

SOEPpapers

on Multidisciplinary Panel Data Research

189

Carsten Sauer • Katrin Auspurg • Thomas Hinz • Stefan Liebig • Jürgen Schupp

**Die Bewertung von Erwerbseinkommen -
Methodische und inhaltliche Analysen zu einer
Vignettenstudie im Rahmen des SOEP-Pretest 2008**

Berlin, Mai 2009

SOEPPapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPPapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPPapers are available at
<http://www.diw.de/soeppapers>

Editors:

Georg **Meran** (Dean DIW Graduate Center)

Gert G. **Wagner** (Social Sciences)

Joachim R. **Frick** (Empirical Economics)

Jürgen **Schupp** (Sociology)

Conchita **D'Ambrosio** (Public Economics)

Christoph **Breuer** (Sport Science, DIW Research Professor)

Anita I. **Drever** (Geography)

Elke **Holst** (Gender Studies)

Frieder R. **Lang** (Psychology, DIW Research Professor)

Jörg-Peter **Schräpler** (Survey Methodology)

C. Katharina **Spieß** (Educational Science)

Martin **Spieß** (Survey Methodology, DIW Research Professor)

Alan S. **Zuckerman** (Political Science, DIW Research Professor)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | urahmann@diw.de

Die Bewertung von Erwerbseinkommen -
Methodische und inhaltliche Analysen zu einer
Vignettenstudie im Rahmen des SOEP-Pretest 2008

Carsten Sauer^{}, Katrin Auspurg^{**}, Thomas Hinz^{**}, Stefan Liebig^{***} und Jürgen
Schupp^{****}*

Mai 2009

* Fakultät für Soziologie, Universität Bielefeld

** Fachbereich für Geschichte und Soziologie, Universität Konstanz

*** Fakultät für Soziologie, Universität Bielefeld und DIW Berlin

**** SOEP am DIW Berlin und Institut für Soziologie, Freie Universität Berlin

Berlin, Mai 2009

Zusammenfassung

Im Pretest 2008 wurde erstmals für die Längsschnittstudie Sozio-oekonomisches Panel (SOEP) die Erhebungsmethode des Faktoriellen Surveys erprobt. Es werden Aufbau und die Umsetzung des Vignettenmoduls beschrieben, das inhaltlich auf dem Gebiet der empirischen Gerechtigkeitsforschung, speziell Messung von Einkommensgerechtigkeit, angesiedelt ist. Es wird gezeigt, dass der Faktorielle Survey bei Beachtung einiger Randbedingungen auch in der großflächigen Umfrageforschung einsetzbar ist und Befragte verschiedener Alters- und Bildungsgruppen mit der Beantwortung hinreichend gut zurecht kommen. Die erzielten Ergebnisse deuten darauf hin, dass vor allem ältere Befragte weniger Dimensionen zu Beurteilung heranziehen. Ob dies als Hinweis für eine für diese Befragten zu komplexe Urteilsaufgabe und damit einen methodischer Effekt zu deuten ist, oder aber ein inhaltlich valides Ergebnis darstellt, wäre in künftigen Studien zu klären. Die inhaltlichen Ergebnisse zeigen beispielhaft, dass neben dem Beruf, der Ausbildung und der Leistung – also Faktoren die im direkten Bezug zur Erwerbstätigkeit stehen – ebenso familiäre Aspekte, wie der Familienstand, die Erwerbstätigkeit des Partners und die Anzahl der Kinder relevante Kriterien für die Einkommensgerechtigkeit darstellen. Der Faktorielle Survey im SOEP Pretest 2008 bietet sowohl in methodischer als auch inhaltlicher Hinsicht vielzählige Analysemöglichkeiten. Die positiven Erfahrungen des SOEP-Pretest 2008 ermutigen dazu, auch in einer künftigen Haupterhebung des SOEP Vignettenmodule einzusetzen.

Abstract

In the 2008 Socio-Economic Panel Study (SOEP) Pretest, the factorial survey method was tested for the first time for use in the SOEP longitudinal study. In this paper, we describe the construction and application of the vignette module, which has its origins in the field of justice research and is used in particular in the measurement of income justice. We show that the factorial survey method is applicable in large-scale survey research when taking certain constraints into account, and that respondents of varying ages and educational groups are able to deal sufficiently well with answering the questions. The results obtained suggest that older respondents tend to take fewer dimensions into consideration in forming their opinions. Further studies will be needed to determine whether this is evidence that the evaluation tasks were too complex for these respondents and should thus be interpreted as a method effect, or whether it represents a valid substantive result. The results of the study demonstrate convincingly that alongside occupation, education, and performance—factors relating directly to employment—familial aspects such as civil status, the partner's employment status, and number of children constitute important criteria for determining what constitutes a "fair" income. The factor survey in the 2008 SOEP Pretest offers diverse analytical potential, both from a methodological point of view and in terms of the empirical results obtained. The positive experience with the 2008 SOEP Pretest suggests that the SOEP vignette module can be used effectively in a future wave of the main SOEP survey.

JEL: C81, D63, J31

Inhaltsverzeichnis

1	Einleitung.....	1
2	Der Faktorielle Survey als Erhebungsinstrument in großen Umfragen.....	4
3	Konstruktion des Faktoriellen Surveys für den SOEP-Pretest 2008.....	6
3.1	Entwicklung der Vignettendimensionen und Ausprägungen	6
3.2	Bildung des Vignettenuniversums und Stichprobenziehung	7
3.3	Bewertungsmodus und Präsentation der Vignetten.....	8
3.4	Befragten- und Vignettenstichprobe (SOEP-Pretest 2008).....	9
4	Methodische Ergebnisse	11
4.1	Befragten- und Interviewereindrücke	11
4.2	Antwortverhalten	15
4.2.1	Befragungsdauer.....	16
4.2.2	Nutzung der Skala.....	17
4.2.3	Konsistenz der Urteile	21
5	Inhaltliche Fragestellungen.....	26
6	Fazit.....	32
	Literatur	34
	Anhang	37

Verzeichnis der Abbildungen und Tabellen

Tabellen

Tab. 1: Dimensionen und Ausprägungen des Vignettenmoduls	7
Tab. 2: Häufigkeit der Decks	11
Tab. 3: Kommentare.....	12
Tab. 4: Verständlichkeit (Angaben in Prozent).....	12
Tab. 5: Antwortbereitschaft (Angaben in Prozent)	14
Tab. 6: Antwortbereitschaft im Vignettenteil nach Altersgruppen (Angaben in Pro.....	15
Tab. 7: Antwortbereitschaft im Vignettenteil nach Schulabschluss (Angaben in.....	15
Tab. 8: Dauer der Befragung: arithmetisches Mittel und 5-Punkte-Statistik (in Minuten).....	16
Tab. 9: Verteilung der Variable „Gerechtigkeitsurteil“	17
Tab. 10: Häufigkeit der Kategorie <i>gerecht</i> nach Schulabschluss	18
Tab. 11: Häufigkeit der Kategorie <i>gerecht</i> nach Altersgruppen	19
Tab. 12: Häufigkeit unterschiedlicher Zahlenangaben nach Schulabschluss	20
Tab. 13: Häufigkeit unterschiedlicher Zahlenangaben nach Altersgruppen.....	20
Tab. 14: Verwendete Zahlenwerte bzw. Skalierungen.....	21
Tab. 15: Modellgüte und signifikante Koeffizienten nach Schulabschluss und Altersgruppe	24
Tab. 16: Determinanten der Einkommensgerechtigkeit (multinomiales Logit-Modell), Marginaleffekte	27

Abbildungen

Abb. 1: Beispielvignette.....	9
Abb. 2: Verständlichkeit nach Altersgruppen (Angaben in Prozent)	13
Abb. 3: Verständlichkeit nach Bildung (Angaben in Prozent).....	14
Abb. 4: Verteilung der Befragungszeiten für das Vignettenmodul nach Bildungs- und Altersgruppen (in Minuten)	17
Abb. 5: Verteilung der Gerechtigkeitsurteile.....	18
Abb. 6: Häufigkeit der Kategorie <i>gerecht</i> nach Vignettenposition	19
Abb. 7: Modellgüte in den sechs Phasen des Vignettenmoduls	22
Abb. 8: Anzahl signifikanter Koeffizienten und Pseudo-R ² in den sechs Phasen des Vignettenmoduls.....	23
Abb. 9: Gerechtigkeitsbewertung nach Berufen der Vignettenpersonen und Befragengruppen.....	29
Abb. 10: Gerechtigkeitsbewertung nach Leistung (Mittelwertunterschiede)	30
Abb. 11: Gerechtigkeitsbewertung nach Geschlecht der Vignettenpersonen	31

Anhang

Tab. A1: Stichprobenstruktur und Gewichtung.....	37
Tab. A2: Korrelationen der Vignettendimensionen.....	38
Tab. A3a: Korrelationen der Vignettendimensionen für die Vignetten 1 bis 4.....	38
Tab. A3b: Korrelationen der Vignettendimensionen für die Vignetten 5 bis 8.....	39
Tab. A3c: Korrelationen der Vignettendimensionen für die Vignetten 9 bis 12	39
Tab. A3d: Korrelationen der Vignettendimensionen für die Vignetten 13 bis 16.....	39
Tab. A3e: Korrelationen der Vignettendimensionen für die Vignetten 17 bis 20.....	39
Tab. A3f: Korrelationen der Vignettendimensionen für die Vignetten 21 bis 24.....	39
Tab. A3g: Varianzen der Vignettendimensionen für die einzelnen Vignettenabschnitte	40

1 Einleitung¹

In den letzten Jahren war in Deutschland eine Zunahme der Einkommensungleichheit zu beobachten (Giesecke und Verwiebe 2008; Grabka und Frick 2008, OECD 2008). Diese Entwicklung wird in öffentlichen Debatten als Ausdruck zunehmender sozialer Ungerechtigkeit interpretiert. Ob diese Einschätzung mit den Einstellungen der Bürgerinnen und Bürger übereinstimmt, kann bislang aber von Seiten der empirischen Sozialforschung nicht hinreichend beantwortet werden. Denn während die tatsächliche Einkommensentwicklung durch die verfügbaren Daten der empirischen Sozial- und Wirtschaftsforschung und amtlichen Statistik sehr gut abgebildet werden kann, fehlen bislang detaillierte Informationen dazu, was in der Bevölkerung als ein gerechtes Einkommen angesehen wird und wovon gerechterweise die Höhe eines Einkommens abhängen sollte. Dies liegt vor allem daran, dass es bislang keine geeigneten Erhebungsinstrumente gibt, mit denen die Gerechtigkeitsvorstellungen der Menschen differenziert erhoben werden können.

Die dominierende Forschung bedient sich vor allem der klassischen Itemabfrage. Die Befragten sollen beispielsweise angeben, ob sie die Einkommensunterschiede in unserer Gesellschaft als zu groß oder zu klein einschätzen und ob sie meinen, ihren gerechten Anteil zu erhalten (Prüfer et al. 2003). Daneben werden sie gefragt, was bedeutsame Bestimmungsgrößen für eine gerechte Entlohnung sein sollten. Das individuelle Einkommen ist faktisch von sehr unterschiedlichen Faktoren abhängig, etwa der Bildung und Berufserfahrung, dem Beruf und der Branche, den aktuellen Leistungen am Arbeitsplatz oder auch der Größe des Haushalts, den jemand zu versorgen hat. Mit Item-Abfragen wird üblicherweise erfasst ob überhaupt, jedoch nicht wie stark und in welcher Relation diese Merkmale auf das gerechte Einkommen einwirken sollten. Zudem lassen sich nur ansatzweise Wechselwirkungen zwischen verschiedenen Merkmalen erheben – sollte sich beispielsweise eine längere Berufserfahrung speziell für höher Gebildete oder bestimmte Berufsgruppen auszahlen? Ähnliche Probleme betreffen eine alternative Erhebungsstrategie, Befragte ihr eigenes Einkommen evaluieren zu lassen und sie offen danach zu fragen, welches Einkommen für sie selbst gerecht wäre (Schwarze 2007; Liebigh und Schupp 2005,2008b). Auch diese direkten Abfragen erlauben kaum Rückschlüsse darauf, welche Einkommensverteilung als gerecht angesehen wird und auf welcher Grundlage die Gerechtigkeitsbewertungen erfolgen.

In den Sozialwissenschaften wird deshalb seit einigen Jahren zunehmend ein Verfahren zur Messung von Normen und Wertvorstellungen angewandt, das diese Defizite klassischer Itemabfragen überwindet und eine weitaus differenziertere Abbildung individueller Einstellungen erlaubt: der *Faktorielle Survey* bzw. die *Vignettenanalyse*. Die Grundidee des Faktoriellen Surveys besteht darin, die Befragten hypothetische Personen-, Objekt- oder Situationsbeschreibungen – so genannte „Vignetten“ – beurteilen zu lassen. Es handelt sich also um Qua-

¹ Die Autoren danken Andy Donaubaue für wertvolle Kritik und Anmerkungen.

si-Experimente, bei denen es – im Gegensatz zu den üblichen Verhaltensexperimenten der experimentellen Ökonomie – für Teilnehmer nichts zu verdienen gibt.

Kennzeichen der Vignetten ist, dass in ihnen Merkmale (*Dimensionen*) unabhängig voneinander in ihren Ausprägungen (*Levels*) variieren (Beck und Opp 2001; Jasso 2006). Dies erlaubt es, die genaue Urteilsrelevanz dieser einzelnen Merkmale zu identifizieren – ziehen die Variationen der Merkmale entsprechende Variationen der Urteile nach sich? Die Bedeutung der unterschiedlichen Personen-, Objekt- oder Situationsmerkmale wird also nicht direkt und einzeln abgefragt, sondern die Befragten müssen sie in eine gemeinsame Bewertung integrieren. Die Relevanz der einzelnen Merkmale wird dann indirekt aus ihren Antworten rekonstruiert. Üblicherweise werden einzelnen Befragten gleich mehrere solcher Personen-, Objekt- oder Situationsbeschreibungen vorgelegt.

Als wesentliche Vorteile des Verfahrens gegenüber itembasierten Abfragen gelten:

- (1) Experimenteller Charakter des Designs. Die Befragten bewerten – wie dargelegt – Personen, Objekt- oder Situationsbeschreibungen, die in ihren Merkmalen (möglichst) unabhängig variieren. Damit wird ermöglicht, einen kausalen Effekt der Variation der einzelnen Merkmale auf die Reaktionen der Befragten zu ermitteln.
- (2) Präsentation einer möglichst realistischen Entscheidungs- oder Bewertungsaufgabe. In der Regel werden Urteile, Einschätzungen und Entscheidungen auf der Basis von mehreren Informationen abgegeben bzw. getroffen. Genau dies wird bei der Konstruktion eines Faktoriellen Survey bedacht, wenn die zu evaluierenden Beschreibungen mehrdimensional sind.

Die Bandbreite der bisherigen Studien, welche den Faktoriellen Survey verwenden, ist sehr weit gefächert. Neben der hier im Mittelpunkt stehenden Forschung zur Einkommensgerechtigkeit (Alves und Rossi 1978; Hermkens und Boerman 1989; Jasso 1978; Jasso und Rossi 1977; Jann 2003; Jasso 1994; Jasso und Webster 1997, 1999) gibt es Studien zu gerechter Besteuerung (Liebig und Mau 2005) oder Bestrafung (Berk und Rossi 1977; Miller et al. 1986). Weiterhin werden Normen und Werte (Jasso und Opp 1997; Beck und Opp 2001; Mäs et al. 2005) oder der Grad der Zufriedenheit (Kapteyn et al. 2008) thematisiert. Daneben wird in einigen Untersuchungen nach Entscheidungen in Paarhaushalten oder zur Vertrauenswürdigkeit gefragt (Auspurg und Abraham 2007; Auspurg et al. 2009b; Barrera und Buskens 2007). Bei den Teilnehmern handelt es sich in diesen Studien überwiegend um weitgehend homogene Spezialpopulationen wie Studierende.

Mit der Einschaltung eines entsprechenden Fragemoduls in den Pretest des Sozio-oekonomischen Panels (SOEP) 2008 [nachfolgend: SOEP-Pretest 2008] sollte die *methodische Fragestellung* überprüft werden, ob dieses Verfahren auch in allgemeinen Bevölkerungsumfragen eingesetzt werden kann und welche methodischen Implikationen sich dabei ergeben. Dies zu klären ist vor allem deshalb dringlich, weil Faktorielle Surveys bislang hauptsächlich in kleineren Labor- oder Feldstudien eingesetzt wurden (Jasso und Meyersson Milgrom 2008; Auspurg et al. 2008). Die Anwendung in größeren Bevölkerungsumfragen und insbesondere

die Akzeptanz bei Befragten und Interviewern wurden bislang noch nicht systematisch untersucht.

