

methoden daten analysen

ZEITSCHRIFT FÜR EMPIRISCHE SOZIALFORSCHUNG

mda

2011, Jahrgang 5, Heft 2



Dorothee Behr und Evi Scholz Questionnaire Translation in Cross-National Survey Research. On the Types and Value of Annotations

Britta Gauckler und Thomas Körner Measuring the Employment Status in the Labour Force Survey and the German Census 2011. Insights from Recent Research at Destatis

Gerrit Müller Wechsel des Erhebungsinstituts und Adressweitergabe mit Widerspruchsverfahren im IAB-Haushaltspanel PASS. Ein Praxisbericht

Herausgegeben von *Christof Wolf*
Marek Fuchs
Bärbel Knäuper
Petra Stein

Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung

Die Zeitschrift wird herausgegeben von GESIS – Leibniz-Institut für Sozialwissenschaften.

Herausgeber: Christof **Wolf** (Mannheim, geschäftsführend), Marek **Fuchs** (Darmstadt), Bärbel **Knäuper** (Montreal), Petra **Stein** (Duisburg-Essen)

Wissenschaftlicher

Beirat: Hans-Jürgen **Andreß** (Köln), Andreas **Diekmann** (Zürich), Sabine **Häder** (Mannheim), Udo **Kelle** (Hamburg), Dagmar **Krebs** (Gießen), Frauke **Kreuter** (College Park, Maryland), Edith **de Leeuw** (Utrecht), Norbert **Schwarz** (Ann Arbor)

Redaktion: Paul **Lüttinger**
GESIS – Leibniz-Institut für Sozialwissenschaften
Postfach 12 21 55
68072 Mannheim
Tel.: 0621 – 1246-268
E-Mail: mda@gesis.org
Internet: www.gesis.org/MDA/

Die MDA deckt alle Fragestellungen aus dem Bereich der Empirischen Sozialforschung ab, insbesondere aus dem Bereich der Umfragemethodik. Im Vordergrund stehen Artikel, die die methodischen und/oder statistischen Kenntnisse der Profession erweitern, sowie Beiträge, die sich mit der Anwendung der Methoden der Empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder solche, in denen ein statistisches Verfahren exemplarisch angewandt wird. Obwohl der Schwerpunkt auf Umfragemethoden liegt, sind Beiträge zu anderen methodischen Bereichen willkommen.

Alle Beiträge, die zur Veröffentlichung in der MDA eingereicht werden, werden von mindestens zwei unabhängigen Gutachtern blind begutachtet.

Der Nachdruck von Beiträgen ist nach Absprache möglich. Die MDA erscheint zweimal im Jahr und steht als Printversion und online zur Verfügung. Die Registrierung für den Bezug der MDA erfolgt über die Web-Seiten von GESIS:

<http://www.gesis.org/publikationen/zeitschriften/mda/bestellung/>

Druck: Verlag Pfälzische Post GmbH, Neustadt
Gedruckt auf chlorfrei gebleichtem Papier.

ISSN 1864-6956

5. Jahrgang 2011 © GESIS, Mannheim, Dezember 2011

Inhalt

155 Editorial

FORSCHUNGSBERICHTE

157 Questionnaire Translation in Cross-National Survey Research.
On the Types and Value of Annotations
Dorothee Behr und Evi Scholz

181 Measuring the Employment Status in the Labour Force
Survey and the German Census 2011.
Insights from Recent Research at Destatis
Britta Gauckler und Thomas Körner

PRAXISBERICHT

207 Wechsel des Erhebungsinstituts und Adressweitergabe
mit Widerspruchsverfahren im IAB-Haushaltspanel PASS.
Ein Praxisbericht
Gerrit Müller

REZENSION

227 Spieltheorie. Einführung, Beispiele, Experimente.
2. überarbeitete Auflage. Andreas Diekmann, 2009
Thomas Gautschi

A N K Ü N D I G U N G E N

- 231 Ausschreibung: ALLBUS-Preis 2012
- 232 GOR 12. International GOR Conference. General Online Research. Duale Hochschule Baden-Württemberg Mannheim, 5. - 7. März 2012
- 233 AutorInnen, RezensentInnen, GutachterInnen 2011
- 234 Hinweise für unsere Autorinnen und Autoren

Editorial

Liebe Leserin, lieber Leser,

dieses Heft der MDA beschließt den fünften Jahrgang der Zeitschrift. Insgesamt wurden in diesen fünf Jahren ca. 100 Manuskripte eingereicht, von denen knapp 40 veröffentlicht wurden. Im laufenden Jahr erreichten die Redaktion jedoch deutlich weniger Manuskripte als in den Vorjahren, eine Entwicklung, die sich bei den Abonnenten nicht widerspiegelt. Die haben von 1.139 auf 1.171 zugenommen; die Zeitschrift stößt also weiterhin auf großes Interesse.

Tabelle 1 Manuskripteingang und Verbleib, 2007-2011

	Eingegangen	Angenommen	Abgelehnt	Noch offen	Annahme- quote
2007	28	9	19	-	32 %
2008	14	8	6	-	57 %
2009	25	11	14	-	44 %
2010	22	9	13	-	41 %
2011	10	2	4	4	
Gesamt	99	39	56	4	

Das aktuelle Heft wird eröffnet mit einem interessanten Beitrag, der sich intensiv mit Übersetzungsanweisungen beschäftigt, einer bisher wenig beachteten Facette der international vergleichenden Umfrageforschung. Solche Anweisungen, die sich in allen anspruchsvollen akademischen Fragebögen der vergleichenden Sozialforschung finden, sollen sicherstellen, dass die übersetzten Fragen in allen Sprachen dieselbe Bedeutung haben, also von Befragten in gleicher Weise interpretiert werden. Dorothee Behr und Evi Scholz analysieren Übersetzungsanweisungen, hauptsächlich aus dem ISSP und dem ESS, und entwickeln daraus eine Typologie. Die Autorinnen zeigen den Nutzen von Übersetzungsanweisungen auf, machen aber auch deutlich, wann solche Anweisungen problematisch sind. Aus dieser Diskussion entwickeln sie Richtlinien sowohl für das Schreiben von Übersetzungsanweisungen als auch für deren Gebrauch.

Auf ein anderes Problem international vergleichender Umfrageforschung gehen Britta Gauckler und Thomas Körner im nächsten Beitrag ein. Sie verdeutlichen, welche Schwierigkeiten entstehen, wenn ein international verein-

bartes Konzept wie der ILO Erwerbsstatus in einer nationalen Erhebung umgesetzt werden muss. Das Meistern dieser Schwierigkeiten und die Dokumentation der deutschen Adaption ist umso wichtiger, als es sich bei der von den Autoren bearbeiteten Datenquelle um den Zensus 2011 handelt. Wegen der großen Bedeutung dieser Erhebung und der Herstellung internationaler Vergleichbarkeit wurden beim Statistischen Bundesamt umfangreiche Untersuchungen durchgeführt. Diese hatten das Ziel, die Erfassung der Erwerbstätigkeit zu verbessern. Dabei kommt es auf viele Details an; der Beitrag konzentriert sich auf den Einfluss der Fragebogengestaltung.

Im dritten Beitrag dieses Hefts berichtet Gerrit Müller von den Schwierigkeiten, die ein Wechsel zwischen Erhebungsinstituten für eine laufende Panelerhebung mit sich bringt. Die möglichen Effekte eines solchen Institutswechsels sind bisher kaum untersucht. Hinzu kommt, dass Bestimmungen des Datenschutzes einen Institutswechsel erheblich erschweren können. Im konkreten Fall, dem Panel „Arbeitsmarkt und Soziale Sicherung“, wurde den zu befragenden Personen zugesagt, dass ihr Name und ihre Anschrift nicht an Dritte weitergegeben werden. Um nicht gegen diese Zusage zu verstoßen, wurden alle teilnehmenden Haushalte über den anstehenden Institutswechsel informiert, und es wurde ihnen eine Widerspruchsmöglichkeit gegen die Weitergabe ihrer Daten an das neue Erhebungsinstitut eingeräumt. Müller beschreibt in seinem Beitrag eingehend, welche Erfahrungen bei diesem Vorgehen gesammelt wurden und analysiert das Ausmaß und die Struktur der Panelausfälle, die durch den Widerspruch gegen eine Adressweitergabe entstanden sind.

Abgerundet wird das Heft durch die Besprechung des von Andreas Diekmann vorgelegten Lehrbuches zur Spieltheorie, welche Thomas Gautschi verfasst hat.

Im Namen der Herausgeber danke ich allen Autoren, Rezensenten und Gutachtern, die an diesem Jahrgang der Zeitschrift mitgewirkt haben. Außerdem gilt mein Dank der Redaktion der MDA, insbesondere Paul Lüttinger und Christa von Briel, ohne deren tatkräftige Unterstützung kein Heft der MDA erschienen wäre.

Questionnaire Translation in Cross-National Survey Research

*On the Types and Value
of Annotations*

Fragebogen- übersetzung in der international- vergleichenden Umfrageforschung

*Typologisierung und Nutzen
von fragespezifischen
Übersetzungsanweisungen*

Dorothee Behr und Evi Scholz

Abstract

Good questionnaire design and high quality questionnaire translations are vital for data comparability in cross-national survey research. With the aim to ensure comparability and prevent unintended deviations, major academically-driven studies such as the International Social Survey Programme (ISSP) or the European Social Survey (ESS) annotate the source questionnaire specifically for translation, thus providing guidance on what needs to be considered in translation. This paper studies these *translation annotations*, a topic having received scant attention in research so far. The goal of this paper is to raise awareness on this special support structure in comparative research as well as on potential pitfalls in questionnaire translation. To this end, first, translation annotations, mainly from the ESS and the ISSP, are analyzed with a view to setting up a classification of translation annotations. Second, examples of annotation types are presented together with what

Zusammenfassung

Um die Vergleichbarkeit von Daten in der international-vergleichenden Umfrageforschung zu gewährleisten, sind ein gutes Fragebogendesign beim Ausgangsfragebogen und eine hohe Qualität der Übersetzung des Fragebogens unerlässlich. Mit dem Ziel, Vergleichbarkeit zu sichern und Abweichungen zu vermeiden, kommentieren bedeutende akademische Umfrageprogramme wie das International Social Survey Programme (ISSP) oder der European Social Survey (ESS) ihre Ausgangsfragebogen speziell für die Übersetzung. Dieser Artikel untersucht diese fragespezifischen Übersetzungsanweisungen, die bisher in der Forschung kaum behandelt wurden. Ziel des Artikels ist es, diese besondere Form der Unterstützung in der vergleichenden Forschung ins Bewusstsein zu rufen sowie auf mögliche Probleme oder Fehlerquellen in der Fragebogenübersetzung hinzuweisen. Zu diesem Zweck werden erstens Übersetzungsanweisungen, hauptsächlich aus dem ISSP und

they meant for translation. Third, merits of annotations and potential criticism are discussed. Fourth, guidelines are ventured for writing translation annotations as well as for working with them. Fifth, research fields are listed in order to further explore the issue of translation annotations and its impact on translation and comparability.

dem ESS, analysiert. Hieraus resultiert eine Typologisierung von Übersetzungsanweisungen. Zweitens werden Beispiele für verschiedene Typen von Übersetzungsanweisungen zusammen mit ihrer Bedeutung für die Übersetzung präsentiert. Drittens wird der allgemeine Nutzen von Übersetzungsanweisungen, aber auch mögliche Kritik diskutiert. Viertens werden Richtlinien sowohl für das Schreiben von Übersetzungsanweisungen als auch für deren Gebrauch vorgeschlagen. Und fünftens werden Forschungsfelder aufgezeigt, die sich zur weiteren Untersuchung der Thematik der Übersetzungsanweisungen sowie zu deren Einfluss auf Übersetzung und Vergleichbarkeit anbieten.

1 Introduction

In cross-national surveys that are based on the "ask-the-same question" approach (Harkness 2003: 35), that is, where across all participating countries the same questions are asked, both a high quality source questionnaire and high quality translations are central to the overall comparability of survey data. During source text design, items should ideally be developed that match scientific needs, cross-national relevance, and translatability. To reach these goals, cross-cultural collaboration and input in formulating and selecting items for the source questionnaire is the method of choice (different forms of this type of collaboration and input are outlined in Dean et al. 2007; Fitzgerald et al. 2009; Harkness 2008a; Harkness et al. 2003; Mapi n. d.; Smith 2003). During translation then, the items should – despite the change of culture and language – maintain the intended meaning and the measurement properties of the source questionnaire. Special translation and assessment methods have been developed over the past decades in order to ensure this form of translation quality (Behr 2009; Harkness 2003; Harkness/Pennell/Schoua-Glusberg 2004; Harkness/Schoua-Glusberg 1998). With the aim to aid and "standardize" translation and to do this *before* the actual translation begins, major studies, notably the academically-driven programmes International Social Survey Programme (ISSP) and European Social Survey (ESS) or the more policy-driven Programme for International Student Assessment (PISA), have come to use *translation annotations*, that is, written aids in various forms supplementing the source ques-

tionnaire and supporting translation.¹ These translation annotations are mainly developed within the context of collaborative cross-cultural questionnaire design and draw their content from the various methods employed, such as expert review, advance translation, translatability assessment, or cross-cultural pretesting (e. g., Harkness 2008b; Jowell 1998). The overall goal of the annotations is to ensure that the translated items measure what they are supposed to measure and that equivalence to the source questionnaire is not lost because of a flawed translation.

Translation annotations, even though used in large-scale projects, have hitherto received scant attention in research. An exception being Pan, Kleiner, and Bouic (2007) and Kleiner, Pan, and Bouic (2009), respectively², who conducted an experimental study on the use of translation annotations in connection with more comprehensive translation instructions, and who concluded that annotations and instructions do indeed impact on translation but differently across target languages and to a yet unknown effect. In the light of these findings, Kleiner, Pan, and Bouic call for careful considerations when providing such guidance.

This article aims at raising awareness on translation annotations as a special support structure in comparative research as well as on potential pitfalls in questionnaire translation. To this end, (1) the article provides a classification of translation annotations, largely based on the ESS and the ISSP; (2) it looks into specific annotations and what they meant for ESS and ISSP translations; (3) it discusses the merits of annotations and potential criticism linked to them; (4) it ventures guidelines for questionnaire designers on how to write translation annotations and for translators on how to work with translation annotations; (5) it concludes with further research needs regarding translation annotations.

Before going into detail, a note on the terminology seems apt. What are generically called *translation annotations* in this article may have different names, and possibly different scope, in actual use. The ESS uses the term *translation annotation* (Harkness 2008b), thus lending its terminology to this paper. At the same time, the ISSP speaks of *translation or clarification notes*³ (e. g., ISSP), and also

- 1 Also the Comparative Study of Electoral Systems (CSES) (<http://www.cses.org/>) uses so-called *notes* in their questionnaires. These to some extent cover what we focus on in this paper. They also go beyond this paper such as when, within the realm of output harmonization, concepts are defined for which countries are to produce their own wordings. In these cases we rather speak of country-specific translation of concepts into items rather than translation within the scope of the "ask-the-same question" approach. Overall, in the CSES there is a stronger emphasis on country-specific design which is expressed in the *notes*.
- 2 In most cases, reference is made to the 2007 report rather than the 2009 paper, since the report is more detailed and contains also appendixes with the study material.
- 3 These two names already suggest that the scope of notes in the ISSP may be larger than "just" pertaining to translation.

in PISA reference is made to *translation notes* (Organisation for Economic Co-operation and Development 2005, 2009). Pan et al. (2007) refer to *question-by-question explanations* (QxQs). Regardless of the name, what information do all these annotations carry?

2 A Classification of Translation Annotations

To find out what is covered by the different types of annotations, translation annotations from the major academically-driven surveys ISSP and ESS were analyzed and categorized.⁴ The fact that these surveys are at the forefront of developments in comparative questionnaire design makes them a suitable ground for the analyses. In the ESS, the three core modules A, B, and C from the Round 2 questionnaire⁵ (European Social Survey 2004) as well as the rotating modules D and E from the Round 4 questionnaire (European Social Survey 2008) were analyzed in terms of their content. In the ISSP, four surveys were chosen for analysis of their translation annotations. These are the 2004 survey on *Citizenship* (first module of its kind), the 2005 survey on *Work Orientation* (3rd partial replication), the 2006 survey on the *Role of Government* (4th partial replication of module), and the 2007 survey on *Leisure Time and Sports* (first module of its kind). Overall, we assume that knowledge and experiences from previous rounds or waves have found their way into the translation annotations of these recent ESS and ISSP surveys.

What is the goal of the annotations in these studies? On its cover page of the Round 2 questionnaire, the *European Social Survey* provides a definition of ESS translation annotations (2004: 2):

"Throughout the questionnaire, annotations are provided to aid translation. These attempt to avoid ambiguity by providing definition and clarifications of the concept behind questions, especially where the words themselves are unlikely to have direct equivalents in other languages. Annotations are also used to mark new questions and questions from the Family Work and Well-being module that are located in the core."

4 The published PISA questionnaires (<http://nces.ed.gov/surveys/pisa/questionnaire.asp>) do not contain translation notes; therefore, PISA questionnaires could not be consulted for this paper.

5 The core modules are repeated in each ESS round. The ESS Round 2 questionnaire was chosen in view of subsequent analysis of translations of annotated items: For Round 2, four French and four German translations are available, whereas for Round 1, where the core questionnaire was set up, only three German translations are available. The Round 2 annotations are equal to the Round 1 annotations.

Learning from previous rounds, these introductory remarks have slightly changed over ESS surveys, but they still give some general information on the role of translation annotations in the ESS. To obtain a more detailed picture of translation annotations, the following classification was set up, mainly based on analysis of annotations within the above mentioned studies:

- Semantic-Pragmatic Annotations: Providing any of the following information/instructions for a *specific* term or phrase:
 - a synonym, a paraphrase, a definition, examples, implied or intended meaning components;
 - (additional) information on what is *not* included in or *not* meant by a given term or phrase;
 - explicit instruction on what should be taken into account when translating (e. g., that certain words are central in an item, that certain words should be avoided);
 - explicit instruction that countries can translate a term x by translating the alternative term y, if appropriate or more suitable in their language.
- Adaptation Annotations: Providing an instruction on permissible or required adaptations (including elements that – at the discretion of the individual countries – can either be translated or left out).
- Consistency Annotations: Providing an instruction in one of the following forms:
 - reference establishing link *within* a round or wave (e. g., to ensure consistent use of response scales within a module);
 - reference establishing link *between* rounds or waves of a survey (e. g., former rotating module question now core question; key term or phrase used already in previous rounds; item battery already used in previous rounds)⁶.
- Routing Annotations: Clarifying routing (e. g., through indicating that respondents who have answered question x with 'yes' now get the following set of questions).
- Conceptual Annotations: Providing an *overall* concept definition or providing the rationale for asking the question.

6 However, attention should be paid to the fact that what was working in a previous context may not work in a new context for various reasons. This especially refers to individual key terms or phrases. Embedded in a different survey question and a different questionnaire context, they may need to be translated differently compared to the previous use.

2.1 Semantic-Pragmatic Annotations

Semantic-pragmatic annotations are those that occur most often. Given the goal of questionnaire translation, namely to maintain the intended meaning of the source questionnaire, this comes as no surprise. Semantic-pragmatic annotations indicate what is meant in survey measurement terms by a given word or phrase (Harkness 2008b), what the intended and most salient reading of a question is (Harkness/Schoua-Glusberg 1998), or what the intended meaning of an item is (Harkness/Schoua-Glusberg 1998).

Work – should it be understood as paid work or any kind of work including housework and care of children? *Holidays* – should it be understood in the sense of Christmas and All Hallows or in the sense of vacation? *Concerns* – meant in the sense of worries or in a more neutral sense? Depending on what is meant translators will most often need to choose among several translation options. Semantic-pragmatic annotations can help to make the right choice, granted that a problem has been anticipated, of course. Furthermore, some semantic-pragmatic annotations may indicate the degree of freedom permitted in translation, as will be shown below.

We will now turn to different types of semantic-pragmatic annotations, first to those which offer synonyms, paraphrases, definitions, or implied or intended meaning components and which are most typical. We will thereby examine what they (can) mean for translation. The first example stems from the ESS, item C4:

Compared to other people of your age, how often would you say you take part in social activities (27)? Please use this card.

Annotation (27): Events/encounters with other people, by choice and for enjoyment rather than for reasons of work or duty.

What is the scope of "social activities"? If such a term were embedded in a normal text, its meaning could easily be identified by translators because of the surrounding context. The lack of context in questionnaires, however, poses a particular challenge for translators and so the annotation is welcome, especially given that comparable measurement depends on the right interpretation of the term. The annotation specifies the kind of activities involved, together with the contextual meaning of the adjective "social". French translations from Belgium, France, Luxembourg, and Switzerland thus render the term by "*activités sociales*" [*social activities*] or "*vie sociale*" [*social life*]⁷. The German translations of Austria, Germany, Luxembourg, and Switzerland⁸ use the phrase "gesellige Ereignisse oder Treffen"

7 The translated questionnaires of the ESS can be accessed at <http://ess.nsd.uib.no/>.

8 It is not clear, however, to what extent they collaborated and jointly produced their translations.

[*sociable events or meetings*]. The specification of events and meetings in the German wording may have been triggered by the translation annotation. Likely as it may be, however, such a link can only be hypothesized because translation process data revealing how the translation came about are not available (see for translation process research Behr 2009). The vague term "Aktivitäten" [*activities*] may have been considered to be inappropriate or misleading in the German context.

A special category of semantic-pragmatic annotations are those which, in addition to what *is* meant, make explicit what *is not* meant. By providing this comparison the actual meaning can be made even more explicit. Besides, the reason for adding the annotation can be traced back by translators, which can also aid translation and avoid misunderstanding of the annotation's content itself. The following example is taken from the ISSP 2004, item 2:

There are different opinions as to what it takes to be a good citizen. As far as you are concerned personally on a scale of [...], how important is it:

Never to try to evade taxes

Annotation: **Evade** has the connotation of illegality in not paying taxes owed, and does not mean "avoid", since "tax avoidance" is not illegal.

A potential misunderstanding of "evade" was anticipated. The provision of the intended meaning helps to focus the translations on the illegal aspect.

Another type of annotation makes clear what should be taken into account when translating, e. g. in terms of terminology or register. Such annotations fulfill a highlighting function. For instance, in the ISSP 2004, item 18 asks whether the following activity was done in a given time span:

Contacted, or attempted to contact, a politician or a civil servant to express your views.

It is supplemented by the following annotation:

Civil servant should be translated with the appropriate term for the public service. Do not use the term "bureaucrat."

This annotation suggests that the word "civil servant" (≠ "bureaucrat") had been chosen deliberately and that this deliberate choice was felt necessary to be conveyed to translators. If during design certain word choices in the English language are clearly preferred over others, this knowledge can be communicated to translators so that, if appropriate for the target language context, they can replicate these choices. After all, the use of differently connotated words can make a difference in measurement. For example, attitudes differ depending on whether the wording "people on welfare" or "the poor" is used in an item (Weisberg 2005: 103). French- and German-language translations of the above ISSP item refer to

"Beamte" and "haut fonctionnaire", respectively. Certainly, in this and all other cases we cannot say that only because of the annotation the translation rendering is correct. Especially in the case of "social activities", however, our experience with questionnaire translation in large international studies tells us that meaning problems and potentially biased data would have arisen without the annotation.

