

Plagiate in studentischen Arbeiten

Eine empirische Untersuchung unter Anwendung des Triangular Modells

Plagiarism in Student Papers

An Empirical Study Using the Triangular Model

Julia Jerke und Ivar Krumpal

Zusammenfassung

Sensitive Fragen in sozialwissenschaftlichen Erhebungen stellen Forscher immer wieder vor Probleme. Um beispielsweise niedrige Response-Raten und verzerrte Ergebnisse zu überwinden, schlagen Yu, Tian und Tang (2008) zwei vielversprechende neue Befragungstechniken vor: das Crosswise Modell (CM) und das Triangular Modell (TM). Diese beiden Modelle können als Alternative zur bekannten Randomized Response Technik (RRT) eingesetzt werden und überwinden die meisten Nachteile der RRT. Während das CM bereits empirisch erforscht wurde (Jann et al. 2012) liegen für das TM bisher noch keine empirischen Befunde vor. Wir stellen daher Ergebnisse aus einer an der Universität Leipzig durchgeführten Studie vor, in der das TM angewandt und der direkten Befragung gegenübergestellt wurde. Hierfür befragten wir Studenten, ob sie schon einmal plagiiert haben. Die Ergebnisse zeigen, dass sich in der TM-Kondition mehr

Abstract

Sensitive questions in social surveys often yield problems such as item nonresponse and response bias. Yu, Tian and Tang (2008) proposed two new models to overcome these problems: the crosswise model (CM) and the triangular model (TM). Both models can be seen as an alternative to the well-known randomized response technique (RRT) and are meant to overcome most of its drawbacks. While the CM has already been applied empirically (Jann et al. 2012), no evidence for the TM exists. We therefore present first results of an empirical survey conducted at the University of Leipzig comparing direct questioning and the TM. Students were asked to self-report whether they ever have plagiarised in one of their term papers. The results show that under the TM condition a higher percentage of students admit plagiarism than under direct questioning. This and the fact that the TM is very easy to implement show that the TM is



Studenten zu einem Plagiat bekennen als in der direkten Befragung. Dieses Resultat und die einfache Implementierung machen das TM zu einer vielversprechenden Alternative zur direkten Befragung, sowie zu anderen indirekten Befragungsmethoden, wie beispielsweise der RRT.

a promising alternative to direct as well as indirect questioning such as the RRT.

1 Einleitung

In sozialwissenschaftlichen oder medizinischen Befragungen werden häufig sensitive bzw. heikle Fragen, – etwa Fragen zu kriminellen Aktivitäten, zu sexuellen Präferenzen, zum illegalen Drogenkonsum, zum Einkommen oder zur Wahlbeteiligung, – gestellt. Das Stellen solcher Fragen verletzt zunächst die Norm der Privatheit. So werden bei Fragen nach der sexuellen Orientierung private Persönlichkeitsbereiche berührt, die ein Befragter nicht unbedingt mit anderen Personen teilen möchte. Zudem können heikle Fragen aber auch illegale oder strafrechtlich relevante Verhaltensweisen (z.B. Steuerhinterziehung, Plagiiere) des Befragten offenlegen. Bei einer ehrliche Beantwortung und Offenlegung von illegalen Verhaltensweisen gegenüber Dritten müsste der Befragte die Möglichkeit von unangenehmen Konsequenzen befürchten. Neben der Dimension der Privatheit und der strafrechtlichen Relevanz, tangieren heikle Fragen schließlich eine dritte Dimension, nämlich die der sozialen Erwünschtheit (Tourangeau/Yan 2007: 860)¹. Beispielsweise sind Fragen nach Einstellungen zu gesellschaftlichen Minderheiten wie Ausländern und Juden stark normativ konnotiert und führen häufig zu sozial erwünschtem Antwortverhalten bzw. Antwortverweigerungen (Krumpal 2012, 2013). Dabei ist das Problem der sozialen Erwünschtheit umso gravierender, je weniger anonym die Befragungssituation ist. Zusammenfassend lassen sich somit Fragen als ‚heikel‘ oder ‚sensitiv‘ definieren, wenn sie persönliche Details oder sozial unerwünschte Merkmale (Ver-

1 Auf Grundlage der bestehenden Forschungsliteratur können ‚Impression Management‘ sowie ‚Self-Deception‘ als mögliche Mechanismen angenommen werden, die sozial erwünschtes Antwortverhalten hervorbringen (Paulhus 1984; Paulhus/John 1998; Krumpal/Näher 2012). Während der Impression Management Mechanismus annimmt, dass der Befragte bewusst falsch antwortet, um vor einer tatsächlichen oder eingebildeten Öffentlichkeit ein vorteilhaftes Bild abzugeben, nimmt der Self-Deception Mechanismus an, dass der Befragte unbewusst falsch antwortet, indem er oder sie negative Aspekte des Selbst verdrängt und selbst an seine (falschen) positiven Selbstdarstellungen glaubt. In der vorliegenden Arbeit nehmen wir an, dass dem sozial erwünschten Antwortverhalten beide psychologischen Mechanismen zugrunde liegen.

haltensweisen oder Einstellungen), die gegen soziale Normen verstoßen und somit sanktioniert werden, nicht ausreichend anonym erfragen.

Wenn Befragte die Beantwortung unangenehmer Fragen verweigern oder keine wahrheitsgemäßen Antworten geben, dann kann das für die Validität der erhobenen Daten negative Konsequenzen haben. So leiden Umfragestudien zu sozial erwünschten Verhaltensweisen oder Einstellungen (aufgrund von ‚Overreporting‘) häufig unter systematischer Überschätzung und Studien zu sozial unerwünschten Verhaltensweisen oder Einstellungen (aufgrund von ‚Underreporting‘) unter systematischer Unterschätzung selbiger: *„Respondents consistently underreport some behaviors (the socially undesirable ones) and consistently overreport others (the desirable ones)”* (Tourangeau/Yan 2007: 863). Zudem können Antwortverweigerungen ähnliche Verzerrungen hervorrufen, nämlich dann wenn Befragte, die die Antwort verweigern, sich bezüglich des sensitiven Merkmals systematisch von denen unterscheiden, die an der Befragung teilnehmen.

Da direkte Befragungen zu sensitiven Themen häufig unter verzerrten Ergebnissen leiden, wurden in der Survey Forschung verschiedene Verfahren zur Reduzierung dieser Verzerrungen, wie etwa die Randomized Response Technik (RRT; Warner 1965), vorgeschlagen. Hierbei sollen über einen Zufallsmechanismus die Antworten anonymisiert und ein ehrlicheres Antwortverhalten erzielt werden. In einer Vielzahl von experimentellen Studien entlockte die RRT den Befragten mehr sozial unerwünschte Antworten als die konventionelle direkte Befragung (Lensvelt-Mulders et al. 2005). Allerdings wurde auch immer wieder die hohe Komplexität der RRT, wie etwa die Notwendigkeit eines Randomisierungsinstruments (z.B. Münzen, Würfel, etc.) und die direkte Konfrontation mit der heiklen Frage, moniert. Yu et al. (2008) haben deshalb zwei alternative Methoden für die Erhebung sensitiver Merkmale vorgeschlagen: das Crosswise Modell (CM) bzw. das Triangular Modell (TM). Diese neuen Befragungsverfahren sollen die Schwächen der RRT kompensieren und validere Resultate bei der Messung sensitiver Merkmale erzielen. Während das Crosswise Modell bereits empirisch evaluiert wurde und, im Vergleich zur direkten Befragung, gute Resultate erzielt hat (Coutts 2011; Jann et al. 2012), liegt zur Effektivität des TM bislang keine empirische Studie vor. Der nachfolgende Beitrag schließt diese Forschungslücke. In einem experimentellen Design wird das TM mit der direkten Befragungsmethode verglichen. Hierbei werden Studierende der Universität Leipzig zum heiklen Thema „Plagiate in Hausarbeiten“ befragt. Bisherige Befragungsstudien zu Plagiaten im universitären Kontext führten im deutschsprachigen Raum (Deutschland oder Schweiz) zu teilweise recht unterschiedlichen Prävalenzschätzungen. So reichten bei direkten Selbstauskünften die Anteile derjenigen Studierenden, die zugaben eine Haus- oder Abschlussarbeit teilweise plagiiert