Das *inhaltliche Interesse* leitet sich aus dem eingangs skizzierten Defizit bisheriger Gerechtigkeitsforschung ab. Gerade die in vielen öffentlichen Debatten zur Einkommensungleichheit vorgenommene Gleichsetzung von Ungleichheit mit Ungerechtigkeit basiert auf einem spezifischen Gerechtigkeitsverständnis, das nicht notwendigerweise mit den in der Bevölkerung vertretenen Ansichten übereinstimmen muss. So spricht sich in internationalen Umfragen regelmäßig eine überwältigende Mehrheit an Befragten für eine Differenzierung von Einkommen nach der individuellen Leistung am Arbeitsplatz aus, was einer Gleichverteilung von Einkommen widerspricht (so etwa ISJP 1991, 1996, 2001, 2006). Vor diesem Hintergrund lautet die leitende inhaltliche Fragestellung dieser Studie: Welche Faktoren sind für ein gerechtes Einkommen maßgeblich? Gemäß der Humankapitaltheorie und Arbeitsmarktforschung führen individuelle, familiäre und arbeitsplatzbezogene Merkmale zu einer Differenzierung von Erwerbseinkommen. Individuelle Fähigkeiten und Bildungsabschlüsse gehen mit unterschiedlichen Einkommen einher, daneben beeinflussen familiäre Kontexte wie der Ehestand die Höhe des Verdienstes und zusätzlich existieren Lohnunterschiede zwischen Frauen und Männern, die zumindest theoretisch auf Diskriminierung zurückgeführt werden. Neben diesen personenbezogenen Eigenschaften spielen stärker arbeitsplatzbezogene Merkmale eine wichtige Rolle. So sind die Einkommen von Beschäftigten in größeren Betrieben höher als in kleineren Betrieben. Zugleich wird die ökonomische Situation von Betrieben oftmals als Rechtfertigung für Entlassungen herangezogen (Struck et al. 2006) und steht damit ebenso in engem Zusammenhang zur Entlohnung. In der vorliegenden Studie interessiert nun, ob diese faktischen Entlohnungsfaktoren mit dem Gerechtigkeitsempfinden in der Bevölkerung übereinstimmen.

Die Gliederung ist wie folgt: In Abschnitt 2 wird zunächst das Erhebungsinstrument des faktoriellen Surveys allgemein dargestellt. Abschnitt 3 beschreibt die konkrete Umsetzung im SOEP-Pretest 2008 sowie die Befragten- und Vignettenstichprobe. In Abschnitt 4 wird dann die Tauglichkeit des Instruments analysiert, wozu die Rückmeldungen der Befragten und Interviewer sowie das Antwortverhalten (Beantwortungsdauer, Nutzung der Antwortskala und Urteilkonsistenz) herangezogen werden. Abschnitt 5 beinhaltet ausgewählte inhaltliche Ergebnisse. Der Text schließt mit einer Zusammenfassung der wichtigsten methodischen Erkenntnisse dieser Studie.

2 Der Faktorielle Survey als Erhebungsinstrument in großen Umfragen

Die Konstruktion der hypothetischen Personen-, Objekt- und Situationsbeschreibungen ist der wichtigste Schritt bei der Planung eines Faktoriellen Survey. Je nach inhaltlicher Fragestellung sind die Dimensionen (Merkmale) der Vignetten und ihre Ausprägungen aus dem Stand der theoretischen Diskussion zu entwickeln (Alves 1982; Jasso 2006). Dabei ist große Sorgfalt angesagt, weil selbst relativ belanglos anmutende Entscheidungen (etwa zur Zahl der Ausprägungen bei metrischen Dimensionen) für die Konzeption und Auswertung beachtliche Konsequenzen haben. Die Hauptaufgabe bei der Auswahl von Dimensionen besteht darin, möglichst entscheidungsrelevante Merkmale zu finden. Darüber hinaus sollten die angegebenen Dimensionen die Objekte, Personen oder Situationen so genau abbilden, dass keine für die Urteilsaufgabe wesentliche Information fehlt.

Wir stehen vor der Aufgabe, Beschreibungen von Vollzeitbeschäftigten anzufertigen, die es ermöglichen, die Bruttoeinkommen der geschilderten Personen hinsichtlich ihrer Gerechtigkeit zu bewerten. Während einige Dimensionen (wie zum Beispiel das Geschlecht) bereits eine vorgegebene Anzahl möglicher Ausprägungen haben, muss bei anderen Dimensionen die Anzahl der Ausprägungen bestimmt werden. Die Dimension Alter kann beispielsweise mit vier Kategorien von 30 bis 60 Jahre variiert werden (30, 40, 50 und 60 Jahre), alternativ lassen sich aber ebenso sieben Kategorien einsetzen (30, 35, 40, 45, 50, 55, 60). Wie viele unterschiedliche Berufskategorien und Bildungsabschlüsse sind erforderlich? Welche Berufe werden ausgewählt? Bei der Festlegung der Ausprägungen muss bedacht werden, dass mit ihrer Anzahl ebenso die Menge der in den Auswertungen zu schätzenden Parameter einhergeht (Alves 1982; Jasso 2006). Außerdem steigt mit der Anzahl von Ausprägungen exponentiell die Grundgesamtheit möglicher Merkmalskombinationen.

Diese Grundgesamtheit bezeichnet man als *Vignettenuniversum*. Es berechnet sich als das kartesische Produkt der Ausprägungen aller Dimensionen. Drei Dimensionen mit jeweils fünf Ausprägungen ergeben beispielsweise ein Vignettenuniversum von $5 \cdot 5 \cdot 5 = 125$ Vignetten.² Im Regelfall ist das Vignettenuniversum zu groß, um es durch einzelne Befragte vollständig bewerten zu lassen. Deshalb werden Stichproben gebildet (ähnlich wie beim Matrix-Sampling; vgl. Thomas et al. 2006) und den Befragten vorgelegt. Man kann entweder für jeden Befragten eine eigene Auswahl zusammenstellen (eine solche Auswahl bezeichnet man als *Vignettenset*) oder, falls man mehrere Bewertungen der gleichen Vignetten gewinnen möchte, einige wenige Vignettensets ziehen und diese von jeweils mehreren Befragten beantworten lassen (Jasso 2007). Jedes einmalige Vignettenset wird als *Deck* bezeichnet. Es gibt

² In solch einem vollen faktoriellen Design ist die Korrelation zwischen den einzelnen Dimensionen null. In manchen Fällen sind Kombinationen jedoch sachlogisch nicht möglich (beispielsweise ein Arzt ohne Schulabschluss). Solche Kombinationen werden als unlogische oder unplausible Fälle bezeichnet und aus der Grundgesamtheit ausgeschlossen. Dadurch wird allerdings die Korrelation zwischen einzelnen Dimensionen größer null.

einen *trade off* zwischen beiden Vorgehensweisen: Im ersten Fall hat man das Vignettenuniversum stärker ausgeschöpft. Im letzten Fall kann das Antwortverhalten verschiedener Subgruppen von Befragten (zum Beispiel Männer und Frauen) besser verglichen werden, da sie die identischen Vignetten beantwortet haben.

Die Vignetten können per Zufall (Jasso 2006) oder durch eine Quotenauswahl (*fraktionalisierte Designs*) ausgesucht werden (Dulmer 2007; Steiner und Atzmüller 2006; Kuhfeld 2005; Kuhfeld et al. 1994). Unabhängig vom Stichprobenverfahren ist es gewünscht, möglichst unkorrelierte Dimensionen zu erhalten, da dies eine Grundvoraussetzung dafür ist, die Effekte einzelner Dimensionen unabhängig voneinander modellgestützt schätzen zu können. Nur wenn die Merkmale (möglichst) orthogonal zueinander stehen, lässt sich ihr Einfluss auf das Urteilsverhalten unabhängig von den Effekten anderer Attribute analysieren. Die Quotenauswahl ist diesbezüglich (und insbesondere bei kleinen Stichproben) dem Zufallsverfahren überlegen. Sie basiert auf einer systematischen Zusammenstellung der Vignetten mit dem expliziten Ziel, möglichst alle Kombinationen gleichmäßig abzudecken. Damit wird im Regelfall ein maximal „effizientes“ Design ermittelt. *Effizienz* bedeutet Unkorreliertheit sowie gleichzeitig größtmögliche Varianz der Dimensionen³. Neben den fraktionalisierten Quotenstichproben, die sich ausschließlich am Kriterium der Orthogonalität der Dimensionen orientieren, werden auch D-effiziente Stichproben gebildet. In diesen wird die Forderung nach Orthogonalität der Dimensionen etwas gelöst und zusätzlich versucht eine maximale Varianz der Attributausprägungen zu erreichen. Insbesondere in experimentellen Designs mit unplausiblen Kombinationen sind D-effiziente Designs sinnvoll. Umgesetzt wird die Quotenauswahl durch bestehende Versuchspläne in der Literatur (Addelman 1962b,a) oder mit Hilfe entsprechender Software (z.B. SAS, JMP, Sawtooth Software). Die ausgewählten Vignetten werden im Idealfall wiederum fraktionalisiert auf die einzelnen Decks verteilt, um auch innerhalb der Decks möglichst unabhängig variierende Merkmale präsentieren zu können.

Die so zusammengestellten Vignetten werden dann den Befragten vorgelegt. Diese sollen alle Vignetten der ihnen zugewiesenen Vignettensets beantworten. Zur Vermeidung von Reihenfolgeeffekten ist die Reihenfolge der einzelnen Vignetten pro Befragten möglichst jeweils zufällig zu variieren.

Ein letzter wichtiger Planungsschritt ist die Festlegung der Antwortskala. Häufig werden Ratingskalen mit fünf bis 15 Kategorien verwendet (Mäs et al. 2005; Schulte 2002; Thurman et al. 1988; Dülmer 2001), in einigen Studien werden aber auch Magnitudeskalen (Liebig und Mau 2002) eingesetzt.

³ Beides sind wiederum Voraussetzungen für eine möglichst effiziente Schätzbarkeit des Einflusses der Dimensionen auf die Urteile.

3 Konstruktion des Faktoriellen Surveys für den SOEP-Pretest 2008

In der vorliegenden Studie geht es inhaltlich darum, fiktive Erwerbspersonen im Hinblick auf die Gerechtigkeit ihres Bruttoeinkommens zu bewerten. Die Annahme ist, dass die Höhe eines gerechten Einkommens von den bereits eingangs erwähnten Faktoren abhängt (siehe dazu auch nachfolgenden Abschnitt 3.1). Dementsprechend werden fiktive Einkommensbezieher anhand von zehn Dimensionen beschrieben.

Insgesamt wurden den einzelnen Befragten des SOEP-Pretest 2008 jeweils 24 Beschreibungen erwerbstätiger Personen mit einer vereinbarten Wochenarbeitszeit von 40 Stunden vorgestellt.⁴ Die Aufgabe bestand darin, in einem mehrstufigen Bewertungsprozess für jede der vorgestellten, Vollzeit abhängig beschäftigten Personen die Gerechtigkeit des angegebenen Bruttoeinkommens einzuschätzen, und gegebenenfalls (bei einer Bewertung als ungerecht) auch das jeweilige Ausmaß an Ungerechtigkeit anzugeben.

3.1 Entwicklung der Vignettendimensionen und Ausprägungen

Da der inhaltliche Schwerpunkt der Befragung auf der Einkommensgerechtigkeit liegt, wurden solche Dimensionen ausgewählt, welche relevante Faktoren bei der Beurteilung von Einkommen darstellen (vgl. Tabelle 1). Orientierungspunkte waren dabei hauptsächlich vorliegende Vignettenstudien zu Einkommensgerechtigkeit (Jann 2003; Jasso 1978; Jasso und Webster 1999; Jasso und Rossi 1977; Jasso und Webster 1997; Alves 1982; Alves und Rossi 1978). Die Dimensionen Alter, Geschlecht, Anzahl an Kindern, Beruf und Ausbildungsabschluss haben in diesen Studien allesamt einen signifikanten Einfluss auf die Bewertung des Bruttoeinkommens. Ergänzt wurden diese Merkmale durch weitere Dimensionen, die entweder a) in der Gerechtigkeitsforschung als bedeutsam gelten: Leistung und Familienstand (Liebig und Schupp 2005,2008b,a; Struck et al. 2006), und/oder b) für die reale Entlohnung relevant sind: Betriebsgröße und ökonomische Situation des Betriebs (Abraham und Hinz 2005b; Abraham und Hinz 2005a).

⁴ Allen Befragten wurde zusätzlich eine „erste“ identische Beispielvignette präsentiert, die der Interviewer mit dem Befragten gemeinsam durchging. Der Text dieser Vignette lautete: „Ein 35 jähriger, allein stehender Mann hat eine abgeschlossene Berufsausbildung und arbeitet derzeit als Friseur in einem Kleinbetrieb, der hohe Gewinne macht. Er erbringt an seinem Arbeitsplatz überdurchschnittliche Leistungen und verdient im Monat 350 Euro brutto. Ist das monatliche Brutto-Einkommen für diese Person Ihrer Meinung nach gerecht oder ungerecht?“ Außerdem wurde nach dem Vignettenmodul mit 24 Vignetten noch eine weitere Vignette mit zwei zusätzlichen Dimensionen (Nationalität, Aufenthaltsdauer) getestet. Wir beschränken uns in diesem Methodenbericht auf den Kern des Vignettenmoduls, die 24 Vignetten mit zehn Dimensionen.

Tabelle1

Dimensionen und Ausprägungen des Vignettenmoduls

Dimension	Ausprägungen
Alter	25/ 35/ 45/ 55 Jahre
Geschlecht	Mann/ Frau
Familienstand	Alleinverdiener verheiratet/ Zweiverdiener verheiratet/ alleinstehend
Ausbildungsabschluss	ohne Berufsabschluss/ mit abgeschlossener Berufsausbildung/ mit Hochschulabschluss
Beruf	Hilfsarbeiter(in)/ Pförtner(in)/ Lokführer(in)/ Verwaltungsfachkraft/ Friseur(in)/ Sozialarbeiter(in)/ Programmierer(in)/ Elektroingenieur(in)/ Leitende(r) Manager(in)/ Arzt, Ärztin
Einkommen	500€/ 950€/ 1200€/ 1500€/ 2500€/ 3800€/ 5400€/ 6800€/ 10000€/ 15000€ ⁵
Kinder	kein Kind/ 1 Kind/ 2 Kinder/ 3 Kinder/ 4 Kinder
Leistung	unter-/ über-/ durchschnittlich
Ökonomische Situation des Betriebs	hohe Gewinne/ vom Konkurs bedroht/ wirtschaftlich stabil
Betriebsgröße	Klein-/ Mittel-/ Großbetrieb

3.2 *Bildung des Vignettenuniversums und Stichprobenziehung*

Das Vignettenuniversum (d.h. das Universum aller möglichen Merkmalskombinationen) wird wie geschildert – durch vollständige Kreuzung aller Dimensionen erstellt. In dieser Studie umfasst es 972.000 verschiedene Vignetten. Bei der Auswahl der Vignetten für die Befragten wurden aus sachlogischen Überlegungen heraus die folgenden unplausiblen Fälle ausgeschlossen:

- Bruttoeinkommen von über 3800 Euro für Hilfsarbeiter
- Bruttoeinkommen von über 5400 Euro für Pförtner und Lokführer
- Bruttoeinkommen von über 6800 Euro für Verwaltungsfachkräfte, Friseure und Sozialarbeiter
- Bruttoeinkommen unter 1200 Euro für Elektroingenieure
- Bruttoeinkommen unter 2500 Euro für leitende Manager und Ärzte.

⁵ Die Stufen orientieren sich an den Perzentilen der Einkommensverteilung des SOEP 2007 für Vollzeit Beschäftigte. Die obere und untere Kategorie wurden als Extremwerte ergänzt.

Ausgeschlossene unrealistische Fälle aus der Kombination von Ausbildungsstand und Beruf sind:

- Elektroingenieure ohne Berufsabschluss
- Ärzte ohne Hochschulabschluss.

Die Stichprobenauswahl erfolgte über ein fraktionalisiertes Design, indem die D-Effizienz unter Beachtung unplausibler, daher zu vermeidender Kombinationen, maximiert wurde (Kuhfeld 2005; Kuhfeld et al. 1994). Das Zielkriterium der D-Effizienz umfasst, wie erwähnt, neben einem möglichst orthogonalen Design zugleich auch eine größtmögliche Varianz der Ausprägungen. Für die 24 Vignetten mit zehn Dimensionen wurden zunächst 240 Vignetten mit einer D-Effizienz von *über 90* ausgewählt, welche anschließend auf zehn Decks fraktionalisiert wurden.⁶

3.3 Bewertungsmodus und Präsentation der Vignetten

Zur Bewertung der Vignetten wurde ein dreistufiges, geschlossenes Antwortverfahren eingesetzt. Im ersten Schritt beantworteten die Befragten, ob (1) sie das angegebene Einkommen der beschriebenen Person als gerecht oder ungerecht empfinden. Falls sie das Einkommen als gerecht bewerteten, wurde ihnen die nächste Vignette vorgelegt.

Falls sie das Einkommen als ungerecht ansahen, gaben sie in zwei weiteren Schritten an, ob (2) das Einkommen ungerechterweise zu hoch oder zu niedrig ist und (3) wie hoch das Ausmaß an empfundener Ungerechtigkeit ist. Dieses Ausmaß durften die Befragten dabei mit einer frei wählbaren Zahl von 1 bis 100 ausdrücken.