Annotations do not only provide definitions in the form of synonyms, paraphrases or example lists; they also *explicitly* offer alternative source wording for translation. This is illustrated by the following example from the ISSP 2006, item 3b:

There are some people whose views are considered extreme by the majority. Consider people who want to overthrow the government by revolution. Do you think such people should be allowed to ... publish books expressing their views?

Annotation: In Q3b, 'publish books' can be translated as 'have their books published'.

Equally in the ISSP 2006, a scale is annotated:

Almost none/a few/some/quite a lot/almost all

Annotation: Precode: if necessary, the difference between 'a few' and 'some' can be clarified by using a term such as 'only a few'.

Scales are a particularly critical issue in translation. It is often difficult to closely match the different dimensions (agreement, satisfaction, etc.), negations (disagree, unsatisfied, etc.), or quantifications (strongly, somewhat, etc.) on the semantic level and at the same time maintain the measurement properties of the questionnaire (Behr 2009; Harkness 2003). By providing a suggestion on how to translate, the above annotation supports countries in producing a translation with appropriate distances between the scale points, especially between "a few" and "some". A very close semantic translation of the scale might end up in non-differentiation between these scale points and thus in measurement problems.⁹

Written aids *explicitly* offering alternative source formulations fulfill two functions. On the one hand, they further define meaning. On the other hand, they provide countries with a second or alternative source thereby clarifying the leeway possible in translation. This second reason has probably influenced the choice of instrument design in PISA where two source versions, one in English and one in

9 See also Fitzgerald et al. (2009) and Behr et al. (2008) on the issue of 'some' versus 'a few' – a difficult issue in scale translation!

French, have simultaneously been used since PISA 2000.¹⁰ In this context, the PISA technical report can be quoted as follows:

- Many translation problems are due to idiosyncrasies: words, idioms, or syntactic structures in one language appear untranslatable into a target language. In many cases, the opportunity to consult the other source version may provide hints at solutions.
- The desirable or acceptable degree of translation freedom is very difficult to determine. A translation that is too faithful may appear awkward; if it is too free or too literary it is very likely to jeopardise equivalence. Having two source versions in different languages (for which the translation fidelity/freedom has been carefully calibrated and approved by consortium experts) provides national reconcilers with accurate benchmarks in this respect, and that neither back translation nor double translation from a single language could provide. (OECD 2009: 88)

Since striking a balance between faithfulness and fluency, and since identifying the needed degree of closeness (in form and content) and the possible degree of freedom (in form and content) is challenging, help in the form of approved alternative source versions in English might be useful where otherwise problems may arise. As to the issue of what is permissible in translation, Harkness/Schoua-Glusberg shall equally be quoted who already back in 1998 state that documentation could or should include, amongst others, "what is possible in terms of translation versus other forms of adaptation" (97). This statement leads us straight to the next type of annotations.

2.2 Adaptation Annotations

This group of annotations includes instructions for cultural adaptation. Intercultural questionnaire design teams should work towards preparing the source questionnaire to such an extent that anticipated permissible adaptations are marked. In this context, item 6g from ISSP 2006, received the following annotation:

6g refers to unemployment benefits. If there are no such benefits within a country (this applied to the Philippines in 1996), the question should not be asked.

The more cross-cultural input is received during development of the source questionnaire, the better either the source questionnaire is made to match diverse cul-

10 The principle of two source versions has in fact also been followed by the *Eurobarometer* of the EU since the 1970s.

tures or, if this does not work, the better the source questionnaire can be prepared in terms of permissible adaptations. This does not exclude the possibility that translators identify the need for further adaptations once the source text is finalized and translation is actually conducted.

The following types of annotations were partly identified not by looking at the source questionnaire documents themselves (as downloaded from the website) but by analyzing additional study material, such as specifically prepared translation templates (e. g., European Social Survey 2009) or background information on questionnaires. Equally, own translation experiences in the ESS, the ISSP, and further international studies were a helpful source.

2.3 Consistency Annotations

Consistency annotations refer to more "mechanical" aspects of the translation process. They support translators with regard to consistency *within* and *across* survey rounds. Translation annotations *within* a questionnaire can signal the repetition of scales or that of key terms. However, countries and translators must be aware of the fact that where one language can use the same term or wording throughout different linguistic contexts, other languages cannot. To take a simple example, the scale labels "good" and "bad" can be applied in the English language to all nouns irrespective of gender or number. In other languages, the adjectives need to be adapted in terms of gender and number to the corresponding noun. Or let us take a different example: A response scale may apply to a battery of 5 items. It is translated in relation to the first item and then not considered any more in the knowledge that it is the same scale for the rest of the items. The translation of the scale chosen may not fit with the other items, though. Re-translating is needed taking into account the entire item battery context; or, if needed and approved by the designers or central coordinating office of a study, a slightly different response scale is necessary for some of the items. On the whole, it should be well considered where, if at all, consistency notes *within* a questionnaire are offered. If offered, they should in no case lead to false security and thoughtless copy-paste activities among translators.

Consistency aids are in fact a requirement when it comes to replicating entire modules or parts of these *across* survey rounds or waves. The goal to measure change over time makes it necessary that translations are re-used unless serious mistakes are found. Otherwise it is not clear whether changes in results are due to changes in the stimulus or due to real changes in attitudes, behavior, etc. Besides, countries need to be provided with clear instructions on whether a trans-

lation is to be re-used unchanged (in the case of a real item replication) or whether it needs slight modifications (if the source item itself has undergone slight modifications). There are different ways available as to how replication across rounds or waves can be signaled. Separate documentation, prefixes or similar methods are possible.¹¹

2.4 Routing Annotations

In questionnaires where rather complex routing is used (e. g., due to the work situation of respondents), references to preceding questions may ensure understanding in context and thus facilitate translation. It is easy to imagine different scenarios, in which routing takes place, such as: 'questions 1-4 are addressed to self-employed respondents whereas questions 5-12 are addressed to employed respondents' or 'question 10-12 are asked when question 1 is answered with 'no''. Knowing about this will help translators to maintain consistency in terminology and to ensure the questionnaire flow for respondents.

2.5 Conceptual Annotations

Conceptual annotations, as understood in this paper, refer to the overall theoretical concept or dimension to be measured, or the rationale for asking specific questions. With rare exceptions, this information is not included in the source questionnaire documents themselves which were looked at for this analysis. Conceptual information may overwhelm translators, especially if they are not knowledgeable of social research. It shall be argued here that it makes sense to provide this information, if not to translators, than at least to cross-national research teams who are responsible for overseeing and signing off the translations, and who are most likely to be social scientists.

First, conceptual information enables cross-national teams to assess whether a given question can be translated following the general or current understanding of an *ask-the-same question* approach (see e. g. Behr 2009 or Harkness et al. 2004 for more specific information) or whether the question needs to be adapted and rewritten in order to measure the defined concept in the target culture. Harkness (2008a: 74) provides the following example: "Do you have difficulty reading a newspaper, even with spectacles?" If newspapers or vision aids are not readily available in certain cultures, the most salient reading (in a close

11 E. g. from the ISSP 2006: http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/issp/excerpts/excerpt_gmr_2006.pdf.

translation) will probably diverge from the intended reading: rather than collecting information on whether or not the respondent's vision is impaired, his or her reading skills may be assessed. In such cases, knowledge of the underlying concept helps to identify the need for reframing and adapting a question to make it suit the target culture. It needs to be said, however, that if a survey is designed as a comparative survey right from the start and if *asking-the-same-question* is the preferred approach in producing the measurement instruments in all countries and cultures, issues such as the one above by and large should ideally have been identified during design and either improved for cross-cultural application or earmarked for country-specific adaptation.

Second, conceptual knowledge may lead to more effective review processes. Behr (2009) shows that quite some time may be spent in team review processes on questioning the concept behind a given item. These discussions are certainly more likely if subject matter or design experts are involved in review processes, that is, those knowledgeable of the relationship between concepts and questions. A concept description, if available, can shorten conceptual discussions during review processes. However, a note of caution is also apt: Having a concept description at one's disposal can equally lead to lengthy discussions since review participants might not agree with the operationalization from concept into questions (notwithstanding cross-cultural concerns). If the source text is already finalized and cannot undergo change anymore, these discussions mostly cannot result in any changes, though. Thorough documentation of concerns and feedback to the questionnaire designers should then be the preferred approach to give designers and analysts the chance to learn from this.

While the above sections focused on a classification of annotations, the following section discusses more in general the value of providing annotations but also addresses potential criticism related to annotations.

3 Discussion

Critical voices may reject the use of annotations, particularly that of semantic-pragmatic annotations. They may say that the source question should be clear enough in the first place because otherwise English-language respondents will not know either what is meant, and so validity is put at risk in any case. Such a statement can fully be endorsed, but there are also qualifications to be made.

First, the source questions may be as well designed as possible in the sense of reducing ambiguity of meaning. However, different readings can always occur

because meaning is not static but determined by context, personal characteristics, etc. (Harkness/Schoua-Glusberg 1998). To assure that indeed the intended and most salient meaning is conveyed in the translation, an annotation may help.

Second, despite good foreign language skills, translators may not always perceive what English native speakers perceive as the most salient reading of an item (Harkness et al. 2004). Using the scenes-and-frames-semantics by Fillmore as an explanatory framework for this (e. g., Kußmaul 1995, 2007), one can also say that the scenes activated in the minds of non-English native speakers when reading or hearing a given item may not be equivalent to the scenes activated by English native speakers. The better the linguistic and cultural knowledge of translators, the closer their understanding is to native-speaker understanding or the more at least they are in a position to identify interpretation problems. In order to avoid any form of misunderstanding and, consequently, loss of item comparability, a translation annotation can be provided.

Third, it seems that often people inexperienced in translation translate source questionnaires into the target language (e. g., Hambleton 2005; Harkness et al. 2004; see also Jowell 1998). There is evidence available in translation research (e. g., Krings 1986; Kußmaul 1995) that inexperienced translators often follow too closely source text structures or tend to provide only fixed one-to-one equivalencies (those that they have learned as direct equivalences in school, for instance) that may not be appropriate in the given context. For those people, translation annotations are especially helpful since they prevent inexperienced translators from choosing potentially wrong automatic word-for-word replacements. For instance, if an item reads "failing to declare your income", then this "failing" should not be understood as "trying but not succeeding", which might be the most prominent reading of "fail" for inexperienced translators who are not native speakers of English. It should rather be understood as "not doing what is expected or needed". An annotation can clarify this meaning and thus prevent misunderstanding. However, needless to say translation annotations in general should not become the means to make up for incompetent translators. Highly skilled practitioners should be the rule in survey translation in order to ensure translation quality, and ultimately, quality of substantive conclusions drawn from cross-cultural survey data.

We would now like to address potential criticism of translation annotations or potential limitations. First, cases have occurred where the annotations have led to quite explicit translations through full integration of the annotation into the item – a fact which has led to concerns over translation annotations (see also Pan et al. 2007 on this issue). The explicit integration can be illustrated with the following example on religious identification that is measured in the ESS with item C9:

Do you consider yourself as belonging to (32) any particular religion or denomination?

Annotation (32): Identification is meant, not official membership.

The German and Austrian team both integrated the annotation's content into the item by asking:

Unabhängig davon, ob Sie Mitglied oder Angehöriger einer Kirche oder Religionsgemeinschaft sind, fühlen Sie sich einer bestimmten Religion oder Konfession zugehörig? [*Regardless of whether or not you are member (two words in German for "member") of a church or denomination, do you consider yourself as belonging to a certain religion or denomination?*]

Integrating the annotation into the item can make the intended meaning of this item clearer and may thus be truly respondent- and researcher-friendly. At the same time, this example raises the question of whether this happens at the expense of reducing measurement error in the German context compared to measurement error in the UK context. Was a close translation along the lines of the English source text considered to be confusing or unclear in German? Did the translators decide on purpose to make the non-membership aspect explicit? The Swiss-German and Luxemburg-German versions do not add the information on membership and neither do the different French versions, but this finding shall not lead to automatically rejecting one translation and favoring one over the other simply because of different levels of closeness to the source text. It may be that the more explicit German translation is indeed closer to the source text in terms of respondent's understanding than are the other translations. Effects of the difference between the various German versions and between these German versions and the English version can only be assessed empirically, though. Documentation on the difference between source text and translation would in any case have helped to understand the choices made.

The ESS is not alone when it comes to different uses of translations notes. In ISSP 2007, item 1f reads:

Get together with relatives

Annotation: who do not live in your household

Across the German-language versions¹², the translations chosen for "get together" ("besuche", "treffen") imply that the meetings occur with relatives that one does not see regularly at home; hence, an explicit additional sentence is unnecessary.

The situation is different with the French-language versions. The translators in the two French-language translations from France and Switzerland¹³ opted for translating the annotation itself and adding it in parentheses to the item. For instance, the French translation from Switzerland reads "Voir des membres de votre famille (qui ne vivent pas avec vous)" [*Seeing members of your family (who do not live with you)*]. Here, the additional parentheses seem almost imperative because seeing members of your family may include daily interaction with parents or siblings. In the South African English version, the annotation was not added in parentheses but fully integrated into the item: "Get together with relatives who do not live in your household." This addition might even be required within the cultural context of South Africa. Regarding the addition of parentheses or of further information, the presented country versions differ. This might give rise to concern, at least if one expects a translation that closely mirrors the source text. However, the different country solutions may indeed cover the intended meaning in a suitable way in all the countries. What would have been the translation outcome if no clarification had been given? Would some countries have produced a translation where regular seeing of one's family at home had been (implicitly) included? This would certainly not have been the goal of the item.

In sum, quite some changes or (formal) differences compared to the source text item can happen when the content from annotations is taken into account in translation. These changes or (formal) differences should not automatically be rejected, though. They can very well be justified and identified as the best possible solution to cover the intended meaning. This being said, there may certainly be some mishandling or misunderstanding on the use of annotations, too. Instructions on how to use translation annotations when translating are thus vital to the concept of annotating the source questionnaire for translation. More on such instructions can be found further below.

Second, annotations do not automatically lead to translations conveying the intended meaning of an item. They should thus not be misunderstood as a cure-all in view of comparability. The following example from the ISSP 2004, item 4, can be used as a case in point. The item reads:

To keep watch on the actions of government.

Annotation: Keep watch means exercise vigilance in observing government, with a view to pointing out unwarranted actions or ensuring that proper actions are conducted.

The idiomatic expression "keep watch" in relation to government actions is difficult to render in many languages, especially in terms of direct or formal equivalence. The vigilance aspect seems to have been expressed in the German-language translations with an adverb, for instance: "sehr aufmerksam verfolgen", "immer genau verfolgen" [*to (always) follow attentively/meticulously what the government does*]. Does this really tap the key ideas of "keep watch", though? One of the French-language translations reads: "Se tenir au courant des actions du gouvernement" [*keep up to date with what the government does*]. Although there is certainly a large meaning overlap, the vigilance aspect has probably not sufficiently been represented in the French translation, and possibly also not in the German translations. Something different may be measured by the translated items than intended. Could it be that the idiomatic expression is on the whole quite difficult to render in other languages, especially if one aims at a translation that is as concise as possible? A change in the source text rather than the annotation might have been the better choice to assure equivalence in meaning. This also brings us to the next point.

Third, annotations might be misleading if they contain information or clarification that is not – directly or indirectly – conveyed by the source text item itself. By taking into account the annotation content, the translation might well cover the concept but it might not be comparable to the source item any more. Any improvements should therefore be done on the source items themselves.

In the following two sections, we want to venture some suggestions for both writing translation annotations and using them. Hopefully these suggestions will encourage researchers to further explore the issue of translation guidance in comparative survey research. It is hoped as well that they can foster further discussion about best practices in comparative questionnaire design and questionnaire translation. As Kleiner et al. (2009) have already shown – there is much in survey translation and guidance that deserves further attention.

4 On Writing Translation Annotations

One should consider in the first place which help to provide to translators, e. g., if information on theoretical concepts is to be provided; if primarily aids clarifying meaning of individual words or phrases should be given; if consistency guidance is offered; or if routing help is needed. One may also want to point to overall design and layout issues, although the target group should be questionnaire designers then rather than actual translators in the latter cases. Overall, a proper balance needs to be found between useful information and information overflow. In this

context also comments on translation notes from evaluators participating in the study of Pan et al. (2007) are revealing. Questioned about their beliefs regarding question-by-question explanations to translators, a translation evaluator considered them "to be very helpful to translators, but they should be provided selectively and not for every question" (19).

At what stage in the process should the annotations be drafted? This depends on what is desired. Guiding information pertaining to theoretical concepts should be assembled within the process of good questionnaire design in general. It does not, therefore, require any drafting specifically for translation. The same generally also applies to routing.

Guiding information related to consistency on the "macro-level" (that is, consistency *between* survey waves or rounds) is a general issue of questionnaire design, too, because here decisions are made as to which items to replicate, which items to modify, etc.¹⁴ Therefore, this information will also already be available.

Guiding information related to consistency on the "micro-level" (that is, consistency *within* a survey such as repetition of response scales or important reoccurring terminology within a questionnaire) is more translation-specific. It may be added to the questionnaire at the latest *after* the questionnaire itself has been drafted and *before* it is sent out for translation. Alternatively, this guidance can be produced simultaneously with the design process. Care should be taken, though, that the chosen level of consistency notes, if they are used at all, is appropriate, otherwise translation when following these notes might end up in piecemeal or bad translation. Good translators will certainly look out for consistency with or without annotations, especially if the importance of consistency, such as consistent translation of "agree-disagree" scales for measurement purposes, has been pointed out to them.

Annotations pertaining to the meaning of items, to word choices, or to adaptations should ideally be prepared simultaneously with the source questionnaire. The more intensive cross-cultural input there is during the design phase (international design and expert review teams, advance translations, etc.), the better can translation annotations in terms of meaning, word choice, or adaptations be taken into account. In the same context, knowledge of translation problems in similar or preceding studies will be helpful in annotating the source text, especially

14 The use of computer-aided translation software, including translation memories and terminology databases, is not taken into account in these descriptions on consistency. This software can support consistency within and across survey waves as well as standardization of translation and facilitates or even makes unnecessary some of the more manual checks of translation consistency. See also Bowker (2002).

when it is known that certain terms again and again pose problems in cross-cultural implementation. When clarifying meaning, it surely cannot be the goal to rephrase the entire questionnaire for people who do not properly speak the source language or have never attempted a translation. In this context it should rather be stressed that researchers or translation commissioners need to select translators with high proficiency language and translation skills and a good expertise in translating questionnaires and familiarity with social research, or to assemble translation teams that, on the whole, meet the needed expertise (see Harkness 2003 on team approaches in translation).

All in all, a good overview of translation processes and common translation mistakes is needed in order to make appropriate decisions as to when annotations are useful. Each survey has its own needs (due to different countries participating, due to the topic, the translation actors, etc.) and so the needed degree of guidance will necessarily differ between surveys.

Eventually, one should also consider in which form information should be provided. Any information offered should be both easily accessible and assembled in the least documents as possible in order to allow easy overview of translation documentation.

5 On Instructing Translators

Since translation annotations are a special text feature of questionnaires, instructions to translators on how to work with these are needed. Pan et al.'s (2007) study shows that, in addition to the actual translation annotations, instructions have an effect, so one should carefully outline how one envisages the use of annotations. Instructions handed out to translators in experimental groups in Pan et al. (2007: 6-7) are:

"We ask that you translate the survey items with respect to their **intended meaning**. Before attempting to translate each item, read carefully the explanation that follows the item in order to get a better sense of what is intended. (The explanations are all in italics – do not translate these.)"

and

"There are culturally different ways of using language in interaction. Please translate the survey items so that they sound as natural as possible in the context of a telephone interview. This means that the questions should not sound awkward to the survey respondents, and the questions should be phrased in a culturally appropriate way. Feel free to make whatever changes necessary to accomplish this."

In our view, these instructions do not fully tell translators what they should do with the annotations other than not translating them. For the experimental group that gets both of the instructions above (other experimental groups get different sets or combinations of instructions), the second instruction may even be read as the approval for all types of changes in translation ("Feel free to make ..."). Here, of course, one must not wonder if the translation almost becomes a text on its own with few links to the source item left. Problems with the latter instruction are in fact also discussed by Kleiner et al. (2009).

It seems pressing to address the following questions when providing translators with instructions on how to use translation annotations:

- a) Are translation annotations meant as translation aid only or should the respective information be added to the questionnaire as interviewer or respondent notes, if necessary?
- b) Should they "merely" help to trigger the correct translation or should they themselves in terms of their entire content, that is, literally, be integrated into the item itself, if necessary?

These questions are not trivial considering that in both the ESS and ISSP countries have dealt differently with translation annotations. These different dealings can be due to different understandings of translation annotations, to different conceptions of what a good and comparable survey translation is, to different approaches from question designers as to what annotations involve and how efficient they are, but also to the different linguistic and cultural systems involved. What works in some languages (e. g., a close translation of the source) does not work in other languages. To convey the intended meaning, some languages need to produce more elaborate wordings than other languages, thereby even taking up examples from the translation annotation. An instruction compendium that applies to all types of annotations, items, or language combinations, therefore, cannot be given. Some approximations shall be ventured, however:

Semantic-Pragmatic Annotations: In an *ask-the-same-question* approach where the same question is asked in several countries in different languages, a principal message for the use of translation annotations, following current expectations, could be to mirror the intended meaning and to do this without spelling out all the examples or meaning features that are given in an annotation. The annotation for the question

"During the past 12 months, did (CHILD) receive all the routine preventive care that (he/she) needed?" may read: "*Preventive care*" is defined as measures taken in advance by health care providers that emphasize prevention, early detection, and early treatment of illness, injury, or long term health problems." (Pan et al. 2007, B-5)

How should translators deal with such an annotation? We would suggest that if there is a word or phrase in the target language that is usually used to denote this type of preventive care, this word should be used in translation. If, however, such a word does not exist (which may even indicate that the concept of preventive care does not exist!) or if it is not understandable by the target respondents, a paraphrase taking up some of the information from the annotation might be the only option. Still, the task should be to be as comparable *and* concise as possible *and* to help the respondent to understand and answer the question with the minimum of effort. Basic guidance such as this should be part of the general instructions to translators, though; it does not need to be included in an annotation itself. On the whole, insights into comparability requirements within cross-national survey research and into questionnaire design are deemed necessary in order to translate appropriately – and to assess annotations and their rational adequately.

Consistency Annotations: Consistency annotations can facilitate translation work. Translators should still assess whether the micro-consistency level (e. g., reoccurring terminology) from the source text, as indicated in annotations, can be maintained in the target text or whether target language grammar and idiom make a deviation unavoidable. Annotations signaling the replication of previously used translations should not prevent researchers from regularly checking that the translation wording is still up to date and in line with current developments in society.