zu haben von 3 % (Höglinger et al. 2013: 23) bzw. 8.1 % (Coutts et al. 2011: 755) bis zu 15,9 % (Preisendörfer 2008: 22) bzw. 22,9 % (Sattler 2007: 149). Direkte Fragen nach Vollplagiaten führten zu geschätzten Anteilen von 1,4 % (Preisendörfer 2008: 22) bis 2 % (Coutts et al. 2011: 755; Höglinger et al. 2013: 23). Diese variierenden Befunde sind aufgrund von Unterschieden in den Frageformulierungen, Stichprobenszusammensetzungen oder Erhebungsmodi nur eingeschränkt miteinander vergleichbar, können jedoch als grober Orientierungsrahmen dienen um die Ergebnisse der vorliegenden Studie besser in die Forschungslandschaft einzuordnen. Bei der vorliegenden Studie ist es von zentralem Interesse, ob unter Anwendung des TM mehr Studenten ein Plagiat offenbaren als bei einer klassischen direkten Befragung. Zudem sollen die Ergebnisse auch mit der bereits existierenden Studie zum CM (Jann et al. 2012) kontrastiert werden.

2 Befragungsmethoden im sensitiven Kontext

2.1 Die Randomized Response Technik (RRT)

Die RRT wurde erstmals im Jahr 1965 durch Stanley Warner vorgestellt und später vielfach weiterentwickelt (siehe beispielsweise Horvitz et al. 1967, Greenberg et al. 1969, Boruch 1971, Kuk 1990). Warners grundlegendes Ziel war es, die Anonymität des Befragten zu erhöhen. Dieser antwortet dazu nur mit einer vorher festgelegten Wahrscheinlichkeit auf die sensitive Frage: *„Essentially the method involves the device that [...] the interviewee responds with answers that furnish information only on a probability basis“* (Warner 1965: 63). Es wird also eine probabilistische Beziehung zwischen einer gegebenen Antwort und der heiklen Frage hergestellt. Die Umsetzung dieses Prinzips geschieht in Form eines Zufallsexperiments.

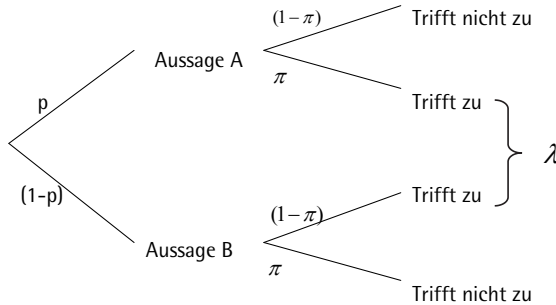
Die Individuen einer Grundgesamtheit lassen sich dahingehend unterscheiden, ob sie Träger eines sensitiven Merkmals X sind oder nicht. Gesucht ist dann der Anteil π der Merkmalsträger in der Grundgesamtheit, wobei $\pi = P(X = 1)$ auch als die Wahrscheinlichkeit, das sensitive Merkmal zu tragen, beschrieben werden kann. Dem Befragten werden dabei nachstehendem Schema folgend zwei Aussagen vorgelegt:

Aussage A: *Ich bin Träger des Merkmals X .*

Aussage B: *Ich bin nicht Träger des Merkmals X .*

Mithilfe eines Randomisierungsinstruments – durchgesetzt haben sich vor allem Würfel und Münze (Lensvelt-Mulders et al. 2005) – wird zufällig eine der beiden

Abbildung 1 Darstellung der Randomized Response Technik nach Warner (1965)



Anmerkung. p ist die Wahrscheinlichkeit mit der die Aussage A gewählt wird; π ist die gesuchte Wahrscheinlichkeit, das sensitive Merkmal X zu tragen; λ ist die Wahrscheinlichkeit, dass eine auf den Befragten zutreffende Aussage (ja' Antwort) gewählt wird

Aussagen ausgewählt. Die Auswahl im Warner Modell muss dabei zwei Bedingungen erfüllen: (1) die Auswahlwahrscheinlichkeiten der beiden Aussagen sind im Vorfeld bekannt und jeweils ungleich 0.5. (2) der Interviewer kennt den Ausgang des Zufallsexperimentes nicht, d.h. nur dem Befragten ist bekannt, welche der beiden Aussagen ausgewählt wurde. Er gibt dann lediglich an, ob die ausgewählte Aussage auf ihn zutrifft oder nicht. In Abbildung 1 ist das Befragungsprinzip schematisch dargestellt.

Aus diesem Baumdiagramm kann die Schätzformel für den Populationsparameter π über elementare wahrscheinlichkeitstheoretische Umformungen leicht abgeleitet werden. Wird der erwartete Anteil zutreffender Aussagen $\lambda = p\pi + (1 - p)(1 - \pi)$ durch $\hat{\lambda}$ – den beobachteten Anteil zutreffender Aussagen – ersetzt und umgeformt, dann ergibt sich der Schätzer von π (Warner 1965):

$$\hat{\pi} = \frac{\hat{\lambda} + p - 1}{2p - 1}$$

Die Varianz des Schätzers ist dann gegeben durch:

$$Var(\hat{\pi}) = \frac{\pi \cdot (1 - \pi)}{n} + \frac{p \cdot (1 - p)}{n \cdot (2p - 1)^2}$$

Die Varianz wird also in einen durch die Stichprobenziehung und einen durch die Randomisierung verursachten Anteil zerlegt (Warner 1965).

Lensvelt-Mulders et al. (2005) zeigen in einer Meta-Analyse, dass die RRT in sensitiven Befragungssituationen bessere Schätzungen sensitiver Merkmale liefert als die direkte Befragung. In sechs der ausgewerteten Studien lagen externe Validierungsdaten bezüglich des sensitiven Merkmals vor. Obwohl auch die RRT die tatsächliche Prävalenz der sensitiven Merkmale unterschätzte, lieferte sie deutlich validere Schätzungen als konventionelle direkte Befragungsmethoden. In weiteren von Lensvelt-Mulders et al. (2005) untersuchten Studien lagen keine Validierungsdaten vor und die RRT wurde nach dem Prinzip der „more is better“-Annahme (Umesh/Peterson 1991) anderen Befragungsmethoden gegenübergestellt. Diesem Prinzip folgend wird bei erwartetem ‚Underreporting‘ eine Methode umso valider bezeichnet, je höher die durch sie ermittelten Prävalenzraten des sensitiven Merkmals sind². Auch hier wurden mit der RRT validere Schätzungen ermittelt.

