartphone-Nutzer wiesen eine signifikant längere Bearbeitungsdauer auf als PC-Nutzer, und auch die Completion Rate lag in der mobilen Gruppe deutlich niedriger.

Die Kontrolle und Untersuchung solcher durch das Gerät verursachter Effekte wird in künftigen Befragungen eine nicht unerhebliche Rolle spielen, ganz analog zu der aktuellen Analyse möglicher Mode-Effekte zwischen der Papier- und der Online-Befragung. Durch die Adaption des Instruments für mobile Endgeräte müssen verschiedene Indikatoren zwischen den einzelnen Befragungsgruppen kontrolliert werden, um zu überprüfen, ob es sich um Mode-Effekte handelt, oder ob die Unterschiede auf die Befragten zurückgehen. Gleichzeitig muss ein Kompromiss erarbeitet werden: Einerseits soll das Instrument die Möglichkeiten des jeweiligen Geräts nutzen und Beschränkungen sowie Gestaltungsrichtlinien berücksichtigen. Andererseits sollen die daraus resultierenden individuellen Fragebogen zu möglichst vergleichbaren Ergebnissen führen, oder in den Analysen müssen Unterschiede zwischen den Befragungsgruppen kontrolliert werden (Mode-Effekte), damit Aussagen über die gesamte Befragungsgruppe hinweg möglich sind.

Wird den Teilnehmern die Entscheidung überlassen, welches Gerät er/ sie zur Bearbeitung der Befragung nutzt, steigt das Risiko für eine Verzerrung durch Selbstselektion (Self Selection Bias). Wie in Kapitel fünf gezeigt, existieren zwischen den Befragten der randomisierten und selbstselektierten Papier-/ Onlinegruppe Unterschiede in zentralen demographischen Eigenschaften. Zudem traten auch bei den inhaltlichen Indikatoren vereinzelt Differenzen auf (vgl. Kapitel sieben), wenn auch nur sehr geringe. Darum gilt es auch in zukünftigen Befragungen Verzerrungen dieser Art zu kontrollieren.

¹⁸ In der Studie wurden neben Smartphones auch so genannte feature phones untersucht. Feature phones weisen im Vergleich zu Smartphones gewisse Beschränkungen auf. Sie verfügen meist über eine eingeschränkte Eingabefunktion (kein Touchscreen) und verwenden andere Betriebssysteme als die gängigen (Android, iOS, etc), weshalb auch nur begrenzt Apps verwendet werden können.

9.3.2 Barrierefreiheit

Eng mit dem responsive (Web-)design verbunden ist das Thema Barrierefreiheit. Personen mit Behinderungen stoßen im Internet auf verschiedene Hindernisse: Es fehlen zum Beispiel Navigationshilfen (Sprachausgabe bzw. Sprachsteuerung nicht implementiert), die Navigation ist nur mittels Maus, aber nicht per Tastatur möglich, es existieren keine Textalternativen zu Bild- oder Videoinhalten und Timeouts (bei der Bearbeitung einer Aufgabe, dem Ausfüllen eines Formulars, etc.) sind zu kurz bemessen¹⁹.

Hierbei handelt es sich nur um eine Auswahl an Hindernissen, die bei der Programmierung der Online-Befragung berücksichtigt werden müssen. Ernsthafte Lösungswege erfordern weitreichende Anpassungen des Instruments. Die Web Content Accessibility Guidelines (WCAG) 2.0 geben Empfehlungen für eine barrierefreie Gestaltung von Webseiten, die sich so auch für Online-Befragungen übernehmen lassen (Caldwell, Cooper, Guarino Reid, Vanderheiden, 2008):

- Der Ton in einem Video sollte bspw. auch durch Untertitel angezeigt werden, und für alle Bildinhalte sollten schriftliche Alternativen bestehen. Videos sollten außerdem einen Kommentar in Gebärdensprache enthalten.
- Zudem muss es sehbehinderten Befragten möglich sein, die Bildschirmdarstellung anzupassen, also zum Beispiel Texte/ Abbildungen zu vergrößern (Bildschirmlupe).
- Beim Farbschema ist auf einen starken Kontrast zwischen Vorder- und Hintergrund zu achten bzw. den Teilnehmern die Möglichkeit zu bieten, das Farbschema anzupassen.
- Teilnehmer sollten sich Texte mittels geeigneter Software (Screenreader) vorlesen lassen können. Alternativ wäre auch die Ausgabe der Textinhalte in Brailleschrift denkbar.
- Für Personen mit Hörbehinderung empfiehlt sich eine Übersetzung der Fragen und Antworten in Gebärdensprache und die Einbindung eines entsprechenden Videos auf der jeweiligen Fragebogenseite.
- Befragte mit sensomotorischen Einschränkungen benötigen Alternativen zur Bearbeitung mittels Tastatur und Maus (Benutzung nur über die Maus mittels Bildschirmtastatur, Mund-Maus, Joystick).

Anhand dieser (nicht erschöpfenden) Liste wird deutlich, wie viel Entwicklungsarbeit allein für die Programmierung der Software anfällt. Darüber hinaus ist auch eine intensive Einarbeitung in das Themenfeld notwendig, um überhaupt konkrete Lösungen für die Befragung zu erarbeiten. Werden die Anforderungen der Barrierefreiheit mit jenen des responsive design kombiniert, müssten zahlreiche Fragebogenversionen entwickelt werden, um beiden Prinzipien wirklich gerecht zu werden. Diese Entwicklung sollte durchdacht erfolgen und wird daher sehr viel Zeit in Anspruch nehmen.

9.4 Technische Anforderungen

Ein weiterer nicht zu unterschätzender Punkt sind gestiegene technische Anforderungen. Forscher (innen) müssen sichergehen, dass sie ihre Fragebögen für eine Vielzahl an Gerätetypen, Betriebssystemen, Browsertypen und Bildschirmauflösungen programmieren bzw. dass die von ihnen verwendete Software alle gängigen Kombinationen berücksichtigt und unterstützt. Dazu ist eine Einarbeitung in die Materie nötig. Die Verwendung und Auswertung der user agent strings sowie anderer Paradata erfordern ebenfalls eine gewisse Expertise, und auch die Entwicklung spezieller Befragung-Apps präsentiert sich als Alternative, die allerdings mit sehr spezifischem Fachwissen verbunden ist.

¹⁹ Für eine erste Übersicht über das Thema: <http://www.einfach-barrierefrei.net/start/willkommen.html> (Letzter Zugriff: 19.07.2017).

Mit der parallel immer weiter voranschreitenden Durchdringung der Gesellschaft durch das Internet und der steigenden Verfügbarkeit von Smartphones und anderer internetfähiger mobiler Geräte stellen sorgfältig geplante „multi-device-surveys“ die einzige logische Strategie dar. Wie bereits in Abschnitt 9.3 erläutert wurde, sind damit umfangreiche Anpassungen des Instruments verbunden. Zudem müssen zahlreiche technische Aspekte berücksichtigt werden, wie etwa gerätespezifische Betriebssysteme oder die Spannweite möglicher Verbindungsarten und damit verbundenen Downloadraten (EDGE, UMTS, HSPA, HSPA+, LTE, LTE-Advanced), etc.

Weil zum jetzigen Zeitpunkt keine Informationen über das durchschnittliche Download-Volumen oder durchschnittliche Verbindungsgeschwindigkeiten der von Studierenden verwendeten Smartphones vorliegen, sind bei der (Weiter-) Entwicklung der Online-Befragung vor allem zwei Punkte zu beachten: 1) sollten möglichst keine oder wenige Inhalte während der Befragung heruntergeladen werden müssen (Bilder, Video, Tonaufnahme), 2) sollte die Gesamtdauer der Befragung insbesondere für Nutzer mobiler Geräte gering gehalten werden. Beide Maßnahmen sollen eine übermäßige Belastung des Datenvolumens der Befragten verhindern.

Des Weiteren sehen sich alle groß angelegten Studierendenbefragungen mit dem Vorwurf mangelnder Vergleichbarkeit konfrontiert. Eine Herausforderung liegt darin, den Studierendensurvey stärker mit anderen Befragungen (Sozialerhebung, best-Studie, NEPS) zu verschränken und die Anschlussfähigkeit des Instruments zu erhöhen. Dazu sind vielfältige Maßnahmen nötig wie zum Beispiel die Abstimmung der Sozialdaten, der inhaltlichen Themenbereiche, der Samplestruktur und der Stichprobenziehung zwischen verschiedenen Befragungen. Die technische Schwierigkeit liegt darin, diese verschiedenen Befragungen auch tatsächlich zusammenzuführen: Entweder über die Integration der verschiedenen Fragebögen in einer gemeinsamen Umfrage mit derselben Software, oder aber durch an die eigentliche Erhebung anschließende Maßnahmen (Gewichtung, Matching).