Routing Annotations: These annotations are particularly helpful when routing is complex since they allow translation in context where, on the paper surface, the normal linear flow has been interrupted.

Conceptual Annotations: These highlight the theoretical framework behind questions. Especially for those familiar with social research this information helps to understand the purpose of the items. If, for cultural reasons, the relationship between the item (in close translation) and the concept is questioned, this should be raised with the designers, ideally before the source questionnaire is finalized. If (professional) translators are to use these annotations, they will need training on how to relate them to the items and their translations.

Despite all the value of annotations, we wish to stress that translation annotations on the whole cannot act as a substitute for subject matter knowledge, knowledge of comparability requirements, and expertise in questionnaire design and translation. This knowledge and expertise is needed in the end in order to draw the line between closeness in translation and allowed freedom in translation. And this line will always depend on which language pairs are involved in translation.

6 Conclusions

Current practice in methodological forerunner studies suggest that annotations – or whatever their name – are on the whole helpful and comparability-enhancing. Otherwise, these written aids would not still be in use in these surveys. However, some annotations are surely better than others, and this also applies to translations which draw on the help of these annotations.

Annotations mainly arise from cross-cultural collaboration. On the one hand, this collaboration makes it possible to phrase and select items that are most suitable for international comparison. On the other hand, this collaboration permits identification of issues that, despite a well-designed source questionnaire, may be problematic in translation or could potentially be misunderstood. These issues can then be “translated” into translation annotations. The overall goal of these annotations is to ensure equivalence and to ensure that data is not excluded from international comparison only because of flawed or biased translations. In any case, the better cross-cultural collaboration is integrated into the design process, the better translation and comparability needs can be catered for.

When designing translation annotations, a balance needs to be found between offering too little guidance and too much guidance. The topic, the languages and cultures involved in a survey, as well as the actual questionnaire will determine in each single case which guidance is needed. There is no universal recommendation to be made. Despite this, however, we hope that we have succeeded in raising awareness on what can be done in comparative survey research to enhance comparability and where potential pitfalls lay in translation (that hopefully can be circumvented by annotations).

To empirically determine the impact of annotations on comparability and respondent comprehension, further studies are needed where translation is done with and without these notes, where it is done by experts from different fields with or without these notes, and where subsequent assessment and cross-cultural cognitive pretesting of translated versions rounds up these studies. As Kleiner et al. (2009) rightly point out, the ultimate test on how comparable and good translations are is not a purely linguistic test based on formal differences between source questionnaires and their translations, but it rests with actual respondents. In addition, it may be worthwhile investigating systematically the impact on how much annotations lead to a greater explicitness in translation compared to the source text. Research in translation studies of translations being more explicit than the source text could be explored for these purposes (e. g., Englund Dimitrova 2005; Klaudy 1998). Knowledge of the challenging task of translation can greatly be enhanced by these studies.

References

- Behr, D., 2009: Translationswissenschaft und international vergleichende Umfrageforschung: Qualitätssicherung bei Fragebogenübersetzungen als Gegenstand einer Prozessanalyse. Bonn: GESIS.
- Behr, D., J. Harkness, R. Fitzgerald, and S. Widdop, 2008: ESS Round 4 Questionnaire and Translation Queries. London: European Social Survey, Centre for Comparative Social Surveys, City University.
- Bowker, L., 2002: Computer-Aided Translation Technology: A Practical Introduction. Ottawa: University of Ottawa Press.
- Dean, E., R. Caspar, G. McAvinchey, L. Reed, and R. Quiroz, 2007: Developing A Low-Cost Technique For Parallel Cross-Cultural Instrument Development: The Question Appraisal System (QAS-04). *International Journal of Social Research Methodology* 10 (3): 227-241.
- Englund Dimitrova, B., 2005: Expertise and Explication in the Translation Process. Amsterdam: John Benjamins.
- European Social Survey, 2004: Source Questionnaire, Round 2, 2004/5 Final Version, Amendment 03, 21.07.2004. London: Centre for Comparative Social Surveys. Retrieved December, 07, 2009, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=63&Itemid=356.
- European Social Survey, 2008: Source Questionnaire Amendment 03 (Round 4, 2008/9). London: Centre for Comparative Social Surveys. Retrieved December, 07, 2009, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=63&Itemid=356.
- European Social Survey, 2009: ESS R4 Translation Templates. London: Centre for Comparative Social Surveys. Retrieved August, 23, 2011, from http://www.europeansocialsurvey.org/index.php?option=com_docman&task=cat_view&gid=92&Itemid=80.
- Fitzgerald, R., S. Widdop, M. Gray, and D. Collins, 2009: Testing for Equivalence Using Cross-National Cognitive Interviewing. City University London, Centre for Comparative Social Surveys. Working Paper Series 1. Retrieved June, 04, 2009, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=148&Itemid=340.
- Hambleton, R. K., 2005: Issues, Designs, and Technical Guidelines for Adapting Tests into Multiple Languages and Cultures. Pp. 3-38 in: R. K. Hambleton, P. F. Merenda, and C. D. Spielberger (Eds.): *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Harkness, J., 2003: Questionnaire Translation. Pp. 35-56 in: J. Harkness, F. J. R. van de Vijver, and P. Ph. Mohler (Eds.): *Cross-Cultural Survey Methods*. Hoboken, NJ: Wiley.
- Harkness, J. A., 2008a: Comparative Survey Research: Goal and Challenges. Pp. 56-77 in: E. de Leeuw, J. J. Hox, and D. A. Dillman (Eds.): *International Handbook of Survey Methodology*. New York: Lawrence Erlbaum Associates.
- Harkness, J. A., 2008b: Round 4 ESS Translation Strategies and Procedures. Retrieved September, 12, 2008, from http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=66&Itemid=112.
- Harkness, J., B.-E. Pennell, and A. Schoua-Glusberg, 2004: Survey Questionnaire Translation and Assessment. Pp. 453-473 in: S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, and E. Singer (Eds.): *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: Wiley.
- Harkness, J. A. and A. Schoua-Glusberg, 1998: Questionnaires in Translation. Pp. 87-127 in: J. Harkness (Ed.): *ZUMA-Nachrichten Spezial 3: Cross-Cultural Survey Equivalence*. Mannheim: ZUMA. Retrieved September, 19, 2011, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial3.pdf.
- Harkness, J., F. J. R. van de Vijver, and T. P. Johnson, 2003: Questionnaire Design in Comparative Research. Pp. 19-34 in: J. Harkness, F. J. R. van de Vijver, and P. Ph. Mohler (Eds.): *Cross-Cross-Cultural Survey Methods*. Hoboken, NJ: Wiley.

- ISSP Research Group: International Social Survey Programme 2004: Citizenship (ISSP 2004). GESIS Data Archive, Cologne, Germany, ZA3950_bq Basic (Source) Questionnaire; doi:10.4232/1.10078.
- ISSP Research Group: International Social Survey Programme 2005: Work Orientation III (ISSP 2005). GESIS Data Archive, Cologne, Germany, ZA4350_bq Basic (Source) Questionnaire; doi:10.4232/1.4350.
- ISSP Research Group: International Social Survey Programme 2006: Role of Government IV (ISSP 2006). GESIS Data Archive, Cologne, Germany, ZA4700_bq Basic (Source) Questionnaire; doi:10.4232/1.4700.
- ISSP Research Group: International Social Survey Programme 2007: Leisure Time and Sports (ISSP 2007). GESIS Data Archive, Cologne, Germany, ZA4850_bq Basic (Source) Questionnaire; doi:10.4232/1.10079.
- Jowell, R., 1998: How Comparative is Comparative Research? *American Behavioural Scientist* 42 (2): 169-177.
- Klaudy, K., 1998: Explication. Pp. 80-84 in: M. Baker and K. Malmkjær (Eds.): *Routledge Encyclopedia of Translation Studies*. London: Routledge.
- Kleiner, B., Y. Pan, and J. Bouic, 2009: The Impact of Instructions on Survey Translation: An Experimental Study. *Survey Research Methods* 3 (3): 113-122.
- Krings, H. P., 1986: *Was in den Köpfen von Übersetzern vorgeht: eine empirische Untersuchung zur Struktur des Übersetzungsprozesses an fortgeschrittenen Französischlern*. Tübingen: Narr.
- Kußmaul, P., 1995: *Training the Translator*. Amsterdam: John Benjamins.
- Kußmaul, P., 2007: *Verstehen und Übersetzen: ein Lehr- und Arbeitsbuch*. Tübingen: Narr.
- Mapi Research Institute (n. d.). *Translatability AssessmentSM*. Retrieved August, 2, 2009, from <http://www.mapi-institute.com/linguistic-validation/services/translatabilityassessment>.
- Organisation for Economic Co-operation and Development. 2005: *PISA 2003 Technical Report*. Paris: OECD.
- Organisation for Economic Co-operation and Development. 2009: *PISA 2006 Technical Report*. Paris: OECD.
- Pan, Y., B. Kleiner, and J. Bouic, 2007: The Impact of Instructions on Survey Translation: An Experimental Study. *Survey Methodology* 18. Retrieved December, 01, 2009, <http://www.census.gov/srd/papers/pdf/rsm2007-18.pdf>.
- Smith, T. W., 2003: Developing Comparable Questions in Cross-National Surveys. Pp. 69-91 in: J. Harkness, F. J. R. van de Vijver, and P. Ph. Mohler (Eds.): *Cross-Cultural Survey Methods*. Hoboken, NJ: Wiley.
- Weisberg, H. F., 2005: *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.

Addresses of the Authors

Dr. Dorothee Behr
GESIS – Leibniz-Institut für Sozialwissenschaften
B 2, 1
68159 Mannheim
Dorothee.Behr@gesis.org

Dr. Evi Scholz
GESIS – Leibniz-Institut für Sozialwissenschaften
B 2, 1
68159 Mannheim
Evi.Scholz@gesis.org

Measuring the Employment Status in the Labour Force Survey and the German Census 2011

Insights from Recent Research at Destatis

Zur Erfassung der Erwerbstätigkeit in der Arbeitskräfteerhebung und im Zensus 2011

Erkenntnisse neuerer Forschung bei Destatis

Britta Gauckler und Thomas Körner

Abstract

Measuring the employment status according to the labour force concept of the International Labour Organization (ILO concept) is far from being straightforward. By defining employment as any economic activity of at least one hour per week, the ILO guidelines apply a strictly economic concept which risks to conflict with everyday life perception. Consequently, small and informal jobs are likely to be overlooked in household surveys.

The employment status according to the ILO concept is the conceptual backbone of the Labour Force Survey (LFS) and at the same time it is required as a compulsory variable for the European census round. Due to its particular importance, the German Federal Statistical Office (Destatis) carried out extensive research in order to improve the survey measurement in the LFS and at the same time to develop a suitable approach for the requirements of the household survey carried out under the census 2011. The paper will focus on the impact of the questionnaire design. It is based upon the results

Zusammenfassung

Die Erfassung des Erwerbsstatus gemäß dem international vereinbarten Labour-Force-Konzept (ILO Konzept) kann in Erhebungen nicht ohne Weiteres umgesetzt werden. Insbesondere das Ein-Stundenkriterium, nach dem jede ökonomische Aktivität ab einer Stunde pro Woche als Erwerbstätigkeit definiert wird, steht nicht immer im Einklang mit einem Alltagsverständnis von Erwerbsarbeit und kann daher bei Befragungen zu Schwierigkeiten führen. Infolgedessen sind kleinere oder informelle Tätigkeiten in Haushaltsbefragungen schwieriger zu erfassen. Der Erwerbsstatus gemäß dem ILO Konzept ist nicht nur das zentrale Merkmal der Arbeitskräfteerhebung, sondern auch ein Pflichtmerkmal der europaweiten Zensusrunde in 2011. Gemessen an der Bedeutung wurden beim Statistischen Bundesamt (Destatis) umfangreiche Untersuchungen durchgeführt, um die Erfassung der Erwerbstätigkeit in der Arbeitskräfteerhebung zu verbessern und um einen geeigneten Ansatz für die Haushaltestichprobe des Zensus 2011 zu entwickeln. Dieser Beitrag konzentriert

of a follow-up survey which was carried out in the Microcensus in 2008, a large-scale field test for the preparation of the census as well as a number of cognitive laboratory tests.

sich auf den Einfluss der Fragebogengestaltung. Er baut auf den Ergebnissen einer im Jahr 2008 durchgeführten Nachbefragung im Mikrozensus, den Ergebnissen eines Feldtests im Rahmen der Vorbereitung für den Zensus 2011 sowie verschiedenen kognitiven Tests im Pretestlabor auf.

1 Introduction^{1,2}

Statistical measurement can be described as a sequence of translation processes (Radermacher/Körner 2006). In the case of household surveys, the translation from the language of the survey researcher into the language of the respondent plays a key role. The translation problem stems from the fact that the respondents are usually not familiar with the underlying concepts of the researchers. Nevertheless, they have to provide information that fits the need of the researches. Getting this translation problem under control is a basic prerequisite for achieving accurate and valid results.

There are many reasons why statistical concepts will normally deviate from everyday life's perception. Research concepts need to be more systematic, more differentiated, and universally applicable than necessary for (or even viable in) everyday life. Furthermore, the concepts of statistical measurement will always depend upon the specific research problem whereas everyday life's concepts tend to focus on practical viability and therefore lack the methodological rigour that characterises analytical concepts. Also the discipline of the researcher has eminent implications: Sociologists usually focus on different aspects of a phenomenon than economists or lawyers do, for instance.

The specific conceptual details are not easily taken into account by the researcher's most valuable source of information – the respondent. It is the task of the questionnaire to translate what the researcher wants to know into a language that the respondent understands. As often demonstrated in questionnaire testing studies – and shown in the results presented in this paper – this process, or operationalisation, is cumbersome, imperfect, and usually involves pragmatic compromises.

1 An earlier version of this paper was presented at the European Conference on Quality in Official Statistics (Q2010), Helsinki, Finland, 3-6 May 2010. The views expressed in this paper are those of the authors and do not necessarily represent the views of their agency.

2 The authors thank the two anonymous referees for their constructive and helpful comments.

Measurement of the employment status is a particularly striking case in point. At first sight not being a particularly complex phenomenon, employment seems to be a familiar notion to almost everybody. Nevertheless, respondents regularly have deviating definitions from researchers – be it in economy, sociology or other areas. Measuring the number of employed persons therefore requires particular methodological attention in order to adequately portray what statisticians are supposed to measure.

This is especially important for European Union surveys, like the Labour Force Survey or Censuses; statistics will not be relevant for European Union policies, unless they are internationally comparable. The aim of international comparison has important implications: For international comparisons, concepts have to be as independent as possible regarding their specific institutional and cultural context. In the case of employment, for example, all connotations with national social or labour legislation have to be avoided. Normally this leads to an increased level of abstractness and accessibility for the respondents is further decreased. International concepts therefore often require special efforts in the operationalisation process.

At the same time and quite paradoxically, international concepts are often more directly relevant for specific policies. At the European level, the Excessive Deficit Procedure, but also the Open Method of Coordination (OMC) for social policies are frameworks in which statistical results are being used as target values for policy programmes. Employment figures are at the forefront of variables tailored for such purposes.³ In the Europe 2020 strategy, the employment rate has been defined as one out of only five target indicators for “smart, sustainable and inclusive growth”. These indicators are going to be strictly monitored under this programme, requiring a solid and reliable statistical basis.

Against this background, it becomes evident why the European Union Labour Force Survey and the decennial censuses mandated by EU legislation are subject to strict quality control and have led to intensive methodological research. Many of these activities focussed on the problem of how to design the survey questionnaire in order to achieve a maximum of accuracy in the number of employed persons. Currently survey researches have ample tools at their command to calibrate measuring instruments to fit their requirements (for a recent overview

3 In a recent communication, the European Council for Economic and Financial affairs underlines this point by stating that the council “considers it important that the statistical indicators and underlying data to be used for the enhanced economic policy coordination are firmly based on a sound statistical methodological framework, compiled in accordance with the principles laid out in the European Statistics Code of Practice” (Press communication issued on 17 November 2010; see http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ecofin/117762.pdf (23.11.2010).

from official statistics, see Brancato et al. 2006). Regarding employment, the main research problem has been how people, for whom the statistical definition and the everyday life's perception are in conflict, can nevertheless be accurately captured in the survey interview.

This paper summarises the main insights from recent research carried out at the Federal Statistical Office Germany (FSO). It first outlines the main features of the internationally agreed concept for the measurement of employment – the Labour Force Concept of the ILO (chapter 1). Chapter 2 presents the problems of implementing the Labour Force Concept in the German Labour Force Survey and summarises the results of a follow-up survey in chapter 3. The findings of this research revealed some important problems and resulted in a revised operationalisation of employment. This revised operationalisation are being used both in the questionnaire of the Labour Force Survey but also in the questionnaire of the household sample within the German Census in 2011. The operationalisation for the census and the specific aspects in the context of a population census are presented in chapter 4.

2 The Labour Force Concept of the ILO – a Challenge for Statistical Measurement

In the field of employment and unemployment, the definition of the labour force concept of the ILO (ILO concept) is the conceptual backbone of both the EU Labour Force Survey (LFS) and the 2011 Census of the European Statistical System (ESS). While unemployment and employment are by far the most important variables within the LFS, they are also prominent – albeit to a lesser degree – for the census.

The labour force concept exclusively and exhaustively divides the population of a given country into two broad classes: Those who are considered "economically active" (or in the labour force) and those who are considered "economically inactive" (or outside the labour force). Persons in the labour force are additionally subdivided in employed persons and unemployed persons. Figure 1 outlines the main criteria of the ILO concept. It should be noted that for certain groups of employed persons special rules apply, which are however not relevant in the context of this paper.

The labour force concept was first agreed as an international standard at the Sixth International Conference of Labour Statisticians (ICLS) in 1947 (see Rengers 2004). The last major changes have been endorsed by the 13th ICLS in 1982 and the 16th ICLS in 1998 (ILO 1982; 1998). The UNECE recommendations for the 2010

population censuses (UNECE 2006) are largely built around the ICLS resolutions and guidelines, which are almost literally reproduced in many parts. It should be noted that in order to measure the employment status in household surveys, the labour force concept needs to be further operationalised. In the case of the LFS, a series of regulations guarantees a large degree of conceptual harmonisation (namely EU Council Regulation no. 577/98 and the EU Commission regulations no. 1897/2000 as well as no. 1575/2000). For censuses, similar operationalisations are laid down in the recommendations of the United Nations Economic Commission for Europe (UNECE), but with slightly more flexibility regarding the practical implementation (EU Council Regulation no 763/2008; EU Commission Regulation 1201/2009).

The basis for the distinction between the economically active and inactive parts of the population is the notion of work: The economically active population comprises all persons who provide the supply of labour, as employed or as unemployed, for the production of goods and services. The background of this basic definition is the close connection of the labour force concept with the System of National Accounts (SNA): In order to be consistent with National Accounts, all people who provide the supply of labour for the production of goods and services have to be included, even if the input towards the production is as small as one hour per week.⁴

Figure 1 The Labour Force Concept of the ILO (as operationalised in the EU-LFS)

Employed	Unemployed	
<ul style="list-style-type: none"> • ≥ 15 years old and either • at work for at least one hour (as employee or self-employed) or • with a job but not at work (formal job attachment) 	<ul style="list-style-type: none"> • 15 – 74 years old and • without work (or less than one hour) and • actively seeking job in the last four weeks and • currently available for work (2 weeks) 	<ul style="list-style-type: none"> • not employed and • not unemployed
Economically active population		Inactive population

4 This close link to the National Accounts also has consequences for the data analysis. As often argued, the employment status according to the ILO concept is more relevant for economists and has some limitations in social sciences and social policy analysis. In these areas it is helpful to supplement the results on the ILO employment status by additional indicators (see, e. g., Eurostat 2010, Rengers 2010).

Measuring the ILO employment status in household surveys and censuses is challenging in several respects. As mentioned in the introduction of this paper, translation is one of the main problems. The ILO defines employment in the broadest term, whereby one hour per work counts as being employed. A small job of one hour per week is enough. Such a definition will sometimes be in conflict with the respondent's everyday life perception. A student or a pensioner with a small side-job will often not consider themselves as being employed. Consequently, questionnaire design and fieldwork have to find ways to detect such small jobs as well.⁵

Already back in the 1980s, these problems have inspired some research regarding the measurement of the ILO employment status (see, e.g., Schwarz 1987; Hussmanns/Mehran/Verma 1990; van Bastelaer 1994). The latter publication even culminated in an EU resolution (no. 1897/2000) which stipulates that a whole sequence of questions is needed for a proper measurement within the LFS. The regulation also provides detailed practical principles for question wording and question order. Despite these common rules, the results of the LFS remained incoherent compared with other data sources in Germany, like administrative data or data from the National Accounts. These incoherences lead to further research with the German Labour Force Survey and the Mikrozensus. The results of this research revealed some new problems and resulted in a revised operationalisation of employment. This revised operationalisation will be used both in the Labour Force Survey but also in the household sample within the next German Census. The operationalisation for the census and the specific aspects in the context of a population census are presented in section 4.

3 Measuring Employment in the German LFS

The German Labour Force Survey is currently integrated in a large scale household survey covering one percent of the population every year, the "Mikrozensus" (German for "micro census"). The LFS and the Mikrozensus share the largest part of the survey variables, but have some important differences regarding the survey objectives (the LFS emphasising more on short-term labour market reporting). Survey participation to the Mikrozensus is required by law. As a result of the mandatory participation annual response rates of about 95 % are being achieved. The majority

5 Further problems arise for persons with a job, but not working during the reference week. However, these problems, as well as the whole area of the measurement of unemployment are beyond the scope of this paper.

of the interviews, 77 %, are carried out as computer assisted personal interviews (CAPI), whereas about 20 % of the interviews are still a self-administered paper-and-pencil questionnaire (for an overview on the methodological characteristics of the Mikrozensus, see Statistisches Bundesamt 2009). In contrast to the census, the Mikrozensus has a vast survey programme of more than 150 questions.

In recent years, the number of employed persons in the Mikrozensus and the LFS was subject to intensive debate. Users were confused about deviations of the results from Mikrozensus and LFS compared to other data sources. For example, in 2009 there were 40.17 million employed persons according to National Accounts, whereas the Mikrozensus had only 38.66 million persons, a difference of 1.5 million employed or 3.8 %. The incoherence largely varied regarding specific groups of the employed. While there were only minor differences for the self-employed, the Mikrozensus shows 2.3 % more (non-marginal) employees and 37 % less marginal employees. The reasons for these differences are manifold and concern not only the survey measurement in the Mikrozensus, but also the specific opportunities of estimation within the National Accounts (see Körner/Puch 2009; Körner/Puch 2011).

This paper focuses on questionnaire-related measurement problems in household surveys as one major source of these incoherences. It should be noted that the questionnaire design and wording is not the only factor impacting upon data quality. Further important sources of error include interviewer and mode effects, selection bias, bias due to the weighting scheme as well as non-response bias. This might pose a problem as the various sources of error do not stand side-by-side, but tend to interact with one another. For example, effects due to the wording of questionnaire will to some extent vary according to the data collection mode chosen. However, these other sources of error are not within the scope of this study.