Durch die Einführung eines Randomisierungsinstruments in den Befragungsprozess folgen jedoch zahlreiche Probleme (Lensvelt-Mulders et al. 2005). Mit der erhöhten Varianz geht ein im Vergleich zur direkten Befragung größerer Standardfehler der RRT bei der Schätzung der Prävalenz sensibler Merkmale einher. Größere Stichproben können dieses Problem lösen, führen jedoch zu erhöhten Datenerhebungskosten. Ein weiterer Nachteil der RRT liegt in der Komplexität des Befragungsprozesses. Der Befragte muss nun nicht nur eine Antwort auf die ihm gestellte Frage finden, er muss sich auch stets die Regeln der RRT ins Gedächtnis rufen und zudem ein Randomisierungsinstrument (z.B. Münzen oder Würfel) zur Hand haben. Das kann zur kognitiven Überlastung und schließlich zu Beantwortungsfehlern führen. Ein Überblick über empirische Studien, bei denen die RRT im Vergleich zur direkten Befragung gleich hohe oder sogar deutlich niedrigere Anteilsschätzungen sozial unerwünschter Verhaltensweisen lieferte, findet sich bei Holbrook/Krosnick (2010).

- 2 Wenn Studien mit externem Außenkriterium aufgrund datenschutzrechtlicher Aspekte oder fehlender Existenz externer Validierungsdaten nicht möglich sind, dann werden häufig comparative, experimentelle Studien ohne externes Außenkriterium durchgeführt. In letzteren werden Hilfsannahmen getroffen, um Aussagen über die Validität einer Methode machen zu können. So wird angenommen, dass Befragte fälschlicherweise ein Verhalten leugnen können, aber niemals fälschlicherweise ein Verhalten zugeben werden was sie nicht begangen haben. Zudem wird in Abwesenheit von Validierungsdaten a priori angenommen, dass von der Norm abweichendes Verhalten systematisch unterberichtet wird. Gegeben diese Annahmen, werden höhere Anteilsschätzungen als valider eingestuft („more-is-better“ Annahme). Der Begriff der „Validität“ wird im Folgenden genau in diesem Sinne verwendet (vgl. auch Lensvelt-Mulders et al. 2005). Obwohl diese Annahmen in der vorliegenden Studie hoch plausibel sind, sind es letztendlich nur Annahmen und können eine „echte“ Validierungsstudie, also einen Vergleich von Antworten mit bekannten „wahren“ Werten, nicht ersetzen. Letzteres ist und bleibt der Goldstandard.

2.2 Eine Alternative zur RRT: Das Crosswise Modell (CM) und das Triangular Modell (TM)

Es werden zwei neue Befragungsmethoden vorgestellt, die einige Nachteile der RRT überwinden können (Yu et al. 2008): das Crosswise Modell (CM) bzw. das Triangular Modell (TM), welche in Anlehnung an Tian et al. (2009) im Folgenden unter dem Begriff der Non-Randomized Response (NRR) - Modelle zusammengefasst werden. Mathematisch weichen sie nur kaum von der RRT ab³; in ihrer Umsetzung unterscheiden sie sich jedoch erheblich von dieser. Maßgeblich ist dabei der Verzicht auf ein Randomisierungsinstrument (Tian et al. 2009). Yu et al. (2008: 255) fassen die Stärken der NRR-Modelle folgendermaßen zusammen: *„Neither model requires randomizing device, the models are easy to be implemented for both interviewer and interviewee, the interviewee does not face any sensitive questions, and both models can be applied to both face-to-face personal interviews and mail questionnaires.“* Analog zur zuvor vorgestellten RRT liegt auch bei den NRR-Modellen das Ziel darin, validere Schätzungen für ein sensitives Merkmal zu generieren. Dazu ist es notwendig, die Anonymität der Befragten zu erhöhen. Auf diese Weise sollen Vertrauen und Kooperationsbereitschaft der Befragten gesteigert werden.

Das Grundprinzip beider Modelle besteht darin, die Randomisierung mit Hilfe einer zweiten nicht-sensitiven, dichotomen Frage zu gewährleisten (Yu et al. 2008). Man betrachte erneut ein sensitives Merkmal X , wobei die Wahrscheinlichkeit für ein Individuum einer Population, dieses Merkmal zu tragen mit $\pi = P(X = 1)$ bezeichnet wird. Hinzu kommt nun ein zweites nicht-sensitives Merkmal Y , bei dem Unabhängigkeit von X vorausgesetzt wird. Mit $p = P(Y = 1)$ wird demnach die Wahrscheinlichkeit, das nicht-sensitive Merkmal zu tragen, bezeichnet. Diese Wahrscheinlichkeit wird als bekannt vorausgesetzt. Typischerweise wird dazu auf eine Geburtstagsfrage zurückgegriffen (wie z.B. „Sind Sie im Januar, Februar oder März geboren?“). Nachfolgend wird die Operationalisierung der NRR-Modelle in einer Befragungssituation mit Fragebogen erläutert. Ob und wie dies auf andere Befragungsmodalitäten (z.B. Befragungen via Online-Access-Panel oder telefonische Befragungen) übertragen werden kann, ist aufgrund mangelnder empirischer Basis in zukünftigen Studien zu untersuchen.

3 Um genau zu sein, ist der CM Schätzer mit dem klassischen RRT Modell von Warner (1965) formal identisch. Eine formale Darstellung des CM und des TM (d.h. Herleitung der Schätzformeln, Diskussion von asymptotischen Eigenschaften und Effizienzanalysen) mit Bezügen zu bisherigen RRT Designs findet sich in Yu et al. (2008).

Abbildung 2 Antwortschema des Crosswise Modells (Yu et al. 2008)

Crosswise		
	X=0	X=1
Y=0	□	■
Y=1	■	□

□ = {X = 0, Y = 0} ∪ {X = 1, Y = 1}

■ = {X = 0, Y = 1} ∪ {X = 1, Y = 0}

Anmerkung: X - sensitiv, Y - nicht-sensitiv; □ entspricht Antwortoption A, ■ entspricht Antwortoption B

2.2.1 Das Crosswise Modell (CM)

Dem Befragten werden die sensitive sowie die nicht-sensitive Frage in einem Block vorgelegt und er wird instruiert, eine kombinierte Antwort auf beide Fragen zu geben. Die dabei zu befolgenden Antwortregeln sind in Abbildung 2 dargestellt.

Der Befragte wird demnach gebeten Antwortoption A anzukreuzen, falls er beide Fragen gleichermaßen beantwortet, d.h. entweder beide mit „Ja“ oder beide mit „Nein“. Andernfalls, d.h. wenn er eine der beiden Fragen mit „Ja“, die andere mit „Nein“ beantwortet, soll er Antwortoption B ankreuzen. Zu keinem Zeitpunkt der Befragung sollen die beiden Fragen explizit einzeln beantwortet werden. Im CM sind beide Antwortoptionen als nicht-sensitiv zu betrachten. Da die Ausprägungen {X=0, Y=0} bei Option A bzw. {X=0, Y=1} bei Option B implizieren, dass der Befragte nicht Träger des sensitiven Merkmals ist, lassen sowohl Option A als auch Option B keinen eindeutigen Rückschluss darauf zu, welche Ausprägung der Befragte bezüglich des sensitiven Merkmals X hat. Yu et al. (2008) gehen davon aus, dass ein Erkennen dieses Mechanismus die Befragten motiviert, ehrlich auf die Fragen zu antworten, denn ihre Privatsphäre bleibt dadurch vollständig geschützt.