Dieser weite Zahlenraum sollte den Befragten mehr Spielraum bieten als die sonst üblichen Ratingskalen mit fünf bis elf Ausprägungen. In Abbildung 2 ist eine Beispielvignette, bestehend aus den zehn Dimensionen, zusammen mit den einzelnen Bewertungsschritten dargestellt.

Die Vignetten wurden den Befragten auf einem Computerbildschirm präsentiert⁷, wobei die Befragten ihre Einschätzungen selbst in den Laptop des Interviewers eingaben. Es handelte sich also um eine CAPI-Befragung (Computer Assisted Personal Interview). Ein einführender Text erläuterte die Verwendung der Skala und wies darauf hin, dass 24 Vignetten zu beantworten sind. Anschließend wurde anhand einer Übungsvignette das Vorgehen geprobt, mit der Möglichkeit, Fragen an den Interviewer zu stellen. Die Befragten wurden dann zufällig auf eines der Decks mit jeweils 24 Vignetten zugewiesen. Da die Vignetten der einzelnen Decks in einer festen Reihenfolge vorgegeben wurden, sind bei weitergehenden Analysen mögliche Reihenfolgeeffekte zu berücksichtigen. Das Instrument war so programmiert, dass

⁶ Die maximal mögliche D-Effizienz liegt in einem symmetrischen und orthogonalen Design bei 100. Dieser Wert wird als Referenz für die Gütebewertung des Designs herangezogen. Eine Effizienz von über 90 gilt als zufriedenstellend.

⁷ Wir danken Andreas Stocker, TNS Infratest Sozialforschung, München, für die Sorgfalt bei der Umsetzung und Programmierung der Vignettenvorlagen in die computerunterstützte Fragebogenfassung.

das Überspringen einer Vignette ausgeschlossen war.⁸ Die Befragten mussten alle Vignetten bewerten. Es gab auch keinen „Fortschrittsbalken“ oder ähnliches.

Abbildung 1: Beispielvignette

Eine 45 jährige, verheiratete Frau mit zwei Kindern,
deren Mann kein eigenes Einkommen bezieht,
hat eine abgeschlossene Berufsausbildung und
arbeitet derzeit als Friseurin in einem Großbetrieb, der kurz vor dem Konkurs steht.
Sie erbringt an ihrem Arbeitsplatz eher unterdurchschnittliche Leistungen
und verdient im Monat **1200 Euro brutto**, d.h. vor Abzug von Steuern und Abgaben.

Ihre Bewertung:

F 1:
Ist das monatliche Brutto-Einkommen für diese Person Ihrer Meinung nach gerecht oder ungerecht?

- Einkommen ist **gerecht** (→ Weiter mit der nächsten Personenbeschreibung)
- Einkommen ist **ungerecht** (→ Weiter mit F 2)

F 2:
Ist das Einkommen ungerechterweise zu hoch oder zu niedrig?

- Ungerechterweise **zu hoch** (→ Weiter mit F 3)
- Ungerechterweise **zu niedrig** (→ Weiter mit F 3)

F 3:
Welche Zahl zwischen 1 und 100 drückt das Ausmaß an Ungerechtigkeit, das Sie persönlich empfinden, am besten aus?

3.4 Befragten- und Vignettenstichprobe (SOEP-Pretest 2008)

Das Frageprogramm des SOEP für die nächsten Erhebungswellen wird immer wieder durch Pretests vorbereitet. Vorrangiges Ziel ist dabei, neu aufgenommene Fragenkomplexe oder Modifikationen von Fragen zu erproben. Der SOEP-Pretest geht seit mehreren Jahren deutlich über das bei Befragungen übliche Format hinaus. Seit 2002 werden in seinem Kontext eigenständige Bevölkerungsbefragungen mit etwa 1.000 Befragten und einem für wissenschaftliche Forschungszwecke voll auswertbaren Datenbestand durchgeführt (Siegel et al. 2009). Die Stichprobe ist repräsentativ für die Wohnbevölkerung Deutschlands ab 16 Jahren angelegt. Die Durchführung obliegt TNS Infratest Sozialforschung.

⁸ Dies ist sicherlich ungewöhnlich, gerade wenn man die Akzeptanz einer Befragungsart untersuchen will. Allerdings lassen sich Verweigerungen auch über extrem kurze Antwortzeiten rekonstruieren (vgl. Abschnitt 5).

Die von TNS Infratest Sozialforschung durchgeführten Pretest-Befragungen erfolgten in den letzten Jahren durchgängig in Form von CAPI-Interviews (während in der Haupterhebung der größere Teil der Interviews mit herkömmlichen Papierfragebögen durchgeführt wird). Damit bestehen erheblich verbesserte Möglichkeiten zum Einsatz experimenteller Testverfahren.⁹

Während in den Haupterhebungen des SOEP alle Personen ab 16 Jahren in den Teilnehmerhaushalten befragt werden, ist der SOEP-Pretest einfacher angelegt. Es gibt *einen* Fragebogen, der von *einer* Person im Haushalt beantwortet wird. Dabei handelt sich um eine von der Hauptbefragung unabhängige Pretest-Stichprobe, das heißt, es sind keine Personen aus der Hauptbefragung enthalten. Die Befragung des SOEP-Pretest 2008 wurde im Zeitraum 1. August bis 31. August 2008 durchgeführt. Die Interviewdauer war auf 45 Minuten angelegt und der realisierte Medianwert liegt auch exakt bei diesen 45 Minuten. Insgesamt wurden 1.066 Personen befragt.

Die von TNS Infratest ausgewiesene Ausschöpfungsquote liegt bei etwas unter 50 Prozent (Siegel et al. 2009). Dabei ist zu beachten, dass bei dem zum Einsatz gekommenen Random-Route-Verfahren weniger strikte Vorgaben an die Interviewer definiert werden, als bei einem Random-Route mit getrenntem Adressvorlauf. Die Ausschöpfungsquote ist nur sehr bedingt mit Antwortraten „härterer“ Verfahren wie einer Einwohnermeldeamtsstichprobe vergleichbar.

Die realisierte Stichprobe wurde einer Gewichtung nach regionalen und demographischen Verteilungen unterzogen. Damit wird gewährleistet, dass die gewichtete Stichprobe mit wesentlichen Strukturdaten der Grundgesamtheit der amtlichen Bevölkerungsstatistik übereinstimmt. Tabelle A1 (im Anhang) enthält wichtige Strukturinformationen der realisierten Stichprobe, vor und nach der Gewichtung (Personengewichtungsfaktor). Zu beachten ist, dass die Auswertungen dieses Berichts ausschließlich auf ungewichteten Daten basieren.

Im Vignettenmodul verteilen sich die Befragten trotz der Zufallszuweisung relativ gleichmäßig auf die unterschiedlichen Decks (Tabelle 2). Die Zahl der pro Deck realisierten Befragten schwankt zwischen 96 (Decks 7 und 2) und 127 Personen (Deck 9).¹⁰ Die Korrelationen zwischen den einzelnen Dimensionen sind insgesamt und auch auf der Ebene der einzelnen Decks gering (anders gesagt: das Design ist also tatsächlich sehr effizient). Eine Übersicht hierzu findet sich im Anhang (Tabelle A2).

⁹ Im Jahr 2008 lag der Schwerpunkt bei der Erprobung innovativer Befragungskonzepte und -instrumente im Wesentlichen auf vier Themenbereichen: (1) Subjektive Bewertung von Einkommensgerechtigkeit, konkret die Bewertung von Erwerbseinkommen unter Einsatz von Vignetten (faktorieller Survey), (2) Alltagsstimmung: Selbsteinschätzung der Befragten nach der Verteilung von „Stimmungslagen“ in einer „typischen Woche“ (habituelle Stimmung), (3) Fragen zu „Charakterstärken“, wobei es sich um eine erstmals eingesetzte Übersetzung und Modifizierung der deutschen Fassung der *Values in Action (VIA)* – „*Classification of Strengths*“ handelt, (4) Neue Fragen zur Erfassung von (chronischen) Krankheiten (vgl. Siegel et al. 2009).

¹⁰ Die Schwankungen weichen nicht vom Zufall ab (Chi-Quadrat = 9,0; df = 9; p = 0,436).

Tabelle 2: Häufigkeit der Decks

Deck	Häufigkeit
1	110
2	96
3	99
4	104
5	102
6	121
7	96
8	108
9	127
10	103

4 Methodische Ergebnisse

In diesem Abschnitt wird analysiert, ob der Faktorielle Survey im Rahmen allgemeiner Bevölkerungsumfragen eingesetzt werden kann. Dafür werden drei Informationsquellen herangezogen: (1) Befragteindrücke, (2) Interviewereindrücke und (3) das Antwortverhalten der Befragten. Nach einer kurzen Darstellung der offen abgefragten Befragteindrücke folgen tieferegehende Auswertungen der Interviewer-Einschätzungen und des Antwortverhaltens. Hinsichtlich der Interviewereindrücke wird die Bearbeitung des Vignettenmoduls mit der des gesamten Fragebogens verglichen. Um festzustellen, ob verschiedene Befragtengruppen gleichsweise gut oder schlecht mit dem Vignettenmodul umgehen können, werden drei Alters- (16 bis 39 Jahre, 40 bis 65 Jahre, über 65 Jahre)¹¹ und Bildungsgruppen (Schulabschluss: Hauptschule, Realschule, Abitur) gebildet und ihr Antwortverhalten jeweils miteinander verglichen.

4.1 Befragten- und Interviewereindrücke

Die Befragten konnten Anmerkungen und Kritik am Vignettenmodul in einer offenen Antwort nennen. Die meisten Befragten (81 Prozent) haben diese Möglichkeit nicht genutzt, wobei man nicht weiß, ob sie keine Kritik zu äußern hatten oder die Mühe einer offenen Antwort scheuten. Von 191 Befragten liegt ein Kommentar vor. Die häufigsten Nennungen sind in Tabelle 3 dargestellt.

Von den Befragten, die einen Kommentar abgegeben haben, stören sich 36 Prozent an teilweise unrealistischen Fällen. Knapp 35 Prozent fanden das Modul zu lang; in Bezug auf alle Befragten waren dies etwa sechs Prozent. Verständnisprobleme hatten zwölf Prozent. Schwierigkeiten mit der Zuordnung der Einkommen als gerecht oder ungerecht äußerten gut neun Prozent.

¹¹ Die Gruppen werden so festgelegt, damit einerseits die Besetzung der Gruppen gleichmäßig ist und andererseits die Intervalle eine ähnliche Anzahl von Lebensjahren (ca. 25) umfassen. Während die äußeren Kategorien etwa gleich besetzt sind (303 und 325 Befragte), befinden sich in der Mittelkategorie etwas mehr Befragte (438).

Tabelle 3: Kommentare

Inhalt der Anmerkungen	Anteil der Nennungen in Prozent	Anteil an der Gesamtstichprobe in Prozent
Formulierung der Vignetten ungünstig oder unrealistisch	36,2	6,9
Fragenset zu lang	34,7	6,1
Verständnisprobleme	11,7	2,0
Zuordnung zu gerecht und ungerecht problematisch	9,2	1,6
Sonstiges	8,2	1,4
N (Anzahl abgegebener Kommentare)	191	1066

Die Verständlichkeit des Vignettenmoduls für die Befragten wurde darüber hinaus von den Interviewern bewertet.¹² Dabei spielt natürlich die subjektive Einschätzung der Interviewer eine große und schwer zu kontrollierende Rolle. Dennoch sind solche Abfragen sinnvoll, um eventuell schwerwiegende Probleme von Befragungsmodulen aufzudecken. Über 80 Prozent der Befragten verstehen demnach die Aufgabe sehr gut oder gut (Tabelle 4). Für den gesamten Fragebogen liegt die Verständlichkeit für etwa 90 Prozent der Befragten in einem guten oder sehr guten Bereich. Diese Differenz von zehn Prozentpunkten zeigt, dass es sich bei den Vignetten um ein vergleichsweise komplexes Fragebogenmodul handelt. Die Verständnisschwierigkeiten übersteigen – wie weitere, nicht dargestellte Auswertungen belegen – allerdings keinesfalls das Ausmaß, das in anderen komplexeren Modulen des SOEP-Pretest 2008 erreicht wird.

Tabelle 4: **Verständlichkeit** (Angaben in Prozent)

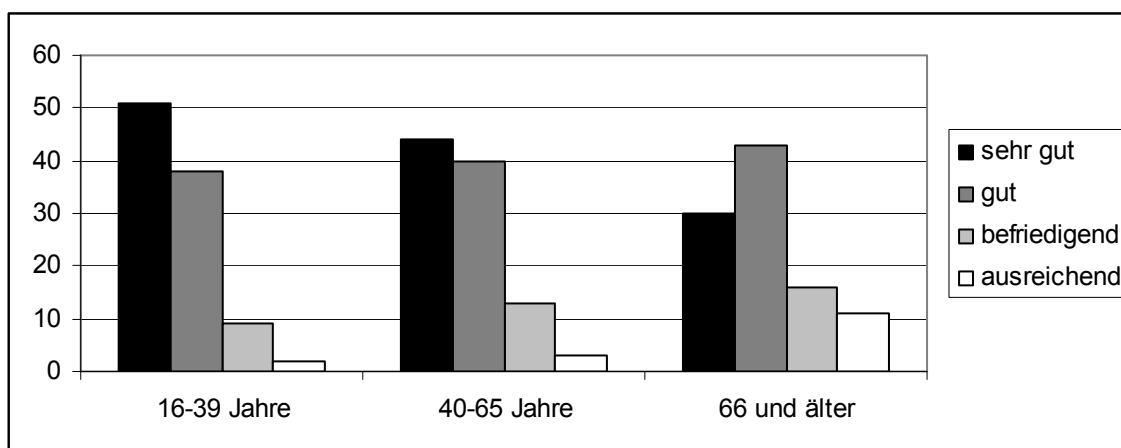
Bewertung	Vignetten	Gesamter Fragebogen
sehr gut	41,74	51,13
gut	40,34	40,15
befriedigend	12,95	6,75
ausreichend	3,19	1,41
mangelhaft	0,75	0,28
ungenügend	1,03	0,28
Total	1.066	1.066

Die Aufschlüsselung nach *Altersgruppen* ergibt Unterschiede im Verständnis der Vignetten zwischen der jüngsten Befragtengruppe der 16- bis 39-Jährigen und den Personen ab 66 Jah-

¹² Die Frageformulierung an die Interviewer lautete: „Bitte geben Sie uns zur letzten Frage bzw. Fragengruppe „Einkommensgerechtigkeit“ an, wie verständlich die Frage für die Zielperson formuliert ist, und wie Sie die Antwortbereitschaft der Zielperson einschätzen würden.“ (Geschlossene Kategorien: sehr gut, gut, befriedigend, ausreichend, mangelhaft, ungenügend).

ren (Abbildung 2). Über 50 Prozent der jüngeren Befragten haben die Aufgabe aus Sicht der Interviewer sehr gut verstanden, im Vergleich zu 30 Prozent der Älteren. Allerdings haben über 40 Prozent der älteren Befragten die Aufgabe immerhin gut verstanden. Zehn Prozent der Befragten über 65 Jahre haben das Vignettenmodul lediglich ausreichend, mangelhaft oder ungenügend begriffen. Ein vergleichender (nicht dargestellter) Blick auf den gesamten Fragebogen und andere Fragebogenmodule zeigt, dass diese Unterschiede zwischen den Altersgruppen keine Besonderheit des Vignettenmoduls darstellen.

Abbildung 2: Verständlichkeit nach Altersgruppen (Angaben in Prozent)

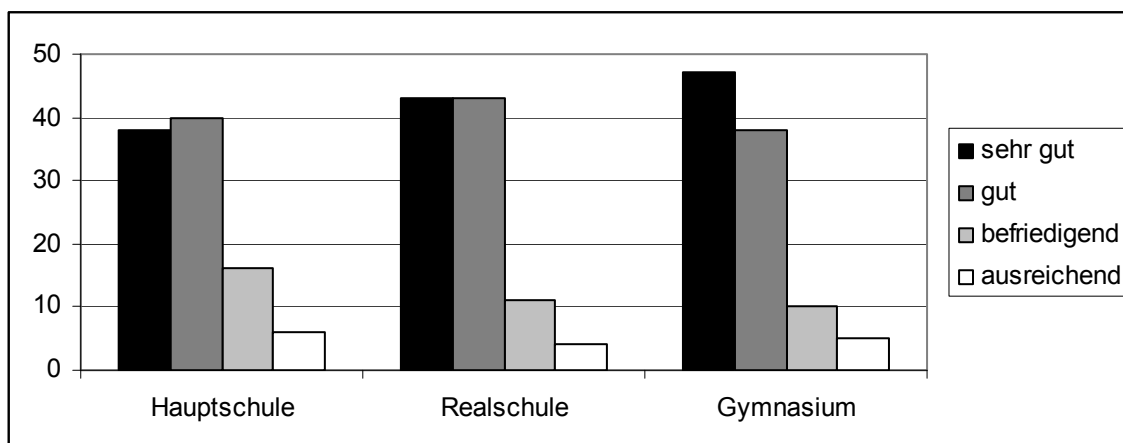


Anmerkung: Die Kategorien *ausreichend*, *mangelhaft* und *ungenügend* sind hier zusammengefasst.