Diese und weitere Überlegungen stellen hohe Anforderungen an die Programmierungsmöglichkeiten der Befragungssoftware sowie die gesamte technische Infrastruktur eines Projekts. Das in der aktuellen Studie verwendete System des Zofar-Teams des DZHW weist bereits eine sehr gute Flexibilität in der Programmierung/ Gestaltung der Fragen auf, ist aber bei der Erhebung zusätzlicher Prozessdaten noch entwicklungsfähig. Zudem muss geprüft werden, ob die nötigen Anpassungen des Designs, insbesondere betreffend der Barrierefreiheit und des responsive designs im Allgemeinen, vollumfänglich umgesetzt werden können.

10 Fazit und Ausblick

Der vorliegende Methodenbericht hat das vorrangige Ziel zu prüfen, ob die in der 13. Erhebung des Studierendensurveys gewonnenen Daten aus der Online-Befragung und die aus der gleichzeitig durchgeführten Papierbefragung übereinstimmen. Hintergrund war die Frage, ob bei einem zukünftigen Umstieg auf Online-Befragungen die Datenqualität beibehalten werden kann, ob die inhaltlichen Aussagen der Befragten der Online-Befragung mit jenen der schriftlichen Befragung vergleichbar sind und ob die Zeitreihen weiterhin fortgeführt werden können.

Dazu musste untersucht werden, inwieweit sich die Samplestruktur, die Teilnahme und die Befunde zwischen den beiden Gruppen gleichen bzw. unterscheiden. Als Ergebnis kann festgehalten werden, dass die gefundenen Unterschiede zwischen den beiden Gruppen vernachlässigt werden können, die beiden Datenquellen also zusammen analysiert werden dürfen. Das bedeutet, dass ein Umstieg auf eine Online-Befragung die Ergebnisse des Studierendensurveys nicht nachhaltig verändern wird und die Zeitreihen somit weitergeführt werden können.

Mit Blick auf die Zusammensetzung der gezogenen Stichproben weist die Gruppe der online Befragten sogar eine größere Nähe zur Grundgesamtheit der deutschen Studierenden an Universitäten und Fachhochschulen auf. Die im Studierendensurvey bisher immer auftretende Überrepräsentanz weiblicher Teilnehmer gleicht sich in der Online-Befragung aus, so dass das Geschlechterverhältnis fast 50:50 beträgt. Auch die Verteilung der Studierenden über die Fächergruppen hinweg bewegt sich näher an der Verteilung in der Grundgesamtheit.

Mit dem vorliegenden Bericht sollte vorrangig das praktische Vorgehen dokumentiert und die Breite der notwendig zu untersuchenden Aspekte herausgestellt werden. Die dazu einsetzbaren statistischen Verfahren ließen sich sicherlich erweitern und ebenso ein differenzierteres Methodenarsenal verwenden. Doch sollte dieser Bericht keine methodische Abhandlung über statistische Verfahren darstellen, sondern einzig die praktische Frage klären, ob die beiden Datenquellen gemeinsam nutzbar sind. Das Ergebnis stimmt positiv: Die wenigen gefundenen Unterschiede sprechen nicht ernsthaft gegen einen Wechsel des Erhebungsmodus. Auf ein paar wichtige Ergebnisse und ihre Konsequenzen, vor allem in Hinblick auf zukünftige Erhebungen, soll im Folgenden noch einmal kurz eingegangen werden.

Rücklauf und Nonresponse

Zwei Themen, die in zukünftigen Erhebungen besondere Aufmerksamkeit verdienen, sind der Rücklauf und die itembezogene Antwortverweigerung. Auch wenn der Rücklauf der schriftlichen Befragung in der aktuellen Erhebung mit 16,1 % ebenfalls deutlich hinter den Erwartungen zurückblieb, schneidet die randomisierte Onlinegruppe mit gerade einmal zehn Prozent noch deutlich schlechter ab. Zwar reicht die absolute Gruppengröße für die Analysen aus und lässt sich durch die Ausweitung des Samples auch weiter erhöhen. Trotzdem stellt sich angesichts des hohen Unit Non-Response die Frage, ob sich die wenigen kooperativen Studierenden nicht doch in ihren Einstellungen und Orientierungen gravierend von ihren unkooperativen Kommilitonen unterscheiden. Diese Unsicherheit besteht und muss in künftigen Erhebungen aktiv angegangen werden. Zum einen muss mittels unterstützender Maßnahmen (verstärkte Werbemaßnahmen, ggf. Incentives, Kommunikation des Nutzens einer Teilnahme, etc.) die Teilnahmebereitschaft der Studierenden wieder erhöht und damit der Rücklauf verbessert werden. Andererseits müssen Methoden entwickelt werden, um die Gründe für die Verweigerungshaltung bei den unkooperativen Studierenden zu erfassen. Gleiches gilt für Studierende, die die Befragung abbrechen. Nur wenn grundlegende Informationen über diese Personengruppen vorliegen, können Unterschiede zu den kooperativen Studierenden analysiert werden.

Ein anderes Problem ist der erhöhte Item Non-Response in der Gruppe der online Befragten. Wird die weite Definition des Unit Non-Response gewählt, beträgt der IN in der schriftlichen Befragungsgruppe gerade einmal zwei Prozent, in der Onlinegruppe zwischen 15 % (selbstselektiert) und 20 % (randomisiert). Auch nach der Einschränkung auf Personen mit mindestens 50 % der Angaben liegt der IN in der Onlinegruppe noch zwischen fünf bzw. vier Prozent²⁰, zugleich wird damit aber auch die Anzahl verwertbarer Fälle weiter reduziert (von 1.433 auf 1.165 Fälle). Der erhöhte Anteil fehlender Antworten in den Gruppen der Online-Befragung ist ein Problem, für das es keine einfache Lösung gibt. Fehlende Angaben können aufgrund von Befragungsabbrüchen oder gezielter Verweigerung entstehen. Beides kann verschiedene Ursachen haben (Ermüdung, problematische Fragen bzw. problematisches Design der Fragen etc.). Besondere Aufmerksamkeit verdienen offene Fragen, die eine Zahleneingabe erfordern, weil sich gezeigt hat, dass der IN in der Onlinegruppe bei diesen Fragen besonders hoch ist.

Zwar existieren verschiedenste Imputationsverfahren für die Ergänzung fehlender Werte, die Voraussetzungen für deren Einsatz (fehlende Werte sind „missing at random“ oder „missing completely at random“) dürften aber in den allermeisten Fällen nicht zutreffen. Hier sind weitere Tests und Analysen notwendig um zu prüfen, inwiefern Ergänzungs- und Imputationsverfahren ggf. genutzt werden können. Andere Maßnahmen zur Vermeidung von Item Non-Response sind leichter umzusetzen und beinhalten einerseits die gezielte Überarbeitung des Fragebogens, um dessen Bearbeitung allgemein sowie die Entscheidungsfindung der Teilnehmer zu erleichtern. Andererseits können so genannte „Soft Checks“ programmiert werden: Gibt ein Befragter keine Antwort, erscheint ein Pop-Up Fenster das ihn/ sie auf die fehlende Antwort hinweist. Soft Checks können vom Befragten ignoriert werden, er/ sie wird nicht zu einer Antwort gezwungen. Allerdings können mehrere dieser Hinweise auch Abbrüche provozieren, weshalb Vor- und Nachteile vorab geprüft werden müssen.

Effekte der Erhebungsform und der Selbstselektion

Die Ergebnisse zum Mode-Effekte haben gezeigt, dass Unterschiede zwischen den schriftlich und online Befragten Studierenden bestehen: Studierende der Onlinegruppe sind älter, der Frauenanteil in der Gruppe ist geringer, Studierende der Onlinegruppe studieren häufiger an Fachhochschulen und befinden sich eher in Bachelorstudiengängen. In der selbstselektierten Gruppe zeigt sich außerdem ein Mode-Effekt bei den Fächergruppen: In der Onlinegruppe sind Studierende der Wirtschafts-, Natur- und Ingenieurwissenschaften stärker repräsentiert als in der Papiergruppe. Die Selbstselektionseffekte haben sich sowohl in der Papiergruppe als auch in der Gruppe der online Befragten als gering herausgestellt. Trotzdem bleibt die Selbstselektion bestimmter Personengruppen in die Online-Erhebung problematisch, weil nicht ausgeschlossen werden kann, dass noch weitere hier nicht berücksichtigte Faktoren eine Rolle bei der Wahl der Erhebungsform gespielt haben. Weil die Effekte aber insgesamt nur gering ausgeprägt sind und die Verteilung zentraler soziodemographischer Merkmale in der Gruppe der online Befragten auch stärker mit jener der Grundgesamtheit übereinstimmt, ist der Wechsel in der Erhebungsform von der Papier- zur Online-Befragung möglich.