3.1 Existing Measurement Problems

Measuring employment according to the guidelines of the ILO is not straightforward for several reasons. As already mentioned, the basic problem is that the definition of employment according to the ILO concept deviates from everyday life perception. For instance, people with side jobs still tend to perceive themselves as pensioner, pupil, or student first, while not always considering themselves as "employed". These people sometimes will be guided by their main social status and not report small side jobs (this phenomenon is referred to as main status effect in the following). Therefore, the questionnaire has to translate the ILO concept into everyday life concepts, so the respondents can understand.

In the Mikrozensus (and consequently in the LFS), there is a sequence of several questions trying to cover all sorts of small jobs (see figure 2). After an initial question asking whether the respondent has worked for pay or profit in the reference week, further questions ask for employment situations which are likely to be overlooked. These employment situations include working as an unpaid family member, being in marginal employment, or having a job without being at work in the reference week. A final question tries to catch all those employed yet still not covered by all these questions by asking "Even if employment is not the main activity, one can still work to earn some extra money. Did you have such a paid activity last week?" Since the year 2003, this sequence of questions has been gradually enlarged in the German LFS up to this present operationalisation, with the result that the difference between the Mikrozensus and the National Accounts decreased from 2.8 million in 2004 to 1.5 million in 2007, 2008, and 2009. Consequently, the comparability over time is restricted for instance in the year 2005 (for a detailed discussion see Körner/Puch 2009; Körner/Puch 2011).

Nevertheless, the sheer number of questions was not easy to deal with in the fieldwork as a number of respondents had the feeling that there was some redundancy in this block of questions (which in fact is true). It was concluded that further progress could only be made by developing a new approach of measuring the employment status. This new approach tried to take into account everyday life's perception of the respondents and tailoring the questions to the situation of different groups of respondents, without giving up the objective of strictly applying the criteria laid down in the resolutions of the ILO as well as the relevant EU regulations.

3.2 Results of a Follow-Up Survey

In order to analyse the measurement of the employment status with the current operationalisation, the Federal Statistical Office (together with several regional statistical offices) carried out a follow-up survey to the Mikrozensus in 2008. The follow-up survey had the objective to quantify the impact of (1) the main status effect, (2) the effect due to proxy interviews (the proxy rate in the Mikrozensus is about 25 %) and (3) misclassifications regarding marginal employment. The hypotheses regarding these effects had been developed during a telephone survey carried out on a monthly basis from 2005 to 2007 and were to be tested in the follow-up survey. In order to carry out this analysis, a new operationalisation was developed, which was deemed to measure particularly small jobs in a more comprehensive way.

The households that had participated in the Mikrozensus were recruited for the follow-up survey after the regular interview; about 20 % of the households agreed to participate. Between April and August 2008, about 4,000 persons from the age of 15 to 74 years were re-interviewed about their employment status using CATI and the revised operationalisation. Information was to be provided for the same reference week as in the regular interview. The survey instrument of the follow-up survey was specifically optimised with the aim to cover smaller secondary jobs and revised after a cognitive pretest at the Leibniz Institute for the Social Sciences (GESIS). When respondents had given their consent, the micro data from the follow-up survey were linked with the original data from the latest Mikrozensus interview. This link enabled comparative analyses and indicated groups of people for whom it is particularly difficult to record the ILO employment status in the Mikrozensus. In addition, the direct interviews of both the respondents who had participated personally in the initial interview and the proxy respondents made it possible to form control and reference groups. This way proxy effects and main status effects could be studied separately (for more details, see Statistisches Bundesamt 2008; for an English summary, see Köhne-Finster/Körner 2009).

Table 1 Employed in the Follow-Up Survey and in the Mikrozensus Interview (Adapted from Statistisches Bundesamt 2008)

	Cases in the follow-up survey	Thereof employed ...		Employment rate (in %)	
		... in the LFS	... in the follow-up survey	LFS (respondents in follow-up survey)	Follow-up survey
Pupils	210	41	64	20	31
Students	140	62	70	44	50
Pensioners	734	27	62	4	8
Registered unemployed	217	42	70	19	32

The follow-up survey provided many useful insights in various areas. Regarding the employment status, the most important points are as follows (see Statistisches Bundesamt 2008; Köhne-Finster/Lingnau 2008):

- For 8 % of the respondents a different employment status was identified compared to the regular interview. In most cases the status from the regular interview to the follow-up survey switched from unemployment or inactivity to employment: 20 % of the unemployed and 8 % of those economically not active indicated a kind of employment in the follow-up survey, but not in the Mikrozensus. There were also employed who no longer said to be employed

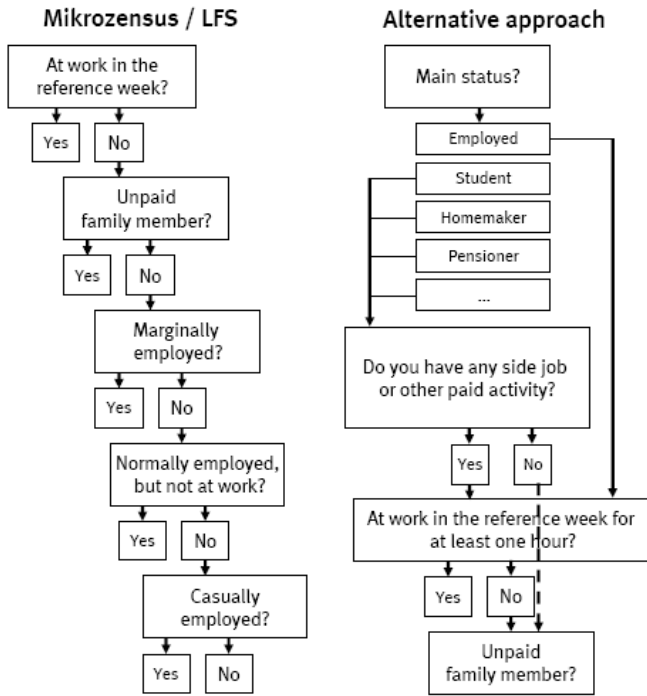
in the follow-up survey, but to a much lesser degree. In total, 3 % of the respondents indicated an employment in the follow-up survey, which they did not indicate in the initial interview.

- Not surprisingly, the largest parts of employment not indicated in the regular interview were small jobs or side jobs. For instance, the share of employed persons in all respondents (employment rate) was higher in the follow-up survey compared to the Mikrozensus interview (see table 1). The employment rate of pupils rose from 20 % to 31 %, of students from 44 % to 50 %, of pensioners from 4 % to 8 %, and that of registered unemployed from 19 % to 32 %. Note that, despite the small number of cases, all these differences are statistically significant.
- Due to the modest sample size of the follow-up survey, the effect of the differences on the total number of employed persons allowed a broad estimation only. Nevertheless, if one takes the differences for the various population groups as a starting point, one might conclude that the Mikrozensus currently understates the number of employed persons by about 900.000 persons (Statistisches Bundesamt 2008: 58).
- The analysis from the follow-up survey concluded that the differences in the results were to the largest part due to the main status effect, i. e. people with a minor job tend not to indicate this job in the survey, when their main status is not employment. In comparison, it was possible to show that the effects due to proxy interviews were not equally important. Whereas for the proxy interviews, in 11 % of the cases an employment was not detected in the Mikrozensus, it was not detected in the 7 % of the cases of the direct interviews. This confirms findings of other studies that proxy interviews slightly underestimate employment rates (see, e. g., Kleven/Lagerstrøm/Thomsen 2008; Thomsen/Villund 2011).

The follow-up survey came to the conclusion that the measurement of the employment status could be improved by an alternative questionnaire design. It should however be noted that the differences might partly also be due to diverging methodological choices in the follow-up survey and the Mikrozensus: The follow-up survey was carried out by CATI, while most Mikrozensus interviews use CAPI as a mode. Furthermore in the Mikrozensus, the interviews are simultaneously carried out for the entire household, while the follow-up survey focussed on individual people.⁶

6 An exact quantification of the effects will be possible in the Mikrozensus 2011. Here the introduction of the new operationalisation of the employment status is being accompanied by a split-ballot experiment. One tenth of the sample is randomly assigned as a control group. This group is continued to be interviewed with the old operationalisation. With this experimental design it will be possible for the first time to get a highly reliable estimate regarding the resulting break in the time series of the Mikrozensus.

Figure 2 Two Operationalisations of the Employment Status in the German LFS / German Census (Adapted from Köhne-Finster/Lingnau 2008)



3.3 A Revised Operationalisation of Employment

As the main effect of the differences was attributed to the changes in the questionnaire design, it was decided to adopt the new approach in the operationalisation of the employment status in the regular Mikrozensus and LFS from 2011 onwards and to use a similar approach in the German population census in 2011. The new approach radically changes the sequence of question. Instead of starting with a question regarding any paid work in the reference week, it uses the main status of the respondent as point of departure.⁷ The rationale for this change is to start the questions on employment with a question that the respondent can answer using

7 A similar approach was used in the 2006 round of the European Social Survey (see Erikson/Jonsson).

everyday life's perception. The respondents are asked to say what characterizes their current situation best (right side of figure 2, which gives a simplified outline of the sequence of the employment questions). All those who answer another item than any kind of employment are filtered to a further question, which (in the case of computer assisted interviewing) asks for various types of small jobs, casual jobs, jobs in the context of workfare programmes and side jobs. In contrast to the follow-up survey, the Mikrozensus could not tailor the questions according to the specific situation of various groups of respondents at this stage. For example, pensioners have been asked in the follow-up survey "Even as a pensioner, one can still have a side job to earn some extra money or to have an assignment. Do you have such a job?" The students were asked "Even as a student one can work in addition to studying to earn some money. Do you have such a job?" In the Mikrozensus these kind of questions tailored to specific groups of respondents is technically feasible, but has not been implemented in the revised questionnaire in 2011. Similar to the census approach (see below) all main status groups are now being asked an identical question concerning small jobs or any other kind of paid activity.

Only after having measured the main status and the existence of a side job, respondents are asked the traditional questions on the ILO employment status, namely whether the job was carried out in the reference week or not. Traditionally, it is assumed (although only rarely backed up by empirical evidence) that respondents tend to not answer questions regarding the ILO status correctly when they have previously been asked about their main status: "When the general question on the main activity is immediately followed by the question about currently working or having a job, the respondent is likely to interpret the latter question as also referring to the usual situation." (van Bastelaer 1994: 283). The results from the follow-up survey show that this is not necessarily the case. On the contrary, the distinct questions on the main status and small side jobs obviously clarify both concepts for the respondent and even lead to a more comprehensive measurement of employment.

The new operationalisation used in the follow-up survey necessitates a computer-assisted questionnaire as it would otherwise lead to very complex skip instructions. In preparing the implementation of the Mikrozensus, a simplified version was developed, which now can also be applied in a self-administered questionnaire. A cognitive pretest in the context of a register survey on marginal employment has led to encouraging results: Among 20 test persons (who were all known to have marginal or informal jobs), only one was not correctly classified. This test person was a registered unemployed with a full-time job in the context of a workfare programme ("Ein-Euro-Job"). This specific arrangement caused a misunderstanding in the questionnaire which was subsequently corrected.

4 Measuring Employment in the Next German Census

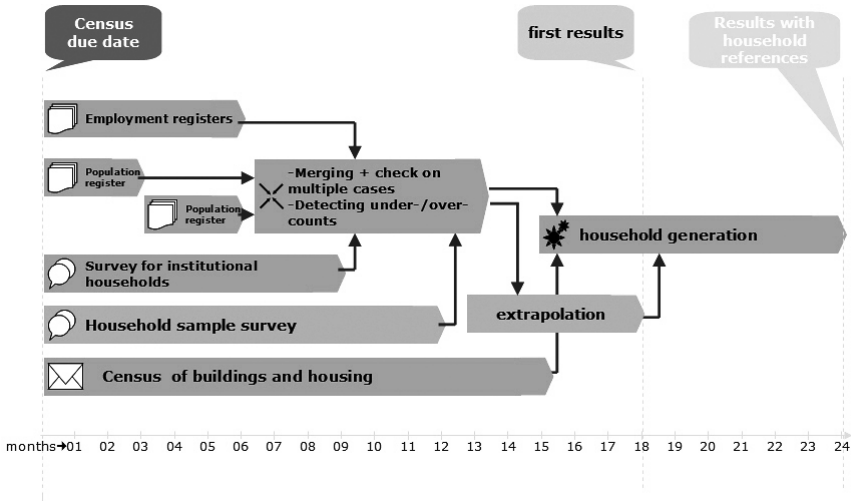
The intensive testing and ongoing work towards an improved set of questions for the measurement of employment for the Mikrozensus had crucial synergy effects also for the census questionnaire within the household sample survey. The measurement of employment according to ILO's labour force concept is one of the core topics within the European requirements and therefore within the German census (EU Commission Regulation no 1201/2009).

The 2011 German census is not traditional in the sense of a complete enumeration of the whole population via face-to-face interviews. In order to reduce cost and lower the burden on citizens, the German census is a combination of a register-based census, a conventional census and a sample survey. The information for the register-based census is drawn from different existing administrative registers, mainly the population registers of the municipalities and the registers of the Federal Employment Agency, combined with primary sources mainly the postal census of buildings and housing and the household sample survey (for more detailed information see Statistische Ämter des Bundes und der Länder 2011). For the organisational preparation a register of addresses and buildings was built. This register functions as a link between the different data sources and as a basis for buildings with living space which supports the primary surveys (Meder 2009: 2).

Data that cannot be drawn from registers are complemented by primary surveys: The census of buildings and housing gathers information via a mail questionnaire of all 17.5 million property owners and administrators of buildings with residential space. The household sample survey interviews a maximum of 10 % of the population via a face-to-face interview. Moreover, information about residents in special buildings (such as institutions or buildings providing collective accommodation) is also collected via face-to-face interviews. As information on household relationships is not directly available from population registers, the bodies of official statistics have developed a new approach, the household generation procedure. Here, resident household relationships are determined by means of characteristics from the population register and the census of buildings and housing which can be evaluated statistically (see Vorndran 2004).

The household sample survey (where the employment questions are being asked) is needed for the estimation of over- and undercounts in the population registers, and is mainly used for collecting information that is otherwise not available. The focus of the household sample survey is – besides the collection of socio-demographic information and information about religion and migration – the collection of information about education and employment.

Figure 3 Structure and Release Dates of the German Census 2011 (Overview)



4.1 Operationalising the Labour Force Concept for a Population Census

Within the questionnaire for the household survey the collection of employment data is the lengthiest and among the more complicated parts of the questionnaire. Measuring the employment status according to the labour force concept is particularly challenging for a census. Due to the large number of interviews, the questionnaire has to be kept as short as possible. Moreover, the response burden for the respondents has to be kept as low as possible. At the same time, a high quality standard has to be achieved during the data collection.

There are certain important methodological differences between labour force surveys and population censuses, which have to be taken into account while designing the employment questions. One difference and advantage is that the census questionnaire in general is quite short compared to the Mikrozensus questionnaire. Nevertheless, there are also several methodological challenges. The results of the follow-up survey proposed a more tailored design for the employment questions, adapting the wording to specific situations of different population groups. However, for the census 2011 household survey it will not be possible to introduce customised questions, for instance for pensioners or students due to the

fact that the household survey is conducted without computer assistance.⁸ Therefore, the possible difficulties with certain groups had to be solved in the questionnaire itself and by intensive interviewer training and by raising the awareness of the interviewers towards arising problems.

Moreover, the paper questionnaire had to be designed in a way that it can be done without the assistance of interviewers in a self administered way, e. g. via Internet. This is a difficult task from the questionnaire design point of view. Firstly, there are two different survey modes to consider in parallel. Secondly, it is not straightforward which instructions and other additional information (definitions, examples etc.) should be provided to the respondent in the questionnaire. The main task is keeping the balance between providing information necessary, while limiting the amount of information to a level that could still be taken in by the respondents. Otherwise, explanations in the questionnaire might lose their effect and even become misleading.

Furthermore, the follow-up survey showed no major effects of proxy interviews, which is in favour for a census where it would be very difficult, if not impossible to prevent all proxy interviews. Consequently, the employment questions were conducted in a similar way to the new implementation within the Mikrozensus. The approach is to gather the correct information about the respondents' main status first and in doing so to ensure that in this first question about employment the respondent is able to express his own everyday life's perception of his/her self-perceived main status.⁹

4.2 Testing the Questionnaire

In order to test the household sample questionnaire a two-stage pretest has been conducted. In a first step 18 qualitative, cognitive interviews have been carried out during summer 2009 in the cognitive laboratory of the FSO. The aim of this qualitative step was to test the overall comprehension of questions, filters and the design. Concerning the employment questions, the issue was to check whether the slightly shortened questions from the Mikrozensus questions worked in the intended way.¹⁰

Following the completion of the qualitative test, the questions were reworked where necessary and subjected to a second, quantitative field test from

8 The option to fill in the questionnaire via internet is also being provided.

9 For the final question wording and order please see figure 2.

10 In 2010, a further laboratory test focussed on the usability of the web option of the household sample questionnaire.

December 2009 until February 2010 with about 8.000 respondents in five federal states within Germany. The voluntary field test consisted of face-to-face interviews (17 %) and self-administered completion (83 %) where the respondents filled in the questionnaires at home without interviewer assistance. The sample was – with one exception – drawn from the access panel of German official statistics. The access panel ("Dauerstichprobe befragungsbereiter Haushalte") consists of former respondents from the Mikrozensus who are being asked after the completion of the final participating year whether they are willing to take part in further voluntary surveys of official statistics (Körner et al. 2006). One federal state (Bremen) used address data from administrative registers of the municipality. The face-to-face interviews were mainly conducted by experienced Mikrozensus interviewers. Only one federal state used prior unexperienced interviewers for the test in order to gain information for the later interviewing process and the practical fieldwork for the census 2011.

After the data collection for the field test, 60 interviewers were asked to participate in focus groups. Moreover, for the 17 % face-to-face interviews the interviewers filled in an additional short questionnaire where they have been asked about problems with the interview. Furthermore, to gather more in-depth information some selected interviewers conducted 94 additional interviews about how the respondents perceived the questionnaire and where problems occurred. The aim of the field test was to test some organisational aspects such as the automatic questionnaire image coding and to test the questions and the design on a larger scale. Apart from the Mikrozensus follow-up survey, the field test is the first larger field test of the new question design as described earlier in the paper.¹¹ This paper focuses on the test results that concern the employment questions of the census household sample questionnaire.

4.3 Laboratory Test Results

The laboratory test largely showed plausible results for the employment questions. In general, the intended, subjective decision process of respondents works when asking the main status question first. However, some groups such as people in parental leave (Elternzeit) or partial retirement (Altersteilzeit) or unemployed participating in a workfare programme (Ein-Euro-Job) are still unsure whether they should classify themselves as employed. For this reason in the quantitative field

11 A split ballot experiment would have been desirable but for organisational reasons was not feasible.

test questionnaire the people on parental leave and partial retirement were especially mentioned in the category in brackets. Consequently, there were no signs that the employment situation is not being accurately captured by the employment questions.

Moreover, to clarify the subject for people who interrupt their occupation (e. g. on parental leave or people in partial retirement) additional instructions were given under the question of the one-hour criteria. Those promising results validated the previous findings from the Mikrozensus follow-up survey and other cognitive tests once more. Nevertheless, in order to finally evaluate the new design the slightly adjusted questionnaire underwent the necessary field test.

4.4 Field Test Results

The objectives of the field test were manifold. Apart from a real life test of the technical and organisational procedure, the employment part mainly focussed on the effectiveness of the skip instructions and the accuracy of the measurement of the level and structure of employment. In the following, we present the key findings in these two areas.

For every questionnaire skip instructions are a challenge for the respondents as well as for the questionnaire designer. Because they entail complex cognitive processes (see, e. g., Redline/Dillman 2002), the aim is to reduce them as much as possible. Unfortunately, for the measurement of employment status, it is not possible to avoid skip instructions entirely. Against this background, during the development of the household sample questionnaire, skip instructions have been limited to a minimum. One of the objectives of the field test has been to investigate whether these skip instructions are functional and where they needed further improvement. When respondents disregard skip instructions, the flow of the interview is disturbed in any case. Regarding the accuracy of the data, one has to distinguish two cases: (1) When the respondents disregard skip instructions where they should have been regarded they respond to questions which do not apply to them. The result in this case is redundant information, but no missing data (item non-response). (2) The more problematic case are those where people have not answered questions which they should have (missing information) as they are compulsively needed for determining the employment status. The analysis of the skip instructions concentrated mainly on the latter problematic case. Nevertheless, even the extent of redundant information is an important indicator for the functionality of the questionnaire.

For the employment questions in general, the errors resulting from skip instructions are between 2 % and 5 %, where redundant information and missing information is about even.¹² This result can be considered as very satisfactory especially for postal surveys (83 % were mail out-mail in answers). Consequently, there are only minor problems for the calculation of the employment status due to missing data. However, it was quite problematic to guide employed respondents to the question for the occupational status. From 3.873 employed persons only 3.527 provided the necessary information, which led to 9 % non-respondents (346 cases). We concluded that the main reason for the high item non-response rate, in comparison with other questions was due to the specific questionnaire layout; the question on the occupational status was positioned at the beginning of a new page, without a separate heading. At the same time, the vast majority of the respondents had to skip a varying number of questions. As a solution for the final questionnaire there are now skip instructions indicated for all response categories, thus meaning that not only for the items where respondents have to skip, but also when the respondents are required to continue with the subsequent question. The idea is to provide help to those respondents who do not have to skip; an approach which is usually not applied in questionnaires of official statistics. Moreover, to clarify the topic on the following page, a clear and explicit heading was introduced as further guidance.

The results for the main status question showed multiple responses in some 1.5 % of the cases. In comparison with other questions in the questionnaire the 1.5 % multiple responses are not very large. Nevertheless, further analysis showed some explanation for possible reasons why respondents had difficulties to provide the information. The additional questionnaires (which the interviewer filled out after each interview) indicated that some respondents had problems deciding what their predominant status is. For instance, a retired woman was not sure whether she should tick housewife or retiree. Moreover, the additional interviews showed that some respondents missed additional categories like occupational disability, short-time work, self-employed or sickness-pay. Those cases were only rare exceptions and cannot be generalised, therefore they did not result in major changes. The only exception was a change providing further clarification for people with occupational disabilities, which have now been added in brackets for the category "none of the above". Another difficulty could be found in the data, which never-

12 As the focus in the field test was on methodological analysis, no plausibility checks were performed, i. e. the rates of item non-response and (for instance) redundant information presented in this paper could have been considerably reduced during the data processing.

theless remains inherent to the main status question itself: The respondents were asked to decide for themselves which their most important status was. The results show that for certain groups of respondents this choice does not always seem obvious. Results of the focus groups, for example, point out those respondents with more than one job have problems deciding which job they should refer to in the main status. Despite these problems, the item non-response rate for the main status question was not very large either (1.5 %).