Betrachtet man die drei *Bildungsgruppen*, sind ebenfalls Unterschiede festzustellen (Abbildung 3). Befragte mit Abitur verstehen die im Vignettenmodul verlangte Bewertungsaufgabe nach Einschätzung der Interviewer in fast 50 Prozent der Fälle sehr gut und in knapp 40 Prozent gut. Die Gruppe der Personen mit Realschulabschluss kommt dicht dahinter, im sehr guten und guten Bereich liegen hier ebenfalls fast 90 Prozent. Befragte mit Hauptschulabschluss haben in knapp 80 Prozent der Fälle ein zumindest gutes Verständnis der Aufgabe. Damit fallen die Unterschiede zwischen den Bildungsgruppen geringer aus als die zwischen den Altersgruppen.¹³ Bemerkenswert ist, dass sich die Muster nach Bildungsgruppen ähnlich für den gesamten Fragebogen zeigen (hier nicht dargestellte Analysen). Wiederum finden sich also keine Anzeichen dafür, dass das Vignettenmodul einzelnen Befragtengruppen verstärkt Schwierigkeiten bereitet. Dies ist deshalb wichtig, weil alters- oder bildungsabhängige Methodeneffekte leicht zu inhaltlichen Fehlinterpretationen verleiten (unentdeckte Alters- oder Bildungsunterschiede könnten fälschlicherweise inhaltlich interpretiert werden, siehe z.B. Schwarz und Knäuper 2006).

¹³ Dies mag mitunter daran liegen, dass das Alter mit der Einbindung in den Arbeitsmarkt kovariert, Rentner und Pensionisten haben möglicherweise auch aufgrund ihrer stärkeren „Arbeitsmarktfremde“ größere Bewertungsschwierigkeiten bei den Einkommensvignetten als jüngere Befragte.

Abbildung 3: Verständlichkeit nach Bildung (Angaben in Prozent)



Anmerkung: Die Kategorien *ausreichend*, *mangelhaft* und *ungenügend* sind hier zusammengefasst.

Das zweite Kriterium, nach dem die Interviewer den Befragten einschätzen sollten, ist die Antwortbereitschaft. Diese bezieht sich nicht auf die generelle Teilnahmebereitschaft am Interview, sondern ausschließlich auf die Situation im Interview selbst (vgl. Fußnote 11). Aus Tabelle 5 wird ersichtlich, dass aus Sicht der Interviewer das Vignettenmodul von einem Großteil der Befragten bereitwillig bearbeitet wurde. Im Vergleich zum gesamten Fragebogen zeigt sich eine etwas geringere Antwortbereitschaft bei den Vignetten.

Tabelle 5: **Antwortbereitschaft** (Angaben in Prozent)

Bewertung	Vignetten	Gesamter Fragebogen
sehr gut	44,18	51,69
gut	37,52	37,43
befriedigend	11,35	7,41
ausreichend	5,35	2,53
mangelhaft	1,22	0,66
ungenügend	0,38	0,28
Total	1.066	1.066

Dabei fällt die Antwortbereitschaft in den drei Altersgruppen unterschiedlich aus. Während in der jüngsten Befragtengruppe die Antwortbereitschaft in fast 90 Prozent der Interviews mindestens als gut eingestuft wird, gilt dies nur für knapp über 70 Prozent der Befragten ab dem Alter von 66 Jahren (Tabelle 6). Auch dies ist kein vignettentypisches Problem.

Tabelle 6: **Antwortbereitschaft im Vignettenteil nach Altersgruppen** (Angaben in Prozent)

Bewertung	Altersgruppe (Jahre)		
	16-39	40-65	66-
sehr gut	52,48	46,58	33,23
gut	37,29	36,76	38,77
befriedigend	6,60	11,19	16,00
ausreichend	3,30	4,34	8,62
mangelhaft	0,00	0,68	3,08
ungenügend	0,33	0,46	0,31
Total	303	438	325

Die Differenzierung nach Schulabschluss deckt vor allem Unterschiede zwischen Personen mit Hauptschulabschluss gegenüber den anderen beiden Bildungsgruppen auf (Tabelle 7). Die Teilnahmebereitschaft liegt aber bei 78 Prozent der Befragten mit Hauptschulabschluss noch mindestens im guten Bereich, im Vergleich zu 87 und 83 Prozent bei den Befragten mit Realschulabschluss oder Abitur. Wiederum finden sich sehr ähnliche Bildungseffekte für den gesamten Fragebogen und die anderen Fragebogenmodule.

Tabelle 7: **Antwortbereitschaft im Vignettenteil nach Schulabschluss** (Angaben in Prozent)

Bewertung	Schulabschluss		
	Hauptschule	Realschule	Abitur
sehr gut	40,98	45,32	48,98
gut	36,68	41,69	33,47
befriedigend	13,32	9,06	10,61
ausreichend	6,76	3,93	4,49
mangelhaft	1,84	0,00	1,63
ungenügend	0,41	0,00	0,82
Total	488	331	245

4.2 Antwortverhalten

Das Antwortverhalten der Befragten lässt detaillierte Rückschlüsse auf die Tauglichkeit des Verfahrens zu. Die zentralen Parameter der hier vorgestellten Analysen sind die Befragungsdauer, die Nutzung der Antwortskala und die Konsistenz der Urteile.

4.2.1 Befragungsdauer

Die Befragungsdauer liegt nur für das komplette Vignettenmodul vor. Die Bearbeitungszeit einzelner Vignetten wurde nicht erfasst. Die Analyse prozessproduzierter Befragungsdaten ist nicht unproblematisch, da diese oft wichtige Kontextinformationen vernachlässigen, etwa ob die Befragten während des Vignettenmoduls die Befragung unterbrochen haben. Dennoch liefert die Verteilung der Befragungsdauer wichtige Hinweise – etwa zu *faktischen* Verweigerungen. Die Umsetzung der CAPI-Programmierung schloss „Abbrecher“ oder „Verweigerer“ aus, da es (abgesehen von Totalabbrüchen, die nicht auftraten) innerhalb des Vignettenmoduls keine Möglichkeit zur Antwortverweigerung gab (vgl. oben Abschnitt 3.3). Allerdings ist beispielsweise die gemessene Bearbeitungszeit von 20 Sekunden für das komplette Vignettenmodul als Quasi-Verweigerung zu interpretieren. Etwa fünf Prozent der Befragten benötigen zur Bearbeitung des Vignettenmoduls weniger als 3,5 Minuten, was einer durchschnittlichen Vignettenbewertungszeit von maximal acht Sekunden entspricht. Am anderen Ende der Verteilung sind die zwei herausstechenden Werte von 137 und 139 Minuten (damit durchschnittlich gut fünf Minuten pro Vignette) möglicherweise durch nicht gemessene Unterbrechungen bspw. durch angeregte Diskussionen mit dem Interviewer nach oben verzerrt. Im Mittel benötigen die Befragten für die Beantwortung des Vignettenmoduls (also die 24 Vignetten und die Beispielvignette zusammen) insgesamt 13,5 Minuten, der Median liegt bei 12,4 Minuten. Tabelle 8 informiert über wichtige Kennzeichen der Verteilung dieser Befragungsdauer. Zu erwähnen ist noch, dass das Vignettenmodul im Durchschnitt nach 25 Minuten Befragungszeit begonnen wurde.

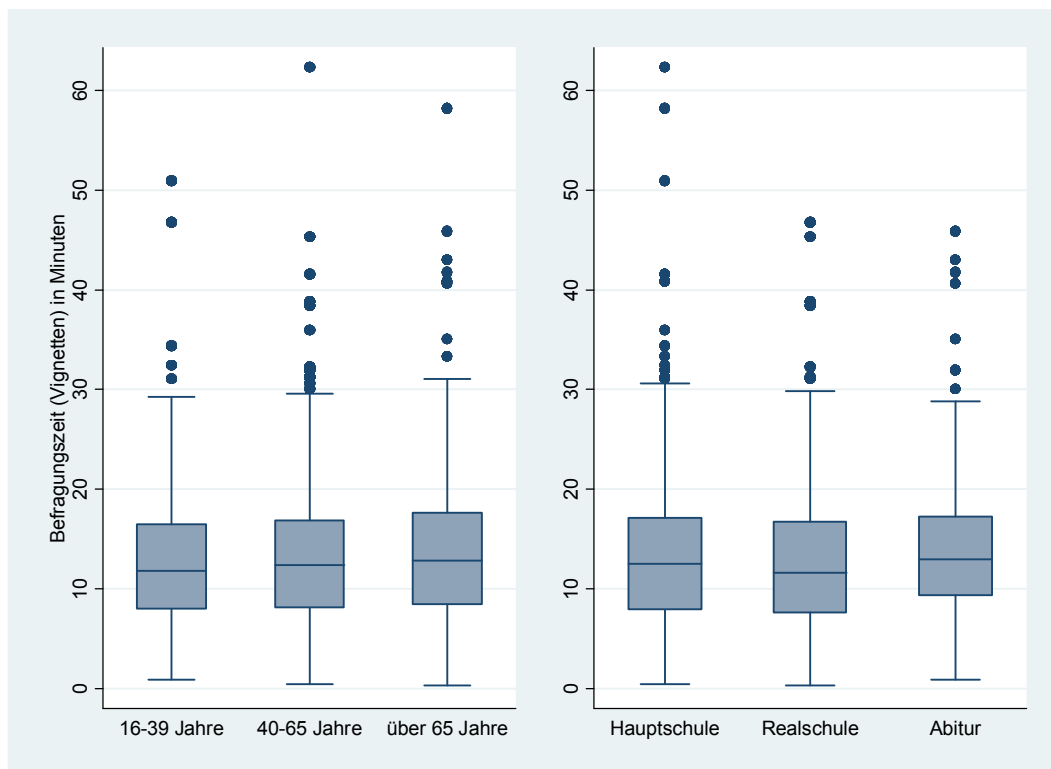
Tabelle 8: Dauer der Befragung: arithmetisches Mittel und 5-Punkte-Statistik (in Minuten)

	N	Mittelwert	S.D.	Min	,25	,5	,75	Max
Vignettenmodul	1.066	13,52	9,26	0,30	8,10	12,38	17,02	138,97
Fragebogen	1.063	50,65	24,69	18,97	36,68	45,27	57,92	341,22

In Abbildung 4 sind die Verteilungen der Befragungszeit in Abhängigkeit von Alter (linke Boxplots) und Bildung der Befragten (rechte Boxplots) dargestellt. Aussagekräftige Vergleiche ermöglichen vor allem die gegenüber Extremwerten bzw. „Ausreißern“ recht robusten Mediane (Mittellinie der Plots) sowie 25- und 75-Prozentperzentile (obere und untere Begrenzungen). Man erkennt keine dramatischen Unterschiede, die Werte liegen bei allen Gruppen in etwa gleich auf, lediglich Befragte mit höherer Schulbildung weisen eine geringfügig höhere Bearbeitungsdauer auf. Ebenso benötigen ältere Befragte im Hinblick auf den Median eine Minute länger als jüngere Befragte. Wiederum finden sich diese Unterschiede für die anderen Befragungsmodule in ähnlicher Weise.

Ziehen wir ein Zwischenfazit: Die Analyse der Bearbeitungszeit zeigt soweit, dass sich die Vignetten relativ zügig beantworten lassen, es werden ungefähr 30 Sekunden pro Vignette benötigt (Medianwert). Die Unterschiede zwischen Alters- und Bildungsgruppen sind gering. Die Bearbeitung ist also für alle Befragten in einer ähnlichen Zeitspanne machbar.

Abbildung 4: Verteilung der Befragungszeiten für das Vignettenmodul nach Bildungs- und Altersgruppen (in Minuten)



Anmerkung: Die beiden besonders großen Ausreißer (137 und 139 Minuten Bearbeitungszeit) sind nicht in der Abbildung enthalten.

4.2.2 Nutzung der Skala

Wie nutzen die Befragten die angebotene Skala, die sich ja aus einer dreistufigen Fragefolge ergibt? Der Wertebereich von [-100; -1] umfasst Urteile, bei denen das präsentierte Einkommen als ungerechterweise zu gering eingeschätzt wird. Die Null zeigt an, dass das in der Vignette vorgegebene Einkommen als gerecht empfunden wird. Der Wertebereich von [1; +100] enthält hingegen Urteile, bei denen die Befragten das präsentierte Einkommen als ungerechterweise zu hoch einschätzten. Fasst man die Urteile zunächst zu diesen drei groben Kategorien zusammen (zu gering, gerecht, zu hoch), sind diese ungefähr gleich besetzt (Tabelle 9).

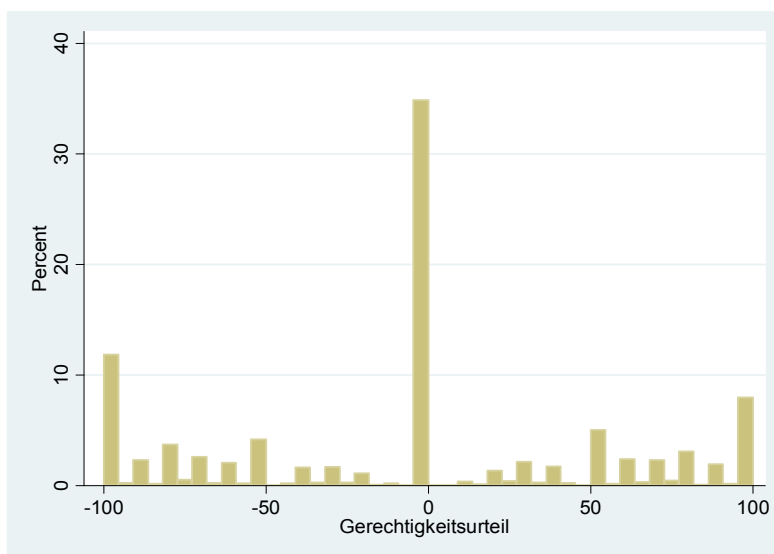
Tabelle 9: Verteilung der Variable „Gerechtigkeitsurteil“

Verteilung	Urteile (N)
Einkommen zu gering	8.759
Einkommen gerecht	8.897
Einkommen zu hoch	7.928
Total	25.584

Es ist also eine relative Häufung gerechter Urteile festzustellen, die vermutlich auf den Abfragemodus zurückgeht. Außerdem sind durch die erzwungenen Urteile potentielle Verweigerer enthalten (siehe Abschnitte 3.3 und 4.2.1). Wie sehr die Null („gerecht“) andere Ska-

lenwerte dominiert, ist in Abbildung 5 ersichtlich. Die Ballung an den Rändern der Skala deutet auf einen „Ceiling“-Effekt (Deckeneffekt) hin, vor allem im negativen Bereich. Weiterhin fällt auf, dass bestimmte Werte des Skalenbereichs besonders oft angegeben wurden (-100, -50, 0, 50, 100). Dies ist ein Hinweis darauf, dass insgesamt von einer Ordinalskala auszugehen ist. Wir berücksichtigen die offenkundigen Schwierigkeiten, metrisches Messniveau zu erreichen, indem wir in den multivariaten Modellen nur noch drei Kategorien unterscheiden („ungerechterweise zu gering“, „ungerechterweise zu hoch“ und „gerecht“ – als Referenz).

Abbildung 5: Verteilung der Gerechtigkeitsurteile



Um genauer feststellen zu können, wie die Skala von den Befragten verwendet wurde, werden zwei Parameter näher betrachtet. Zum einen ist die Häufung der Kategorie *gerecht* bemerkenswert. Gibt es hier Bildungs- und Alterseffekte oder einen Hinweis dafür, ob Befragte durch die Einordnung der Vignetteneinkommen als gerecht, damit Überspringen der Antwortschritte zwei und drei (vgl. oben Abschnitt 1) Befragungszeit einsparen wollten? Zum anderen ist interessant, wie viele unterschiedliche Werte von den Befragten angegeben wurden (im Maximalfall können es 24 sein, die „Übungsvignette“ wird hier und bei den folgenden Auswertungen nicht weiter berücksichtigt) bzw. welche Skalierung sie eingesetzt haben.

Im Durchschnitt werden 8,3 von 24 Vignetten (also ein gutes Drittel) als gerecht eingestuft, wobei lediglich marginale Unterschiede zwischen den Bildungsgruppen bestehen (Tabelle 10).

Tabelle 10: Häufigkeit der Kategorie *gerecht* nach Schulabschluss

Schulabschluss	Mittelwert	S.D.	N
Hauptschule	8,26	4,53	488
Realschule	8,29	4,30	331
Gymnasium	8,66	4,24	245
Total	8,33	4,36	1.064

Bei den Altersgruppen gibt es hingegen einen signifikanten Unterschied: Die Gruppe der 16- bis 39-Jährigen verwendet die Kategorie „gerecht“ signifikant ($p < 0,01$) häufiger als die mittlere Altersgruppe der 40- bis 65-Jährigen (Tabelle 11). Die älteste Befragtengruppe liegt mit einer mittleren Anzahl von 8,5 gerechten Urteilen dazwischen, sie unterscheidet sich nicht signifikant von den beiden anderen Gruppen.