Auch die Effekte, die bei der Analyse inhaltlicher Unterschiede zwischen den Designgruppen auftreten, sind nur schwach ausgeprägt. Die wenigen gefundenen Differenzen sind so gering, dass keine ernsthaften Bedenken gegen eine gemeinsame Analyse der Daten sprechen und dem Umstieg auf eine reine Online-Befragung entsprochen werden kann.

²⁰ Werden nur jene Personen berücksichtigt, die die Befragung bis zum Schluss bearbeitet haben, sinkt der Item Non-Response auf vier bzw. drei Prozent. Damit nähert sich der IN noch einmal jenem in der schriftlichen Befragungsgruppe an; jedoch würde eine solche Begrenzung die verwertbare Stichprobe auf gerade einmal 1.088 Fälle reduzieren (online Wahlgruppe und randomisierte Gruppe zusammen).

Die Möglichkeit, den Erhebungsmodus selbst zu wählen, hat auch weitere Effekte zur Folge. So entscheiden sich deutlich mehr Studierende, vor allem Studentinnen, für den schriftlichen Fragebogen wenn sie die Wahl haben. In der selbstselektierten Onlinegruppe sind die Antworten weniger durch Item Non-Response verzerrt und es brechen weniger Teilnehmer die Befragung vorzeitig ab. Die Wahlmöglichkeit bewirkt also einerseits vollständigere Angaben der Befragten und wirkt rücklaufsteigernd.

Effekte unterstützender Maßnahmen

Die Ergebnisse zu den unterstützenden Maßnahmen sind gemischt. Das Einblenden von Motivationsseiten führt insgesamt nicht zu einer Verringerung der Befragungsabbrüche; lediglich im letzten Befragungsabschnitt lassen sich weniger Abbrüche verzeichnen als in der Gruppe ohne Motivationsseiten. Es zeigt sich jedoch ein Effekt auf die Vollständigkeit der Antworten. Die Angaben in der Gruppe mit Motivationsseiten sind weniger durch Item Non-Response verzerrt.

Während die Vorankündigung der Befragung keinen Einfluss hat, scheint eine zweite Erinnerung per E-Mail einen positiven Effekt auf den Rücklauf zu haben. Der Befund spricht für Überlegungen zur gezielteren, wiederholten Ansprache der Befragten in künftigen Erhebungen.

Ausblick für den Studierendensurvey

Der Studierendensurvey kann somit zukünftig online über das Internet durchgeführt werden. Dies hat mehrere Vorteile, zum einen Entlastungen und zum anderen neue Perspektiven. Zu Entlastungen kommt es vorrangig durch den Wegfall der Papierfragebogen. Damit kann auf den aufwendigen Druck, die Versendung über den Postweg, das Porto sowie die postalischen Rückläufe verzichtet werden, was ökonomische und ökologische Vorteile verspricht. Zusätzlich entfallen die Versendungsarbeiten an den Hochschulen, die für die dortigen Mitarbeiter umständlich und zeitraubend waren. Mit dem Wegfall des Papierfragebogens wird auch die Datenerfassung überflüssig, welche die schriftlichen Antworten in die EDV übertragen musste, was ebenfalls Einsparungen verspricht.

Die Verwendung von Onlinezugängen bietet mehrere Vorteile, die bislang nicht in dieser Form genutzt werden konnten.

- Die Stichprobe an zu befragenden Studierenden (und Hochschulen) kann deutlich erhöht werden, da solch eine Erhöhung im Vergleich zu Papierbefragungen weder auffällige zusätzliche Kosten noch übermäßige zusätzliche Arbeitsbelastungen produziert.
- Durch diese Erhöhung der Stichprobe können bislang nicht berücksichtigte Spezialfälle bzw. spezifische Gruppen analysiert werden, seien es spezielle Hochschultypen oder bestimmte Studierendengruppen. So werden z.B. spezielle Auswertungen von „kleinen“ Einzelfächern (z.B. Philosophie), von Studierenden mit Kindern oder von Fernhochschulen möglich.
- Die Onlinezugänge liefern zusätzliche (Para-)Daten, die mit den Befragungsdaten kombiniert analysiert werden können. So kann z.B. eine genaue Zeitmessung wichtige und hilfreiche Informationen zur Durchführbarkeit des Instruments und der einzelnen Fragen sowie zur möglichen Länge der Befragung liefern.
- Die Onlinezugänge bieten die Möglichkeit zur zufälligen Verteilung der Stichproben in unterschiedliche Module, wie sie in der 13. Erhebung durchgeführt wurden. Damit lassen sich einerseits kurze Teilbefragungen durchführen und somit die große Bandbreite an Themen des Studierendensurveys aufteilen, andererseits lassen sich aber auch neue Fragen, Inhalte oder Gestaltungen austesten und mit bestehenden Fragen des Surveys kombinieren. Insofern ist die Ausweitung der Themenbereiche sogar einfacher als in Papierbefragungen. Vor allem

diese empirischen Möglichkeiten sind für die weitere Surveyforschung bzw. Studierendenbefragungen von besonderem Interesse.

- Des Weiteren lassen sich zusätzliche Maßnahmen mit den Online-Befragungen kombinieren, um spezifische Fragen auszutesten und empirisch zu prüfen. In der 13. Erhebung wurden so eine zweite Erinnerung sowie die Ankündigung der Erhebung in den sozialen Medien geprüft. Hier lassen sich vielfältige weitere Maßnahmen kreieren.
- Erinnerungen an die befragten Studierenden können gezielter als bislang auch mehrfach eingesetzt werden. Es werden beispielsweise nur Personen erneut zur Teilnahme aufgefordert, deren Zugangscode noch nicht verbraucht wurde. So kann es vermieden werden, Studierende zu erinnern, die bereits an der Umfrage teilgenommen haben.

Weitere Vorteile der Online-Befragung sind:

- Online-Befragungen bieten eine größere Flexibilität. Die komplexere Filterführung ermöglicht es, für den einzelnen Befragten unzutreffende Fragen zu überspringen und Module, die nur für spezielle Untergruppen relevant sind, der entsprechenden Personengruppe zuzuspielen.
- Durch die Möglichkeit der Randomisierung von Themenblöcken und/ oder der Fragereihenfolge in Online-Umfragen können bspw. Positionseffekte oder Reihenfolgeeffekte vermieden bzw. getestet werden.
- Die Online-Umfrage, sollte sie für alle gängigen mobilen Endgeräte passend umgesetzt werden, bietet den Studierenden eine größere räumlich-zeitliche Flexibilität. Kombiniert mit der Möglichkeit, sich erneut in die Umfrage einzuloggen und diese an der zuletzt bearbeiteten Stelle fortsetzen zu können, sind die Studierenden bei der Beantwortung völlig flexibel.

Allerdings bestehen neben den Vorteilen auch Herausforderungen, die für weitere Online-Erhebungen zum Tragen kommen:

- Überarbeitung der textlichen Fragen- und Itemformulierungen: In Online-Erhebungen sollten diese kurz und knapp sein, einzelne Fragen sind in ihrer jetzigen Form möglicherweise nicht mehr aktuell und auch das Frageformat muss in vielen Fällen angepasst werden. Daher gilt es zu prüfen, inwieweit textliche Veränderungen inhaltlich vergleichbar bleiben.
- Anpassung der Antwortskalen: Sollten Änderungen in den Antwortskalen anfallen, gilt zu prüfen, inwieweit diese in den Aussagen vergleichbar zu früheren Formulierungen bzw. Abfragen bleiben.
- Aufteilung langer Itembatterien: Matrixfragen sind in Online-Erhebungen problematisch. Da vor allem bei kleinen Bildschirmen lange Itemlisten nicht mehr problemlos abgebildet werden können, muss geprüft werden, ob Aufteilungen, Änderungen in der Reihenfolge oder Auslassung bestimmter Items Auswirkungen auf die Beantwortung der dargestellten Items haben.
- Neue Gestaltung der Fragen: Wird das bisherige Design deutlich verändert, gilt es zu prüfen, ob die Neudarstellung zu anderem Antwortverhalten führt.
- Dateneingabe und Datenplausibilisierung: Die unmittelbare Verfügbarkeit der Daten birgt nicht nur Vorteile; vielmehr muss bereits bei der Planung des Instruments und dessen Programmierung darauf geachtet werden, dass fehlerhafte Eingaben der Teilnehmer von Vornherein unterbunden werden, etwa durch die Einführung von Plausibilitätschecks.
- responsive design: Online-Befragungen werden zunehmend häufiger auf unterschiedlichen Eingabegeräten durchgeführt, was hohe Anforderungen an die Programmierung erfordert,

damit die Nutzer an ihre Geräte angepasste Designs erhalten, die eine überschaubare und akzeptable Bearbeitung ermöglichen.