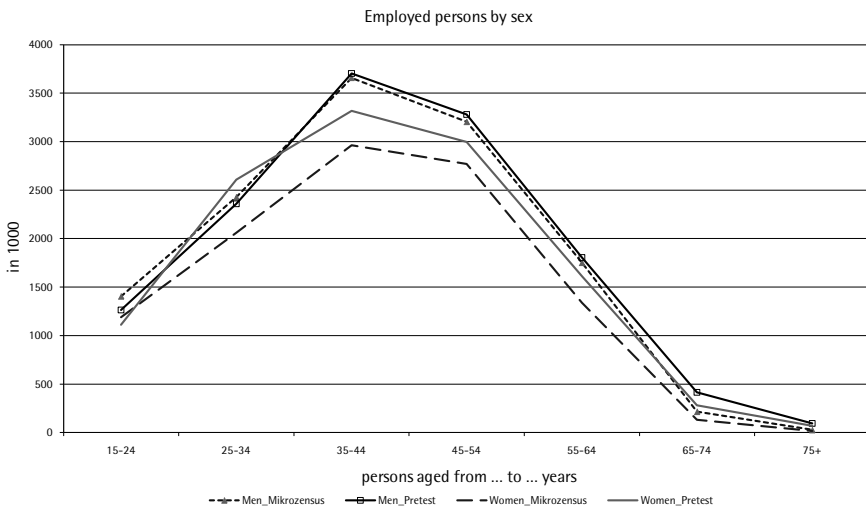
Nevertheless, another result is interesting: only 90 % of the employed respondents could be captured by the first question itself. This proves that the following question is definitely needed in order to capture especially the marginally employed persons. Analysing the following questions, it showed that there was still potential for improvement. Interestingly, the results of the focus groups also demonstrated that especially for people in marginal or small jobs there is a tendency not to declare the job. For instance, parents do not consider newspaper delivery of their children as paid work or pensioners do not declare their caretaker job. Some interviewers noted that a certain, albeit small, number of people can not be convinced to declare their small jobs. Therefore, in order to further improve the questionnaire, additional examples for jobs have been added for the question on small and side jobs. Concerning additional information given for the paper questionnaire in terms of the employment questions it was decided that it is best to reduce additional information and instructions to the minimum. Of course, additional information which could not be included in the questionnaire was provided to the interviewer during the training and also in an interviewer guidebook. With respect to the employment question, the main area focused on raising awareness for the reference date 9 May 2011, the reference week for employment, the importance of completely capturing minor jobs, and the explanation of terms and definitions.

For the evaluation of the new approach, the final results were extrapolated (using a simple post-stratification) and compared to other surveys such as the Mikrozensus and the Employment Statistics Register of the Federal Employment Agency (Beschäftigungsstatistik).¹³ Noting that the field test was a voluntary test and with respect to the modest sample size of 8.000 persons as well as the fact that no randomised experiment was conducted, the comparisons and the results have to be interpreted cautiously. Nevertheless, they at least show some directions regarding the effects of the new operationalisation. Through all age groups

13 Due to time restrictions, the post-stratification was based on results from the Mikrozensus 2008. This does however not significantly restrict the analysis as the population size and age structure does not show major changes from the annual average of 2008 to the turn-of-the-year 2009/2010.

there were – as expected – more employed persons captured through the field test than through the Mikrozensus 2008.¹⁴ Interestingly, there were relatively more employed in the age group of 65 to 74 years. This result corresponds with the Employment Statistics Register. By contrast the employment among the 15 to 24 year old was lower than expected. Generally and across all age groups, the field test indicated higher numbers of employed women, whereas the differences for employed males are less clear (see figure 4). The only exception is the age group from 65 to 74 years for which the field test yielded higher results in both cases.

Figure 4 Comparison of Persons in Employment by Sex for the Field Test and the Mikrozensus



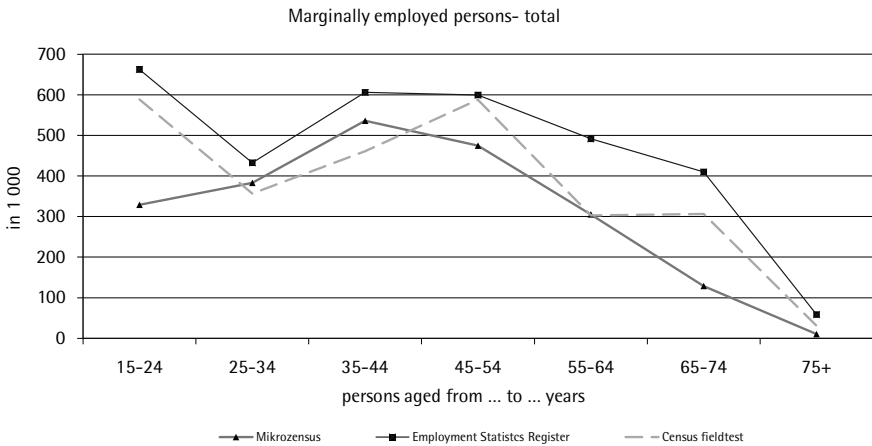
Further analysis revealed that the difference among the 15 to 24 year old group was due mainly to full-time workers, yet the field test showed more 15 to 24 year old employed part-time. This result confirms the observation for the marginally

14 It should be noted that the reference period of the field test was in December 2009 and January 2010, while the data available from the Mikrozensus at the time of the analysis refer to an annual average of 2008 (and 30 June 2008 for the Employment Statistics Register). This does however not cause major restrictions, as the overall number of employed persons does not largely differ between these reference periods. According to the results from the monthly LFS, there were 38.5 million employed persons in December 2009 and January 2010 compared to 38.7 million in the year 2008. Consequently, the results from the field test slightly "underestimate" the number of employed persons in this comparison. Due to the large sampling error of the monthly results, it was however decided not to use the monthly Mikrozensus results for the comparisons by age group and status in employment or the extrapolation.

employed. The numbers of marginal employees are much lower in the Mikrozensus than in the Employment Statistics Register (see Körner/Puch 2009). In general, capturing the employment status of the age cohort in question fared better with the field test than the Mikrozensus and it is in accordance with the numbers of the Employment Statistics Register.¹⁵

This result is strengthened through further outcomes from the field test, showing the rate of persons holding side/marginal jobs beside their (higher) education or retired people, was higher than in the Mikrozensus. Concretely, the field test showed 21 % of pupils as being employed compared with the Mikrozensus showing 8 % only for this group.¹⁶ Likewise the situation for students, unemployed and retirees: here the field test showed 44 % versus 29 % in the Mikrozensus pertaining to employment in conjunction with their main status. Finally, the field test indicated that 27 % of unemployed and 8 % of retirees had side jobs while the Mikrozensus accounted for only 14 % and 4 % respectively.

Figure 5 Comparison of Persons in Marginal Jobs by Age Group



15 The Employment Statistics Register is not necessarily free of error. However, as the follow-up survey of the Mikrozensus has shown, it is likely that a major part of the difference between both sources might be attributed to measurement errors in the Mikrozensus. An ongoing study carried out in co-operation between the FSO and the Federal Labour Agency focuses in detail on the quantification of measurement errors in the Employment Statistics Register.

16 Note that the Mikrozensus includes the main status question from the year 2011 onwards only. For the comparisons presented in this contribution, the analysis had to draw on different other variables, which can be compared only to an approximate level.

Overall the results of the field test confirm the results of the qualitative test as well as for the follow-up study of the Mikrozensus. The chosen approach combines a respondent-friendly question technique with a better coverage of employed persons with marginal jobs. It should be mentioned that the conclusions drawn from the field test are somewhat limited due the design of the test. Further evidence will become available after the survey year 2011 of the Mikrozensus has been finished: In 2011, a randomised experiment is being integrated into the regular Mikrozensus in order to quantify the effects of the questionnaire change.

Only small adjustments had to be made concerning the question wording or given additional explanations. The field test findings certainly also pointed out that it was very important and a great challenge for the census to train the quite large number of interviewers in an effective and comprehensible way. The field test was conducted with probably more experienced respondents than the average respondent in a census and more importantly with trained Mikrozensus interviewers who know about the hints and potential pitfalls.

5 Conclusions

The results from the research carried out in the context of the Mikrozensus and the census have clearly shown that the measurement of the employment status according to the ILO labour force concept requires special efforts. The concept's deviation from everyday life's perception hampers the comprehensive coverage of small and minor jobs. Nevertheless, at least for international comparison, there is a broad consensus that there is no alternative to the main pillars of the labour force concept. Any other choice regarding the one-hour-criterion would be even more arbitrary and presumably less straightforward to measure (see Eurostat 2009; ILO 2009). Furthermore, the labour market statistics have to be adjusted to reliably portray the fringes of the labour market. In recent years many important changes pertain to the marginal employment sector, like the rise of atypical employment in the case of Germany.

Nevertheless, this paper demonstrates that there is an enormous potential for improvement while engaging with questionnaire design and working on quality issues. The operationalisation to be used in Mikrozensus/LFS and in the next German census reflect the requirement to speak the language of the respondents (by asking for the main status first) and at the same time achieving a more comprehensive picture of small side jobs as well. This paper also shows that simplistic rules like "never ask for the main status before you ask about work in the reference

week" are not appropriate for questionnaire design. Seemingly similar questions might provide radically different results and any solution needs thorough testing before its implementation. Our paper also gives some striking examples about the large potential of both laboratory and field testing.

Finally, the survey mode is important. Our results indicate that paper-and-pencil questionnaires are difficult to handle in combination with complex concepts like the employment status. First of all, the skip instructions needed to measure the employment status necessarily are quite complex. Our results show that an important share of respondents (about ten percent in our case) seems to be puzzled even by the modest skip instruction of the census household survey questionnaire. Secondly, only computer assisted questionnaires offer the possibility to tailor questions to the situation of various population groups and to use other information which has already been provided during the interview.

References

- Van Bastelaer, A., 1994: Differences in the Measurement of Employment in the Labour Force Surveys in the European Community. *Journal of Official Statistics* 10: 277-305.
- Brancato, G. et al., 2006: Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System. European Commission Grant Agreement 200410300002. Eurostat, Luxembourg. http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/Handbook_questionnaire_development_2006.pdf (6.12.2010).
- Erikson, R. and J. O. Jonsson: European Social Survey Questionnaire Development Report. Chapter 2: How to ascertain the socio-structural position of the individual in Society. http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=62&Itemid=96 (7.7.2011).
- Eurostat, 2009: Task Force on the Quality of the Labour Force Survey. Final report. Luxembourg: European Commission. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-09-020/EN/KS-RA-09-020-EN.PDF (7.7.2011).
- Eurostat, 2010: Task Force on Indicators to Supplement the ILO Unemployment Rate. Final Report, Luxembourg: European Commission. http://epp.eurostat.ec.europa.eu/portal/page/portal/labour_market/documents/Report_TF_indicators.pdf (7.7.2011).
- Hussmanns, R., M. Fahrhad and V. Vijay, 1990: Surveys of the Economically Active Population, Employment, Unemployment, and Underemployment. An ILO Manual on Concepts and Methods. Geneva: ILO.
- ILO, 1982: Resolution concerning statistics of the economically active population, employment, unemployment and underemployment. Adopted by the 13th International Conference of Labour Statisticians, Geneva: ILO.
- ILO, 1998: Guidelines concerning treatment in employment and unemployment statistics of persons on extended absences from work. Endorsed by the Sixteenth International Conference of Labour Statisticians, Geneva: ILO.
- ILO, 2009: Report of the Seminar "Employment and unemployment: Revisiting the relevance and conceptual basis of the statistics", Geneva, 4-5 December 2008. Geneva: ILO.

- Kleven, Ø., B. O. Lagerstrøm and I. Thomsen, 2008: A Simple Model for Studying the Effects of Proxy Interviewing: The Norwegian Labour Force Survey as a Case Study. *Statistics Norway* 2008. http://www.ssb.no/emner/06/90/doc_200802/doc_200802.pdf (7.7.2011).
- Köhne-Finster, S. and T. Körner, 2009: Microcensus Follow-up Survey – A Project of the Federation and the Länder to Measure the ILO Employment Status. *Methods, Approaches, Developments* 1: 3-6.
- Köhne-Finster, S. and A. Lingnau, 2008: Untersuchung der Datenqualität erwerbsstatistischer Angaben im Mikrozensus. Ergebnisse des Projekts „Nachbefragung im Mikrozensus/LFS“. In: *Wirtschaft und Statistik* 12: 1067-1088.
- Körner, T. and K. Puch, 2009: Der Mikrozensus im Kontext anderer Arbeitsmarktstatistiken. Ergebnisunterschiede und ihre Hintergründe. *Wirtschaft und Statistik* 6: 528-552.
- Körner, T. and K. Puch, 2011: Coherence of German Labour Market Statistics. The German Labour Force Survey in Comparison with the Employment Accounts and the Unemployment Register. *Statistics and Science* 19, Wiesbaden 2011 (forthcoming).
- Körner, T., A. Nimmergut, J. Nökel and S. Rohloff 2006: Die Dauerstichprobe befragungsbereiter Haushalte. Die neue Auswahlgrundlage für freiwillige Haushaltsbefragungen. *Wirtschaft und Statistik* 5: 451-467.
- Meder, N., 2009: Grundzüge des Zensus 2011. *DAGStat-Bulletin* Nr. 4: 2-3.
- Radermacher, W. and T. Körner, 2006: Fehlende und fehlerhafte Daten in der amtlichen Statistik. Neue Herausforderungen und Lösungsansätze. *Allgemeines Statistisches Archiv* 90: 553-576.
- Redline, C. D. and D. A. Dillman, 2002: The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires. In: Groves, Robert et al. (Eds.): *Survey Nonresponse*. New York: John Wiley.
- Rengers, M., 2004: Das international vereinbarte Labour-Force-Konzept. *Wirtschaft und Statistik*, 12: 1369-1383.
- Rengers, M., 2010: Some nine million people wishing to work (more). *StatMagazin* published on 29 June 2010. Wiesbaden. http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/EN/Navigation/Publications/STATmagazin/2010/LabourMarket2010__06,templateId=renderPrint.psm!__nnn=true (7.7.2011).
- Schwarz, N., 1987: Cognitive Issues of Labour Force Surveys in a Multinational Context: Issues and Findings. Paper prepared for the OECD Working Party on Employment and Unemployment Statistics. Paris, 14-16 April 1987.
- Statistische Ämter des Bundes und der Länder, 2011. <https://www.zensus2011.de/> (7.7.2011).
- Statistisches Bundesamt, 2008: Ergebnisse der Nachbefragungen im Mikrozensus/Labour Force Survey. Ein Gemeinschaftsprojekt der Statistischen Ämter des Bundes und der Länder zur Qualitätssicherung im Mikrozensus/Labour Force Survey. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt, 2009: Monthly Unemployment Statistics According to the ILO Concept. Quality Report. Wiesbaden. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/EN/Content/Publikationen/QualityReports/Labourmarket/MonthlyILOConcept.property=file.pdf> (7.7.2011).
- Thomsen, I. and O. Villund, 2011: Using Register Data to Evaluate the Effects of Proxy Interviews in the Norwegian Labour Force Survey. *Journal of Official Statistics* 27: 87-98.
- UNECE, 2006: Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing. Prepared in Cooperation with the Statistical Office of the European Communities (Eurostat). Geneva and New York: UNECE.
- Vorndran, I., 2004: Zensusstest-Verfahren und Ergebnisse der Haushaltgenerierung. *Wirtschaft und Statistik* 11: 1245-1255.

Legal Acts

- Council regulation (EC) No 577/98 of 9 March 1998 on the organisation of a labour force sample survey in the Community. Official Journal of the European Union, 14.3.1998, L 77/3.
- Commission regulation (EC) No 1897/2000 of 7 September 2000 implementing Council Regulation (EC) No 577/98 on the organisation of a labour force sample survey in the Community concerning the operational definition of unemployment. Official Journal of the European Union, 8.9.2000, L 228/18.
- Commission regulation (EC) No 1575/2000 of 19 July 2000 implementing Council Regulation (EC) No 577/98 on the organisation of a labour force sample survey in the Community concerning the codification to be used for data transmission from 2001 onwards. Official Journal of the European Union, 20.7.2000, L 181/16.
- Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses. Official Journal of the European Union, 13.8.2008, L 218/14.
- Commission regulation (EC) No 1201/2009 of 30 November 2009 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns. Official Journal of the European Union, 15.12.2009, L 329/29.

Addresses of the Authors Britta Gauckler
Statistisches Bundesamt
Zensus
Gustav-Stresemann-Ring 11
65189 Wiesbaden
britta.gauckler@destatis.de

Thomas Körner
Statistisches Bundesamt
Arbeitsmarkt
Gustav-Stresemann-Ring 11
65189 Wiesbaden
thomas.koerner@destatis.de

Wechsel des Erhebungsinstituts und Adressweitergabe mit Widerspruchsverfahren im IAB-Haushaltspanel PASS

Ein Praxisbericht

Change of the Fieldwork Agency in the IAB Household Panel Study PASS

A Practical Report

Gerrit Müller

Zusammenfassung

Wechsel des Erhebungsinstituts zwischen den Erhebungswellen einer wissenschaftlichen Panelstudie sind selten. Die möglichen Effekte eines solchen Institutswechsels sind im Vorhinein schwer abzuschätzen und methodisch kaum untersucht. Zudem erschweren in der Praxis datenschutzrechtliche Anforderungen einen Institutswechsel. So wurde, wie bei den meisten akademischen Studien in der Bundesrepublik üblich, auch den in PASS (Panel „Arbeitsmarkt und Soziale Sicherung“) zu befragenden Haushalten in einem vorab versandten Merkblatt zum Datenschutz zugesagt, dass Name und Anschrift der Befragten nicht an Dritte weitergegeben werden. Um nicht gegen diese Zusage zu verstoßen, war es geboten, alle teilnehmenden Haushalte über den anstehenden Institutswechsel zu informieren und diesen eine explizite, für sie kostenlose, Widerspruchsmöglichkeit gegen die Weitergabe ihrer Daten an das neue Erhebungsinstitut einzuräumen. Diesbezüglich werden

Abstract

Ongoing longitudinal surveys seldom switch fieldwork agencies between waves. The possible effects of such a change are difficult to estimate in advance and have, until now, rarely been explored. In addition, data privacy statements hamper the switch between fieldwork agencies in practice. As in most other academic surveys in Germany, the households to be surveyed in PASS (Panel Study “Labour Market and Social Security”) were assured in advance in writing that their name and address data would not be passed on to third parties. To comply with this data privacy statement, the PASS survey needed to inform all the participating households of the upcoming change in fieldwork agency and to give them the opportunity to explicitly object to the transfer of their address data. This article reports the effects of this notification, which may be of relevance for other panel studies in future. Moreover, it attempts to quantify the costs relating to a change of fieldwork agency in

im Folgenden Vorgehensweise und Erfahrungen beschrieben, die unter Umständen auch für andere Panelstudien bei zukünftigen Institutswechseln von Relevanz sein könnten. Dabei werden insbesondere Ausmaß und Struktur der Panelausfälle durch Widerspruch gegen eine Adressweitergabe näher untersucht.

terms of magnitude and selectivity of the panel attrition arising in particular from the objection to the transfer of address data.

1 Einleitung¹

Mit dem Panel „Arbeitsmarkt und Soziale Sicherung“ (PASS) hat das Institut für Arbeitsmarkt- und Berufsforschung (IAB) in den letzten Jahren einen neuen Datensatz für die Arbeitsmarkt-, Sozialstaats- und Armutsforschung in Deutschland aufgebaut. Die Studie ist Teil der gesetzlichen Wirkungsforschung des IAB nach § 55 Sozialgesetzbuch Zweites Buch (SGB II) und wird entsprechend aus dem Bundeshaushalt (Bundesministerium für Arbeit und Soziales, BMAS) finanziert. Die ersten drei Wellen der Befragung wurden in den Jahren 2006/07, 2007/08 und 2008/09 gemeinsam mit dem Erhebungsinstitut TNS Infratest Sozialforschung (München) durchgeführt. Im Jahr 2009 wurde zwischen dem IAB und dem BMAS vereinbart, PASS zunächst mit drei weiteren Wellen (4–6) in den Jahren 2010, 2011 und 2012 fortzuführen. Im Rahmen des Ausschreibungs- und Vergabeverfahrens für die Wellen 4–6 hat mit infas Sozialforschung (Bonn) ein neues Erhebungsinstitut den Auftrag erhalten, die Feldarbeit sowie die Datenaufbereitung für das laufende Panel zu übernehmen.

Zur Fortführung des Panels wurde daher die Weitergabe von Adress- und Befragungsdaten vom bisherigen an das neue Erhebungsinstitut notwendig. In den Wellen 1–3 war jedoch allen erstmals zu befragenden Haushalten ein Merkblatt zum Datenschutz, zusammen mit einem Ankündigungsschreiben, zugestellt worden. Darin wurde unter anderem zugesichert, dass Name und Anschrift nicht an Dritte (außer den namentlich genannten Instituten) weitergegeben werden. Aufgrund dieser Zusage war es geboten, alle teilnehmenden Haushalte über den anstehenden Institutswechsel vorab zu informieren und eine Widerspruchsmöglichkeit gegen die Weitergabe ihrer Adressdaten einzuräumen. Auch andere wissenschaft-

1 Für wertvolle Hinweise und Verbesserungsvorschläge zu früheren Versionen dieses Textes danke ich zwei anonymen Gutachtern, Mark Trappmann, Stefanie Gundert, Frauke Kreuter sowie den Herausgebern der MDA. Für das Korrekturlesen des Textes danke ich Beatrix Hillen.

liche Panelstudien in Deutschland haben in der Vergangenheit ihren Befragten fast wortgleiche Zusagen gemacht und würden sich, so denn die Absicht bzw. Notwendigkeit zu einem Institutswechsel bestünde, in einer ähnlichen Ausgangslage befinden.²

Da Wechsel des Erhebungsinstituts zwischen zwei Wellen einer laufenden Panelstudie sehr selten und zudem kaum dokumentiert sind (vgl. Schnell 2004: 1-3, 17), bietet der vorliegende Beitrag einen Erfahrungsbericht zu einem wichtigen Teilaspekt eines Institutswechsels: dem Ausmaß und Struktur der Panelausfälle durch Widerspruch gegen eine Adressweitergabe. Gerade in einer Panelstudie stellt der durch ein solches Widerspruchsverfahren verursachte Verlust von bereits mehrfach befragten Beobachtungseinheiten einen nicht zu vernachlässigenden Kostenaspekt dar, und ist daher bei geplanten Institutswechseln in das Kalkül mit einzubeziehen. Hierbei sind sowohl die Reduktion des (effektiven) Stichprobenumfanges als auch mögliche Stichprobenverzerrungen durch systematischen Ausfall zu berücksichtigen. Inhaltlich knüpft der vorliegende Beitrag eng an die Literatur zur wiederholten Teilnahme bzw. Panelausfall bei Haushaltsbefragungen an (z. B. Watson/Wooden 2009; Groves/Couper 1998; Schnell 1997), untersucht diesbezüglich aber einen Sonderfall mit eigenen Mechanismen.