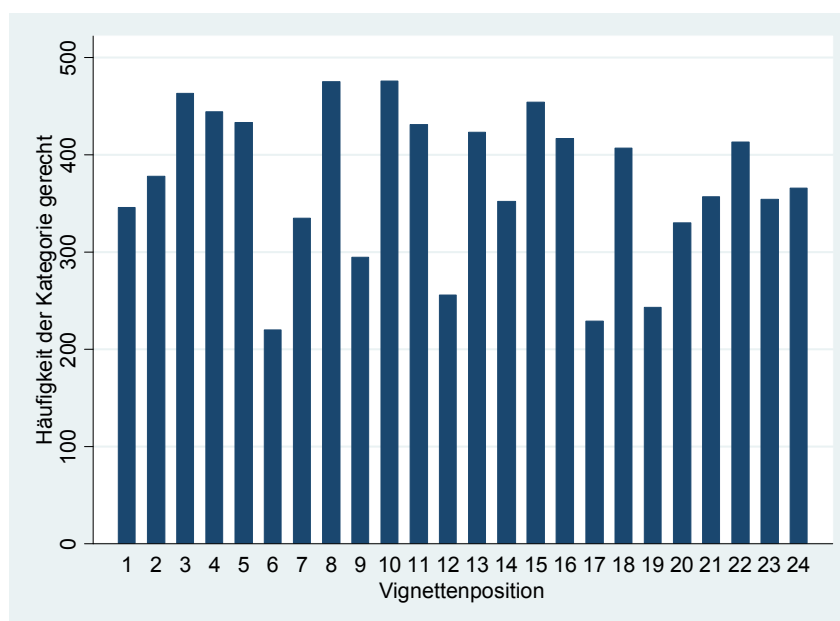
Tabelle 11: Häufigkeit der Kategorie *gerecht* nach Altersgruppen

Altersgruppe	Mittelwert	S.D.	N
16-39	8,87 ^a	4,27	303
40-65	7,86 ^a	4,14	438
66-	8,50	4,69	325
Total	8,35	4,37	1.066

a: Gruppenvergleich sig. ($p < 0,01$)

Die Korrelation zwischen der Vignettenposition im Fragebogen und der Bewertung der Vignette als gerecht ist in allen Bildungs- und Altersgruppen sehr gering (siehe dazu auch Abbildung 6). Es gibt somit keine Hinweise darauf, dass mit zunehmender Vignettenposition die Einschätzung „gerecht“ häufiger auftritt (was ein deutliches Indiz für einen rein methodischen Effekt wäre; *inhaltlich* unterscheiden sich die Vignettenmerkmale nicht stark nach der Bearbeitungsposition, vgl. dazu die in den Anhangstabellen A3 ausgewiesenen Statistiken).

Abbildung 6: Häufigkeit der Kategorie *gerecht* nach Vignettenposition



Als weiteres Kriterium dafür, wie die Urteile pro Befragten abgestuft werden, wird die Anzahl unterschiedlicher Zahlenangaben pro Befragten betrachtet. Die Frage ist, ob die Befragten ihre Urteile so stark variieren, dass man mit einer kleineren Skala (z.B. 7-stufigen Ratingska-

la) die Befragten in ihrem Urteilsverhalten einengen würde. Der Durchschnittswert unterschiedlicher Zahlenangaben beträgt 8,53, der Median liegt bei 8. Unterschiede gibt es zwischen den Bildungsgruppen (Tabelle 12). Befragte mit Hauptschulabschluss verwenden signifikant weniger Zahlenwerte zur Einkommensbeurteilung als Befragte mit Abitur. Ähnliche Befunde liegen auch zu Magnitudeskalen in Conjointanalysen vor (Steenkamp und Wittink 1994; Teas 1987) vor. Die Altersgruppen unterscheiden sich dagegen untereinander nicht signifikant (Tabelle 13).

Tabelle 12: Häufigkeit unterschiedlicher Zahlenangaben nach Schulabschluss

Schulabschluss	Mittelwert	S.D.	N
Hauptschule	8,33 ^a	3,22	488
Realschule	8,52	2,98	331
Gymnasium	8,96 ^a	3,06	245
Total	8,53	3,12	1.064
a: Gruppenunterschied sig. ($p < 0,05$)			

Tabelle 13: Häufigkeit unterschiedlicher Zahlenangaben nach Altersgruppen

Altersgruppe	Mittelwert	S.D.	N
16-39	8,66	3,15	303
40-65	8,69	3,08	438
66-	8,19	3,11	325
Total	8,53	3,11	1.066

Die Anzahl unterschiedlicher Zahlenangaben ist allerdings nur ein erster Hinweis auf die von den Befragten verwendeten Skalierungen. Tabelle 14 gibt genaueren Aufschluss darüber, welche Zahlen bzw. Skalen von den Befragten gewählt wurden. Knapp acht Prozent der Befragten haben für ihre 24 Vignettenurteile ausschließlich Zahlen im Abstand von 25 verwendet (25, 50, 75, 100).

Für zwei Drittel der Befragten hätte ebenso gut eine 10er Skala ausgereicht, sie nutzen ausschließlich auf Zehner-Werte gerundete Zahlen. Zusammen mit denjenigen, welche zusätzliche Abstufungen nach „Fünfer-Kategorien“ (5, 15, ...) vorgenommen haben (das ergibt insgesamt 20 Stufen), sind bereits 90 Prozent aller Befragten erfasst. Anders ausgedrückt nutzen nur knapp zehn Prozent aller Befragten weitere Zahlen, und somit wäre nur eine kleine Minderheit bei der Vorgabe einer 20-stufigen Skala (im Vergleich zu der verwendeten 100er Skala) in ihrem Urteilsverhalten eingengt.

Tabelle 14: Verwendete Zahlenwerte bzw. Skalierungen

Skala (Zahlenwerte)	Prozent	N
4er Skala (25, 50, 75, 100)	7,69	82
10er Skala (10, 20, ..., 90, 100)	65,01	693
20er Skala (5, 10, 15, ..., 95, 100)	90,34	963

4.2.3 Konsistenz der Urteile

Um die Konsistenz der Antworten in Vignettenbefragungen zu prüfen, werden üblicherweise zwei Strategien verfolgt. Zum einen wird die Modellgüte im Bearbeitungsablauf betrachtet. Gibt es eine Eingewöhnungsphase bei den ersten Urteilen, deren Schlüssigkeit daher nicht mit den späteren vergleichbar ist? Gibt es eine Phase im Vignettenmodul, in der die Befragten die „konsistentesten“ Urteile abgeben? Finden sich Hinweise auf Ermüdungseffekte am Ende des Vignettenmoduls?

Um diese Fragen zu beantworten, werden die Antwortkonsistenzen unterschiedlicher Phasen des Vignettenmoduls miteinander verglichen (in OLS-Modellen die Varianzaufklärung). Zum anderen interessiert, inwiefern die Antwortkonsistenz von Befragtenmerkmalen abhängig ist. Sind alte und junge, gut und weniger gut gebildete Befragte gleichermaßen zu konsistenten Antworten in der Lage? Angesichts der gegenüber herkömmlichen Itemabfragen komplexeren Urteilsaufgabe bei Vignetten ist hiervon nicht unbedingt auszugehen (zumal hier pro Vignette eine relativ komplexe Beantwortung in drei Schritten abverlangt wird). Wir vergleichen also wiederum das Antwortverhalten einzelner Alters- und Bildungsgruppen.

Die Konsistenz der Urteile, gemessen durch die Modellgüte, ist allerdings kein ausreichendes Kriterium, um die Güte von Vignettenstudien oder gar von einzelnen Dimensionen zu begründen (Auspurg et al. 2009a). Befragte können auch dann konsistente Antworten liefern, wenn sie einzelne Dimensionen ausblenden, etwa weil sie sich die Urteilsaufgabe vereinfachen oder durch Ermüdung nicht mehr in der Lage sind, umfassendere Urteilen abzugeben. Deshalb werden zusätzlich die Signifikanzen der einzelnen Koeffizienten (bei standardisierter Fallzahl) betrachtet.

Als Grundlage für diese Auswertungen dient jeweils ein multinomiales Logit-Modell, welches das erreichte Skalenniveau in konservativer Weise berücksichtigt.¹⁴ Als abhängige Variable dient das von -100 bis +100 reichende Gerechtigkeitsurteil (vgl. Abb. 5). Diese Urteile werden in drei Gruppen – ungerechterweise zu niedrig (-1 bis -100), gerecht (0) und ungerechterweise zu hoch (1 bis 100) – zusammengefasst. Angesichts der extremen Häufung des Urteils „gerecht“ sowie der ebenfalls deutlichen Häufungen auf den Maximalwerten von +100

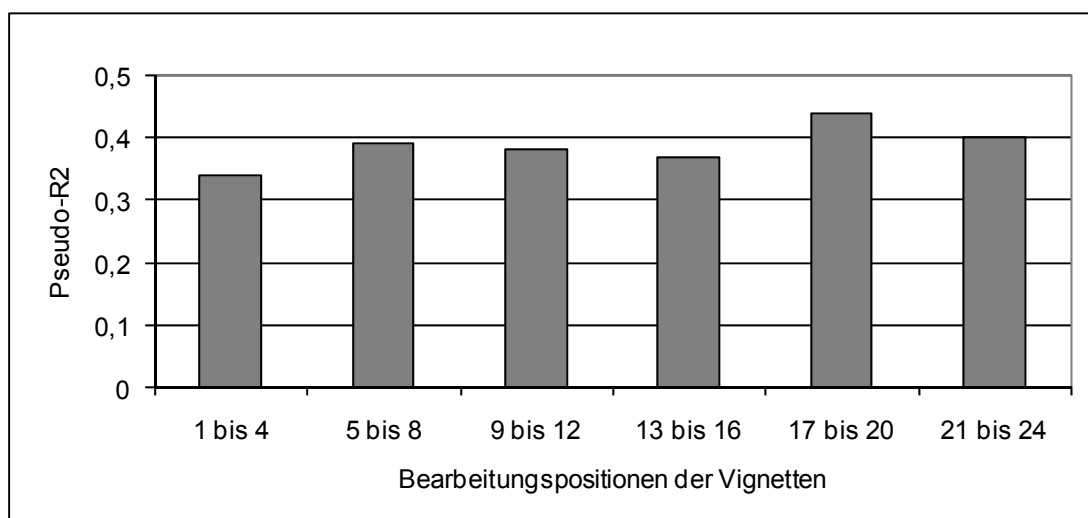
¹⁴ Alternativ lassen sich Zensierungen von Daten mit Tobit-Regressionen modellieren. Im vorliegenden Fall führt allerdings die Null mit extrem vielen Ausprägungen zu einem zusätzlichen Modellierungsproblem, weshalb sich für ein Logitmodell entschieden wurde.

und -100 ist ein lineares OLS-Modell (wie es bislang in Auswertungen von Faktoriellen Surveys gängig ist) nicht adäquat.

Als Gütekriterium wird das Pseudo R^2 nach McFadden herangezogen (Long und Freese 2006; Long 1997). Dieses stellt kein Maß für die „Varianzaufklärung“ dar (anders als das R^2 in OLS-Modellen), bietet jedoch gleichwohl Aufschluss über die Anpassungsgüte des Modells (und damit zugleich die Konsistenz des Antwortverhaltens). Als unabhängige Variable gehen alle Vignettendimensionen in das Modell ein (siehe oben Tabelle 1).

In Abbildung 7 sind die Pseudo R^2 -Werte von sechs Phasen des Vignettenmoduls dargestellt (multinomiale Logitmodelle über jeweils vier Vignetten unter Berücksichtigung der Klumpung der Stichprobe).¹⁵ Alle Urteile der jeweiligen Phase werden dabei gepoolt, d.h. es handelt sich um Berechnungen über alle Befragten hinweg. Es zeigt sich insgesamt, dass die konsistentesten Urteile in der fünften Phase (17te bis 20te Vignette) mit Pseudo R^2 -Werten von über 0,4 abgegeben werden. Nur im ersten Teil liegen die Pseudo R^2 -Werte deutlicher unter 0,4. Die Konsistenz der Urteile nimmt nach diesen ersten vier Vignetten leicht zu. Insgesamt sind die Unterschiede in der Modellgüte zwischen den einzelnen Phasen des Vignettenmoduls aber gering.

Abbildung 7: Modellgüte in den sechs Phasen des Vignettenmoduls



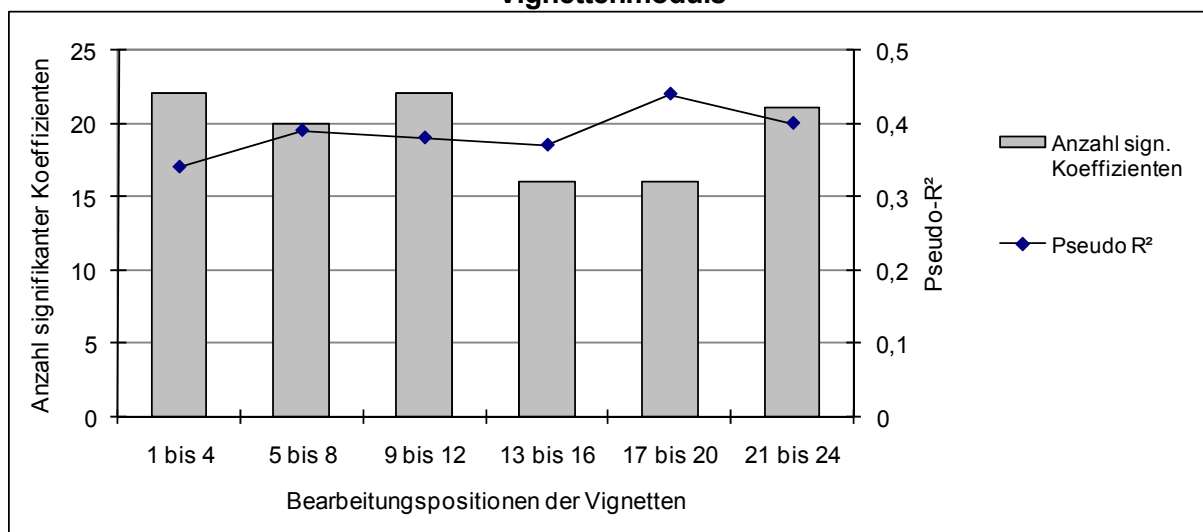
Diese Ergebnisse könnten zu dem Schluss verleiten, dass es keine Ermüdungseffekte gibt. Allerdings können die Befragten eine Überforderung oder Ermüdung ebenso vermeiden, indem sie Heuristiken anwenden, bei denen sie ihre Urteile auf wenige oder gar nur eine einzige Dimension beschränken (sie ziehen beispielsweise nur den Beruf oder das Einkommen heran). Dadurch würde man ebenso konsistente Urteile über alle Befragten hinweg beobachten (hohe Pseudo R^2 -Werte), obwohl einzelne Dimensionen vernachlässigt sind. Deshalb wird

¹⁵ Die Stichprobe ist insofern „geklumpt“, als einzelne Befragte jeweils mehrere Urteile abgeben. Die Urteile sind daher nicht unabhängig voneinander, sondern es liegt eine Datenstruktur ähnlich zu der von Paneldaten vor. Dies wird hier durch die Schätzung von robusten Standardfehlern berücksichtigt (Huber-White-Korrektur).

noch zusätzlich betrachtet, wie viele Dimensionen jeweils einen signifikanten Einfluss auf das Gerechtigkeitsurteil haben (unter Zugrundelegung eines Signifikanzniveaus von fünf Prozent).

Ein multinomiales Modell, welches alle zehn Vignettenmerkmale berücksichtigt, ergibt (aufgrund der Dummy-Splits für einige Variablen) eine Gesamtzahl von 15 unabhängigen Variablen. Die abhängige Variable hat drei Kategorien. Es werden ohne Konstanten $15 \times 2 = 30$ Koeffizienten geschätzt. Abbildung 8 zeigt einen Effekt, der im Spannungsverhältnis zu den obigen Befunden steht. In der ersten, dritten und letzten Phase des Vignettenmoduls sind mehr als 20 Koeffizienten signifikant. In der zweiten, vierten und fünften Phase sind jeweils weniger signifikante Faktoren auszumachen (etwa 16). Die fünfte Phase zeigte – gemessen an den Pseudo R²-Werten, die hier ebenfalls nochmals dargestellt sind – die höchste Konsistenz. Die Anzahl signifikanter Koeffizienten deutet darauf hin, dass die höhere Konsistenz mitunter einer Heuristik geschuldet ist. Die Befragten erreichen konsistentere Urteile, in dem sie weniger Dimensionen berücksichtigen.