Ausgehend von der Annahme, dass alle Befragten die Methode korrekt anwenden, lässt sich durch einfache mathematische Überlegungen ein Schätzer für $\pi = P(X = 1)$, den Populationsanteil der Befragten, die das sensitive Merkmal tragen, herleiten. Es sei mit λ der unbeobachtete Anteil an „Option A“ - Antworten einer Grundgesamtheit bezeichnet. Dieser lässt sich innerhalb einer hinreichend großen Stichprobe durch den beobachteten Stichprobenanteil schätzen, also: $\hat{\lambda} = \frac{a}{n}$, wobei n die Anzahl aller Befragten in der Stichprobe und a die Anzahl der Befragten, die „Option A“ wählen, darstellen. Für den Erwartungswert des Anteils an „Option A“ - Antworten gilt: $\lambda = p\pi + (1 - p) \cdot (1 - \pi)$, wobei mit $p = P(Y = 1)$ die als bekannt vorausgesetzte Wahrscheinlichkeit, das nicht-sensitive Merkmal Y zu tragen, bezeichnet wird. Nun ergibt sich als Schätzer für π :

$$\hat{\pi}_C = \frac{\hat{\lambda} + p - 1}{2p - 1}$$

Offensichtlich muss p die Bedingung $p \neq \frac{1}{2}$ erfüllen. Für die Varianz des Schätzers gilt:

$$\text{Var}(\hat{\pi}_C) = \frac{\pi \cdot (1 - \pi)}{n} + \frac{p \cdot (1 - p)}{n \cdot (2p - 1)^2},$$

wobei folgende Formel für eine unverzerrte Schätzung der Varianz herangezogen werden kann:

$$\text{Var}(\hat{\pi}_C) = \frac{\hat{\pi}_C \cdot (1 - \hat{\pi}_C)}{n - 1} + \frac{p \cdot (1 - p)}{(n - 1) \cdot (2p - 1)^2}$$

An dieser Stelle wird der formale Bezug zur RRT nach Warner (1965) ersichtlich: Auch wenn beide Techniken in ihrer Implementierung grundverschieden sind, sind die statistischen Berechnungsformeln äquivalent.

Yu et al. (2008) liefern in ihrem Artikel theoretische Analysen und leiten statistische Schätzer für das CM her, führen jedoch keine eigene empirische Studie durch. In der ersten bisher bekannten Studie mit experimentellem Design zeigen Jann et al. (2012), dass das CM deutlich höhere Schätzungen sensitiven Verhaltens liefert als die direkte Befragung. Die Autoren befragten 474 schweizerische und deutsche Universitätsstudenten mittels selbstadministrierter Fragebögen zu Plagiaten in Hausarbeiten. Dazu wurde zwischen Teil- und Vollplagiaten unterschieden. Per Zufallsprinzip wurden die Studenten der direkten oder der CM-Befragung zugewiesen. Die Ergebnisse der Befragung zeigen ein klares Bild: 22,3% der CM-Befragten gaben demnach an, bereits einzelne Textstellen plagiiert zu haben. Dieser Wert liegt signifikant über dem der direkten Befragung: nur 7,3% der direkt befragten Studenten gaben ein Teilplagiat zu. Bei der Frage nach Vollplagiaten ergab sich ebenfalls eine Differenz zu Gunsten des CM (CM: 1,6 %, direkt: 1,0%), allerdings erwies sich diese nicht als signifikant. Die Autoren verweisen dabei auf die generell niedrige Prävalenz von Vollplagiaten in der Population. Insgesamt entlockte das CM in dieser Studie deutlich mehr sozial unerwünschte Antworten als eine direkte Befragung, allerdings müssen zukünftige Studien noch zeigen, ob sich die Ergebnisse auch auf andere Befragungskontexte und nicht-studentische Befragungspopulationen übertragen lassen (Jann et al. 2012).

Abbildung 3 Antwoortschema des Triangular Modells (Yu et al. 2008)

Triangular			
	X=0	X=1	
Y=0	□	■	□ = {X = 0, Y = 0} ■ = {X = 0, Y = 1} ∪ {X = 1, Y = 0} ∪ {X = 1, Y = 1}
Y=1	■	■	

Anmerkung: X - sensitiv, Y - nicht-sensitiv; □ entspricht Antwortoption A, ■ entspricht Antwortoption B.

2.2.2 Das Triangular Modell (TM)

Analog zum eben erläuterten Crosswise Modell werden dem Befragten sowohl eine dichotome sensitive als auch eine nicht-sensitive Frage in einem Block vorgelegt. Auch hier sieht es die Methode vor, eine kombinierte Antwort auf beide Fragen zu geben. An dieser Stelle unterscheiden sich die beiden Modelle, wie aus Abbildung 3 hervorgeht.

Der Befragte wählt Antwortoption A, wenn er beide Fragen mit „Nein“ beantwortet. Lautet seine Antwort auf mindestens eine der beiden Fragen „Ja“, das heißt also auf die sensitive oder auf die nicht-sensitive Frage (oder auf beide), so soll er Antwortoption B anzukreuzen. Auch hier ist es essentiell, dass die beiden Fragen nicht einzeln beantwortet werden. Somit erlauben auch im TM beide Antwortoptionen offensichtlich keine klaren Rückschlüsse darüber, ob der Befragte Träger des sensitiven Merkmals ist; sie sind also als nicht-sensitiv zu betrachten (Yu et al. 2008): Option A wählt der Befragte, wenn er weder Träger des sensitiven noch des nicht-sensitiven Merkmals ist {X=0, Y=0}. Option B enthält die Kategorie {X=0, Y=1}, welche impliziert, dass der Befragte nicht Träger des sensitiven Merkmals ist. Auch sie ist nicht-sensitiv. Somit bleibt die Privatsphäre des Befragten in jedem Fall geschützt.

Sei die Wahrscheinlichkeit $p = P(Y = 1)$, das nicht-sensitive Merkmal Y zu tragen, wieder als bekannt vorausgesetzt. Und sei zusätzlich mit λ der nicht beobachtbare Anteil der „Option A“ – Antworten in einer betrachteten Grundgesamtheit bezeichnet. Dieser kann analog zum CM durch den beobachteten Stichprobenanteil der „Option A“ – Antworten geschätzt werden: $\hat{\lambda} = \frac{a}{n}$, wobei n die Anzahl aller Befragten in der Stichprobe und a die Anzahl der Befragten, die „Option A“ wählen, darstellen. Für den Erwartungswert des Anteils an „Option A“ – Antworten gilt: $\lambda = (1 - p) \cdot (1 - \pi)$. Ein Umstellen der Gleichung führt zum Schätzer für $\pi = P(X = 1)$ (Yu et al. 2008):

$$\hat{\pi}_T = 1 - \frac{\hat{\lambda}}{1-p}$$

Für die Varianz gilt:

$$\text{Var}(\hat{\pi}_T) = \frac{\pi \cdot (1-\pi)}{n} + \frac{p \cdot (1-\pi)}{n \cdot (1-p)},$$

beziehungsweise als erwartungstreuen Varianzschätzer:

$$\text{Var}(\hat{\pi}_T) = \frac{\hat{\pi}_T \cdot (1-\hat{\pi}_T)}{n-1} + \frac{p \cdot (1-\hat{\pi}_T)}{(n-1) \cdot (1-p)}$$

2.2.3 Stärken der NRR-Modelle

Yu et al. (2008) folgend, beruhen sowohl die CM als auch die TM auf dem Randomisierungsprinzip der Paarung einer sensitiven mit einer nicht-sensitiven Frage und überwinden damit offensichtlich die aus der Notwendigkeit eines Randomisierungsinstruments (z.B. Münze, Würfel) resultierenden Probleme der RRT. Komplizierte Instruktionen werden somit überflüssig und Verzerrungen durch kognitive Verständnisprobleme minimiert. Es ist zudem mit geringeren Nonresponse-Raten zu rechnen. Yu et al. (2008) sehen in den NRR-Modellen einen weiteren Vorteil: bei der RRT wird der Befragte trotz Zufallsverschlüsselung weiter direkt mit der sensitiven Frage konfrontiert. Denn zu beachten ist, dass im oben vorgestellten Warner Modell neben der sensitiven Aussage auch ihre Negierung als sensitiv zu betrachten ist. Damit beantwortet der Befragte in jedem Fall eine ihm womöglich unangenehme Frage. Im Falle der NRR-Modelle wird er jedoch gebeten eine kombinierte Antwort auf die sensitive und eine nicht-sensitive Frage zu geben. Damit wird eine direkte Beantwortung des heiklen Items vermieden.