Für die Zukunft des Studierendensurveys ergeben sich damit zwar Veränderungen, aber auch neue Möglichkeiten. Die Onlinenutzung fordert im Einzelnen zwar Notwendigkeiten der Kürzungen, im Gesamten bietet sie aber auch die Chancen auf Erweiterungen durch neue Items und die Möglichkeiten, alle bisher erhobenen Fragen und Items aus allen 13 Wellen modular einzubringen. Damit kann zukünftig gerade ein besonders breites Reservoir an Informationen zur Studiensituation und den studentischen Orientierungen entstehen. Vor allem auch im Hinblick auf konzeptionelle und theoretische Einpassungen bieten Online-Befragungen erweiterte und schnellere Möglichkeiten, bestehende und neue Items zu kreieren und zu prüfen. Ebenso lassen sich methodische Experimente leicht implementieren, die bspw. dazu beitragen, das Verständnis der Bearbeitung des Instruments zu verbessern.

Die weitreichenden Testungen, Änderungen und Neuerungen, die für den Studierendensurvey geplant sind sowie die beobachteten Selektionseffekte bedeuten dennoch einen Bruch in der Zeitreihe. Analysen mit früheren Erhebungen werden dadurch möglicherweise eingeschränkt; zumindest ist noch nicht klar, wie sich der Wechsel des Erhebungsmodus im Detail auf Untersuchungen, die die Zeitreihe nutzen, auswirken wird.

Literatur

- AAPOR. 2011. Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. Unter: http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf (Letzter Zugriff: 20.07.2017).
- Ainsaar, M., L. Lilleoja, K. Lumiste, A. Roots. 2013. ESS Mixed Mode Experiment Results in Estonia (CAWI and CAPI Mode Sequential Design). Report. Unter: http://www.yti.ut.ee/sites/default/files/ssi/ess_dace_mixed_mode_ee_report.pdf (Letzter Zugriff: 20.07.2017).
- Apenburg, E., P. Jurecka, R. Tausendfreund. 1977. Studium und Lehre aus der Sicht von Lehrenden und Lernenden. Universität des Saarlandes. Saarbrücker Studien zur Hochschulentwicklung 24. Saarbrücken.
- Arkes, H.R., P. Ayton. 1999. The Sunk Cost and Concorde Effects: Are Humans Less Rational Than Lower Animals. *Psychological Bulletin*, Vol. 125, No. 5, pp. 591-600.
- Ballerstedt, E., W. Glatzer. 1975. Soziologischer Almanach. Handbuch gesellschaftspolitischer Daten und Indikatoren für die Bundesrepublik Deutschland. Frankfurt/New York.
- Bandilla, W., L. Kaczmirek, M. Blohm, W. Neubarth. 2009. Coverage- und Nonresponse-Effekte bei Online-Bevölkerungsumfragen. In: Jakob N., H. Schoen, T. Zerback (Hrsg.): *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. VS Verlag für Sozialwissenschaften, Wiesbaden. S. 129-143.
- Bargel, T., G. Framhein. 1976. Zur Diskussion von Bildungszielen und zur Leistungsmessung im Hochschulbereich. In: Zapf, W. (Hrsg.): *Gesellschaftspolitische Zielsysteme. Soziale Indikatoren IV*. Frankfurt: Campus, 1976, S. 126-161.
- Bargel, T., G. Framhein, P. Kellermann, H. Peisert, J.-U. Sandberger. 1973. Zweckbestimmung des Studiums als Thema der Hochschulforschung. *Zeitschrift für Pädagogik*, Jg. 19, Heft 6, 1973, S. 943-966.
- Bartsch, S. 2012. Noncoverage und Nonresponse. In: ...würden Sie mir dazu Ihre E-Mail-Adresse verraten? Internetnutzung und Nonresponse beim Aufbau eines Online Access Panels. 1. Auflage Momos, Baden-Baden, pp. 36-63.
- Best, H., C. Wolf. 2012. Modellvergleich und Ergebnisinterpretation in Logit- und Probit-Regressionen. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64: 377-395.
- Blohm, M. & A. Koch. 2009. Ausschöpfungsquoten und Stichprobenqualität am Beispiel des ALLBUS 2008: Führt eine höhere Ausschöpfung zu anderen/besseren Umfrageergebnissen? Vortrag beim Workshop "Qualität von Großerhebungen". Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg, 3.-4. Sept. 2009.
- Bundesministerium für Bildung und Wissenschaft (BMBW) (1976): *Bildungspolitische Zwischenbilanz*. Bonn, o.J.
- Bundesdatenschutzgesetz in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), das zuletzt durch Artikel 7 des Gesetzes vom 30. Juni 2017 (BGBl. I S. 2097) geändert worden ist.

- Bosnjak, M., T.L. Tuten. 2001. Classifying response behaviors in Web-based surveys: An experiment. *Journal of Computer-Mediated Communication*, Vol. 6, No. 3.
- Bosnjak, M., I. Haas, M. Galesic, L. Kaczmirek, W. Bandilla, M.P. Couper. 2013. Sample Composition Discrepancies in Different Stages of a Probability-based Online Panel. *Field Methods*, Vol. 25, No. 4, pp. 339-360.
- Caldwell, B., M. Cooper, L. Guarino Reid, G. Vanderheiden. 2008. Web Content Accessibility Guidelines (WCAG) 2.0. W3C Recommendation. Unter: <https://www.w3.org/TR/2008/REC-WCAG20-20081211/> (Letzter Zugriff : 20. 07. 2017).
- Callegaro, M. 2010. Do you know which device your respondent has used to take your online survey? *Survey Practice*, Vol. 3, No. 6.
- Callegaro, M., K. Lozar-Manfreda, V. Vehovar. 2015. *Web Survey Methodology*. Sage, London.
- Cohen, J. 1988. *Statistical Power and analysis for the behavioral sciences* (2nd ed.), Hillsdale, N.J., Lawrence Erlbaum Associates, Inc.
- Conrad, F.G., M.P. Couper, R. Tourangeau, A. Peytchev. 2010. The impact of progress indicators on task completion. *Interacting with Computers*. Vol. 22, No.5, pp. 417-427.
- Crawford, S.D., M.P. Couper, M.J. Lamias. 2001. Web Surveys. Perception of Burden. *Social Science Computer Review*, Vol. 19, No. 2, pp. 146-162.
- De Bruijne, M., A. Wijnant. 2013. Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment With a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. *Social Science Computer Review*, Vol. 31, No. 4, pp. 482-504.
- De Leuw, E.D., W. de Heer. 2002. Trends in household survey Nonresponse : A longitudinal and international comparison. In: Groves, R.M., D.A. Dillman, J.L. Eltinge, R.J.A. Little (Hrsg.): *Survey Nonresponse*. Wiley, New York, pp. 41-54.
- Deutsche Gesellschaft für Soziologie. 2017. Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS). <http://www.soziologie.de/de/die-dgs/ethik/ethik-kodex.html> (Letzter Zugriff: 23.08.2017)
- Diekmann, A. 2006. *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. 15. Auflage, rohwoit, Reinbek bei Hamburg.
- Dillman, D.A. 1983. Mail and Other Self-Administered Questionnaires. In: Rossi, P.H., J.D. Wright und A. Anderson (Hrsg.), *Handbook of Survey Research*, New York, Academic Press: 359-378.
- Dillman, D.A. 2000. *Internet and Mail Surveys: The Tailored Design Method*. John Wiley: New York.
- Does, K.J. 1978. *Studenten und Politik in der Bildungskrise. Eine Untersuchung über Einstellungen von Studenten in der Bundesrepublik Deutschland zu Politik, Hochschule, Studium und Beruf*. Sozialwissenschaftliches Institut der Konrad-Adenauer-Stiftung. St. Augustin.