Die Arbeit ist hierzu wie folgt aufgebaut: Nach einem kurzen Abschnitt (2) mit relevanten Hintergrundinformationen zum IAB-Haushaltspanel PASS, wird in Abschnitt (3) näher auf die Konzeption und die praktische Durchführung des Widerspruchsverfahrens eingegangen. In den darauf folgenden Abschnitten (4) und (5) wird beschrieben, wie viele Haushalte der Weitergabe ihrer Adressdaten widersprochen haben und im Rahmen eines multivariaten Modells empirisch näher untersucht, mit welchen Faktoren der Ausfall durch Widerspruch zusammenhängt. Abschnitt (6) schließt mit einigen Bemerkungen zu weiteren, mit einem Institutswechsel verbundenen, Kostenaspekten.

2 Siehe bspw. die Erklärung zum Datenschutz der Mitgliedsinstitute des ADM (Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.) unter http://www.adm-ev.de/fileadmin/user_upload/PDFS/Merkblatt.pdf (3.4.2011).

2 Das IAB-Haushaltspanel: Panel „Arbeitsmarkt und Soziale Sicherung“ (PASS)

Ziel des PASS ist die Schaffung einer Längsschnitt-Datenbasis für die ökonomische und soziologische SGB-II-Forschung.³ Im Fokus stehen dabei die Untersuchung der Dynamik des Bezugs von Arbeitslosengeld II (ALGII) und die Erforschung der Konsequenzen des Transferleistungsbezuges für die materielle und soziale Lage der betroffenen Haushalte. Dazu wurden in der ersten Welle mehr als 12.000 Haushalte befragt, die je zur Hälfte aus einer Stichprobe von SGB-II-Leistungsempfängern und einer Bevölkerungsstichprobe mit überproportionaler Berücksichtigung einkommensschwächerer Haushalte stammen. Um Neuzugänge zur Population der SGB-II-Leistungsempfänger mit einzubeziehen, wurde die Welle 1 Prozessdatenstichprobe in den Wellen 2 und 3 mit zusätzlichen Stichproben („Zugangsstichproben“) von jeweils etwa 1.000 Haushalten, die erstmals Leistungen nach SGB II erhalten, ergänzt.⁴

3 Widerspruchsverfahren: Konzeption und praktische Durchführung

In Abstimmung mit der Datenschutzbeauftragten des IAB wurde ein Anschreiben zur Ankündigung des Institutswechsels entwickelt (für Textauszüge siehe Anhang 1) und vom IAB Anfang November 2009 an alle betroffenen Haushalte versendet.⁵ Jedem Anschreiben lag ein an die Projektleitung des PASS voradressierter und freigemachter Rückumschlag bei, den die Haushalte verwenden konnten, um ihren schriftlichen Widerspruch kostenfrei zurückzusenden. Es wurde dabei eine Frist von ungefähr 6 Wochen eingeräumt, wobei Widersprüche, die erst nach Ablauf dieser Frist und vor Feldstart Mitte Februar 2010 am IAB eingingen, trotzdem noch berücksichtigt wurden. Grundsätzlich galt, dass Adressen von Haushalten, die ihren Widerspruch geltend gemacht hatten, nicht an das neue Erhebungs-

3 Nähere Informationen zum Studiendesign, dem Frageprogramm und der Datenstruktur des PASS sind zu finden in Rudolph und Trappmann 2007; Achatz et al. 2007; Trappmann et al. 2009; Christoph et al. 2008; Gebhardt et al. 2009.

4 Der Scientific Use File (SUF), Fragebogen, weitere Arbeitshilfen sowie die Methoden- und Feldberichte von TNS Infratest (Büngeler et al. 2010, 2009; Hartmann et al. 2008), sind über das Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) erhältlich (<http://fdz.iab.de/>).

5 Haushalte der Welle 4 Zugangsstichprobe, die durch das neue Erhebungsinstitut erstmalig zu befragen waren, betraf dieses Verfahren nicht, da in diesen Fällen keine Weitergabe von Adressdaten von einem Erhebungsinstitut an ein anderes stattfand.

institut weitergegeben wurden. Analog wurde mit Haushalten verfahren, bei denen aufgrund unzustellbarer Anschreiben davon ausgegangen werden musste, dass sie die Information über den anstehenden Institutswechsel und die damit verbundene Weitergabe ihrer Adresse nicht erreicht hatte.

Um den Ausfall durch unzustellbare Anschreiben so gering wie möglich zu halten, wurden im Rahmen des Versands sämtliche Briefkuverts mit dem Aufdruck „Wenn unzustellbar, zurück! Bei Umzug mit neuer Anschrift zurück!“ versehen. Hierdurch wurde zum einen sichergestellt, dass im Falle von Umzügen die Auskunft der Post zur Adressaktualisierung und zum erneuten Anschreiben verwendet werden konnte.⁶ Gleichzeitig wurden auf diese Weise alle weiteren Haushalte identifiziert, bei denen das Anschreiben unzustellbar war. Um diese Haushalte aufzufinden und erneut anschreiben zu können, wurden (zum Teil mehrfach) Adressrecherchen sowohl über Register der Bundesagentur für Arbeit als auch über Einwohnermeldeämter vorgenommen.

4 Ergebnis

Insgesamt wurden 10.670 Haushalte angeschrieben, deren Adressdaten an das neue Erhebungsinstitut zur Fortführung der Studie übergeben werden sollten.⁷ Diese Haushalte lassen sich in vier Typen untergliedern, die hinsichtlich Lokalisierbarkeit/Erreichbarkeit und Widerspruch im Folgenden getrennt betrachtet werden.

Der größte Teil der Ausgangsstichprobe besteht aus Panelhaushalten, die bereits in mindestens einer der Vorwellen teilgenommen hatten und auch in Welle 3 wieder erfolgreich interviewt wurden ($n=8.068$). Darüber hinaus gibt es eine kleinere Anzahl von Panelhaushalten, die zuletzt in Welle 2 teilgenommen hatten, in Welle 3 jedoch nicht erfolgreich interviewt wurden ($n=1.004$). Diese Haushalte werden mit Blick auf die Folgewelle als „temporäre“ Ausfälle gewertet, da angestrebt wurde, sie in Welle 4 erneut zu interviewen. Die Gruppe der in Welle 3 erstmalig Befragten ($n=1.306$) besteht zum einen aus 1.150 Haushalten der Welle-3-Zugangsstichprobe von SGB-II-Leistungsempfängern (vgl. Rudolph/Trappmann 2007), zum anderen aus 156 Split-Haushalten, die sich zuvor von bestehen-

6 Dies war allerdings grundsätzlich nur in solchen Fällen möglich, bei denen ein Nachsendeauftrag vorlag.

7 Ausgangspunkt für das Widerspruchsverfahren war die Bruttostichprobe der vierten Welle, d. h. alle Haushalte, die nach zuvor festgelegten Regeln nach Abschluss der Welle 3 Feldarbeit im Panel verbleiben. Dies waren im Wesentlichen alle Haushalte, die in Welle 3 teilgenommen haben bzw. die in Welle 3 nicht teilgenommen haben, bei denen aber durch die Nichtteilnahme keine Befragungslücke von mehr als einem Jahr entstand. Zusätzlich galt, dass Haushalte, die in Welle 3 am Ende des Interviews erklärt hatten, dass sie in Zukunft nicht wieder teilnehmen möchten oder die auf Basis der Kontaktdaten als „harte“ Verweigerer eingestuft wurden, aus der Bruttostichprobe der vierten Welle entfernt wurden.

den Panelhaushalten abgespalten hatten und in Welle 3 erstmalig befragt wurden.⁸ Des Weiteren gibt es noch eine kleine Anzahl von Split-Haushalten (n=292), die sich im Laufe der Welle-3-Feldarbeit von Bestandshaushalten abgespalten hatten, jedoch nicht mehr interviewt werden konnten, und daher in Welle 4 vom neuen Institut erstmalig zu befragen waren.

Tabelle 1 Ergebnis Widerspruchsverfahren

	Anzahl	Adresse unbekannt, Anschreiben unzustellbar (jemale in Recherche)	Widerspruch
		%	%
W1/W2 Panelhaushalte, mit Interview in W3	8068	0,31 (4,04)	9,42
W1/W2 Panelhaushalte, (temporärer) Ausfall in W3	1004	1,79 (13,55)	10,96
W3 Erstbefragte (Zugangsstichprobe, W2/W3 Split-Haushalte)	1306	1,45 (13,09)	5,82
Split-Haushalte, in W4 erstmals zu befragen	292	22,95 (57,88)	7,19
Summe	10670	1,21 (7,52)	9,06

Tabelle 1 zeigt, dass von den Haushalten, die im Rahmen des Verfahrens anzuschreiben waren, insgesamt etwa 10 % für Welle 4ff. verloren wurden. Dabei lassen sich zwei Stufen des Ausfalls unterscheiden: Nichterreichbarkeit und Widerspruch.⁹ Etwas mehr als 1 % der Haushalte konnte, trotz aller Recherchebemühungen, nicht erfolgreich angeschrieben werden, da die aktuell gültige Adresse nicht zu ermitteln bzw. das Anschreiben aus anderen Gründen unzustellbar war. Rund 9 % der Haushalte haben einer Weitergabe ihrer Daten bzw. dem Institutswechsel explizit widersprochen. Mit Blick auf die Lokalisierbarkeit/Erreichbarkeit ist lediglich die Gruppe der Split-Haushalte auffallend. In den meisten dieser Fälle musste als Ausgangspunkt für die Recherche auf die Anschrift des jeweiligen Ursprungshaushalts zurückgegriffen werden, mit entsprechenden Konsequenzen für den Anschreibenversand und Rechercheerfolg.

8 Das Nachverfolgungskonzept von PASS sieht vor, dass im Falle eines Auszugs aus einem Befragungshaushalt beide Haushalte weiterverfolgt werden. Der neue Haushalt ausgezogener Personen wird dabei als Split-Haushalt bezeichnet.

9 In einigen wenigen Fällen (insgesamt 0,16 %) wurde zudem keine Adresse an das neue Erhebungsinstitut übermittelt, da diese Haushalte inzwischen ins Ausland verzogen bzw. verstorben waren und damit nicht mehr zur Grundgesamtheit zählen.

Mit insgesamt 9 von 10 Prozentpunkten macht der Ausfall durch Widerspruch allerdings eindeutig den Hauptteil des Gesamtausfalls aus. Auch hier zeigen sich Unterschiede zwischen den Haushaltstypen, z. B. eine relativ geringe Widerspruchsrate bei Welle-3-Erstbefragten. Bei der Interpretation der in Tabelle 1 dargestellten Widerspruchsrate, und den sich daraus ergebenden Differenzen zwischen den einzelnen Haushaltstypen, muss allerdings stets berücksichtigt werden, dass diese Gruppen sehr unterschiedlich zusammengesetzt sind.¹⁰ Nicht zuletzt um diese Unterschiede zu kontrollieren, vor allem aber um mögliche Ausfallursachen (bzw. Korrelate) zu identifizieren, wird ein multivariates Modell für Ausfall durch Widerspruch notwendig.

5 Ausfallanalysen

Der hier untersuchte Ausfall von Befragungshaushalten kann als ein Sonderfall der „üblichen“, durch Nichtteilnahme an einer weiteren regulären Befragungswelle verursachten, Panelmortalität betrachtet werden. Die nachfolgenden empirischen Analysen orientieren sich daher grundsätzlich an den entsprechenden Untersuchungen zum Panelausfall in Längsschnittstudien, wie beispielsweise dem SOEP, SHP, BHPS, HILDA, und den in diesen Studien gängigen Spezifikationen der multivariaten Ausfallmodelle (vgl. Kroh 2010; Lipps 2007; Uhrig 2008; Watson/Wooden 2009). Mit Blick auf den deskriptiven Charakter des Praxisberichts und der gebotenen Kürze wird hier darauf verzichtet, die einzelnen Theorien zur (wiederholten) Teilnahme an Befragungen ausführlich zu diskutieren.¹¹ Es werden jedoch theoretisch begründete Überlegungen zu *jenen* Faktoren skizziert, die für das Teilnahmeverhalten der Befragten im vorliegenden Kontext *von besonderer Relevanz* sein dürften.

Dabei ist zunächst noch einmal hervorzuheben, dass der Regelfall nicht etwa der Widerspruch, sondern keinerlei Reaktion auf das Anschreiben ist (>90 %). Dieses Verhalten ist im Rahmen von Rational-Choice Theorien erklärbar (vgl. Schnell

10 So besteht beispielsweise die Gruppe der in Welle 3 Erstbefragten fast ausschließlich aus Haushalten der Welle-3-Zugangsstichprobe von SGB-II-Leistungsempfängern. Die Gruppe der Panelhaushalte hingegen besteht knapp zur Hälfte auch aus Haushalten der Stichprobe der Allgemeinbevölkerung. Daraus resultieren Unterschiede in der Merkmalszusammensetzung und möglicherweise im Teilnahmeverhalten.

11 An dieser Stelle sei verwiesen auf Groves et al. 2004; Lepkowski und Couper 2002; Groves, Singer und Corning 2000; Groves und Couper 1998; Schnell 1997; Stoop 2005; Koch und Porst 1998; mit speziellem Fokus auf Populationen im Bereich Niedrigeinkommen und staatlicher Transferleistungen: Ver Ploeg et al. 2002. Aktuelle Zusammenfassungen der empirischen Literatur zum Teilnahmeverhalten finden sich u. a. bei Uhrig 2008; Watson und Wooden 2009.

1997; Esser 1986). Im Kern gehen diese davon aus, dass die Entscheidung für eine Teilnahme bzw. einen Widerspruch als Ergebnis rationalen Handelns zu verstehen ist, bei dem Kosten- und Nutzenaspekte situationsabhängig gegeneinander abgewogen werden. Bei der überwiegenden Zahl der Fälle ist dabei davon auszugehen, dass im Vorhinein keine feste Präferenz – weder für noch gegen eine Teilnahme bzw. Widerspruch – besteht, sondern grundsätzlich eher „Indifferenz“ vorherrscht (vgl. Esser 1986: 44). Das heißt, für die meisten Befragten liegen Kosten und Nutzen eines solchen Widerspruchsverfahrens, auf insgesamt niedrigem Niveau, dicht beieinander. Bereits relativ periphere Kostenaspekte, wie zum Beispiel das Rücksenden eines voradressierten und frankierten Umschlages, sind dann bei vielen ausschlaggebend für die Entscheidung (vgl. Esser 1986: 41). Da die Nichtteilnahme am Widerspruchsverfahren jedoch die weitere Befragung durch ein neues Institut ermöglicht, und damit potentiell den Verbleib im Panel bedeutet, kehren sich die Verhältnisse hier in einer für die Studie günstigen Weise um.

Gleichzeitig ist es allerdings so, dass die Kosten oder „Belastungen“ (vgl. Schnell 1997: 166) eines Institutswechsels mit Adressweitergabe durch ein Anschreiben mit Betonung von Datenschutzaspekten noch einmal explizit gemacht und in der subjektiven Wahrnehmung der Befragten unter Umständen verstärkt werden (vgl. Singer et al. 1993: 466). Gerade Personengruppen, bei denen ohnehin gewisse Befürchtungen über die Verwendung der weitergegebenen Daten durch Dritte oder über eine mögliche Verletzung der Privatsphäre vorherrschen (vgl. Schräpler 2004: 121, 123), dürften sich hier in besonderem Maße angesprochen fühlen; mit dem Ergebnis, dass oftmals widersprochen wird (vgl. Singer et al. 1993; Singer et al. 1992). Ein ähnliches Kalkül könnte auch solche Subgruppen bewegen, die vielleicht keine konkreten Datenschutzbedenken oder sonstige Befürchtungen im Rahmen des Institutswechsels haben, die jedoch einer weiteren Teilnahme zuletzt ablehnend gegenüberstanden. Dies könnten zum einen die Haushalte sein, die bereits in der Welle unmittelbar vor dem Institutswechsel die Teilnahme verweigert hatten, wie z. B. ein Teil der temporären Ausfälle. Zum anderen aber auch solche Haushalte, die zwar in der vorangehenden Befragungswelle noch einmal teilgenommen hatten, dies aber vielleicht mit insgesamt eher geringer Motivation.¹² Das erneute Hervorheben der Kosten einer Teilnahme, ohne die Gelegenheit, beispielsweise über

12 Geringe Motivation bei gleichzeitiger Teilnahme kann sich u. U. widerspiegeln in *fehlenden Angaben* (item-nonresponse), insbesondere bei schwierigen oder sensiblen Fragen, wie bspw. zum Einkommen (vgl. Loosveldt et al. 2002; Schräpler 2004), oder im *partiellen Ausfall* einer Befragungseinheit (partial unit-nonresponse), wie bspw. fehlenden Personeninterviews im Rahmen einer Befragung des gesamten Haushalts (vgl. Watson/Wooden 2009: 173; „part-respondering households“), und zudem prädiaktiv für die Nichtteilnahme in der Folgewelle sein (Ibid.).

Interviewer Einfluss auf die Entscheidung der Befragten nehmen zu können, könnte hier überproportional häufig Widerspruch hervorrufen.

Neben Erklärungsansätzen, die auf Kosten-Nutzen-Abwägungen der Befragten basieren, spielen sicherlich auch die kognitiven Aspekte dieses Verfahrens als mögliche Erklärung für das Teilnahmeverhalten eine Rolle (d. h. das Verstehen des Anschreibens, der Abruf relevanter Informationen, die Urteilsbildung, die kurze schriftliche Rückantwort; vgl. Tourangeau 1984; Tourangeau et al. 2000). So ist aus der Literatur zu postalischen Befragungen bekannt, dass die Teilnahme, neben elementaren Lese- und Schreibfähigkeiten, eine gewisse Vertrautheit mit schriftlich-postalischer Kommunikation erfordert, so dass typischerweise mit einer „Selbstselektierung“ von Personen mit höherem Bildungsniveau zu rechnen ist (Schnell/Hill/Esser 2005: 359; Schnell 1997: 205) bzw. mit überproportionalem Ausfall gering Gebildeter (Lynn 2008: 41; Schnell 2002: 151; Dillman 1983: 369). Es wäre daher zu erwarten, dass sich diese für schriftliche Befragungen typischen Gegebenheiten auch hier zeigen. Allerdings mit umgekehrten Vorzeichen: Weniger Gebildete, oder auch Befragte mit schlechten Deutschkenntnissen, dürften tendenziell seltener widersprechen und damit für das Panel erhalten bleiben.

Um diesen Überlegungen und Hypothesen empirisch nachzugehen, wird ein multivariates Modell für Widerspruch gegeben Erreichbarkeit geschätzt.¹³ Da, mit Ausnahme von Bildung, keine direkten Messungen der diskutierten Konstrukte vorliegen, wird eine Reihe von Proxyvariablen definiert: eine Indikatorvariable, die angibt, ob Befragte es in der Vergangenheit abgelehnt haben, ihre Befragungsdaten mit administrativen Daten der Bundesagentur für Arbeit verknüpfen zu lassen, sowie eine Indikatorvariable für fehlende Einkommensangaben im letzten Interview, als Proxyvariable für konkrete Datenschutzbedenken und/oder Befürchtungen über mögliche Verletzungen der Privatsphäre. Darüber hinaus wird die Anzahl der realisierten bzw. der *fehlenden* Personeninterviews innerhalb eines Haushalts in der Vorwelle als Proxy für geringe Teilnahmemotivation und möglichen Prädiktor für einen Ausfall in der Folgewelle aufgenommen. Für die oben angesprochenen Haushalte, die bereits in der letzten Welle beim bisherigen Institut schon nicht mehr teilgenommen hatten (temporäre Ausfälle), wird zusätzlich eine Variable generiert, die anzeigt, ob dies aufgrund einer Verweigerung oder aus

13 Es sind verschiedene Modell-Spezifikationen geschätzt worden, die auf Anfrage gerne zur Verfügung gestellt werden. Da sich die zentralen Ergebnisse insgesamt jedoch robust gegenüber Änderungen der Spezifikation zeigten, soll im Folgenden nur eines der Modelle vorgestellt und diskutiert werden.

anderen Gründen, z. B. weil nicht erfolgreich kontaktiert, geschah.¹⁴ Darüber hinaus liegen Variablen vor, die die Einschätzungen der Interviewer zum Interesse der Befragten an der Studie bzw. deren generelles Frageverständnis messen. Erstere dient als weitere Proxyvariable für Teilnahmemotivation. Letztere ist wohl eher den Variablen zuzurechnen, die mit kognitiven Fähigkeiten der Befragten korrelieren, wie auch etwa eine Reihe von Indikatoren für den höchsten Bildungsabschluss.

Tabelle A1 zeigt eine Übersicht aller im Modell berücksichtigten unabhängigen Variablen. Neben den bereits beschriebenen, die zumeist die letzte Interviewsituation bzw. die Befragungsgeschichte charakterisieren, ist es im Wesentlichen ein Vektor mit sozio-demografischen Merkmalen der Befragten, sowie Eigenschaften des Studiendesigns. Auch wenn die genaue Variablenauswahl je nach Untersuchung und Survey variiert (vgl. Kroh 2010; Lipps 2007; Uhrig 2008; Watson/Wooden 2009), gehören dazu üblicherweise: Geschlecht, Alter, Angaben zum Familienstand und zu Kindern, bzw. zur Haushaltsgröße und -zusammensetzung, Herkunft und Sprache, Erwerbsstatus, Gesundheitszustand sowie auch Eigenschaften des Wohnortes (Einwohnerzahl, Stadt/Land, Region).¹⁵ Diese sollen auch hier zum einen als Kontrollvariablen bei der Analyse anderer interessierender Einflussgrößen dienen. Zum anderen können diese sozio-demografischen Merkmale aber auch erste Hinweise auf den systematischen Ausfall bestimmter Subgruppen geben.¹⁶ Neben Merkmalen der Befragten gehen zudem Indikatoren für Merkmale des spezifischen Studiendesigns in die Schätzung ein. Beispielsweise wird berücksichtigt, aus welcher Teilstichprobe ein Haushalt stammt (SBG II-Leistungsempfänger oder Allgemeinbevölkerung), in welchem Modus die letzte Befragung stattfand (CATI/CAPI), die Zugehörigkeit zum PASS-Fremdsprachenfeld als Proxy für Deutschkenntnisse der Befragten.

14 Es wäre interessant gewesen, innerhalb der Verweigerungen verschiedene Verweigerungsgründe zu differenzieren, wie z. B. Datenschutzbedenken. Allerdings wurden gerade diese Fälle, wie auch in früheren Wellen üblich, nach Abschluss der Welle 3 Feldarbeit als „harte Verweigerer“ eingestuft und unmittelbar aus der Stichprobe entfernt. Somit sind diese Fälle für weitere Analysen nicht verfügbar.