Im Vergleich der beiden NRR-Modelle miteinander weist das CM dem TM gegenüber einen Vorteil auf. Sowohl beim CM als auch beim TM werden zwar beide Antwortoptionen, Option A und Option B, als nicht-sensitiv betrachtet (Yu et al. 2008). Allerdings beinhaltet das CM für beide Optionen die Möglichkeit, dass der Befragte Träger des sensitiven Merkmals ist. Misstraut ein Befragter der Methode, gibt es für ihn keine offensichtliche Schutzstrategie, um sich mit Sicherheit als Nicht-Träger des sensitiven Merkmals darzustellen. Im Falle des TM findet er diese Strategie jedoch, indem er Option A wählt. Diese bedeutet, dass er weder das sen-

sitive noch das nicht-sensitive Merkmal trägt. Es besteht für den Befragten im TM prinzipiell die Möglichkeit, ausweichend zu antworten. Dagegen zeichnet sich das TM gegenüber dem CM für die meisten Parameterkombinationen durch einen effizienteren Schätzer aus⁴. Dadurch wird eine vergleichsweise geringere Stichprobengröße (und somit geringere Datenerhebungskosten) benötigt um einen gegebenen Standardfehler zu erreichen (Tian et al. 2011). Für das TM liegen bisher keinerlei empirische Befunde vor. Es ist folglich unklar, ob diese Befragungsmethode von den Befragten angenommen und korrekt umgesetzt wird. Diese Lücke schließt nun die vorliegende Studie.

3 Die Studie

In Anlehnung an die CM-Studie von Jann et al. (2012) wurden in der aktuellen Erhebung Leipziger Universitätsstudenten zum heiklen Thema „Plagiate in Hausarbeiten“ befragt. Es wurde sowohl ein Fragebogen mit dem TM als auch ein direkter Fragebogen implementiert. Für den Vergleich der beiden Befragungsmethoden wurde auf die bekannte „more-is-better“ Annahme zurückgegriffen. Unsere Hauptforschungshypothese ist, dass unter Anwendung des TM mehr Studierende ein Plagiat berichten, als in der direkten Befragung.

Nicht nur wegen der medialen Präsenz ist Plagiiere mehr als nur ein akademisches Tabuthema. Ein Student, dem ein Plagiat nachgewiesen wird, muss nicht nur mit sozialen und prüfungsrechtlichen Sanktionen seitens der Dozenten und der Mitstudierenden rechnen, ihm drohen im schlimmsten Fall auch strafrechtliche Konsequenzen, sobald der Betroffene, dessen Werk plagiiert wurde, Anzeige erstattet. Rechtliche Grundlage hierfür ist das Urheberrecht (*Gesetz über Urheberrecht und verwandte Schutzrechte*, Bundesministerium der Justiz 2008). So unterscheidet die „Satzung zur Sicherung guter wissenschaftlicher Praxis“ der Universität Leipzig (2002) umfänglich in arbeits- und dienstrechtliche, zivilrechtliche, akademische und strafrechtliche Konsequenzen. Es ist also zu erwarten, dass ein Student, der diese Konsequenzen ausschließen möchte, in einer direkten Befragung nicht angeben wird, bereits ein Plagiat eingereicht zu haben. Die Frage nach Plagiaten ist somit sensitiv. In unserer Erhebung unterscheiden wir zwischen Teil- und Vollpla-

4 Yu et al. (2008, S. 261) zeigen, dass die Varianz des CM-Schätzers stets größer ist als die des TM-Schätzers, solange die Wahrscheinlichkeit p , das nicht-sensitive Merkmal Y zu tragen, den Wert $2/3$ nicht übersteigt. Bei einem Wert für p von über $2/3$ hängt das Verhältnis der Varianzen zusätzlich von der Wahrscheinlichkeit π , das sensitive Merkmal zu tragen, ab.

giaten. Unter Teilplagiaten ist im Allgemeinen die Übernahme einzelner Textstellen aus einem fremden Werk zu verstehen, ohne die besagten Textstellen als Zitat zu kennzeichnen. Die unrechtmäßige Übernahme eines gesamten Werkes, sei es wörtlich oder paraphrasiert, erfüllt dagegen den Tatbestand des Vollplagiats.

3.1 Der Fragebogen

Im Rahmen eines experimentellen Designs wurden die Studierenden nach dem Zufallsprinzip entweder zur direkten Befragung oder zum TM zugewiesen. Mit Ausnahme des Experimentalteils, der die beiden sensitiven Fragen zu Teil- und Vollplagiaten enthält, sind die beiden Fragebogenversionen identisch. Insgesamt orientierte sich der Fragebogen an der CM-Studie von Jann et al. (2012). Der Wortlaut der beiden sensitiven Fragen ist wie folgt:

- **Teilplagiat:** *„Haben Sie beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Abschlussarbeit, Semesterarbeit etc.) schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne dies als Zitat zu kennzeichnen?“*
- **Vollplagiat:** *„Haben Sie schon einmal einen Großteil einer Arbeit durch eine fremde Person schreiben lassen oder eine fremde Arbeit (z.B. von www.hausarbeiten.de) als Ihre eigene ausgegeben?“*

Im direkten Fragebogen beantworten die Studenten die beiden Fragen direkt, in der TM-Kondition werden sie dagegen nicht direkt beantwortet. Stattdessen sind beide sensitive Fragen jeweils mit einer unabhängigen nicht-sensitiven Frage gepaart und die Befragten werden gebeten eine kombinierte Antwort zu geben. Die nicht-sensitiven Fragen lauten folgendermaßen:

- **Teilplagiat:** *„Hat Ihre Mutter in den Monaten Januar, Februar oder März Geburtstag?“*
- **Vollplagiat:** *„Hat Ihr Vater in den Monaten Oktober, November oder Dezember Geburtstag?“*

Bei den beiden nicht-sensitiven Fragen haben wir einen Näherungswert für die Wahrscheinlichkeit einer „Ja“ Antwort ermittelt: ausgehend von einer Altersspanne von circa 18 bis 25 Jahren der befragten Studierenden und einem durchschnittlichen Alter von 24 Jahren bei der Geburt, sind die Eltern der befragten Studierenden mit hoher Wahrscheinlichkeit zwischen 1962 und 1969 geboren. Für diese Jahrgänge wurden beim Statistischen Bundesamt Deutschland die nach Monaten aufgeschlüsselten Geburtsraten erfragt und eine durchschnittliche Geburtenrate ermittelt. Die so ermittelte Wahrscheinlichkeit p_1 in den Monaten Januar/Februar/März geboren zu sein beträgt 0,26, die Wahrscheinlichkeit p_2 für die Monate

Oktober/November/Dezember hingegen 0,23.⁵ Abbildung 4 visualisiert die Implementierung des TM im Fragebogen der aktuellen Studie.