- El-Menouar, Y., J. Blasius. 2005. Abbrüche bei Online-Befragungen : Ergebnisse einer Befragung von Mediziner*innen. In: ZA-Information / Zentralarchiv für Empirische Sozialforschung, 56, pp. 70-92.
- Felderer, B., A. Birg, F. Kreuter. 2014. Paradata. In: N. Baur, J. Blasius (Hrsg.), Handbuch Methoden der empirischen Sozialforschung, Springer Fachmedien Wiesbaden: 357-365.
- FluidSurveys. 2014. Response Rate Statistics for Online Surveys -What Numbers Should You be Aiming For? Blogbeitrag vom 8. Oktober 2014. Unter: <http://fluidsurveys.com/university/response-rate-statistics-online-surveys-aiming/> (Letzter Zugriff: 20. 07.2017).
- Framhein G., W. Markiewicz, H. Peisert. 1974. Student and Graduates: Their Training at University and their Conception of Life. In: Niessen M., J. Peschar (Hrsg.): Comparative Research on Education. Budapest.
- Framhein G., T. Bargel, B. Dippelhofer-Stiem, H. Peisert, J.-U. Sandberger. 1981. Studium und Hochschulpolitik. Eine Untersuchung über Informationsverhalten, Studiensituation und gesellschaftliche Vorstellungen von Studenten. Bundesministerium für Bildung und Wissenschaft. Schriftenreihe Hochschule 39. München: Gersbach.
- Gabler, S. & S. Häder. 2015. Stichproben in der Theorie. GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_009
- Glatzer W., W. Zapf (Hrsg.). 1984. Lebensqualität in der Bundesrepublik. Objektive Lebensbedingungen und subjektives Wohlbefinden. Frankfurt: Campus.
- Groves R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, R. Tourangeau. 2004. Survey Methodology. New Jersey.
- Habermas J., L. v. Friedeburg, C. Oehler, F. Weltz. 1961. Student und Politik. Neuwied: Luchterhand.
- Häder, S. 2015. Stichproben in der Praxis. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_014
- Hanefeld, U. 1987. Das Sozio-ökonomische Panel - Grundlagen und Konzeption (Sozio-ökonomische Daten und Analysen für die Bundesrepublik Deutschland, Band 1), Frankfurt/M. - New York: Campus.
- Heerwegh, D. 2003. Explaining Response Latencies and Changing Answers Using Client Side Paradata From a Web Survey. Social Science Computer Review, Vol. 21, No. 3, pp. 360-373. DOI: 10.1177/0894439303253985
- Hochschul-Informationssystem (HIS). 1980. Studenten zwischen Hochschule und Arbeitsmarkt. HIS Reihe Hochschulplanung 32. München: Saur.
- Infratest Sozialforschung. 1974. Hochschulbarometer Befragung von Studierenden im WS 1973/74. Zusammenfassung wichtiger Ergebnisse. München, 1974.
- Infratest Sozialforschung. 1979. Ansprachemöglichkeiten für Studenten. Unveröffentlichter Untersuchungsbericht. München, 1979.
- Infratest Sozialforschung. 1982. Thema: Student und Gesellschaft. Bundesministerium für Bildung und Wissenschaft (Hrsg.). Bonn, 1982.

- Infratest Sozialforschung. 1982. Thema: Jugend und Technik. Bundesministerium für Bildung und Wissenschaft (Hrsg.). Bonn, 1982.
- Infratest Wirtschaftsforschung GmbH. 1980. Politischer Protest in der Bundesrepublik Deutschland. Beiträge zur sozioempirischen Untersuchung des Extremismus. Stuttgart: Kohlhammer, 1980.
- Institut für Demoskopie (IfD). 1967. Der deutsche Student. (Eine Umfrage im Auftrag des Spiegels) 1967.
- Institut für Demoskopie (IfD). 1968. Student und Politik Sommer 1967. Allensbach, 1968.
- Institut für Demoskopie (IfD). 1978. Dokumentation einer Studentenumfrage. Allensbach, 1978.
- Institut für Demoskopie (IfD). 1982. Kritik an der Technik und die Zukunft einer Industriena-tion. Demoskopische Nachwuchsanalyse für Ingenieurberufe. Ministerium für Wissen-schaft und Kunst Baden-Württemberg, Schriftreihe Bildung in neuer Sicht 47. Villingen: Neckar-Verlag, 1982.
- Ipsen, D., G. Portele. 1976. Organisation von Forschung und Lehre an westdeutschen Hoch-schulen. HIS, Reihe Hochschulplanung 22. München: Saur.
- Jenne M., M. Krüger, U. Müller-Plantenberg. 1969. Student im Studium. Stuttgart: Klett.
- Kaase, M. 1970. Demokratische Einstellungen in der Bundesrepublik Deutschland. In: Sozial-wissenschaftliches Jahrbuch für Politik, Band II. München: Olzog, 1970, S. 119-326.
- Kaase, M. 1971. Die politische Mobilisierung von Studenten in der BRD. In: Albeck K.R., L. Ro-senmayr (Hrsg.): Aufstand der Jugend? Neue Aspekte der Jugendsoziologie. München: Ju-venta, 1971, S. 155-177.
- Kaczmirek, L., W. Neubarth. 2007. Nicht-reaktive Datenerhebung: Teilnahmeverhalten bei Befragungen mit Paradata evaluiert. In: Welker, M. (Ed.), O. Wenzel (Ed.): *Online-For-schung 2007: Grundlagen und Fallstudien*. Köln : Halem Verl., 2007 (Neue Schriften zur Online-Forschung 1): pp. 293-311.
- Kahneman, D., A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econo-metrica*, Vol. 47, No. 2, pp. 263-292.
- Kaiser, et.al. 1981. Fachhochschulabsolventen beim Übergang vom Studium in den Beruf. Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit. Beitr. AB 24. Nürnberg.
- Kath, G. 1952. Das soziale Bild der Studentenschaft in Westdeutschland und Berlin. Verband Deutscher Studentenwerke (Hg.), Frankfurt.
- Kinder-Kurlanda, K., O. Watteler. 2015. Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungs-daten. GESIS-Papers. 2015/01.

- Knapp, F., M. Heidingsfelder. 1999. Drop-Out-Analyse: Wirkungen des Untersuchungsdesigns. In *Reips, U.* (Hrsg.). Aktuelle Online Forschung – Trends, Techniken, Ergebnisse. Online Press, Zürich. Unter: https://www.gor.de/archive/gor99/tband99/pdfs/i_p/knapp.pdf (Letzter Zugriff: 20.07. 2017).
- Koch, A., M. Blohm. 2015. Nonresponse Bias. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_004.
- Krause C., D. Lehnert, K.-J. Scherer. 1980. Zwischen Revolution und Resignation? Alternativkultur, politische Grundströmungen und Hochschulaktivitäten in der Studentenschaft. Bonn: Neue Gesellschaft.
- Kreutz, H. 1979. Die dreigeteilte Welt der Studenten. Teil I und II. Forschungsprojekt: Sozioökonomische Lage von Studenten in Niedersachsen. Universität Hannover.
- Lewin, K., M. Schacher. 1981. Studienberechtigte 78. Studien- und Berufswahl im Wandel? HIS, Reihe Hochschulplanung 35. Hannover: HIS GmbH.
- Liu, M., L. Wronski. 2017. Examining Completion Rates in Web Surveys via Over 25,000 Real-Worlds Surveys. *Social Science Computer Review*. Online First: Januar 2017.
- Lozar Manfreda, K., M. Bosnjak, J. Berzelak, I. Haas & V. Vehovar. 2008. Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, Vol. 50, No. 1, pp. 79-104.
- Lugtig, P., V. Toepel. 2016. The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey Effects on Survey Measurement Error. *Social Science Computer Review*, Vol. 34, No.1, pp. 78-94.
- Mavletova, A. 2013. Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review*, Vol. 31, No.6, pp. 725-743.
- Mayer, K.U., P. Schmidt (Hrsg.). 1984. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Beiträge zu methodischen Problemen des ALLBUS 1980. (ZUMA Monographien 5). Frankfurt: Campus.
- Menold, N., C. Züll. 2011. Standardisierte Erfassung von Verweigerungsgründen in face-to-face-Umfragen. In: *Methoden, Daten, Analysen (mda)*, Vol. 5, No. 1, pp. 91-108.
- Middendorff E., B. Apolinarski, J. Poskowsky, M. Kandulla, N. Netz. 2013. Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2012. 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das HIS-Institut für Hochschulforschung. Bundesministerium für Bildung und Forschung (Hrsg.), Berlin, Hannover.
- Mohorko, A., E. de Leeuw, J. Hox. 2013. Internet Coverage and Coverage Bias in Europe: Developments Across Countries and Over Time. *Journal of Official Statistics*. Vol. 29, No. 4, pp. 609-622.
- Mood, C. 2010. Logistic Regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*. Vol 26, No. 1, S. 67-82. DOI: 10.1093/esr/jcp006, available online at www.esr.oxfordjournals.org. Online publication 9. March 2009.