15 Sozio-demografische Merkmale der Befragten stellen lediglich Korrelate von Nonresponse bzw. Panelmortalität dar und keine tatsächliche („kausale“) Erklärung der Ausfälle (z. B. Groves/Couper 1998; Schnell 1997; Koch 1997). Sie können eher als „Globalvariablen“ verstanden werden, die relativ schwach mit dem Teilnahmeverhalten zusammenhängen, und „deren einzelne Ausprägungen über eine Vielzahl verschiedener Mechanismen mit einer Vielzahl von Phänomenen gekoppelt sind“ (Schnell 1997: 199).

16 Ob und in welchem Maße solche „selektiven Ausfälle“ für empirische Analysen relevant sind, hängt stets von der jeweiligen inhaltlichen Fragestellung ab (vgl. Schnell 1997: 214ff.). Da sozio-demografische Merkmale der Befragten häufig im Rahmen von „post-survey adjustments“ (wie z. B. „propensity score weighting“) verwendet werden, handelt es sich jedoch um einen wichtigen Variablenvektor (vgl. Watson/Wooden 2009: 164). Zudem kann eine Berücksichtigung dieser Variablen im Ausfallmodell auch erste Hinweise über das Niveau des zu erwartenden Ausfalls in Studien mit anderer Stichprobenzusammensetzung liefern, würde bei diesen ein ähnliches Verfahren durchgeführt.

Tabelle 2 Probit-Modell für Widerspruch gegeben Erreichbarkeit

	$\partial p/\partial x_i$	z
Haushaltstyp		
W1/W2 Panelhaushalte mit W3 Interview, <i>Referenzkategorie</i>		---
W1/W2 Panelhaushalte, (temporärer) Ausfall in W3		
Nicht-Kontaktierte	0,025	(0,93)
Verweigerer	0,066***	(3,15)
Nicht-Interviewte, Sonstige	0,025*	(1,72)
W3 Erstbefragte	0,006	(0,56)
Letztes Interview		
Zuspieldbereitschaft verweigert	0,048***	(3,63)
Einkommensangabe fehlend	0,009	(0,43)
Anz. fehlender P-Int./HH	0,002	(0,73)
Anz. real. P-Int./HH	0,002	(0,45)
Interviewereinschätzung: Interesse	-0,001	(0,21)
Interviewereinschätzung: Frageverständnis	0,001	(0,21)
Merkmale der Befragten		
Ausbildungsabschluss		
Schüler, ohne Abschluss	-0,008	(0,55)
Volks-/Hauptschule	-0,002	(0,25)
Mittlere Reife/Realschule, <i>Referenzkategorie</i>		---
Fach-/Hochschulreife	0,010	(1,10)
Sonstiges	-0,018	(0,68)
Alter in Jahren		
15 bis 24	-0,054***	(6,23)
25 bis 34	-0,035***	(4,36)
35 bis 44, <i>Referenzkategorie</i>		---
45 bis 54	0,024***	(2,75)
55 bis 64	0,028**	(2,36)
65+	0,056***	(3,85)
Männlich	-0,015**	(2,67)
Migrationshintergrund	-0,007	(1,06)
Familienstand		
Ledig	-0,006	(0,82)
Verheiratet, <i>Referenzkategorie</i>		---
Geschieden/Getrennt	-0,037***	(5,93)
Verwitwet	-0,018	(1,50)
Kind unter 4 Jahren im HH	-0,018*	(1,87)
Erwerbstatus		
Nichterwerbspersonen	0,008	(0,83)
Arbeitslos, <i>Referenzkategorie</i>		---
Erwerbstätig, 1-34 Std./W.	0,014	(1,30)
Erwerbstätig, 35-49 Std./W.	-0,004	(0,45)
Erwerbstätig, 50+ Std./W.	-0,018	(1,43)
Schwere gesundheitl. Einschr.	-0,004	(0,57)
Ortsgrößenklasse		
500.000 +, <i>Referenzkategorie</i>		---
100.000 bis unter 500.000	0,004	(0,53)
50.000 bis unter 100.000	-0,000	(0,02)
20.000 bis unter 50.000	0,004	(0,40)
unter 20.000	0,010	(0,86)
Neue Bundesländer	0,021***	(2,59)
Merkmale des Studiendesigns		
Leistungsempfängerstichprobe	-0,042***	(6,71)
CAPI Modus	-0,004	(0,58)
Fremdsprachenfeld	-0,037***	(2,81)
McFadden Pseudo R ²		0,054
N		10,243

*0,10 ≥ p > 0,05; **0,05 ≥ p > 0,01; ***0,01 ≥ p

Die Schätzung der Standardfehler berücksichtigt das komplexe, mehrstufige Ziehungsdesign. Die durchschnittlichen marginalen Effekte bzw. diskreten Änderungen wurden mit dem STATA *ado* *margeff* berechnet.

Das Modell enthält zusätzliche Indikatorvariablen für WN/KA Werte („item-missings“) bei Interviewereinschätzungen, Zuspieldbereitschaft, Migrationshintergrund, Familienstand, Erwerbstatus und gesundheitl. Einschränkung. Keiner der Koeffizienten hiervon war statistisch signifikant und, da ohne inhaltliche Interpretation, in Tabelle 2 nicht dargestellt.

Tabelle 2 zeigt Ergebnisse eines Probit-Modells für Widerspruch, gegeben Erreichbarkeit, mit den gerade beschriebenen unabhängigen Variablen. Dargestellt ist jeweils der durchschnittliche marginale Effekt bzw. die diskrete Änderung (z. B. Cameron/Trivedi 2005: 122, 467), das heißt, der über alle Beobachtungen gemittelte Effekt eines Inkrementes der jeweiligen x-Variablen auf die Wahrscheinlichkeit eines Widerspruchs gegen die Adressweitergabe. Dies hat im Vergleich zur Darstellung von Probit-Koeffizienten unter anderem den Vorteil, dass nicht nur Signifikanz und Vorzeichen, sondern auch die Größenordnung der (durchschnittlichen) Effekte, ähnlich zu linearen Regressionsmodellen, direkt interpretierbar ist.¹⁷

Hierbei zeigt sich, dass unter Kontrolle des gewählten Variablenvektors insbesondere die Haushalte, die bereits in der letzten Welle (d. h. unmittelbar vor dem Institutswechsel) verweigert hatten, eine gegenüber der Referenzkategorie statistisch signifikante und auch in der Größenordnung deutlich höhere Widerspruchsrate aufweisen (ca. 7 Prozentpunkte). Des Weiteren zeigt der Koeffizient der Proxyvariable für mögliche Datenschutzbedenken den erwarteten Zusammenhang. Wurde in der Vergangenheit der Bitte um Erlaubnis, auf individueller Ebene Surveyangaben mit administrativen Daten verknüpfen zu dürfen, nicht zugestimmt, so wird auch einer Weitergabe von Adressdaten häufiger (5 Prozentpunkte) widersprochen. Von den weiteren, zu Beginn als möglicherweise wichtig eingestuften Faktoren, zeigt keiner einen statistisch signifikanten Einfluss. Weder fehlende Angaben bei einem sensiblen Merkmal wie Einkommen, noch fehlende Personeninterviews innerhalb von Haushalten, sind hier prädiktiv für die Wahrscheinlichkeit, einer Adressweitergabe zu widersprechen.

Ähnliches gilt auch für die Interviewereinschätzungen zum Interesse der Befragten am Thema bzw. deren Frageverständnis und überraschenderweise auch für das Ausbildungsniveau. Hier gab es, gerade im Vergleich zu den meisten anderen sozio-demografischen Merkmalen von Befragten, im Vorhinein relativ klare Erwartungen. Man könnte unterstellen, dass dies an der Inklusion einer ganzen Reihe von mit Bildung stark korrelierter Variablen liegt. Dem ist jedoch nicht so. In einer bivariaten Analyse, das heißt in einem Probit-Modell, das außer (einer Konstanten und) den Indikatoren für Ausbildung, bei gleicher Referenzkategorie, keine weiteren Kovariate enthält, ist kein einziger der Ausbildungskoeffizienten signifikant unterschiedlich von Null (5 %-Niveau). Auch ein F-Test für joint-significance zeigt, dass Unterschiede im Ausbildungsniveau keine Rolle spielen ($F=1,36$; $p=0,239$).

Bei den weiteren sozio-demografischen Merkmalen der Befragten fällt vor allem ein stark signifikanter, fast linearer Alterseffekt ins Auge. Jüngere widersprechen deutlich seltener als Ältere, die Differenz zwischen der Gruppe der Jüngsten

und der Ältesten beträgt im Durchschnitt mehr als 10 Prozentpunkte. Aber auch Einflüsse von Geschlecht (männliche Haushaltsvorstände widersprechen seltener als weibliche) und Familienstand (Geschiedene/getrennt Lebende seltener als Verheiratete) sind nachweisbar. Wie zuvor bereits erwähnt, sind dies sozio-demografische Korrelate des Ausfalls und keine substanziellen Erklärungen. Darüber, welche (Mischungen von) Mechanismen genau es sind, die beispielsweise den beobachteten starken Alterseffekt entstehen lassen, lässt sich lediglich spekulieren. Eine mögliche Interpretation wäre, dass Ältere unter Umständen eine stärker ausgeprägte Präferenz für Stabilität und Kontinuität haben und der Veränderung des Erhebungsinstituts und des gewohnten Interviewers eine subjektiv höhere Bedeutsamkeit beimessen.¹⁸ In Bezug auf die seltener auftretenden Widersprüche von Jüngeren könnte vermutet werden, dass Jüngere sich im Gebrauch schriftlicher Medien von Älteren dahingehend unterscheiden, dass ein „Brief“ und eine Antwort darauf (im Vergleich zu E-Mail, Internet, (Mobil-)Telefon) eine weniger gewohnte und akzeptierte Kommunikationsform ist.

Mit Ausnahme eines Indikators für Ost/West (Ostdeutsche widersprechen etwas häufiger als Westdeutsche) weist keines der sonst noch berücksichtigten sozio-demografischen Merkmale auf starke Selektionseffekte hin. Bemerkenswert ist insbesondere der unauffällige Befund bei den Erwerbsstatus-Variablen. Hier wäre, gerade mit Blick auf mögliche Opportunitätskosten, eine gewisse Selektivität zu erwarten gewesen. Die negativen Vorzeichen bei Erwerbstätigen mit 50 und mehr Arbeitsstunden pro Woche deuten in diese Richtung, erreichen in der vorliegenden Stichprobe aber keine Signifikanz. Ähnliches gilt für den Effekt kleiner Kinder im Haushalt. Auch eine von den Befragten in der vorangehenden Welle als „stark“ beschriebene gesundheitliche Einschränkung, scheint im Mittel weder Hemmnis noch Grund für einen Widerspruch zu sein.

Im Hinblick auf Merkmale des Studiendesigns zeigt sich, dass in der Teilstichprobe der SGB-II-Leistungsempfänger die Widerspruchsrate statistisch signifikant und in der Größenordnung deutlich geringer ist (4 Prozentpunkte) als bei Haushalten der Bevölkerungsstichprobe. Frühere Analysen zu Panelausfällen im PASS haben interessanterweise gezeigt, dass Haushalte aus der Leistungsempfängerstichprobe deutlich seltener wiederholt teilnehmen, gegeben dass sie erreicht wurden (z. B.: Büngeler et al. 2009). Im vorliegenden Zusammenhang kann somit eine „Umkehrung“ dieser Selektivität festgestellt werden. Diese für PASS beson-

18 Natürlich handelt es sich bei „Älteren“ oder „Jüngeren“ um Personengruppen innerhalb derer, selbst unter Kontrolle eines umfangreichen Variablenvektors, noch eine beträchtliche (unbeobachtete bzw. unbeobachtbare) Heterogenität vorliegt, die mit deren Teilnahmeverhalten in Zusammenhang steht. Daher treffen die hier angestellten Überlegungen bestenfalls im Mittel zu.

ders wichtige Personengruppe widerspricht also überproportional häufig *nicht* und verbleibt somit häufiger im Panel. Der Modus der letzten Befragung hat auf die Wahrscheinlichkeit, dass ein Haushalt Widerspruch einlegt, keinen Einfluss. Wurden allerdings Haushalte in der Vorwelle mit Hilfe einer der übersetzten Versionen des Instruments befragt, so zeigen diese, auch unter Kontrolle von Ausbildungsniveau, Migrationshintergrund und den anderen Kovariaten, eine im Mittel um 3,5 Prozentpunkte geringere Wahrscheinlichkeit, der Adressweitergabe zu widersprechen.

Insgesamt hat das Ausfallmodell, mit einem McFadden Pseudo- R^2 von 5,4 %, lediglich eine recht geringe Anpassungsgüte; auch eine Reihe von Alternativspezifikationen mit weiteren/anderen Kovariaten zeigte diesbezüglich vergleichbare Werte. Dies ist jedoch für diese Art von Analyse nicht untypisch (vgl. Lipps 2007: 67). Zudem war die Intention auch nicht, eine Modellspezifikation zu finden, die ein bestimmtes Kriterium für Anpassungsgüte maximiert (z. B. Pseudo- R^2 Varianten) oder minimiert (z. B. AIC, BIC). Im Gegenteil, die Vorgehensweise war bewusst explorativ gewählt, um auch „fehlende“ Einflüsse von Faktoren zu dokumentieren, die sich bei empirischen Untersuchungen zu Panelmortalität in anderen Zusammenhängen oftmals als wichtig erwiesen haben.¹⁹

Zusammenfassend kann festgehalten werden, dass zur Bestimmung des Widerspruchsverhaltens als theoretisch wichtig eingestufte Faktoren, insbesondere Datenschutzbedenken, auch empirisch relevant sind. Jedenfalls deuten die signifikanten Koeffizienten der dafür verwendeten Proxyvariablen stark darauf hin. Ein zunächst vermuteter „Bildungs-Bias“, mit weniger (mehr) Widersprüchen von gering (höher) Gebildeten, konnte hingegen nicht bestätigt werden. Auch mit Blick auf eine Reihe weiterer, üblicherweise in Ausfallmodellen berücksichtigter, soziodemografischer Merkmale wurden, mit Ausnahme eines starken Alterseffekts, insgesamt nur wenige signifikante Koeffizienten gefunden.

6 Schlussbemerkungen

Hauptanliegen dieses Beitrages war es, einen in der deutschen Surveypraxis seltenen Vorgang zu dokumentieren: den Wechsel des Erhebungsinstituts zwischen zwei Erhebungswellen einer großen wissenschaftlichen Panelstudie und dem damit

19 Die hier gewählte Spezifikation stellt insofern einen Kompromiss dar zwischen der Inklusion möglichst vieler theoretisch und empirisch relevanter Kovariate und der Vermeidung von Multikollinearitätsproblemen. Der Durchschnitt der Varianzinflationsfaktoren (VIF) über alle verwendeten Regressoren beträgt 1,37; der größte Wert beträgt 2,52.

verbundenen Widerspruchsverfahren zur Adressweitergabe. Die Notwendigkeit für solch ein Verfahren ergab sich aus einer datenschutzrechtlichen Zusage gegenüber den Befragten, dass von den im Merkblatt zum Datenschutz genannten Instituten keine Adresdaten an Dritte übermittelt werden.

Die Ergebnisse des Widerspruchsverfahrens müssen hinsichtlich ihrer Übertragbarkeit auf andere Studien stets entsprechend eingeordnet werden. Nicht nur werden sich Stichprobenzusammensetzung, Befragungsschwerpunkte und andere Aspekte des Surveydesigns von denen des PASS unterscheiden. Auch könnte der Zeitpunkt des Institutswechsels im „Lebenszyklus“ eines Panels eine wichtige Rolle spielen. Gerade durch die wiederholte Befragung durch das gleiche Institut und seine Interviewer werden Beziehungen (Vertrauen, soziales Kapital) zu den Befragten aufgebaut, die durch einen Institutswechsel kurzfristig beendet werden. Bei einem Wechsel in, beispielsweise, Welle 10 oder später, hätten die Ergebnisse daher anders ausfallen können.

Dies wirft generell die Frage nach Kosten und Nutzen von Institutswechseln auf. So betrachtet, hat der vorliegende Beitrag lediglich einen einzelnen, wenn auch wichtigen, Teilaspekt eines Institutswechsels in den Fokus genommen. Neben dem hier untersuchten Panelausfall im Rahmen des Widerspruchsverfahrens wäre bei einer grundsätzlichen Diskussion zur Optimalität von Wechselentscheidungen eine ganze Reihe von Methodeneffekten und finanzieller bzw. personeller Kostenaspekte zu berücksichtigen. Dies in systematischer und detaillierter Weise zu tun, geht jedoch über den Rahmen und die Zielsetzung des vorliegenden Beitrags deutlich hinaus. Einige grundsätzliche Überlegungen hierzu sollen nichtsdestotrotz vorgenommen werden.

In Bezug auf finanzielle bzw. personelle Kosten eines Wechsels ist festzuhalten, dass bei öffentlichen Auftraggebern die Vergabe an klare Richtlinien gebunden ist, insbesondere den Grundsatz, dass der Zuschlag auf das wirtschaftlichste Angebot erteilt wird (§ 97 GWB). Dieser Grundsatz garantiert formal, dass aus einer Reihe von Angeboten bei gegebener Qualität stets die kostengünstigste Alternative gewählt wird. Die Problematik bei einer laufenden Erhebung ist jedoch, dass im Falle eines Institutswechsels ein erheblicher personeller (Zusatz-)Aufwand auf Seiten des Auftraggebers entsteht. Gerade weil bei einem Wechsel sichergestellt werden muss, dass die Anschlussfähigkeit der Datenbasis zu früheren Wellen gewahrt bleibt, muss Hintergrundwissen zu vielen Details einer Studie aufwendig dokumentiert und an den neuen Auftragnehmer übergeben werden. Das heißt, auch wenn das Studiendesign weitgehend unverändert bleibt (wie z. B. das Frageprogramm), fallen durch diese Übertragungsleistung und das Einspielen neuer Prozesse erhebliche Anfangsinvestitionen beim Auftraggeber an, die von der Größenordnung her dem erstmaligen Aufsetzen der Studie nahe kommen.

Neben dem finanziellen und personellen Aufwand gibt es „statistische“ Kosten, beispielsweise durch Methodeneffekte, und die damit verbundene Frage, wie diese im Gesamtzusammenhang zu bewerten sind. Gemeint sind hier mögliche Unterschiede in den Befragungsergebnissen, die sich aus Unterschieden in der Organisation der Feldarbeit zwischen den Instituten, der Zusammensetzung und Arbeitsweise des jeweiligen Interviewerstabes sowie der Datenerfassung und Aufbereitung ergeben könnten, um nur einige Dimensionen zu nennen (vgl. Schnell 2004: 19ff.). Im Moment gibt es leider keine wissenschaftlich fundierte Entscheidungssystematik, die *sämtliche* Kostenaspekte und Methodeneffekte von Institutswechseln quantifizieren und zueinander gewichten würde. Hierzu wären weitere Forschungsarbeiten hilfreich.

Literatur

- Achatz, J., A. Hirsland und M. Promberger, 2007: Rahmenkonzept für das IAB-Panel „Arbeitsmarkt und Soziale Sicherung“. S. 11-32 in: M. Promberger (Hg.): Neue Daten für die Sozialforschung. Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung“. IAB-Forschungsbericht Nr. 12/2007. Nürnberg: IAB.
- Best, H. und C. Wolf, 2010: Logistische Regression. S. 827–854 in: C. Wolf und H. Best (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Büngeler, K., M. Gensicke, J. Hartmann, R. Jäckle und N. Tschersich, 2010: IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 3 (2008/09). Methoden- und Feldbericht von TNS Infratest Sozialforschung, München.
- Büngeler, K., M. Gensicke, J. Hartmann, R. Jäckle und N. Tschersich, 2009: IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 2 (2007/08). Methoden- und Feldbericht von TNS Infratest Sozialforschung, München. FDZ Methodenreport 6/2009. Nürnberg: IAB. http://doku.iab.de/fdz/reporte/2009/MR_08-09.pdf (28.10.2010).
- Cameron, A. C. und P. K. Trivedi, 2005: Microeconometrics. Methods and Applications. New York: Cambridge University Press.
- Christoph, B., G. Müller, D. Gebhardt, C. Wenzig, M. Trappmann, J. Achatz, A. Tisch und C. Gayer, 2008: Codebuch und Dokumentation des „Panel Arbeitsmarkt und Soziale Sicherung“ (PASS) Welle 1 (2006/2007). FDZ Datenreport 5/2008. Nürnberg: IAB. http://doku.iab.de/fdz/reporte/2008/DR_05-08.pdf (28.10.2010).
- Dillman, D. A., 1983: Mail and Other Self-Administered Questionnaires. S. 359–377 in: P. H. Rossi, J. D. Wright und A. B. Anderson (Hg.): Handbook of Survey Research. New York: Academic Press.
- Esser, H., 1986: Über die Teilnahme an Befragungen. ZUMA-Nachrichten 18: 38–47. PID: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-210300>.
- Gebhardt, D., G. Müller, A. Bethmann, M. Trappmann, B. Christoph, C. Gayer, B. Müller, A. Tisch, B. Siflinger, H. Kiesel, B. Huyer-May, J. Achatz, C. Wenzig, H. Rudolph, T. Graf und A. Biedermann, 2009: Codebuch und Dokumentation des „Panel Arbeitsmarkt und Soziale Sicherung“ (PASS) Welle 2 (2007/2008). FDZ Datenreport 6/2009. Nürnberg: IAB. http://doku.iab.de/fdz/reporte/2009/DR_06-09.pdf (28.10.2010).
- Groves, R. M. und M. P. Couper, 1998: Nonresponse in Household Interview Surveys. New York: Wiley.
- Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer und R. Tourangeau, 2004: Survey Methodology. Hoboken, NJ: Wiley.