3.2 Stichprobe und Datenerhebung

Die Befragung fand schriftlich in der Vorlesung „Grundzüge der Soziologie“ statt. Sie wurde in den Wintersemestern 2009/2011 bzw. 2010/2011 durchgeführt. Es handelt sich hierbei um ein Convenience-Sample, d.h. die Vorlesung wurde nicht zufällig aus allen Vorlesungen in Leipzig ausgewählt. Neben weiteren Prüfungsleistungen, schreiben alle Studenten im Rahmen dieser Veranstaltung eine Hausarbeit. Der Abgabetermin der Hausarbeit lag im Semester jeweils einige Wochen vor Vorlesungsende. Die Befragungen fanden immer nach der Abgabe der Hausarbeiten statt. Somit hatte jeder Befragte potentiell eine Möglichkeit zu plagieren. In der Studie wurden alle in der jeweiligen Vorlesung anwesenden Studenten befragt. Es wurden keinerlei monetäre oder sonstige materielle Anreize für die Teilnahme angeboten. Nach einer kurzen Vorstellung der Studie wurden die Fragebögen ausgeteilt. Während der Befragung wurden keinerlei Komplikationen oder Probleme beobachtet.⁶

Die Studenten wurden zufällig einer der beiden Befragungskonditionen – direkt versus TM – zugewiesen und füllten den Fragebogen selbst-administriert aus.⁷ Die Randomisierung fand im Verhältnis 1:3 statt.⁸ Das Verhältnis von 1:3 wurde mit dem Ziel gewählt, auch für die TM-Kondition einen akzeptablen Standardfehler zu erreichen. In der Stichprobe waren im Prinzip alle zur Modulabschlussklausur zugelassenen Studierenden in den betrachteten Jahrgängen (N=621). Von den 621 zur Modulabschlussklausur zugelassenen Studierenden waren 434 zur Befragung anwesend. 382 von ihnen füllten schließlich den Fragebogen aus. In der TM-Kondition wurden 281 Studierende befragt, den direkten Fragebogen füllten 101 Stu-

5 Diese Werte können Schwankungen unterliegen, allerdings sind sie als Näherungswerte besser als ein aus der Annahme einer Gleichverteilung resultierendes p von jeweils 0,25. Zudem wurden auch die Geburtsraten von 1960 bis 1974 überprüft. Es fanden sich die gleichen Wahrscheinlichkeiten.

6 Lediglich eine Studentin fragte nach, ob bei ihrem Fragebogen eine Seite fehlen würde. Sie hatte einen direkten Fragebogen zugeteilt bekommen, der eine Seite kürzer ist als der TM-Fragebogen.

7 Aufgrund der Randomisierung zu einer der beiden Befragungskonditionen ist eine hohe interne Validität sichergestellt. Es wird angenommen, dass sich die Ausfälle (sprich: die am Tag der Befragung nicht anwesenden Studenten) nicht systematisch zwischen den Experimentalgruppen unterscheiden.

8 Hierbei wurden die Fragebögen im Vorfeld dem Verhältnis entsprechend vorsortiert und dann in der Vorlesung der Reihe nach ausgeteilt.

Abbildung 4 Instruktion des Triangular Modells

<p>Im nun folgenden Teil der Befragung wende ich eine neuartige Befragungstechnik an, um Ihre Privatsphäre bei diesen für Sie möglicherweise unangenehmen Fragen zusätzlich zu schützen. Lesen Sie sich bitte zunächst die Instruktionen genau durch und beantworten Sie dann die Fragen.</p> <p>Es werden jeweils zwei Fragen in einem Block gestellt. Überlegen Sie sich bitte zuerst, wie Sie die beiden Fragen einzeln beantworten würden (entweder mit <i>Ja</i> oder mit <i>Nein</i>), schreiben Sie dies aber <u>nicht</u> auf! Je nachdem wie Ihre Antworten auf diese beiden Fragen lauten, kreuzen Sie im Anschluss bitte Möglichkeit (A) oder (B) an, und zwar nach folgenden Regeln:</p> <p>Lautet Ihre Antwort auf beide Fragen <i>Nein</i>, so setzen Sie Ihr Kreuz bei (A).</p> <p>Lautet Ihre Antwort auf mindestens eine der beiden Fragen <i>Ja</i>, so setzen Sie Ihr Kreuz bei (B).</p> <p>Ihre Privatsphäre bleibt geschützt, da ich Ihre Antworten auf die einzelnen Fragen nicht kenne. Mittels statistischer Verfahren kann ich aber berechnen, für wie viele Personen <u>insgesamt</u> die jeweils zweite Frage zutrifft.</p>
12) Block 1
<p>1. Frage: <i>Hat Ihre Mutter in den Monaten Januar, Februar oder März Geburtstag?</i></p> <p>2. Frage: <i>Haben Sie beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Abschlussarbeit, Semesterarbeit, etc.) schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen?</i></p> <hr/> <p>Wie lauten die Antworten auf die beiden Fragen?</p> <p><input type="radio"/> (A) auf beide Fragen <i>Nein</i></p> <p><input type="radio"/> (B) auf mindestens eine der beiden Fragen <i>Ja</i></p>
13) Block 2
<p>1. Frage: <i>Hat Ihr Vater in den Monaten Oktober, November oder Dezember Geburtstag?</i></p> <p>2. Frage: <i>Haben Sie schon einmal einen Großteil einer Arbeit durch eine andere Person schreiben lassen oder eine fremde Arbeit (z.B. von www.hausarbeiten.de) als Ihre eigene ausgegeben?</i></p> <hr/> <p>Wie lauten die Antworten auf die beiden Fragen?</p> <p><input type="radio"/> (A) auf beide Fragen <i>Nein</i></p> <p><input type="radio"/> (B) auf mindestens eine der beiden Fragen <i>Ja</i></p>

dierende aus. Die für beide Jahrgänge aggregierte Ausschöpfungsquote liegt bei 62%. Bezogen auf die anwesenden Studierenden beträgt die Response-Rate 88%.

4 Empirische Befunde

4.1 Soziodemographie in den Befragungskonditionen

Das Durchschnittsalter der befragten Studenten betrug 21,7 Jahre (mit einer Standardabweichung von 2,63 Jahren), 61% von ihnen waren weiblich. Über 50% befanden sich zum Befragungszeitpunkt im ersten Studienjahr. Über 97% der Studenten sind in Studiengängen der neuen Abschlüsse immatrikuliert, von denen sich etwa 43% zurzeit in einem Bachelor-Studium befinden. Alle Befragten haben bereits mindestens eine Hausarbeit geschrieben. Tabelle 1 listet ausgewählte Befragtenmerkmale aufgeschlüsselt nach den beiden Befragungskonditionen auf. Es wurden jeweils paarweise χ^2 -Unabhängigkeitstests zwischen der Befragungskondition und den Befragtenmerkmalen durchgeführt. Keiner der χ^2 -Werte ist auf dem 5% Niveau signifikant. Es bestehen keine systematischen Unterschiede zwischen den beiden Befragungskonditionen.