- Mossig, I. 2012. Stichproben, Stichprobenauswahlverfahren und Berechnung des minimal erforderlichen Stichprobenumfangs. Beiträge zur Wirtschaftsgeographie und Regionalentwicklung, Vol. 1-2012. Pp. 1-28.
- Multrus, F. 2016. Der Studierendensurvey. Konzept, Aufbau und theoretische Grundlagen. Hefte zur Bildungs- und Hochschulforschung, Band 87. Arbeitsgruppe Hochschulforschung. Universität Konstanz. Juni 2016.
- Neller, K. 2005. Kooperation und Verweigerung: eine Non-Response-Studie. In: ZUMA Nachrichten, Vol. 29, No. 57, pp. 9-36.
- Noll, H.-H. 2000. Subjektive Indikatoren. Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Zentrum für Umfragen, Methoden und Analysen (ZUMA), Abteilung Soziale Indikatoren. Mannheim.
- Noll, H.-H. 1989. Indikatoren des subjektiven Wohlbefindens : Instrumente für die gesellschaftliche Dauerbeobachtung und Sozialberichterstattung?. In: ZUMA Nachrichten 13, No. 24, pp. 26-41.
- Oehler, C. 1974. Student und Studienberatung. Frankfurt: Aspekte.
- Oehler, C., L. Birk, F. Blahusch, F. Kazemzadeh, D. Kraft-Krumm. 1976. Studienplanung und Organisation der Lehre. Ergebnisse einer empirischen Untersuchung in den Hochschulregionen Frankfurt und Darmstadt. HIS, Reihe Hochschulplanung 25. München: Saur.
- Peisert H., T. Bargel, G. Framhein. 1981. Untersuchungen zur Entwicklung der Studiensituation. (Konzeptentwurf). (Universität Konstanz, Zentrum I Bildungsforschung/SFB 23). Konstanz.
- Peisert H., T. Bargel, G. Framhein. 1988. Studiensituation und studentische Orientierungen an Universitäten und Fachhochschulen. 2. Erhebung zur Studiensituation im WS 1984/85. Bundesminister für Bildung und Wissenschaft (Hrsg.). Schriftenreihe Studien zu Bildung und Wissenschaft 59. Bonn.
- Peytchev, A. 2009. Survey Breakoff. Public Opinion Quarterly, Vol. 73, No. 1, pp. 74-97.
- Pflüger, A., H. Dobel. 2014. Datenschutz in der Online-Forschung. In: M. Welter, M. Taddicken, J.-H. Schmidt, N. Jakob (Hrsg.). Handbuch Online-Forschung. Sozialwissenschaftliche Datengewinnung und –auswertung in digitalen Netzen. Neue Schriften zur Online-Forschung, Band 12, Halem Verlag, Köln.
- Pötschke, M. 2009. Potentiale von Online-Befragungen: Erfahrungen aus der Hochschulforschung. In N. Jakob, H. Schoen & T. Zerback (Hrsg.), Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung (S. 75–89). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Porst, P. 2014. Fragebogen. Ein Arbeitsbuch. 4. Auflage, Springer VS, Wiesbaden.
- Poskowsky J., T. Weiß, M. Kandulla. 2014. Methodentest im Rahmen der 20. Sozialerhebung. Papierbasierte und webbasierte Befragung im Vergleich. DZHW, Oktober 2014.
- Prensky, M. 2001. Digital Natives, Digital Immigrants. On the Horizon, MCB University Press, Vol. 9 No. 5.

- Ramm, M. 2014. Response, Stichprobe und Repräsentativität . Zwei Dokumentationen zum Deutschen Studierendensurvey (DSS). Hefte zur Bildungs- und Hochschulforschung, Band 72. Arbeitsgruppe Hochschulforschung. Universität Konstanz.
- Rat für Sozial- und Wirtschaftsdaten. 2017. Handreichung Datenschutz. Unter: https://www.ratswd.de/dl/RatSWD_Output5_HandreichungDatenschutz.pdf (Letzter Zugriff: 20. 07. 2017).
- Reissert R., L. Birk. 1982. Studienverlauf, Studienfinanzierung und Berufseintritt von Hochschulabsolventen und Studienabbrechern des Studienjahres 1979. HIS, Reihe Hochschulplanung 41. Hannover: HIS GmbH.
- Sax, L.J., S.K. Gilmartin, A.N. Bryant. 2003. Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, Vol. 44, No. 4, pp. 409-432.
- Sauro, J. 2011. 10 things to know about completion rates. Blogbeitrag vom 6. September 2011. Unter: <https://measuringu.com/completion-rates/> (Letzter Zugriff: 20. 07. 2017).
- Saterdag H., H. Stegmann. 1980. Jugendliche beim Übergang vom Bildungs- in das Beschäftigungssystem. Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, Beitr. AB 41, Nürnberg.
- Schnell, R., P.B. Hill, E. Esser. 2011. Methoden der empirischen Sozialforschung. Oldenburg. München.
- Schnitzer K., W. Isserstedt, M. Leszczensky. 1986. Das soziale Bild der Studentenschaft in der Bundesrepublik Deutschland. 11. Sozialerhebung des Deutschen Studentenwerkes. (BMBW, Studien zu Bildung und Wissenschaft 42). Bad Honnef: Block.
- Shih, T.-H., X. Fan. 2008. Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods*, Vol. 20, No. 3, pp. 249-271.
- Statistisches Bundesamt. 2016. Studierende an Hochschulen – Wintersemester 2015/16. Fachserie 11, Reihe 4.1. Wiesbaden.
- Statistisches Bundesamt. 2016. Ausstattung privater Haushalte mit ausgewählten Gebrauchsgütern. Fachserie 15, Reihe 2. Wiesbaden.
- Struminskaya, B., K. Weyandt, M. Bosnjak. 2015. The Effects of Questionnaire Completion Using Mobile Devices on Data Quality. Evidence from a Probability-based General Population Panel. *methods, data, analyses*, Vol. 9, No.2, pp. 261-292.
- Tessaring M., H. Werner. 1981. Beschäftigung und Arbeitsmarkt für Hochschulabsolventen in den Ländern der Europäischen Gemeinschaft. Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, Beitr. AB 46. Nürnberg.
- Theobald, A. 2000. Das World Wide Web als Befragungsinstrument. Wiesbaden: Deutscher Universitäts-Verlag.
- Vehovar, V., G. Čehovin. 2014: Questionnaire length and breakoffs in web surveys: A meta study. Presented at the 7th Internet Survey Methodology Workshop. Bolzano, Italien.
- Wissenschaftsrat. 2002. Eckdaten und Kennzahlen zur Lage der Hochschulen von 1980 bis 2000. Drs. 5125-02.

Yan, T., F.G. Conrad, R. Tourangeau, M.P. Couper. 2010. Should I Stay or Should I go: The Effects of Progress Feedback, Promised Task Duration, and Length of Questionnaire on Completing Web Surveys. *International Journal of Public Opinion Research*. Vol. 23, No. 2, pp. 131-147.

Zapf, W. (Hrsg.). 1976. Gesellschaftspolitische Zielsysteme. Soziale Indikatoren IV. Frankfurt.

Quellen/ Gesetzestexte:

§3 Abs. 1 BDSG: „Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener).“

§3a BDSG: „Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten und die Auswahl und Gestaltung von Datenverarbeitungssystemen sind an dem Ziel auszurichten, so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen.“

§3 Abs.9 BDSG: „Besondere Arten personenbezogener Daten sind Angaben über die rassische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheit oder Sexualleben.“

§4d BDSG: „[...]es sei denn, dass eine gesetzliche Verpflichtung oder eine Einwilligung des Betroffenen vorliegt oder die Erhebung, Verarbeitung oder Nutzung für die Begründung, Durchführung oder Beendigung eines rechtsgeschäftlichen oder rechtsgeschäftsähnlichen Schuldverhältnisses mit dem Betroffenen erforderlich ist.“

§9 BDSG: „Öffentliche und nicht-öffentliche Stellen, die selbst oder im Auftrag personenbezogene Daten erheben, verarbeiten oder nutzen, haben die technischen und organisatorischen Maßnahmen zu treffen, die erforderlich sind, um die Ausführung der Vorschriften dieses Gesetzes, insbesondere die in der Anlage zu diesem Gesetz genannten Anforderungen, zu gewährleisten. Erforderlich sind Maßnahmen nur, wenn ihr Aufwand in einem angemessenen Verhältnis zu dem angestrebten Schutzzweck steht.“

§3 Abs. 6 BDSG: „Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.“

Internetseiten:

Barrierefreiheit im Internet. Webseite: <http://www.einfach-barrierefrei.net/start/willkommen.html> (Letzter Zugriff: 20. 07. 2017).