Abbildung 8: Anzahl signifikanter Koeffizienten und Pseudo-R² in den sechs Phasen des Vignettenmoduls



Um auszuschließen, dass sich diese Effekte nur durch die hier vorgenommene Einteilung in sechs gleich besetzte Phasen ergeben, wurden alternative Aufteilungen in kürzere oder längere Phasen vorgenommen, die zu gleichen Ergebnissen führen.¹⁶

¹⁶ Bei gesonderter Betrachtung der ersten und zweiten Hälfte des Moduls wird zuerst ein Pseudo R²-Wert von 0,35 bei 20 signifikanten Koeffizienten erreicht, in der zweiten Hälfte sind mit einem Pseudo R²-Wert von 0,37 lediglich 18 Koeffizienten signifikant. Weitere Tests mit drei gleich großen Gruppen bestätigen, dass im Verlauf der Befragung die Anzahl signifikanter Dimensionen etwas abnimmt, bei gleichzeitigem moderaten Anstieg der Pseudo R² Werte. Die unterschiedlichen Anzahlen signifikanter Koeffizienten könnten zudem ein Resultat unterschiedlich starker Korrelationen und Varianzen der unabhängigen Variablen sein (also verschieden „effizienter“ Vignettenstichprobe, vgl. oben Abschnitte 2 und 3.2). Zwar wurden die Decks fraktionalisiert gebildet, die Vignetten innerhalb dieser dann aber zufällig auf die Bearbeitungspositionen zugewiesen. Im Anhang ausgewiesene Statistiken für die einzelnen Bearbeitungsphasen schließen diese Interpretation allerdings aus: Die Korrelationen

Schätzt man Logitmodelle getrennt nach Alters- und Bildungsgruppen, stellt man lediglich geringe Unterschiede in den Modellanpassungen fest (Tabelle 15).¹⁷ Für die Befragten mit Abitur findet sich ein etwas höherer Pseudo R²-Wert als für die anderen beiden Bildungsgruppen. Die Unterschiede sind allerdings insgesamt gering. Die Anzahl der signifikanten Koeffizienten unterscheidet sich kaum nach Schulabschluss. Nach Altersgruppen betrachtet fällt die Modellanpassung in der mittleren Gruppe geringfügig höher aus als in den beiden anderen. Bei der Anzahl der signifikanten Koeffizienten erkennt man einen deutlichen Rückgang mit steigendem Alter der Befragten. Dies könnte ein Hinweis auf in dieser Gruppe wirkende Ermüdungs- oder Überforderungseffekte sein. Analysen für diese Altersgruppe nach den oben beschriebenen sechs Phasen des Moduls (nicht dargestellt) zeigen allerdings ein relativ konstantes Antwortverhalten bezüglich der signifikanten Koeffizienten und der Modellgüte. Lediglich im Bereich zwischen der neunten und zwölften Vignette (ähnlich zu Abbildung 8) werden mehr Koeffizienten signifikant als in den anderen Bereichen. Es kann also anhand der vorliegenden Daten kein Ermüdungseffekt für speziell diese Altersgruppe nachgewiesen werden.

Tabelle 15: Modellgüte und signifikante Koeffizienten nach Schulabschluss und Altersgruppe

Befragtengruppe	Pseudo-R ²	Signifikante Koeffizienten	Anzahl Befragte
Schulabschluss			
Hauptschule	0,35	17,9	245*
Realschule	0,37	17,6	245*
Abitur	0,38	16,0	245
Altersgruppe			
16 bis 39 Jahre	0,36	22,0	303
40 bis 65 Jahre	0,37	18,4	303*
66 Jahre und älter	0,35	14,7	303*

* Mittelwerte aus zehn Stichproben (vgl. Fußnote 16)

Die Analyse der Urteilsconsistenz ergibt somit insgesamt, dass man sich bei der Bewertung der „Antwortgüte“ nicht ausschließlich auf die Modellgüte beziehen sollte. Befragte, welche eine vereinfachende, weniger Dimensionen berücksichtigende Heuristik anwenden, können ebenso hohe Werte bei der Modellgüte erzielen wie Befragte, die keine solchen Antwort-

und Varianzen der Vignettenvariablen unterscheiden sich zwischen den hier gegenübergestellten Phasen nur marginal (vgl. die Tabellen A3).

¹⁷ Hier ist allerdings zu berücksichtigen, dass die Gruppen unterschiedlich stark besetzt sind. Um dadurch bedingte Beeinträchtigungen der Interpretationen zu umgehen, wurden für die Gruppen der Befragten mit Haupt- und Realschulabschluss jeweils zehn Stichproben von der Größe N=245 gezogen und es werden jeweils die Mittelwerte dieser zehn Modelle berichtet. Ebenso wurde bei den Altersgruppen verfahren.

strategien bemühen. Der Blick auf die signifikanten Koeffizienten zeigt, dass sich der „beste“ Bereich (bzw. die umfassendsten Urteile) zwischen der neunten und zwölften Vignette befinden.

5 Inhaltliche Fragestellungen

In diesem Abschnitt werden einige inhaltliche Analysen vorgestellt, die das Analysepotenzial des Vignettenmoduls im SOEP-Pretest 2008 exemplarisch veranschaulichen. Es handelt sich um die Ergebnisse eines multinomialen Logit-Modells, wie es bereits in Abschnitt 4 für die Diskussion methodischer Fragen benutzt wurde.

Abhängige Variable ist die kategoriale Gruppierung der Gerechtigkeitsurteile, unabhängige Variable sind alle (nicht weiter transformierten) Vignettendimensionen. Das Modell wird mit robusten Standardfehlern geschätzt. Analysegrundlage sind die jeweils 24 Urteile der 1.066 befragten Personen, also 25.584 Fälle. Um anschaulichere Interpretationen zu ermöglichen, sind in Tabelle 18 statt der Koeffizientenwerte die Marginaleffekte aufgeführt. Sie geben stets an, um wie viel Prozentpunkte sich die Auswahlwahrscheinlichkeit der jeweiligen Urteilkategorie (ungerecht zu niedrig, gerecht, bzw. ungerecht zu hoch) gegenüber den beiden alternativen Kategorien verändert, wenn sich die betrachtete unabhängige Variable marginal (also um eine Einheit) erhöht bzw. bei Dummy-Variablen die angegebene statt der Referenzkategorie vorliegt. Beispielsweise erhöht sich die Wahrscheinlichkeit, dass das Einkommen als zu niedrig (statt als gerecht oder zu hoch) bewertet wird, um 2,35 Prozentpunkte wenn die Person mit einem Universitätsabschluss statt mit keinem Abschluss beschrieben wird; und verringert sich umgekehrt die Wahrscheinlichkeit einer Bewertung als zu hoch (statt als gerecht oder zu niedrig) um 6,14 Prozentpunkte, wenn die Vignettenperson einen Hochschulabschluss statt keinerlei Abschluss aufweist.

Betrachten wir die Variablen der Reihe nach. Der Geschlechtereffekt ist jeweils signifikant. Konkret besagen die Koeffizienten, dass das Bruttoeinkommen von Männern bei Berücksichtigung aller Kovariaten eher als zu gering angesehen wird als das gleiche Bruttoeinkommen von Frauen. Es liegt also ein „just-gender-wage-gap“ vor (Jasso und Webster 1997, 1999; Jann 2003).

Beim Alter erkennt man, dass das Einkommen von älteren Erwerbstätigen eher als zu gering eingeschätzt wird als das von jüngeren. Insofern wird bei der Gerechtigkeitsbewertung ein Bonus nach Seniorität gewährt.

Die Dimension Ausbildung geht als Dummyset in die Regression ein. Referenzkategorie sind Vignettenpersonen ohne Abschluss. Sowohl eine Lehre als auch ein Universitätsabschluss werden als einkommensrelevant bewertet, und zwar unabhängig von dem ausgeübten Beruf. Umso mehr Ausbildung jemand hat, umso mehr Einkommen wird für ihn als angemessen betrachtet.

Gleiches gilt für das Prestige des Berufs, der in Form eines Magnitude Prestige Scores gemessen ist: Menschen mit einem Beruf mit höherem Prestige wird mehr Gehalt zugesprochen als jenen in einem vergleichsweise weniger prestigeträchtigen Beruf.

Tabelle 16: Determinanten der Einkommensgerechtigkeit (multinomiales Logit-Modell), Marginal-
effekte

	Bruttoeinkommen		
	Zu niedrig	Gerecht	Zu hoch
Geschlecht ¹	0.00840 ^{***} (0.00180)	0.0193 ^{**} (0.00721)	-0.0277 ^{***} (0.00733)
Alter [10 Jahre]	0.00373 ^{***} (0.000729)	0.0150 ^{***} (0.00308)	-0.0188 ^{***} (0.00306)
Lehre ²	0.0155 ^{***} (0.00247)	0.0217 [*] (0.00943)	-0.0373 ^{***} (0.00957)
Universität ²	0.0235 ^{***} (0.00290)	0.0379 ^{***} (0.00979)	-0.0614 ^{***} (0.00961)
Prestige [10 MPS-Scores]	0.00570 ^{***} (0.000618)	0.0183 ^{***} (0.00138)	-0.0240 ^{***} (0.00129)
Bruttoeinkommen [1000 Euro]	-0.0554 ^{***} (0.00470)	-0.0391 ^{***} (0.00474)	0.0945 ^{***} (0.00254)
Leistung: durchschnittlich ³	0.0220 ^{***} (0.00321)	0.0538 ^{***} (0.00894)	-0.0758 ^{***} (0.00865)
Überdurchschnittlich ³	0.0329 ^{***} (0.00402)	0.104 ^{***} (0.00949)	-0.137 ^{***} (0.00901)
Unternehmen stabil ⁴	-0.00477 [*] (0.00194)	0.0307 ^{***} (0.00914)	-0.0260 ^{**} (0.00909)
Vom Konkurs bedroht ⁴	-0.00306 (0.00185)	-0.0301 ^{***} (0.00905)	0.0332 ^{***} (0.00901)
Mittelbetrieb ⁵	0.00124 (0.00199)	0.0193 [*] (0.00836)	-0.0205 [*] (0.00847)
Großbetrieb ⁵	0.00544 [*] (0.00215)	0.0237 [*] (0.00975)	-0.0292 ^{**} (0.00986)
Zweiverdiener ⁶	-0.0131 ^{***} (0.00240)	-0.0239 ^{**} (0.00880)	0.0370 ^{***} (0.00894)
Single ⁶	-0.00684 ^{***} (0.00194)	-0.0244 ^{**} (0.00874)	0.0312 ^{***} (0.00900)
Anzahl Kinder	0.00149 ^{**} (0.000535)	0.00893 ^{***} (0.00254)	-0.0104 ^{***} (0.00260)
Pseudo R ²	0.360		
ll_0	-28074.3		
ll	-17976.9		
# Befragte	1066		
# Vignetten	25584		

Standardfehler in Klammern; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

¹ Referenzkategorie: weiblich

² Referenzkategorie: ohne Abschluss

³ Referenzkategorie: Leistung unterdurchschnittlich

⁴ Referenzkategorie: Unternehmen erwirtschaftet
hohe Gewinne

⁵ Referenzkategorie: Kleinbetrieb

⁶ Referenzkategorie: Alleinverdiener, verheiratet

Das Bruttoeinkommen der „Vignettenperson“, das als Bewertungsgrundlage der Befragten dient, besitzt einen erwartungsgemäß großen Einfluss. Je 1000 Euro, die jemand mehr verdient, erhöht sich die Wahrscheinlichkeit einer Beurteilung als „ungerechterweise zu hoch“ um 9,45 Prozentpunkte. Die am Arbeitsplatz gezeigte Leistung (im Modell als Dummy-Set mit der Referenzkategorie unterdurchschnittlich) hat ebenso einen deutlichen Effekt: Wer mehr leistet, sollte auch entsprechend mehr verdienen. Bei einer unter- statt überdurchschnittliche Leistung wird das Einkommen c.p. um ganze 13,7 Prozentpunkte wahrscheinlicher als zu hoch (statt als gerecht oder zu niedrig) beurteilt.

Die ökonomische Situation des fiktiven Betriebs in dem eine Vignettenperson arbeitet, spielt ebenfalls eine Rolle. Im Vergleich zur Referenzkategorie (Unternehmen erwirtschaftet hohe Gewinne) werden die Bruttoeinkommen von Vignettenpersonen in wirtschaftlichen stabilen Unternehmen signifikant häufiger als gerecht oder ungerechterweise zu niedrig bewertet. Einkommen, die in vom Konkurs bedrohten Unternehmen erzielt werden, werden dagegen im Vergleich zur Referenz wahrscheinlicher als ungerechterweise zu hoch eingeschätzt.

Die Betriebsgröße hat einen ähnlichen, wenngleich schwächer ausgeprägten Effekt: Bei Beschäftigten in Mittel- und Großbetrieben wird das Einkommen mit geringerer Wahrscheinlichkeit als zu hoch (statt als gerecht oder zu niedrig) bewertet als bei Angestellten von Kleinbetrieben.

Die beiden letzten Dimensionen bilden die familiäre Situation ab. Verheiratete Alleinverdiener bilden hier die Referenzgruppe. Dieser Gruppe wird ein höheres Einkommen zugesprochen als Singles oder Zweiverdienern. Die Anzahl der Kinder besitzt einen negativen Einfluss auf das Urteil: Je mehr Kinder eine Person hat, desto mehr sollte sie verdienen.

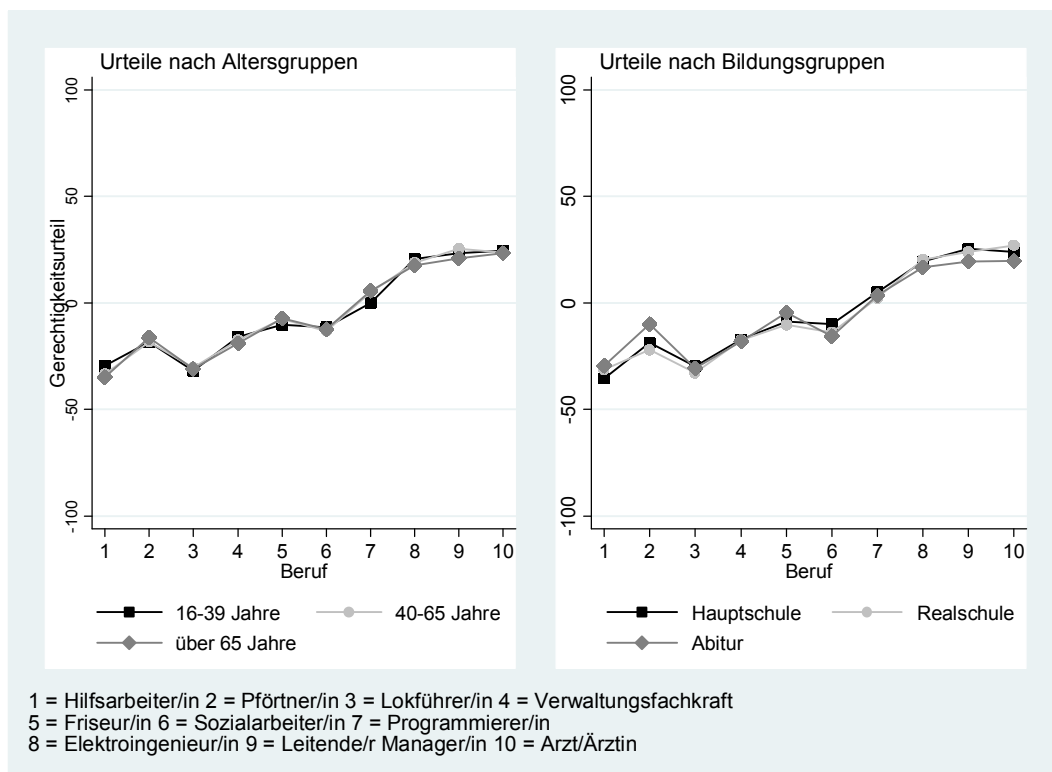
Im Weiteren kann analysiert werden, welche Unterschiede es in den Urteilen verschiedener Befragtengruppen gibt. Hierzu bietet sich diese repräsentative Stichprobe besonders an, sie lässt wesentlich mehr Analysenspielraum als die homogenen Stichproben von Studierenden (oder anderen speziellen Populationen), wie sie bislang bei Faktoriellen Surveys vorherrschend sind.

Zunächst wird betrachtet, welchen Einfluss die in den Vignetten vorgegebenen Berufe auf die Gerechtigkeitsbewertungen haben. Die Befragten werden hierzu nach Alters- und Bildungsgruppen eingeteilt, wie dies bereits bei den methodischen Analysen praktiziert wurde. In Abbildung 9 sind die Mittelwerte der Gerechtigkeitseinschätzungen abgetragen – für die zehn Vignettenpersonen-Berufe, die aufsteigend von niedrigeren zu hohen Prestigewerten sortiert sind.¹⁸ Inhaltlich ist interessant, dass sich diese Kurven nach Bildungs- und Altersgruppen der Befragten kaum unterscheiden. Insofern herrscht also ein sehr großer Konsens über gerechte Einkommen für unterschiedliche Berufsgruppen. Zudem fällt auf, dass Lokführer/innen und

¹⁸ Einkommen für Berufe, die einen höheren Prestigewert besitzen, werden also häufiger als zu hoch bewertet. Dies steht auf den ersten Blick im Widerspruch zu den Ergebnissen des eben dargestellten Multinomialen Logit-Modells. Der Grund ist allerdings einleuchtend: Die Berufe korrelieren durch den Ausschluss unlogischer Fälle bei der Vignettenkonstruktion mit dem Einkommen, was in dem Logit-Modell, nicht aber den hier präsentierten Mittelwertvergleichen kontrolliert wird.