Tabelle 1 Soziodemographie in den beiden Befragungskonditionen

	Direkt (N=101)		Triangular (N=281)		χ^2 (p)
	n	%	n	%	
Geschlecht					
Männlich	38	37,6 %	110	39,2 %	0,063 (0,802)
Weiblich	62	61,4 %	169	60,1 %	
k.A.	1	1,0 %	2	0,7 %	
Alter					
18 – 21	56	55,5 %	151	53,7 %	1,476 (0,688)
22 – 25	39	38,6 %	107	38,1 %	
26 – 29	6	5,9 %	17	6,1 %	
30 und älter	0	0 %	4	1,4 %	
k.A.	0	0 %	2	0,7 %	
Studienjahr					
1	52	51,5 %	161	57,3 %	3,869 (0,276)
2	20	19,8 %	61	21,7 %	
3	28	27,7 %	52	18,5 %	
4	0	0,0 %	4	1,4 %	
6	1	1,0 %	0	0,0 %	
7	0	0,0 %	1	0,4 %	
k.A.	0	0,0 %	2	0,7 %	
Angestrebter Abschluss					
Bachelor	39	38,6 %	118	42,0 %	0,530 (0,767)
Master	58	57,4 %	153	54,4 %	
Diplom	1	1,0 %	3	1,1 %	
Magister	0	0,0 %	1	0,4 %	
Staatsexamen	1	1,0 %	2	0,7 %	
Sonstige	0	0,0 %	2	0,7 %	
k.A.	2	2,0 %	2	0,7 %	
Bisher geschriebene Hausarbeiten					
Keine	0	0,0 %	0	0,0 %	1,908 (0,385)
1 bis 2	57	56,4 %	180	64,0 %	
3 bis 4	19	18,8 %	46	16,4 %	
5 oder mehr	25	24,8 %	55	19,6 %	

Anmerkung: Item Nonresponse führt zu leichten Schwankungen in den Fallzahlen

Tabelle 2 Plagiatsraten nach Befragungskondition

	Direkt (N=101)	Triangular (N=280) ⁹	Differenz
Teilplagiat			
Anteil (in %)	9,9	18,0	8,1 (p = 0,104)
Standardfehler	3,0	4,0	4,9
95 % Konfidenzintervall	[4,1 , 15,7]	[10,2 , 25,7]	[-1,7 , 17,8]
Vollplagiat			
Anteil (in %)	0,0	0,8	0,8 (p = 0,805)
Standardfehler	0,0	3,3	3,3
95 % Konfidenzintervall	[0 , 0]	[-5,7 , 7,3]	[-5,7 , 7,4]

4.2 Plagiatsraten

Tabelle 2 zeigt die Prävalenzraten für Teil- und Vollplagiate aufgeschlüsselt nach den beiden Befragungskonditionen.

Die Differenz der geschätzten Anteile zwischen TM- und direkter Befragung beträgt für die Frage nach Teilplagiaten 8,1%. Es gaben in der TM-Kondition demnach mehr Studenten ein Teilplagiat zu als in der direkten Befragung (18,0% in der TM versus 9,9% in der direkten Version). Allerdings liegt das empirische Signifikanzniveau auf der Grundlage eines zweiseitigen Tests mit $p = 0,104$ über dem konventionellen Niveau von 5%¹⁰. Somit lassen sich in der vorliegenden Studie primär nur deskriptive Aussagen treffen. Für die Frage nach einem Vollplagiat unterscheiden sich die beiden Schätzungen nicht. In der direkten Befragung gab kein Student ein Vollplagiat an. In der TM-Kondition wurde ein Anteilswert von 0,8% geschätzt. Auch wenn dieser Wert leicht über dem der direkten Befragung liegt, weicht er nicht signifikant von 0% ab – leicht erkennbar am Konfidenzintervall. Für die Frage nach Vollplagiaten führt das TM demnach zu keiner nennenswert höheren Prävalenzschätzung als die direkte Befragung.

- 9 In der TM-Kondition verweigerte ein Student komplett die Beantwortung der Fragen nach Plagiaten.
- 10 Für den Test der statistischen Signifikanz wurde als Prüfgröße die Differenz der Plagiatsraten in der direkten und der TM-Kondition, gewichtet mit der Wurzel der gemeinsamen Varianz, gebildet:

$$t = \frac{|\hat{\pi}_d - \hat{\pi}_t|}{\sqrt{\frac{\hat{\pi}_d \cdot (1 - \hat{\pi}_d)}{n_d} + \text{Var}(\hat{\pi}_t)}}$$

wobei $\hat{\pi}_d$ der direkten und $\hat{\pi}_t$ der Triangular-Plagiatsrate entspricht (n_d ist die Stichprobengröße der direkten Befragung). Diese Prüfgröße entstammt einer t-Verteilung mit $n_d + n_t - 2$ Freiheitsgraden.

Tabelle 3 Plagiatsraten der aktuellen Triangular-Studie und der Crosswise-Studie von Jann et al. (2012)

	Crosswise-Studie (Jann et al. 2012)		Triangular-Studie (aktuelle Erhebung)	
	Direkt	Crosswise	Direkt	Triangular
N	96	310	101	280
Teilplagiat	7,3% (2,7)	22,3% (5,5)	9,9% (3,0)	18,0% (4,0)
Vollplagiat	1,0% (1,0)	1,6% (5,0)	0% (0,0)	0,8% (3,3)

Anmerkung: Standardfehler in Klammern.

5 Vergleich des TM und CM

Tabelle 3 kontrastiert die Ergebnisse der vorliegenden TM-Studie mit der CM-Studie von Jann et al. (2012).

Es zeigt sich, dass sowohl das CM mit 22,3% als auch das TM mit 18,0% für die Frage nach Teilplagiaten deutlich höhere Schätzungen erzielen als die direkte Befragung mit 7,3% bzw. 9,9%. Ein ebenso einheitliches Ergebnis zeigt sich für die Frage nach Vollplagiaten. Unter Anwendung dieser beiden Modelle konnten keine nennenswert höheren Raten ermittelt werden. Im direkten Vergleich des CM und des TM wird für die Frage nach Teilplagiaten eine Differenz von 4,3% zugunsten des CM ermittelt. Allerdings zeigt sich mit den deutlich geringeren Standardfehlern des TM ein (verglichen mit dem CM) klarer Vorteil bezüglich der statistischen Effizienz. Insgesamt ist nur ein grober und skizzenhafter Vergleich des CM mit dem TM möglich, da sie in verschiedenen Erhebungen und Stichproben evaluiert wurden¹¹. Auch wird in diesem Zusammenhang auf die Berechnung von Signifikanztests verzichtet, da sie nur wenig sinnvoll erscheinen. In einem nächsten Schritt könnte eine Anschlussstudie die CM und die TM in ein einheitliches Design implementieren und

11 Während in der vorliegenden TM-Studie ausschließlich Studierende der Universität Leipzig befragt wurden, wurden in der CM-Studie neben Leipziger Universitätsstudenten auch Studierende der ETH Zürich und der LMU München befragt. Zudem wurde bei der Ermittlung der Plagiatsraten in der CM-Studie eine Gleichverteilung für die nicht-sensitive Geburtstagsfrage angenommen. Für die Berechnungen lag die Wahrscheinlichkeit p , in den Monaten Januar/Februar/März geboren zu sein, demnach bei 0,25. Somit sind die beiden Studien nur bedingt miteinander vergleichbar und die vorliegende Darstellung dient lediglich einem groben Eindruck.

dann direkt miteinander kontrastieren (mit einer direkten Befragung als Kontrollgruppe).

6 Diskussion

In unserer Studie konnte gezeigt werden, dass unter Anwendung des TM deutlich mehr Studenten ein Teilplagiat zugeben als in einer direkten Befragung. Damit kann unsere Hauptforschungshypothese (mehr sozial unerwünschte Antworten unter anonymen Befragungsbedingungen) zumindest für die Frage nach Teilplagiaten bestätigt werden. Die beobachtete Differenz von 8,1% zwischen den beiden Konditionen verfehlt allerdings das konventionelle 5%-Signifikanzniveau. Somit lassen sich die vorliegenden Befunde nicht generalisieren und sind somit eher als explorativ und vorläufig zu betrachten. Für Vollplagiate konnte dagegen kein nennenswerter Unterschied zwischen den Experimentalkonditionen festgestellt werden. Die in der direkten und in der TM-Befragung ermittelten Anteilsschätzungen weichen statistisch nicht signifikant von Null und auch nicht voneinander ab.