Anhang

1. Anhang zu Kapitel 3
 - a. Abbildung A1: Beispiel für die parallele Darstellung zweier Fragen im Papierfragebogen – Frage 16
 - b. Abbildung A2: Beispiel für die Darstellung zweier Fragen im Onlinefragebogen untereinander – Frage 16
 - c. Abbildung A3: Item Non-Response in Frage 19 – Möglichkeiten der Weiterqualifizierung getrennt nach Papier- und Onlinegruppe
2. Anhang zu Kapitel 4
 - a. Tabelle A1: Tage bis zum Abschluss der Befragung
3. Anhang zu Kapitel 5
 - a. Tabelle A2: Selektionseffekt bei Papiergruppen
 - b. Tabelle A3: Selektionseffekt bei Onlinegruppe
 - c. Tabelle A4: Mode-Effekt zwischen Papier- und Onlinegruppen
 - d. Tabelle A5: Mode-Effekt zwischen Papier- und Onlinegruppe mit Selektionseffekt

Anhänge zu Kapitel 3 – Unterschiede in der Darstellung zwischen Papier- und Onlinefragebogen

Abbildung A1: Beispiel für die parallele Darstellung zweier Fragen im Papierfragebogen – Frage 16

<p>16. Inwieweit scheinen Ihnen die folgenden Dinge für Ihre eigene <i>persönliche Entwicklung</i> nützlich? Und inwieweit scheinen sie Ihnen andererseits nützlich, um Ihre <i>Berufsaussichten</i> zu verbessern?</p> <p>1. an einem Forschungspraktikum/-projekt teilzunehmen.....</p> <p>2. zeitweise im Ausland zu studieren</p> <p>3. während des Studiums Arbeitserfahrungen außerhalb der Hochschule zu gewinnen.....</p> <p>4. sich hochschulpolitisch zu engagieren</p> <p>5. schnell und zielstrebig das Studium abzuschließen.....</p> <p>6. eine möglichst gute Examensnote zu erreichen.....</p> <p>7. nach dem Bachelor ein Masterstudium anzuschließen.....</p>	<p>a) eher nachteilig b) wenig nützlich</p>	<p>c) nützlich d) sehr nützlich</p>	<p>e) kann ich nicht beurteilen</p>																																																																														
	<p>Persönliche Entwicklung:</p> <p>a) b) c) d) e)</p> <table border="1"> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> </table>																																									<p>Berufsaussichten:</p> <p>a) b) c) d) e)</p> <table border="1"> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td></tr> </table>																																							

Abbildung A2: Beispiel für die Darstellung zweier Fragen im Onlinefragebogen untereinander – Frage 16

Arbeitsgruppe
Hochschulforschung

Logout

Fachstudium und Hochschulzugang

Inwieweit scheinen Ihnen die folgenden Dinge für Ihre eigene *persönliche Entwicklung* nützlich?

	<i>Persönliche Entwicklung:</i>				
	eher nachteilig	wenig nützlich	nützlich	sehr nützlich	kann ich nicht beurteilen
an einem Forschungspraktikum/-projekt teilzunehmen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
zeitweise im Ausland zu studieren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
während des Studiums Arbeitserfahrungen außerhalb der Hochschule zu gewinnen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sich hochschulpolitisch zu engagieren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
schnell und zielstrebig das Studium abzuschließen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
eine möglichst gute Examensnote zu erreichen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
nach dem Bachelor ein Masterstudium anzuschließen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Inwieweit scheinen sie Ihnen andererseits nützlich, um Ihre *Berufsaussichten* zu verbessern?

	<i>Berufsaussichten:</i>				
	eher nachteilig	wenig nützlich	nützlich	sehr nützlich	kann ich nicht beurteilen
an einem Forschungspraktikum/-projekt teilzunehmen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
zeitweise im Ausland zu studieren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
während des Studiums Arbeitserfahrungen außerhalb der Hochschule zu gewinnen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sich hochschulpolitisch zu engagieren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
schnell und zielstrebig das Studium abzuschließen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
eine möglichst gute Examensnote zu erreichen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
nach dem Bachelor ein Masterstudium anzuschließen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

powered by

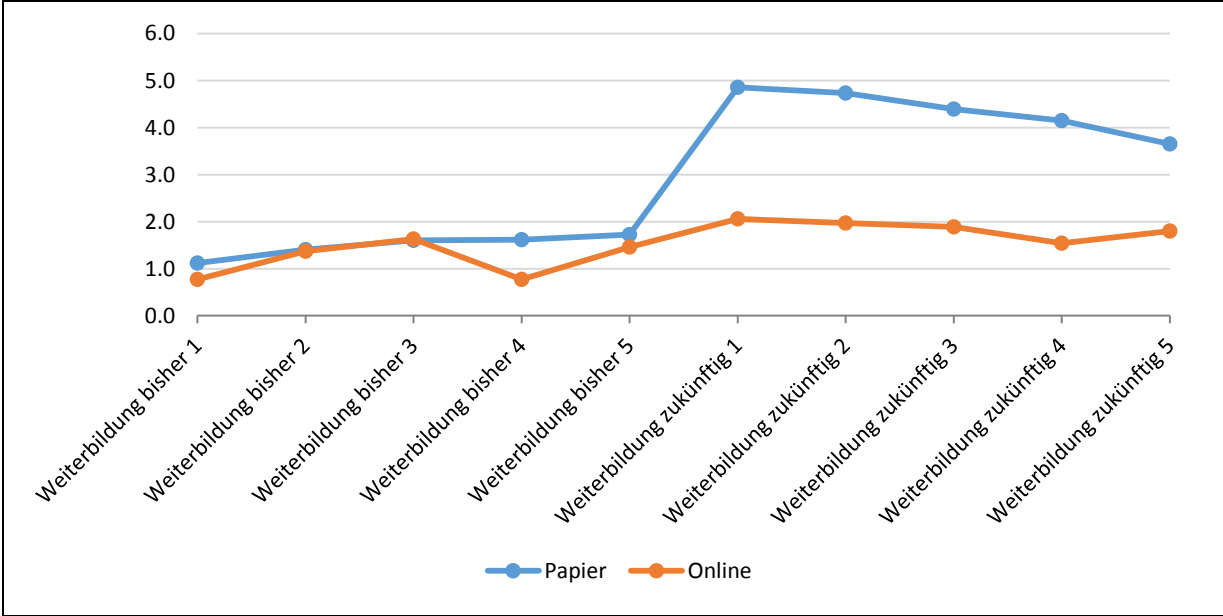
Zurück

Weiter

Abbildung A3

Item Non-Response in Frage 19 – Möglichkeiten der Weiterqualifizierung getrennt nach Papier- und Onlinegruppe (WS 2015/16)

(Angaben in Prozent, Onlinegruppe: nur Personen mit 50 % der Angaben)



Quelle: Studierendensurvey WS 2015/16, AG Hochschulforschung, Universität Konstanz

Anhang zu Kapitel 4 – Tage bis zum Abschluss der Befragung

Tabelle A1

Tage bis zum Abschluss – nur Personen, die die Befragung erfolgreich abgeschlossen haben (WS 2015/16)
(Häufigkeiten und Prozente in Klammern)

Tage bis Abschluss	Wahlgruppe online	Online mit Motivation	Online ohne Motivation
0	193 (87)	360 (85)	373 (84)
1	16 (7,2)	26 (6)	24 (5)
2	2 (0,9)	6 (1,4)	7 (1,6)
3	2 (0,9)	4 (0,9)	7 (1,6)
4	2 (0,9)	5 (1,2)	3 (0,7)
5	2 (0,9)	1 (0,2)	3 (0,7)
6	0	2 (0,5)	2 (0,5)
7	0	2 (0,5)	2 (0,5)
8	0	2 (0,5)	2 (0,5)
9	0	0	3 (0,7)
10	0	0	1 (0,2)
11	0	1 (0,2)	1 (0,2)
13	2 (0,9)	1 (0,2)	1 (0,2)
15	0	1 (0,2)	1 (0,2)
17	0	1 (0,2)	1 (0,2)
18	0	1 (0,2)	1 (0,2)
20	0	1 (0,2)	0
21	0	1 (0,2)	1 (0,2)
22	0	3 (0,7)	0
23	0	1 (0,2)	1 (0,2)
26	0	1 (0,2)	1 (0,2)
28	1 (0,5)	0	0
29	0	1 (0,2)	1 (0,2)
30	1 (0,5)	0	0
31	0	0	1 (0,2)
32	0	1 (0,2)	0
33	0	0	1 (0,2)
35	0	0	1 (0,2)
37	0	0	1 (0,2)
41	0	0	1 (0,2)
42	0	0	1 (0,2)
46	0	2 (0,5)	0
66	0	1 (0,2)	0
Total	221	425	442

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Anhang zu Kapitel 5 – Ergebnisse der logistischen Regression (in STATA)

Tabelle A2

Selektionseffekt bei Papiergruppen (WS 2015/16)

(Odds Ration und marginal effects (dy/dx))