Sozialarbeiter/innen (Kategorie 3 bzw. 6) leicht aus der „Rangfolge“ ausscheren: Diesen beiden Berufsgruppen wird also ein etwas höheres Einkommen zugesprochen (das Urteil fällt eher in den Bereich „zu gering“) als dies nach ihren Prestigewerten zu erwarten wäre.

Abbildung 9: Gerechtigkeitsbewertung nach Berufen der Vignettenpersonen und Befragten Gruppen

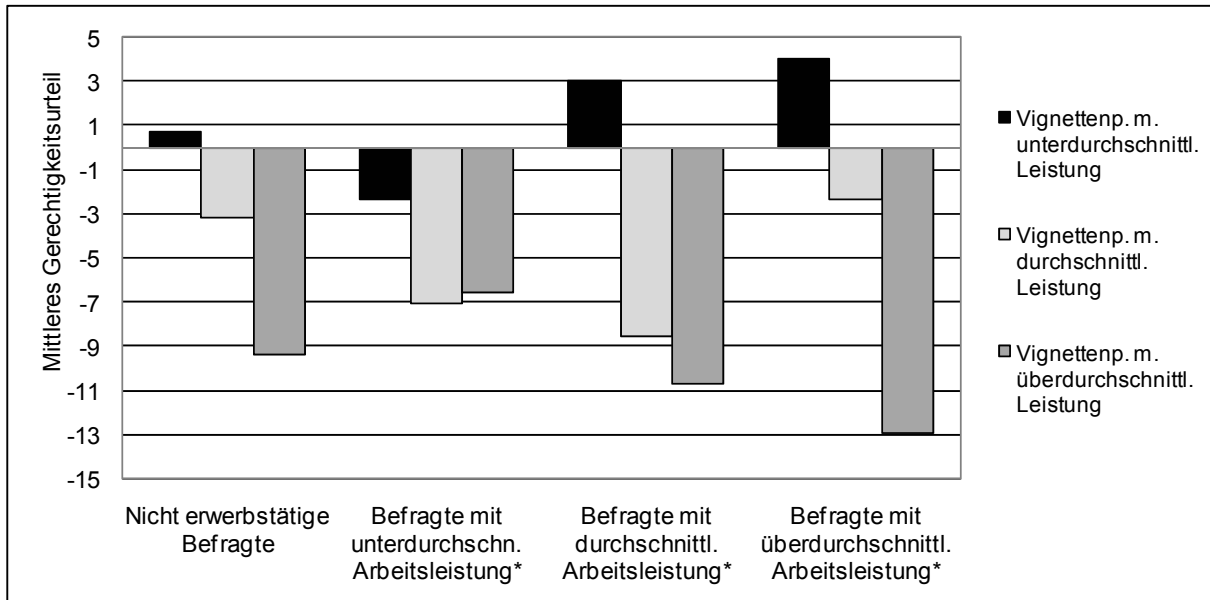


Um das weitere Auswertungspotenzial von Faktoriellen Surveys anzudeuten, werden noch zwei weitere befragtenspezifische Auswertungen vorgestellt. Im Regressionsmodell hatte die von der Vignettenperson erbrachte Leistung eine bedeutende Rolle für die Gerechtigkeitsbewertung gespielt. Um zu prüfen, ob auch dies ein „konsensuales“ Urteil ist, werden die Urteilen von nicht erwerbstätigen Befragten mit denjenigen von drei weiteren Gruppen verglichen. Alle Erwerbstätigen sollten im SOEP-Prestest 2008 ihre eigene Leistung am Arbeitsplatz einstufen, analog zu den drei Kategorien der Vignettendimension Leistung. Aus Abbildung 10 geht hervor, dass die so gruppierten Befragten dieser Vignettendimension unterschiedliche Bedeutung zukommen lassen.

Befragte, die selbst geringe Leistungen am Arbeitsplatz erbringen, stufen ihre Gerechtigkeitsurteil nur wenig nach den Leistungen der Vignettenpersonen ab, Befragte mit mittlerer und

überdurchschnittlicher Leistung am Arbeitsplatz sprechen der Leistungs-Dimension demgegenüber eine wesentlich stärkere Rolle für die Einkommensgerechtigkeit zu.¹⁹

Abbildung 10: Gerechtigkeitsbewertung nach Leistung (Mittelwertunterschiede)

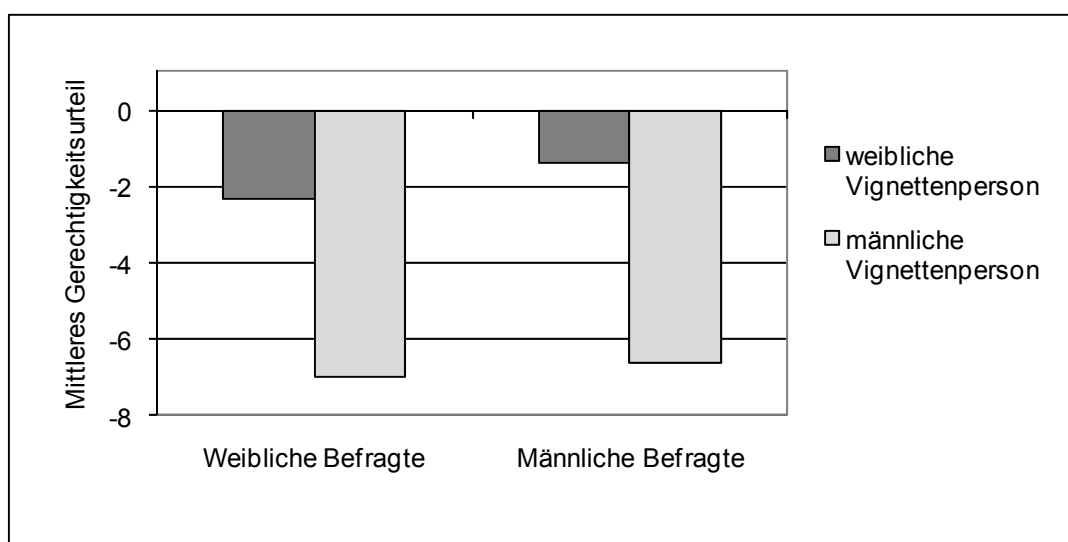


Anmerkung: * = Selbsteinstufung der Befragten.

¹⁹ Nochmals zur Erläuterung: Die negativen Werte bedeuten, dass die Vignettenperson zu wenig, die positiven Werte, dass die Vignettenperson zu viel bekommt.

Eine weitere Dimension, die im Regressionsmodell von Bedeutung war, ist das Geschlecht der Vignettenperson. Betrachtet man sich analog zu der Auswertung der Leistungsunterschiede die Wechselwirkungen zwischen dem Geschlecht der Befragten und dem Geschlecht der Vignettenpersonen (Abbildung 11), stellt man fest, dass Männer wie Frauen männlichen Einkommensbezieher ein höheres gerechtes Einkommen zusprechen als weiblichen. Der Abstand in den geschlechtsspezifischen Gerechtigkeitsurteilen fällt für männliche wie weibliche Befragte ähnlich groß aus.

Abbildung 11: Gerechtigkeitsbewertung nach Geschlecht der Vignettenpersonen



Damit sind einige Auswertungsmöglichkeiten des Vignettenmoduls angedeutet. Für die multivariate Modellierung sind gewiss auch Modelle möglich, die das Skalenniveau der abhängigen Variablen weniger konservativ einschätzen, als wir dies hier getan haben. In inhaltlicher Hinsicht erscheinen vor allem weitere Auswertungen zu subgruppenspezifischen Urteilsregeln interessant (technisch: zu Interaktionen von Vignetten- und Befragtenmerkmalen).

6 Fazit

In dieser Dokumentation wurde die Erhebungsmethode des Faktoriellen Surveys vorgestellt und der Aufbau und die Umsetzung des Vignettenmoduls im SOEP-Pretest 2008 beschrieben. Das methodische Erkenntnisinteresse dieses Beitrags bezieht sich auf die Tauglichkeit von Faktoriellen Surveys in allgemeinen Bevölkerungsumfragen. Der inhaltliche Schwerpunkt liegt auf dem Gebiet der empirischen Gerechtigkeitsforschung, und dort speziell auf der Einkommensgerechtigkeit.

Die wichtigsten methodischen Erkenntnisse dieser Studie sind:

- (1) Der Faktorielle Survey ist bei Beachtung einiger Randbedingungen auch in der großflächigen Umfrageforschung einsetzbar. Befragte verschiedener Alters- und Bildungsgruppen kommen mit der Beantwortung hinreichend gut zurecht. Die höhere Komplexität eines Vignettenmoduls erscheint für die ganz überwiegende Mehrheit der Befragten handhabbar.
- (2) Die Antwortzeit liegt bei ungefähr 30 Sekunden pro Vignette, wobei es hier kaum Bildungs- und Altersunterschiede gibt. Etwa fünf Prozent der Befragten bewältigen die Vignettenbeurteilung so schnell, dass Zweifel an der Ernsthaftigkeit der Bearbeitung angebracht sind – dies mag auch daran liegen, dass die Befragten keine anderen Möglichkeiten zur Antwortverweigerung hatten. Das heißt: ein Vignettenmodul mit einer moderaten Zahl an Vignetten ist in eine allgemeine Bevölkerungsumfrage wie das SOEP gut zeitlich integrierbar. Eine differenzierende Analyse zeigt: bereits die ersten Vignettenurteile liefern vollwertige Daten und sind keinesfalls nur Übungsmaterial für die Befragten. Wegen möglicher Ermüdungseffekte erscheinen insbesondere Vignettenmodule mit bis zu zwölf Vignetten sinnvoll.
- (3) Die Analysen zur Verwendung der Antwortskala zeigen, dass pro Befragten im Durchschnitt nicht mehr als acht bis neun Werte verwendet werden – mit einer klaren Häufung auf runden Werten. Mit einer geschlossenen Ratingskala, die die SOEP-Befragten gewöhnt sind, kann man die Bedürfnisse nach Differenzierung bereits für die Mehrheit der Befragten gut abbilden.
- (4) Die Analyse der Urteilsconsistenz zeigt, dass alle Alters- und Bildungsgruppen mit Vignetten umgehen können, diese also ohne Einschränkungen nutzbar sind. Die Auswertungen haben zusätzlich ergeben, dass eine Orientierung ausschließlich an der Modellgüte zu kurz greift. Befragte können ein konsistentes Urteilsverhalten an den Tag legen und dabei bestimmte Dimensionen (die sie gegebenenfalls aber inhaltlich für relevant erachten) ausblenden. Da im Mittelpunkt des Erkenntnisinteresses meist die Urteilsrelevanz einzelner Dimensionen steht, sollte diese Möglichkeit bei Interpretationen beachtet werden. Die erzielten Ergebnisse deuten darauf hin, dass vor allem ältere Befragte weniger Dimensionen zu Beurteilung heranziehen. Ob dies als Hinweis für eine für diese Befragten zu komplexe Urteilsaufgabe und damit einen methodischer Effekt zu deuten ist, oder aber

ein inhaltlich valides Ergebnis darstellt (ältere Befragten halten möglicherweise tatsächlich weniger Dimensionen als urteilsrelevant), wäre in künftigen Studien zu klären.

Die inhaltlichen Ergebnisse dieses Berichts zeigen beispielhaft, dass neben dem Beruf, der Ausbildung und der Leistung – also Faktoren die im direkten Bezug zur Erwerbstätigkeit stehen – ebenso familiäre Aspekte, wie der Familienstand, die Erwerbstätigkeit des Partners und die Anzahl der Kinder relevante Kriterien für die Einkommensgerechtigkeit darstellen.

Der Faktorielle Survey bietet in methodischer *und* inhaltlicher Hinsicht vielzählige Analyse-möglichkeiten. Die positiven Erfahrungen des SOEP-Pretest 2008 ermutigen dazu, auch in der Haupterhebung des SOEP künftig Vignettenmodule einzusetzen.

7 Literatur

- Abraham, Martin, und Thomas Hinz*, 2005a: Theorien des Arbeitsmarktes: Ein Überblick. S. 17-68 in: *Abraham, Martin/ Hinz, Thomas* (Hg.), Arbeitsmarktsoziologie. Probleme, Theorien, empirische Befunde. Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlag GmbH.
- Abraham, Martin, und Thomas Hinz* (Hg.), 2005b: Arbeitsmarktsoziologie. Probleme, Theorien, empirische Befunde. Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH.
- Addelman, Sidney*, 1962a: Symmetrical and asymmetrical fractional factorial plans. *Technometrics* 4: S. 47-57.
- Addelman, Sidney*, 1962b: Orthogonal Main-Effect Plans for asymmetrical factorial experiments. *Technometrics* 4: S. 21-46.
- Alexander, Cheryl S., und Henry Jay Becker*, 1978: The Use of Vignettes in Survey Research. *Public Opinion Quarterly* 42: S. 93-104.
- Alves, Wayne M.*, 1982: Modeling ditributive justice judgments. S. in: *Rossi, Peter H. , und Steven L. Nock* (Hg.), Measuring social judgments. The factorial survey approach. Beverly Hills: Sage.
- Alves, Wayne M., und Peter Rossi*, 1978: Who Should Get What? Fairness Judgments of the Distribution of Earnings. *American Journal of Sociology* 84: S. 541-564.
- Auspurg, Katrin, und Martin Abraham*, 2007: Die Umzugsentscheidung von Paaren als Verhandlungsproblem. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59: S. 271-293.
- Auspurg, Katrin, Thomas Hinz und Stefan Liebig*, 2009a: Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey. *Methoden, Daten, Analysen im erscheinen: im Erscheinen*.
- Auspurg, Katrin, Martin Abraham und Thomas Hinz*, 2009b: Die Methodik des faktoriellen Surveys in einer Paarbefragung. S. 149-178 in: *Kriwy, Peter, und Christiane Gross* (Hg.), Klein aber fein! Quantitative empirische Sozialforschung mit kleinen Fallzahlen. Wiesbaden: VS.
- Auspurg, Katrin, Thomas Hinz, Stefan Liebig und Carsten Sauer*, 2008: Wer verdient welches Einkommen? Ergebnisse eines faktoriellen Surveys zur Einkommensgerechtigkeit in Deutschland.
- Barrera, Davide, und Vincent Buskens*, 2007: Imitation and Learning under Uncertainty: A Vignette Experiment. *International Sociology* 22: S. 367-396.
- Beck, Michael, und Karl-Dieter Opp*, 2001: Der faktorielle Survey und die Messung von Normen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53: S. 283-306.
- Berk, R. A., und P. H. Rossi*, 1977: Prison reform and state elites. Cambridge, Mass.: Ballinger.
- Dulmer, Hermann*, 2007: Experimental Plans in Factorial Surveys: Random or Quota Design? *Sociological Methods Research* 35: S. 382-409.
- Dülmer, Hermann*, 2001: Bildung und der Einfluss von Argumenten auf das moralische Urteil. Eine empirische Analyse zur moralischen Entwicklungstheorie Kohlbergs. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53: S. 1-27.
- Giesecke, Johannes, und Roland Verwiebe*, 2008: Die Lohnentwicklung in Deutschland zwischen 1998 und 2005 - Wachsende Ungleichheit. *WSI-Mitteilungen*: S. 85-90.
- Grabka, Markus, und Joachim R. Frick*, 2008: Schrumpfende Mittelschicht in Deutschland - Anzeichen einer dauerhaften Polarisierung der verfügbaren Einkommen. *Wochenbericht des DIW Berlin* 75(10): 101-108.