Die Frage ist nun warum sich für das Item zu Vollplagiaten in der TM-Kondition kein nennenswerter Unterschied zur direkten Befragung zeigt. Dies könnte potentiell drei Ursachen haben: (1) Die Abgabe einer vollständig plagiierten Arbeit birgt ein sehr hohes Entdeckungsrisiko und ist in der Regel mit gravierenden Konsequenzen verbunden. Demzufolge sind die tiefen Schätzungen auf Grundlage des TM nicht auf ein Versagen der Methode zurückführbar, sondern auf die generelle Seltenheit von Vollplagiaten. (2) Die Befragten haben sich bei diesem besonders heiklen Item bewusst nicht an die Antwortregeln gehalten. Dies kann zum einen auf Befragte zutreffen, die bereits voll plagiiert haben und nun die sichere Antwortstrategie für den Selbstschutz – „Option A“ – wählen. Zum anderen wären auch Befragte denkbar, die nicht plagiiert haben, auf die jedoch das nicht-sensitive Item zutrifft. Diese könnten möglicherweise befürchten, dass ein Ankreuzen von „Option B“ fälschlicherweise als Vollplagiat interpretiert wird und würden deshalb aus Selbstschutz womöglich „Option A“ wählen. (3) Es kann nicht als gesichert gelten, dass tatsächlich auf 23,3 % der Studierenden das nicht-sensitive Merkmal zutrifft, das heißt, dass sie einen Vater haben, der in den Monaten Oktober, November oder Dezember Geburtstag hat. Der Wert ist daher zufälligen Schwankungen unterlegen, die erst bei sehr großen Stichproben an Einfluss auf die Schätzung der Anteilswerte verlieren. Zudem wird es durch die Einführung eines Randomisierungsprozesses so gut wie unmöglich bei kleinen Prävalenzen nahe Null zuverlässige und effiziente Schätzungen zu erreichen (vgl. Jann et al. 2012: 46).

Welches der oben angeführten Argumente für die Erklärung der kleinen Anteilsschätzungen beim Vollplagiats-Item zutreffend ist kann nicht ad hoc entschieden werden und sollte deshalb in zukünftigen Studien detaillierter untersucht werden. Auch könnte in einem nächsten Schritt die TM mit anderen indirekten Befragungsmethoden, wie beispielsweise der RRT, verglichen werden.

Literatur

- Boruch, R. F., 1971: Assuring confidentiality of responses in social research: a systematic analysis. *The American Psychologist* 26: 413-430.
- Bundesministerium der Justiz, 2008: Das Urheberrecht und verwandte Schutzrechte. Berlin: Bundesministerium der Justiz. <http://www.gesetze-im-internet.de/urhrg> (07.11.2013).
- Coutts, E., B. Jann, I. Krumpal und A.-F. Näher, 2011: Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions. *Jahrbücher für Nationalökonomie und Statistik* 231 (5+6): 749-760.
- Greenberg, B. G., A. A. Abul-Ela, W. R. Simmons und D. G. Horvitz, 1969: The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association* 64: 520-539.
- Holbrook, A. L. und J. A. Krosnick, 2010: Measuring Voter Turnout by Using the Randomized Response Technique: Evidence Calling into Question the Method's Validity. *Public Opinion Quarterly* 74: 328-343.
- Horvitz, D. G., B. V. Shah und W. R. Simmons, 1967: The unrelated question randomized response model. S. 65-72 in: *Proceedings of the Social Statistics Section, American Statistical Association. Washington, D.C.: American Statistical Association.*
- Höglinger, M., B. Jann und A. Diekmann (2013): Sensitive Questions in Online Surveys: An Experimental Comparison of the RRT and the Crosswise Model. Unpublished Working Paper, ETH Zurich and University of Bern.
- Jann, B., J. Jerke und I. Krumpal, 2012: Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism. *Public Opinion Quarterly* 76: 32-49.
- Krumpal, I. 2012: Estimating the Prevalence of Xenophobia and Anti-Semitism in Germany: A Comparison of Randomized Response and Direct Questioning. *Social Science Research* 41(6): 1387-1403.
- Krumpal, I. 2013: Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review. *Quality & Quantity* 47: 2025-2047.
- Krumpal, I. und A.-F. Näher, 2012: Entstehungsbedingungen sozial erwünschten Antwortverhaltens: Eine experimentelle Studie zum Einfluss des Wording und des Kontexts bei unangenehmen Fragen. *Soziale Welt* 63: 65-89.
- Kuk, A.Y.C., 1990: Asking Sensitive Questions Indirectly. *Biometrika* 77: 436-438.
- Lensvelt-Mulders, G., J. Hox, P. van der Heijden und C. Maas, 2005: Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation. *Sociological Methods & Research* 33: 319-348.
- Paulhus, D.L., 1984: Two-component models of socially desirable responding. *Journal of Personality and Social Psychology* 46: 598-609.
- Paulhus, D.L. und O.P. John, 1998: Egoistic and Moralistic Biases in Self-Perception: The Interplay of Self-Deceptive Styles With Basic Traits and Motives. *Journal of Personality* 66: 1025-1060.

- Preisendörfer, P. 2008: Heikle Fragen in mündlichen Interviews: Ergebnisse einer Methodenstudie im studentischen Milieu. *ETH Zurich Sociology Working Papers* 6, ETH Zurich, Chair of Sociology.
http://repec.ethz.ch/ets/papers/preisendoerfer_sensitive_questions.pdf (08.11.2013)
- Sattler, S. 2007: Plagiate in Hausarbeiten – Erklärungsmodelle mit Hilfe der Rational Choice Theorie. Hamburg, Kovač.
- Tian, G.-L., K.Ch. Yuen, M.-L. Tang und M.T. Tan, 2009: Bayesian non-randomized response models for surveys with sensitive questions. *Statistics and its interface* 2: 13-25.
- Tian, G.-L., M.-L. Tang, Z. Liu, M. Tan und N.-S. Tang, 2011: Sample size determination for the non-randomised triangular model for sensitive questions in a survey. *Statistical Methods in Medical Research* 20: 159-173.
- Tourangeau, R. und T. Yan, 2007: Sensitive Questions in Surveys. *Psychological Bulletin* 133: 859-883.
- Umesh, U. N. und R. A. Peterson, 1991: A Critical Evaluation of the Randomized Response Method: Applications, Validation and Research Agenda. *Sociological Methods Et Research* 20: 104-38.
- Universität Leipzig, 2002: *Satzung zur Sicherung guter wissenschaftlicher Praxis*. Leipzig: Universität Leipzig. <http://www.zv.uni-leipzig.de/forschung/satzung>. (07.11.2013).
- Warner, St. L., 1965: Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association* 60: 63-69.
- Yu, J.-W., G.-L. Tian und M.-L. Tang, 2008: Two new models for survey sampling with sensitive characteristic: Design and analysis. *Metrika* 67: 251-263.

Anschrift der Autoren

Julia Jerke (Korrespondenzautorin)
Universität Leipzig
c/o Ivar Krumpal, Institut für Soziologie
Beethovenstrasse 15
04107 Leipzig
E-Mail: julia.jerke@sozio.uni-leipzig.de

Ivar Krumpal
Universität Leipzig
Institut für Soziologie
Beethovenstrasse 15
04107 Leipzig
E-Mail: krumpal@sozio.uni-leipzig.de