Logistic regression		Number of obs	=	4,550		
		LR chi2 (17)	=	29.90		
		Prob > chi2	=	0.0271		
Log likelihood = -1640.563		Pseudo R2	=	0.0090		
papier	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	1.006614	.0120647	0.55	0.582	.9832435	1.030541
sex	1.079112	.1113864	0.74	0.461	.8814657	1.321077
land	.7860299	.0780919	-2.42	0.015	.6469526	.9550052
kid	.9097271	.2301357	-0.37	0.708	.5540907	1.493625
bild	.8507079	.0798756	-1.72	0.085	.7077151	1.022592
school	.8596444	.1152822	-1.13	0.259	.6609504	1.11807
teilzeit	.9923585	.1156344	-0.07	0.948	.7897371	1.246966
mar	1.259716	.2869524	1.01	0.311	.806077	1.968652
zweitst	.8978747	.0994257	-0.97	0.331	.7227005	1.115509
fächer						
1	1.528518	.2651466	2.45	0.014	1.087967	2.147462
2	1.367812	.2145676	2.00	0.046	1.005771	1.860174
3	.9082228	.1496936	-0.58	0.559	.6575019	1.254549
4	1.045492	.2250079	0.21	0.836	.6856893	1.594094
5	1.236477	.2124866	1.24	0.217	.882894	1.731665
zent						
1	.8444503	.2333389	-0.61	0.541	.4913259	1.451371
2	.9597213	.0945041	-0.42	0.676	.7912742	1.164027
comp	1.020445	.0973972	0.21	0.832	.8463423	1.230363
_cons	.1391293	.0461938	-5.94	0.000	.0725775	.2667076
	Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z		
age	.0006839	.0012433	0.55	0.582	-.001753	.0031207
sex	.007898	.0107081	0.74	0.461	-.0130896	.0288855
land	-.0249744	.0103144	-2.42	0.015	-.0451901	-.0047586
kid	-.0098141	.0262416	-0.37	0.708	-.0612467	.0416185
bild	-.0167719	.0097441	-1.72	0.085	-.03587	.0023261
school	-.0156879	.0139138	-1.13	0.260	-.0429585	.0115826
teilzeit	-.0007957	.0120873	-0.07	0.948	-.0244863	.0228949
mar	.0239501	.0236315	1.01	0.311	-.0223667	.070267
zweitst	-.0111744	.0114881	-0.97	0.331	-.0336907	.0113419
fächer						
1	.0458332	.018813	2.44	0.015	.0089605	.0827059
2	.0324116	.0159394	2.03	0.042	.0011709	.0636522
3	-.0084795	.0145596	-0.58	0.560	-.0370157	.0200568
4	.004143	.0201565	0.21	0.837	-.0353629	.043649
5	.0211172	.0171224	1.23	0.217	-.0124421	.0546764
zent						
1	-.016713	.0257626	-0.65	0.517	-.0672069	.0337808
2	-.0042647	.010177	-0.42	0.675	-.0242113	.0156819
comp	.0020994	.0099008	0.21	0.832	-.0173057	.0215046

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Tabelle A3

Selektionseffekt bei Onlinegruppe (WS 2015/16)

(Odds Ration und marginal effects (dy/dx))

Logistic regression		Number of obs		=	840	
		LR chi2(17)		=	19.95	
		Prob > chi2		=	0.2769	
Log likelihood = -401.91121		Pseudo R2		=	0.0242	

online	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.9694492	.0241505	-1.25	0.213	.9232521	1.017958
sex	.8595632	.1658035	-0.78	0.433	.5889619	1.254494
land	1.026582	.2062962	0.13	0.896	.6923742	1.522111
kid	1.26842	.7391396	0.41	0.683	.4048057	3.97447
bild	.8753993	.1604422	-0.73	0.468	.6112219	1.253757
school	1.033138	.2619776	0.13	0.898	.6285141	1.698249
teilzeit	1.021298	.2383611	0.09	0.928	.6463837	1.61367
mar	1.482244	.8375592	0.70	0.486	.4897094	4.486431
zweitst	.982825	.2107713	-0.08	0.936	.6455534	1.496305
fächer						
1	.5710501	.2147585	-1.49	0.136	.2732482	1.193414
2	.6167342	.198434	-1.50	0.133	.3282647	1.158702
3	1.318025	.3789252	0.96	0.337	.750253	2.315472
4	.912985	.3529191	-0.24	0.814	.4279837	1.947601
5	1.148658	.3395289	0.47	0.639	.6435556	2.050195
zent						
1	1.17071	.5055484	0.36	0.715	.502197	2.72913
2	.8592873	.1659893	-0.79	0.432	.588452	1.254775
comp	1.369041	.2472014	1.74	0.082	.9609898	1.950356
_cons	.5428352	.3733677	-0.89	0.374	.1409947	2.089937

	Delta-method		z	P> z	[95% Conf. Interval]	
	dy/dx	Std. Err.				
age	-.004716	.0037803	-1.25	0.212	-.0121253	.0026933
sex	-.0230017	.0292985	-0.79	0.432	-.0804258	.0344224
land	.0039876	.0305435	0.13	0.896	-.0558765	.0638517
kid	.0361403	.0885539	0.41	0.683	-.1374222	.2097028
bild	-.0202269	.0278419	-0.73	0.468	-.074796	.0343423
school	.0049552	.0385417	0.13	0.898	-.0705851	.0804954
teilzeit	.0032032	.0354738	0.09	0.928	-.0663241	.0727306
mar	.0598191	.0858435	0.70	0.486	-.1084311	.2280692
zweitst	-.0026332	.0325957	-0.08	0.936	-.0665197	.0612533
fächer						
1	-.075983	.0493172	-1.54	0.123	-.1726429	.0206768
2	-.06718	.044852	-1.50	0.134	-.1550884	.0207284
3	.0480651	.0498934	0.96	0.335	-.0497242	.1458543
4	-.0142853	.060161	-0.24	0.812	-.1321987	.103628
5	.0232379	.0493591	0.47	0.638	-.0735041	.1199799
zent						
1	.0257658	.0731896	0.35	0.725	-.1176831	.1692147
2	-.0226983	.0286002	-0.79	0.427	-.0787536	.033357
comp	.0477435	.0273515	1.75	0.081	-.0058645	.1013514

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

Tabelle A5

Mode-Effekt zwischen Papier- und Onlinegruppe mit Selektionseffekt (WS 2015/16)

(Odds Ration und marginal effects (dy/dx))

Logistic regression		Number of obs	=	5,390		
		LR chi2(18)	=	97.54		
		Prob > chi2	=	0.0000		
Log likelihood = -2283.56		Pseudo R2	=	0.0209		

group	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Sel	1.784315	.1780089	5.80	0.000	1.467415	2.169652
age	1.044255	.0100986	4.48	0.000	1.024649	1.064237
sex	.7628051	.062391	-3.31	0.001	.6498191	.8954362
land	1.028843	.0871575	0.34	0.737	.8714457	1.214669
kid	.7411317	.171478	-1.29	0.195	.4709233	1.166382
bild	1.059185	.0823553	0.74	0.460	.9094692	1.233547
school	.8416178	.0920823	-1.58	0.115	.6791788	1.042907
teilzeit	.8654436	.0857178	-1.46	0.145	.7127407	1.050863
mar	.5865834	.1335359	-2.34	0.019	.3754512	.9164444
zweitst	.7863086	.0719974	-2.63	0.009	.6571338	.9408757
fächer						
1	.9808196	.1454483	-0.13	0.896	.733436	1.311644
2	1.052049	.1361749	0.39	0.695	.8163158	1.355856
3	.8925989	.1150798	-0.88	0.378	.6932881	1.149209
4	1.181045	.1961277	1.00	0.316	.8529298	1.635382
5	1.462126	.1914771	2.90	0.004	1.131132	1.889977
zent						
1	1.171382	.2355287	0.79	0.431	.7898564	1.737196
2	1.045551	.0846648	0.55	0.582	.8921094	1.225384
comp	1.133946	.0880424	1.62	0.105	.9738739	1.320328
_cons	.0750388	.0204129	-9.52	0.000	.0440283	.1278908

	Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
Sel	.0747097	.0128227	5.83	0.000	.0495776 .0998417
age	.0055873	.0012451	4.49	0.000	.0031469 .0080276
sex	-.0349338	.010545	-3.31	0.001	-.0556015 -.014266
land	.0036688	.0109301	0.34	0.737	-.0177539 .0250915
kid	-.0386528	.0298493	-1.29	0.195	-.0971563 .0198507
bild	.0074189	.0100317	0.74	0.460	-.0122428 .0270806
school	-.0222476	.0141152	-1.58	0.115	-.0499128 .0054176
teilzeit	-.0186457	.0127775	-1.46	0.144	-.0436891 .0063976
mar	-.0688269	.0293646	-2.34	0.019	-.1263805 -.0112734
zweitst	-.0310183	.0118077	-2.63	0.009	-.0541609 -.0078756
fächer					
1	-.002368	.0181194	-0.13	0.896	-.0378814 .0331453
2	.0063547	.0161894	0.39	0.695	-.0253759 .0380853
3	-.0134472	.0152828	-0.88	0.379	-.0434009 .0165065
4	.0216699	.0220178	0.98	0.325	-.0214842 .064824
5	.0530561	.0182543	2.91	0.004	.0172784 .0888337
zent					
1	.0211928	.0281565	0.75	0.452	-.0339929 .0763786
2	.0057483	.0104857	0.55	0.584	-.0148032 .0262999
comp	.0162188	.0100159	1.62	0.105	-.003412 .0358496

Quelle: Studierendensurvey 2015/16, AG Hochschulforschung, Universität Konstanz