- Hermkens, Piet L. J., und Frank A. Boerman*, 1989: Consensus With Respect to the Fairness of Incomes: Differences Between Social Groups. *Social Justice Research* 3: S. 201-215.
- Jann, Ben*, 2003: Lohngerechtigkeit und Geschlechterdiskriminierung: Experimentelle Evidenz. . Unveröffentlichtes Manuskript an der Eidgenössischen Technischen Hochschule Zürich.
- Jasso, Guillermina*, 1978: On the justice of earnings: a new specification of the Justice Evaluation Function. *American Journal of Sociology* 83: S. 1398-1419.
- Jasso, Guillermina*, 1994: Assessing Individual and Group Differences in the Sense of Justice: Framework and Application to Gender Differences in the Justice of Earnings. *Social Science Research* 23: S. 368-406.
- Jasso, Guillermina*, 2006: Factorial Survey Methods for Studying Beliefs and Judgments. *Sociological Methods Research* 34: S. 334-423.
- Jasso, Guillermina*, 2007: Studying Justice: Measurement, Estimation, and Analysis of the actual reward and the just reward. *Forschungsinstitut zur Zukunft der Arbeit*.
- Jasso, Guillermina, und Peter H. Rossi*, 1977: Distributive justice and earned income. *American Sociological Review* 42: S. 639-651.
- Jasso, Guillermina, und Karl-Dieter Opp*, 1997: Probing the Character of Norms: A Factorial Survey Analysis of the Norms of Political Action. *American Sociological Review* 62: S. 947-964.
- Jasso, Guillermina, und Murray Jr. Webster*, 1997: Double Standards in Just Earnings for Male and Female Workers. *Social Psychology Quarterly* 60: S. 66-78.
- Jasso, Guillermina, und Murray Jr. Webster*, 1999: Assessing the Gender Gap in Just Earnings and Its Underlying Mechanisms. *Social Psychology Quarterly* 62: S. 367-380.
- Jasso, Guillermina, und Eva M. Meyersson Milgrom*, 2008: Distributive Justice and CEO Compensation. *Acta Sociologica* 51: S. 123-143.
- Kapteyn, Arie, James P. Smith und Arthur van Soest*, 2008: Are Americans Really Less Happy With Their Incomes?, *RAND Working Paper WR-591*. Santa Monica (CA).
- Kuhfeld, Warren F.*, 2005: .Experimental Design, Efficiency, Coding, and Choice Designs. S. 47-97 in: *Kuhfeld, W. F.* (Hg.), *Marketing Research Methods in SAS: Experimental Design, Choice, Conjoint, and Graphical Techniques*.
- Kuhfeld, Warren F., Randell D. Tobias und Mark Garrett*, 1994: Efficient Experimental Design with Marketing Research Applications. *Journal of Marketing Research* 31: S. 545-557.
- Liebig, Stefan, und Steffen Mau*, 2002: Einstellungen zur sozialen Mindestsicherung. Ein Vorschlag zur differenzierten Erfassung normativer Urteile. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54: S. 109-134.
- Liebig, Stefan, und Steffen Mau*, 2005: Wann ist ein Steuersystem gerecht? *Zeitschrift für Soziologie*. 34: S. 468-491.
- Liebig, Stefan, und Jürgen Schupp*, 2005: Empfinden die Erwerbstätigen in Deutschland ihre Einkommen als gerecht? *DIW-Wochenbericht* 48/2005: S. 721-725.
- Liebig, Stefan, und Jürgen Schupp*, 2008a: Leistungs- oder Bedarfsgerechtigkeit? Über einen normativen Zielkonflikt des Wohlfahrtsstaats und seiner Bedeutung für die Bewertung des eigenen Erwerbseinkommens. *Soziale Welt* 59: S. 7-30.
- Liebig, Stefan, und Jürgen Schupp*, 2008b: Immer mehr Erwerbstätige empfinden ihr Einkommen als ungerecht *DIW Wochenbericht* 2008/31: S. 434-440.
- Long, Scott J.*, 1997: *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.
- Long, Scott J., und Jeremy Freese*, 2006: *Regression Models for Categorical Dependent Variables Using Stata*. College Station: Stata Press.

- Mäs, Michael, Kurt Mühler und Karl-Dieter Opp*, 2005: Wann ist man Deutsch? Empirische Ergebnisse eines faktoriellen Surveys. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 57: S. 112-134.
- Miller, J. L., P. H. Rossi und J.E. Simpson*, 1986: Perceptions of Justice: Race and Gender Differences in Judgment of Appropriate Prison Sentences. *Law & Society Review* 20: S. 313-334.
- OECD*, 2008: Growing Unequal? Income Distribution and Poverty in OECD Countries.
- Prüfer, Peter, Lisa Vazansky und Darius Wystup*, 2003: Antwortskalen im ALLBUS und ISSP. Eine Sammlung. ZUMA-Methodenbericht 2003/11.
- Schulte, Aileen*, 2002: Consensus versus Disagreement in Disease-Related Stigma: A Comparison of Reactions to AIDS and Cancer Patients. *Sociological Perspectives* 45: S. 81-104.
- Schwarz, Norbert, und Bärbel Knäuper*, 2006: Kognitionspsychologie und Umfrageforschung: Altersabhängige Kontexteffekte. S. 203-216 in: *Diekmann, Andreas* (Hg.), Methoden der Sozialforschung. Sonderheft 44 der Kölner Zeitschrift für Soziologie und Sozialpsychologie. Wiesbaden: Verlag für Sozialwissenschaften.
- Schwarze, Johannes*, 2007: Gerechte Löhne? Eine empirische Analyse subjektiver Erwerbseinkommen. S. 80-107 in: Schwarze et al. (Hrsg.) (2007): Arbeitsmarkt- und Sozialpolitikforschung im Wandel – Festschrift für Christof Helberger zum 65. Geburtstag.
- Siegel, Nico A., Andreas Stocker und Sebastian Warnholz*, 2009: SOEP Testerhebung 2008: Persönlichkeit, Gerechtigkeitsempfinden und Alltagsstimmung. Methodenbericht. TNS Infratest Sozialforschung.
- Smith, Tom W.*, 1986: A Study of Non-Response and Negative Values on the Factorial Vignettes on Welfare. GSS Methodological Report No. 44. Chicago. NORC.
- Steenkamp, Jan-Benedict, und Dick R. Wittink*, 1994: The Metric Quality of Full-Profile Judgements and the Number-of-Attribute-Levels-Effect in Conjoint Analysis. *International Journal of Research in Marketing* 11: S. 275-286.
- Steiner, Peter M., und Christiane Atzmüller*, 2006: Experimentelle Vignettendesigns in Faktoriellen Surveys. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58: S. 117-146.
- Struck, Olaf, Gesine Stephan, Christoph Köhler, Alexandra Klause, Christian Pfeifer und Tatjana Sohr* (Hg.), 2006: Arbeit und Gerechtigkeit. Entlassungen und Lohnkürzungen im Urteil der Bevölkerung. Wiesbaden: VS Verlag.
- Teas, R.K.*, 1987: Magnitude scaling of the dependent variable in decompositional multiattributive preference models. *Journal of the Academy of Marketing Science* 15: S. 64-73.
- Thomas, Neal, Trivellore E. Raghunathan, Nathaniel Schenker, Myron J. Katzoff und Clifford L. Johnson*, 2006: An Evaluation of Matrix Sampling Methods Using Data from the National Health and Nutrition Examination Survey. *Survey Methodology* 32(2): 217-231.
- Thurman, Quint C., Julie A. Lam und Peter H. Rossi*, 1988: Sorting out the Cuckoo's Nest: A Factorial Survey Approach to the Study of Popular Conceptions of Mental Illness. *The Sociological Quarterly* 29: S. 565.
- Winkler, Niels, Martin Kroh und Martin Spiess*, 2006: Entwicklung einer deutschen Kurzsкала zur zweidimensionalen Messung von sozialer Erwünschtheit. DIW Discussion Papers Nr 579. Berlin. DIW (German Institute for Economic Research).

8 Anhang

Tabelle A1: Stichprobenstruktur und Gewichtung

Merkmal	Ungewichtet	gewichtet
Geschlecht		
Weiblich	47,3	48,5
Männlich	52,7	51,5
Staatsangehörigkeit		
Deutsch	96,7	96,1
Andere	3,3	3,9
Alter		
16-19	3,8	5,6
20-29	12,6	14,0
30-39	12,0	16,0
40-49	16,0	19,6
50-59	17,4	15,4
60-69	16,4	13,9
70 und älter	21,7	15,5
Bundesland		
Schleswig Holstein	2,3	3,4
Hamburg	2,4	2,2
Niedersachsen	10,1	9,5
Bremen	0,7	,8
Nordrhein-Westfalen	23,7	21,7
Hessen	6,7	7,3
Rheinland-Pfalz	5,7	4,9
Baden-Württemberg	12,2	12,9
Bayern	13,9	15,0
Saarland	0,6	1,3
Berlin	4,6	4,2
Brandenburg	3,9	3,3
Mecklenburg-Vorpommern	2,3	2,2
Sachsen	5,2	5,3
Sachsen-Anhalt	2,9	3,1
Thüringen	2,8	3,0
BIK-Typen		
0	27,3	24,2
1	7,5	8,5
2	15,9	15,3
3	14,2	16,6
4	3,0	2,8
5	9,7	9,0
6	10,7	10,2
7	6,2	7,4
8	3,6	3,6
9	2,0	2,4

Fortsetzung Tabelle A1

Merkmal	Ungewichtet	gewichtet
Schulabschluss		
Schulabschluss in Deutschland	90,2	88,6
davon...		
Volksschul-/ Hauptschulabschluss	42,9	49,4
Mittlere Reife/ Realschulabschluss	30,7	27,4
Fachhoch-/Fachoberschule	5,7	5,2
Abitur	19,8	17,2
Anderer Schulabschluss	,4	,3
Schule ohne Abschluss verlassen	,5	,4
Befragte Person geht noch zur Schule	2,2	3,5
Schulabschluss im Ausland	7,4	7,7
Erwerbstätigkeit		
Voll erwerbstätig	29,0	33,6
In Teilzeit	10,4	11,0
In betriebl. Ausbildung/Lehre/ Umschulung	1,9	1,7
Geringfügig oder unregelmäßig erwerbstätig	2,1	1,9
In Altersteilzeit mit Arbeitszeit Null	,9	,5
Zivildienstleistender/Freiwilliges Soziales Jahr	,1	,1
Nicht erwerbstätig	55,6	51,3
Familienstand		
Verheiratet, mit Ehepartner zusammenlebend	49,5	55,9
Verheiratet, dauernd getrennt lebend	2,3	1,4
Ledig	23,6	25,2
Geschieden	11,2	8,1
Verwitwet	13,5	9,5

Tabelle A2: Korrelationen der Vignettendimensionen

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Geschlecht	1.000									
(2) Alter	0.008	1.000								
(3) Ausbildung	0.005	-0.030	1.000							
(4) MPS	0.028	0.020	0.251	1.000						
(5) Bruttoeink.	-0.007	0.014	0.128	0.586	1.000					
(6) Leistung	0.001	0.003	0.008	-0.008	-0.009	1.000				
(7) Ökon. Situation Betrieb	0.018	0.014	0.020	0.035	0.044	0.008	1.000			
(8) Betriebsgröße	0.017	-0.010	-0.010	-0.005	-0.022	-0.043	0.003	1.000		
(9) Familienstand	-0.000	0.021	-0.037	-0.036	-0.021	0.024	0.004	0.012	1.000	
(10) Kinder	0.011	0.008	-0.011	0.018	-0.005	-0.056	-0.021	0.017	0.000	1.000

Tabelle A3a: Korrelationen der Vignettendimensionen für die Vignetten 1 bis 4

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Geschlecht	1.000									
(2) Alter	0.339	1.000								
(3) Ausbildung	-0.198	0.136	1.000							
(4) MPS	0.084	-0.139	0.317	1.000						
(5) Bruttoeink.	0.045	-0.211	0.054	0.651	1.000					
(6) Leistung	0.174	-0.018	-0.249	-0.153	-0.095	1.000				
(7) Ökon. Situation Betrieb	0.091	-0.003	-0.140	-0.148	-0.043	0.223	1.000			
(8) Betriebsgröße	-0.170	0.103	-0.238	-0.392	-0.379	0.207	-0.034	1.000		
(9) Familienstand	-0.091	0.160	-0.147	0.025	0.100	0.074	-0.061	0.106	1.000	
(10) Kinder	-0.009	0.055	0.145	0.223	-0.039	-0.322	-0.066	-0.031	0.025	1.000

Tabelle A3b: Korrelationen der Vignettendimensionen für die Vignetten 5 bis 8

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Geschlecht	1.000									
(2) Alter	-0.010	1.000								
(3) Ausbildung	-0.149	-0.059	1.000							
(4) MPS	-0.042	0.223	0.335	1.000						
(5) Bruttoeink.	-0.122	0.341	0.020	0.509	1.000					
(6) Leistung	-0.033	0.002	0.086	0.291	0.391	1.000				
(7) Ökon. Situation Betrieb	0.083	0.064	-0.123	0.124	0.052	0.151	1.000			
(8) Betriebsgröße	0.007	-0.036	0.161	-0.113	-0.286	-0.269	-0.111	1.000		
(9) Familienstand	-0.155	0.242	-0.084	-0.016	-0.193	-0.084	0.238	0.055	1.000	
(10) Kinder	0.091	-0.112	0.070	0.073	-0.108	-0.084	0.015	0.287	-0.007	1.000

Tabelle A3c: Korrelationen der Vignettendimensionen für die Vignetten 9 bis 12

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Geschlecht	1.000									
(2) Alter	-0.185	1.000								
(3) Ausbildung	0.382	-0.114	1.000							
(4) MPS	0.217	-0.174	0.375	1.000						
(5) Bruttoeink.	0.195	-0.102	0.003	0.642	1.000					
(6) Leistung	-0.109	0.187	0.073	-0.194	-0.257	1.000				
(7) Ökon. Situation Betrieb	-0.205	-0.138	0.060	0.057	0.070	0.065	1.000			
(8) Betriebsgröße	0.355	-0.206	0.033	0.295	0.324	-0.106	-0.040	1.000		
(9) Familienstand	0.156	0.045	-0.038	0.188	0.027	0.084	-0.163	0.208	1.000	
(10) Kinder	0.049	0.100	-0.127	-0.033	0.113	0.140	0.107	-0.085	0.089	1.000

Tabelle A3d: Korrelationen der Vignettendimensionen für die Vignetten 13 bis 16

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Geschlecht	1.000									
(2) Alter	-0.185	1.000								
(3) Ausbildung	0.382	-0.114	1.000							
(4) MPS	0.217	-0.174	0.375	1.000						
(5) Bruttoeink.	0.195	-0.102	0.003	0.642	1.000					
(6) Leistung	-0.109	0.187	0.073	-0.194	-0.257	1.000				
(7) Ökon. Situation Betrieb	-0.205	-0.138	0.060	0.057	0.070	0.065	1.000			
(8) Betriebsgröße	0.355	-0.206	0.033	0.295	0.324	-0.106	-0.040	1.000		
(9) Familienstand	0.156	0.045	-0.038	0.188	0.027	0.084	-0.163	0.208	1.000	
(10) Kinder	0.049	0.100	-0.127	-0.033	0.113	0.140	0.107	-0.085	0.089	1.000

Tabelle A3e: Korrelationen der Vignettendimensionen für die Vignetten 17 bis 20

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Geschlecht	1.000									
(2) Alter	-0.489	1.000								
(3) Ausbildung	0.184	-0.055	1.000							
(4) MPS	0.076	0.036	0.217	1.000						
(5) Bruttoeink.	0.057	-0.002	0.142	0.617	1.000					
(6) Leistung	0.031	0.056	0.061	0.106	0.134	1.000				
(7) Ökon. Situation Betrieb	-0.022	-0.001	0.233	0.068	0.120	-0.170	1.000			
(8) Betriebsgröße	-0.019	0.026	-0.300	-0.046	-0.128	-0.113	0.116	1.000		
(9) Familienstand	0.006	-0.168	-0.153	-0.301	-0.097	-0.201	-0.020	0.015	1.000	
(10) Kinder	-0.189	0.047	0.045	0.031	0.035	0.113	-0.038	0.069	-0.027	1.000

Tabelle A3f: Korrelationen der Vignettendimensionen für die Vignetten 21 bis 24

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Geschlecht	1.000									
(2) Alter	0.182	1.000								
(3) Ausbildung	-0.157	-0.070	1.000							
(4) MPS	-0.191	-0.033	0.069	1.000						
(5) Bruttoeink.	-0.230	-0.087	0.497	0.558	1.000					
(6) Leistung	-0.052	-0.096	0.127	-0.209	0.040	1.000				
(7) Ökon. Situation Betrieb	0.070	0.059	-0.128	-0.034	-0.058	-0.318	1.000			
(8) Betriebsgröße	0.029	0.137	0.172	0.295	0.414	0.132	0.116	1.000		
(9) Familienstand	-0.045	0.011	0.205	-0.230	0.137	-0.030	-0.063	-0.203	1.000	
(10) Kinder	0.167	0.000	-0.121	-0.066	-0.123	0.115	-0.206	-0.140	0.069	1.000

Tabelle A3g: **Varianzen der Vignettendimensionen für die einzelnen Vignettenabschnitte**

Vignetten- abschnitt	Geschlecht	Alter	Ausbildung	Varianz der Ausprägungen der Dimensionen						
				MPS	Brutto- eink.	Leistung	Ökonom. Situt.	Betriebs- größe	Familien- stand	Kinder
1 bis 4	0.2441	123.0	0.6561	2215	2.41e+07	0.6406	05464	0.6982	0.7974	2.057
5 bis 8	0.2491	145.4	0.6661	2346	1.81e+07	0.5217	0.6912	0.6753	0.6546	2.045
9 bis 12	0.2444	123.9	0.7256	2158	2.39e+07	0.6773	0.8281	0.7166	0.6848	2.495
13 bis 16	0.2452	120.6	0.6081	1764	1.38e+07	0.6745	0.5847	0.4947	0.6463	1.956
17 bis 20	0.2345	122.0	0.6647	1536	1.99e+07	0.6397	0.6786	0.7284	0.7068	2.281
21 bis 24	0.2483	104.1	0.6912	1659	1.71e+07	0.7466	0.5844	0.6735	0.5312	2.088