- Groves, R. M., E. Singer und A. Corning, 2000: Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *Public Opinion Quarterly* 64: 299-308.
- Hartmann, J., K. Brink, R. Jäckle und N. Tschersich, 2008: IAB-Haushaltspanel im Niedrig-einkommensbereich. Methoden- und Feldbericht von TNS Infratest Sozialforschung, München. FDZ Methodenreport 7/2008. Nürnberg: IAB. http://doku.iab.de/fdz/reporte/2008/MR_07-08.pdf (28.10.2010).
- Kish, L., 1992: Weighting for Unequal Pi. *Journal of Official Statistics* 8 (2): 183-200.
- Koch, A., 1997: Teilnahmeverhalten beim ALLBUS 1994. Soziodemographische Determinanten von Erreichbarkeit, Befragungsfähigkeit und Kooperationsbereitschaft. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49 (1): 98-122.
- Koch, A., und R. Porst, 1998 (Hg.): Nonresponse in Survey Research. ZUMA-Nachrichten Spezial Band 4. Mannheim: ZUMA. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial4.pdf (5.7.2011).
- Kroh, M., 2010: Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2009). Data Documentation 50. Berlin: DIW.
- Lepkowski, J. M. und M. P. Couper, 2002: Nonresponse in the Second Wave of Longitudinal Household Surveys. S. 259-272 in: R. M. Groves, D. A. Dillman, J. L. Eltinge und R. J. A. Little (Hg.): *Survey Nonresponse*. New York: Wiley.
- Lipps, O., 2007: Attrition in the Swiss Household Panel. MDA – Methoden, Daten, Analysen. *Zeitschrift für Empirische Sozialforschung* 1 (1): 45-68. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.1_Heft_1/MDA1_Lips.pdf (5.7.2011).
- Loosveldt, G., J. Pickery und J. Billiet, 2002: Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. *Journal of Official Statistics* 18 (4): 545-557.
- Lynn, P., 2008: The Problem of Nonresponse. S. 35-55 in: E. D. de Leeuw, J. J. Hox und D. A. Dillman (Hg.): *International Handbook of Survey Methodology*. New York: Taylor & Francis Group LLC.
- Stoop, I. A. L., 2005: *The Hunt for the Last Respondent*. The Hague, NL: Social and Cultural Planning Office.
- Rudolph, H. und M. Trappmann, 2007: Design und Stichprobe des Panels „Arbeitsmarkt und Soziale Sicherung“ (PASS). S. 60-101 in: M. Promberger (Hg.): *Neue Daten für die Sozialforschung. Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung“*. IAB-Forschungsbericht Nr. 12/2007. Nürnberg: IAB.
- Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske + Budrich.
- Schnell, R., 2002: Anmerkungen zur Publikation „Möglichkeiten und Probleme des Einsatzes postalischer Befragungen“ von Karl-Heinz Reuband in der KZFSS 53(2): 307-333. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54 (1): 147-456.
- Schnell, R., 2004: Effekte des Wechsels des Erhebungsinstituts zwischen den Erhebungswelten bei Panelstudien. Unveröffentlichtes Gutachten. Universität Konstanz.
- Schnell, R., Hill, P. B. und E. Esser, 2005: *Methoden der Empirischen Sozialforschung*. 7. Auflage. München: Oldenbourg.
- Schräpler, J.-P., 2004: Respondent Behavior in Panel Studies: A Case Study for Income Nonresponse by Means of the German Socio-Economic Panel (SOEP). *Sociological Methods and Research* 33 (1): 118-156.
- Singer, E., H.-J. Hippler und N. Schwarz, 1992: Confidentiality Assurances in Surveys: Reassurance or Threat? *International Journal of Public Opinion Research* 4: 256-268.
- Singer, E., N. A. Mathiowetz und M. P. Couper, 1993: The Impact of Privacy and Confidentiality Concerns on Survey Participation: The Case of the 1990 U.S. Census. *Public Opinion Quarterly* 57 (4): 465-482.
- Tourangeau, R., 1984: Cognitive Sciences and Survey Methods. S. 73-100 in: T. Jabine, M. Straf, J. Tanur und R. Tourangeau (Hg.): *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Washington, DC: National Academic Press.
- Tourangeau, R., J. Lance und K. Rasinski, 2000: *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

- Trappmann, M., B. Christoph, J. Achatz, C. Wenzig, G. Müller und D. Gebhardt, 2009: Design and Stratification of PASS. A New Panel Study for Research on Long Term Unemployment. IAB Discussion Paper Nr. 5/2009. Nürnberg: IAB.
- Uhrig, N. S., 2008: The Nature and Causes of Attrition in the British Household Panel Survey. ISER Discussion Paper No. 2008-05. Colchester: University of Essex.
- Ver Ploeg, M., R. A. Moffitt und C. F. Citro, 2002 (Hg.): Studies of Welfare Populations: Data Collection and Research Issues: Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academy Press.
- Watson, N. und M. Wooden, 2009: Identifying Factors Affecting Longitudinal Survey Response. S. 157–181 in: P. Lynn (Hg.): Methodology of Longitudinal Surveys. Chichester: Wiley.

Anschrift des Autors

Dr. Gerrit Müller
Institut für Arbeitsmarkt- und Berufsforschung
(IAB) der Bundesagentur für Arbeit
Regensburger Str.104
90478 Nürnberg
gerrit.mueller@iab.de

Anhang 1 Textauszüge Anschreiben

[...] Bisher haben wir, das Institut für Arbeitsmarkt- und Berufsforschung, die Studie gemeinsam mit TNS Infratest durchgeführt. Mit der Fortführung der Studie und der Wiederholungsbefragung haben wir einen anderen Partner, das infas Institut für angewandte Sozialwissenschaft beauftragt. [...]

Selbstverständlich werden auch in der Kooperation mit infas alle datenschutzrechtlichen Bestimmungen für die Studie strikt eingehalten. Um die Befragung fortsetzen zu können, ist es notwendig, dass wir infas alle notwendigen Informationen, etwa Ihren Namen und Ihre Adresse und – falls sie uns bekannt ist – auch Ihre Telefonnummer übermitteln. Die Übermittlung dieser Angaben erfolgt ausschließlich zum Zweck der weiteren Befragung im Rahmen der Studie [...].

Sie können dieser Weitergabe widersprechen, indem Sie bis zum DATUM eine kurze Mitteilung im beiliegenden Freiumschlag an das Institut für Arbeitsmarkt- und Berufsforschung senden. Bitte geben Sie hierbei Ihren Namen und Ihre Adresse oder folgende Nummer: 123456789 an. [...]

Tabelle A1 Definition und Mittelwerte der erklärenden Variablen
(N=10.243)

Variablenname	Definition	Mittelwert
Zuspätkommensbereitschaft verweigert	1 = HH-Vorstand (HHV) hat Verknüpfung von Befragungs- mit Prozessdaten der BA nicht zugestimmt; 0 = Zugestimmt	0,071
Einkommensangabe fehlend	1 = Angabe zum Haushaltseinkommen WN/KA/"NULL"; 0 = Einkommensangabe	0,020
Anz. fehlender P-Int./HH	(metrisch) Anzahl fehlender Personeninterviews innerhalb eines Haushalts	0,895
Anz. real. P-Int./HH	(metrisch) Anzahl realisierter Personeninterviews innerhalb eines Haushalts	1,409
Interesse	(metrisch, 1-5 Skala) Interviewereinschätzung wie interessant die Befragten das Interview fanden	4,243
Frageverständnis	(metrisch, 1-5 Skala) Interviewereinschätzung wie gut die Befragten die Fragen verstanden haben	4,604
Männlich	1 = HHV männlich; 0 = HHV weiblich	0,437
Ausbildungsabschluss		
Schüler/ohne Abschluss	1 = HHV ist noch Schüler, oder ohne Abschluss	0,051
Volks-/Hauptschule	2 = HHV hat Volks-/Hauptschulabschluss	0,330
Mittlere Reife/Realschule	3 = HHV hat Mittlere Reife/Realschulabschluss, <i>Referenz</i>	0,346
Fach-/Hochschulreife	4 = HHV hat Fach-/Hochschulreife bzw. Abitur	0,234
Sonstiges	5 = HHV hat sonstigen Abschluss	0,010
Alter in Jahren		
15 bis 24	1 = HHV 15 bis 24 Jahre alt	0,070
25 bis 34	2 = HHV 25 bis 34 Jahre alt	0,187
35 bis 44	3 = HHV 35 bis 44 Jahre alt, <i>Referenzkategorie</i>	0,241
45 bis 54	4 = HHV 45 bis 54 Jahre alt	0,245
55 bis 64	5 = HHV 55 bis 64 Jahre alt	0,161
65+	6 = HHV 65 Jahre und älter	0,095
Migrationshintergrund	1 = HHV hat Migrationshintergrund HHV; 0 = HHV hat keinen Migrationshintergrund	0,203
Familienstand		
Ledig	1 = HHV ist ledig	0,327
Verheiratet	2 = HHV ist verheiratet, <i>Referenzkategorie</i>	0,353
Geschieden/Getrennt	3 = HHV ist geschieden/getrennt lebend	0,233
Verwitwet	4 = HHV ist verwitwet	0,052
Kind unter 4 Jahren im HH	1 = Kind unter 4 Jahren im Haushalt; 0 = kein Kind unter 4 Jahren im Haushalt	0,101
Erwerbstatus		
Nichterwerbspersonen	1 = HHV ist Nichterwerbsperson, d. h. weder erwerbstätig noch arbeitslos	0,233
Arbeitslos	2 = HHV ist arbeitslos, <i>Referenzkategorie</i>	0,347
Erwerbstätig, 1-34 Std./Woche	3 = HHV ist erwerbstätig mit einem Stundenumfang von 1-34 Stunden pro Woche	0,110
Erwerbstätig, 35-49 Std./Woche	4 = HHV ist erwerbstätig mit einem Stundenumfang von 35-49 Stunden pro Woche	0,188
Erwerbstätig, 50+ Std./Woche	5 = HHV ist erwerbstätig mit einem Stundenumfang von 50 und mehr Stunden pro Woche	0,058
Schwere gesundheitl. Einschr.	1 = HHV hat schwere gesundheitliche Einschränkung; 0 = keine	0,233

Fortsetzung ...

Fortsetzung Tabelle A1

Variablenname	Definition	Mittelwert
Ortsgrößenklasse		
500.000 +	1 = 500.000 Einwohner und mehr	0,338
100.000 bis unter 500.000	2 = 100.000 bis unter 500.000 Einwohner	0,330
50.000 bis unter 100.000	3 = 50.000 bis unter 100.000 Einwohner	0,111
20.000 bis unter 50.000	4 = 20.000 bis unter 50.000 Einwohner	0,094
unter 20.000	5 = unter 20.000 Einwohner	0,127
Neue Bundesländer	1 = Neue Bundesländer; 0 = Alte Bundesländer	0,291
Leistungsempfängerstichprobe	1 = HH aus Stichprobe von SGB-II-Leistungsempfängern aus Registern der Bundesagentur für Arbeit (BA) 0 = HH aus Bevölkerungsstichprobe	0,600
CAPI	1 = HH-Interview: CAPI face-to-face; 0 = CATI	0,403
Fremdsprachenfeld	0 = Deutsch	0,030

HHV = Haushaltsvorstand

Für die Schätzungen des Ausfallmodells verbleiben 10.243 Haushalte. Zum einen können die 223 Split-Haushalte ohne bisheriges Interview nicht in die Ausfallanalysen einbezogen werden. Zum anderen gab es nachträgliche Löschungen von Interviews im Rahmen der Welle 3 Datenaufbereitung (58 Haushalte), die daher hier ebenfalls nicht berücksichtigt werden können ($N = 10.524 - 223 - 58 = 10.243$).

Rezensionen



ANDREAS DIEKMANN, 2009: Spieltheorie. Einführung, Beispiele, Experimente. 2. überarbeitete Auflage. rowohlt enzyklopädie, roworo Taschenbuch. ISBN 978-3-499-55701-9, 272 Seiten, 13,00 EUR.

Andreas Diekmann (ETH Zürich) präsentiert mit seinem Lehrbuch „Spieltheorie. Einführung, Beispiele, Experimente“ auf rund 230 inhaltlichen Seiten einen umfassenden und klar gegliederten Einstieg in die Thematik. Das Buch ist mittlerweile in der 2. Auflage erschienen, wobei aber lediglich kleinere Korrekturen vorgenommen, inhaltlich jedoch keine Änderungen angebracht wurden. Im Gegensatz zu den meisten Lehrbüchern der Spieltheorie setzt Diekmann kein tiefes mathematisches Wissen voraus. Das Buch kommt ohne komplexe mathematische Notation aus und führt Formeln nur dann ein, wenn es die präzise Definition eines Begriffes oder die Analyse nötig machen. Dies macht die Spieltheorie, ihre Denkweise, ihre Konzepte und ihre Gleichgewichtsanalysen auch für die mathematisch weniger interessierten Sozialwissenschaftler zugänglich. Gleichzeitig verhindert dies aber auch, dass das Buch die Tiefe erreicht, welche für die Analyse interdependenter Situationen vielfach unumgänglich ist. Fairerweise muss man jedoch festhalten, dass Diekmann auch nie den Anspruch erhebt, mehr als eine gut verständliche, kurzweilige und vor allem mit vielen Beispielen aus Soziologie, Politik und Ökonomik versehene Einführung in die Spieltheorie vorzulegen.

Das Buch ist in 10 Kapitel unterteilt. In den ersten beiden Kapiteln werden zentrale Konzepte erläutert. Es wird zwischen der Normalform und der Extensivform unterschieden, der zentrale Begriff der Strategie wird eingeführt und das Konzept des Nash-Gleichgewichts und die Verfeinerung des teilspielperfekten Nash-Gleichgewichts herausgearbeitet. Diekmann tut dies anhand von bekannten 2×2 Spielen wie dem Gefangenendilemma, dem Chicken Spiel oder dem Assurance Spiel. Hilfreich ist dabei die Unterteilung in Koordinationsspiele, Nullsummenspiele und Spiele mit gemischten Interessen. Diese Unterteilung verdeutlicht auch dem Laien klar, wie die Spieltheorie interdependente Situationen strukturiert und entsprechend analysiert. Im dritten Kapitel werden diese Konzepte auf die Analyse von Vertrauenssituationen (z. B. Gebrauchtwagenkauf, Verkauf von Waren auf eBay) angewandt. Kapitel 4 schließt die Grundlagenkapitel ab, in dem die Prinzipien der Entscheidungstheorie diskutiert werden, Begriffe wie „rationales Handeln“ oder „Lernen“ eingeführt und die Grundlagen der Nutzenmessung dargestellt werden.

In den Kapiteln 5 bis 9 werden jeweils spezielle Arten von Spielen dargestellt und die entsprechenden Begrifflichkeiten definiert. In Kapitel 5 werden Nullsummenspiele (d. h. Spiele mit konträren Interessen der Spieler) besprochen sowie gemischte Strategien und Lösungskonzepte wie der Sattelpunkt eingeführt. Im ganzen Buch stellt dieses Kapitel die „höchsten“ mathematischen Anforderungen an den Leser, wenngleich auch dies keine unüberwindbare Hürde darstellt. Kapitel 6 behandelt soziale Dilemmata, also den Umstand, dass individuell-rationales Handeln keine kollektiv-rationalen (Pareto optimalen) Ergebnisse hervorbringen muss. Das Buch nimmt dabei Bezug auf bekannte Spiele (z. B. die Kollektivgutproblematik oder öffentliche Güter), aber auch weniger geläu-

fige Dilemmata wie Spiele zu Paniksituationen, verstopften Straßen oder Shubiks Dollar Auktion und Hofstadters Dilemma der großen Zahl. Ausführlich diskutiert Diekmann auch sein eigenes Freiwilligendilemma. Eine kurze Darstellung von Lösungsvorschlägen leitet zu Kapitel 7 über, in dem wiederholte Spiele als Lösung des Problems sozialer Ordnung vorgestellt werden. Dabei wird einführnd das bekannte Computerturnier von Robert Axelrod besprochen um danach die Konzepte zur Analyse endlich und unendlich wiederholter Spiele einzuführen (Diskontierung, Schatten der Zukunft und Folk Theorem). Das Kapitel endet mit der Besprechung eines alternativen Lösungsvorschlags für soziale Dilemmata, nämlich dem Konzept der Reputation. Implizit verweist Diekmann hier natürlich auf die Effekte von Netzwerken auf Transaktionen in Dyaden. Akteure müssen sich nicht zwingend auf die eigene, vergangene Erfahrung mit einem anderen Akteur verlassen, sondern können auch auf den Erfahrungsschatz anderer (Reputation) zurückgreifen.

Kapitel 8 behandelt die evolutionäre Spieltheorie und das Konzept der evolutionär stabilen Strategien. Dieses Kapitel erscheint dahingehend als ein Bruch in der bis anhin logischen Abfolge der Kapitel, als dass die evolutionstheoretische Basis dieses Ansatzes mit der üblichen spieltheoretischen Annahme nicht vereinbar ist. Verhaltenswahlen finden nicht im Bewusstsein strategischer Interdependenzen statt. Die Theorie evolutionärer Spiele ist für die Sozialwissenschaften aber dennoch von Bedeutung, liefert sie doch alternative Begründungen für bestimmte Einsichten der nichtkooperativen Spieltheorie. Gleichzeitig ergänzt sie die Ergebnisse der traditionellen Analyse oder regt zu deren Kritik und Modifikation an. Logischer wäre jedoch gewesen, die Darstellung der evolutionären Spieltheorie an das Ende des Buches zu setzen.

Kapitel 9 beschließt die Darstellung der Konzepte und Begriffe der klassischen Spieltheorie, indem Situationen mit unvoll-

ständiger Information behandelt werden. Am Beispiel des Vertrauensspiels werden die zentralen Inhalte (rationales Lernen (Satz von Bayes), (perfekte) Bayesianische Nash-Gleichgewichte, Signalspiele) dargestellt und vor dem Hintergrund realer Situationen besprochen, in denen die Präferenzen der Mitspieler ja meistens nicht vollständig bekannt sind.

Dass die Annahmen der Spieltheorie (und das zugrundeliegende Menschenbild) für deskriptive Erklärungen häufig zu restriktiv sind, wird in dem Buch nicht verschwiegen. Das abschließende Kapitel 10 befasst sich daher mit der experimentellen Spieltheorie, welche versucht, die Anomalien in den Vorhersagen der klassischen Spieltheorie durch diverse, auch kritisch zu betrachtende Erweiterungen der Modelle aus dem Weg zu räumen. Der Schwerpunkt in dem Kapitel liegt auf den Begriffen Altruismus, Reziprozität und sozialen Normen sowie entsprechenden experimentellen Ergebnissen dazu.

Wie man es von Andreas Diekmann gewohnt ist (z. B. in seiner Einführung in die empirische Sozialforschung), nimmt auch dieses Buch Bezug auf eine große Anzahl realer und zum Teil historischer Vorkommnisse, welche spieltheoretisch „analysiert“ werden. Dies reicht von bekannten Darstellungen, wie z. B. der Kubakrise, dem Gebrauchtwagenkauf, der Kooperation unter feindlichen Soldaten im 1. Weltkrieg oder von Neumann und Morgensterns spieltheoretischer Darstellung von Sherlock Holmes Flucht vor Professor Moriarty bis zu durchaus süffisanten Anekdoten zum Doping im Fußball, zu der Koordination auf Spurweiten von Eisenbahnen, zu Staus auf Straßen oder zu der Versteigerung von Telefonfrequenzen in den USA. Mit diesen Beispielen gelingt es dem Buch immer wieder, den theoretischen Stoff greifbar zu machen und vor allem, dessen Relevanz für die Analyse interdependenter Situationen hervorzuheben. Wenngleich das Buch keine abschließende Einführung in die Spieltheorie ist, so dürfte es doch viele Leser für die Spieltheorie gewinnen, welche

sich bis anhin von den mathematisch überladenen Lehrbüchern abgeschreckt fühlten. Das Buch beginnt sehr einfach, setzt keine Vorkenntnisse voraus, führt logisch aufbauend auf das Behandelte neue Konzepte und Begriffe ein und steigert die Komplexität nur dort, wo es unumgänglich ist.

Den aus Sicht des Rezensenten vielleicht wichtigsten Beitrag des Buches zu den Sozialwissenschaften – neben der generellen Heranführung an die Spieltheorie – liefert Diekmann auf den Seiten 67 und 68. Er erkennt als einer der Wenigen, dass die Spieltheorie das Mikro-Makro Problem auf eine einfache Weise zu lösen vermag. Die Diskussion um die Aggregation individueller Handlungen zu einem kollektiven Ausgang beschäftigt die Sozialwissenschaften seit langer Zeit, ohne dass wirklich befriedigende Lösungen vorgelegt wurden. Der Hinweis Diekmanns, dass spieltheoretische Lösungskonzepte wie das Nash-Gleichgewicht individuelle Handlungen in kollektive Effekte transformieren, war lange überfällig.

Den inhaltlichen Kapiteln schließt sich, dem Vorbild angelsächsischer Lehrbüchern folgend, ein Anhang an, in dem sich einerseits für einige Begriffe der Spieltheorie ihre formalen Definitionen wiederfinden und andererseits diese und weitere Begriffe nochmals lexikonartig beschrieben werden. Obwohl dieses „Lexikon“ nicht den gesamten Inhalt des Buches mit entsprechenden Begriffsdefinitionen abdeckt, mag es gerade für den Einsteiger in das Themengebiet der Spieltheorie doch sehr hilfreich sein.

Andreas Diekmann ist es gelungen, eine sehr gute Einführung in die Spieltheorie vorzulegen, welche ein sehr breites Zielpublikum ansprechen sollte. Da dies natürlich auf Kosten einer gewissen Breite und Tiefe geht, gilt es auch einige kritische Punkte zu benennen. Wünschenswert wäre auf jeden Fall eine erweiterte Diskussion von Spielen mit unvollständiger Information gewesen (z. B. zu der Bildung von Wahrscheinlichkeiten, wenn die Regel von Bayes

nicht anwendbar ist und Gleichgewichtskonzepten für „out-of-equilibrium“ Handlungen von Spielern (z. B. „trembling-hand“ Gleichgewichte)). Generell fehlt dem Buch die Diskussion zu den mittlerweile gängigen Refinements des Konzeptes des Nash-Gleichgewichts. Auch werden einfache Konzepte wie rationalisierbare Strategien oder korrelierte Strategien – welche vor gut 35 Jahren von Aumann eingeführt und heute von Herbert Gintis als das Lösungskonzept für nichtkooperative Spielsituationen angepriesen werden – nicht besprochen. Ebenso fehlt eine Diskussion der Problematik der Gleichgewichtsauswahl bei Spielen mit multiplen Gleichgewichten, einem nicht unwesentlichen Problem der Spieltheorie. Schließlich muss angemerkt werden, dass das Buch lediglich eine Einführung in die nichtkooperative Spieltheorie ist und keine Lösungskonzepte der kooperativen Spieltheorie (z. B. Kern, Nukleus, Shapley-Wert, Nash-Lösung) diskutiert. Trotz dieser Einwände sollte dieses Buch der Spieltheorie in den deutschsprachigen Sozialwissenschaften zu mehr Popularität verhelfen, zeigt es doch einerseits deren Relevanz für die Analyse interdependenter Situationen und andererseits, dass man kein Mathematiker sein muss, um spieltheoretische Modellierungen vornehmen zu können.

THOMAS GAUTSCHI, MANNHEIM

Ankündigungen

Ausschreibung

ALLBUS-Preis 2012

ALLBUS-Preis 2012

Die seit 1980 alle zwei Jahre durchgeführte Allgemeine Bevölkerungsumfrage der Sozialwissenschaften – ALLBUS – ist ein zentraler Bestandteil der sozialwissenschaftlichen Infrastruktur in Deutschland. Mit dem ALLBUS wird der Profession eine qualitativ hochwertige Datenbasis für Sekundäranalysen zur Verfügung gestellt, die stark genutzt wird. Mittlerweile sind in der ALLBUS-Bibliographie fast 2.000 Arbeiten verzeichnet, in denen ALLBUS-Daten verwendet werden.

Die Verleihung des *ALLBUS-Preises* soll diese besondere Bedeutung des ALLBUS als methodisch anspruchsvolle und inhaltlich reichhaltige Datenquelle für die Sozialwissenschaften unterstreichen.

Mit dem *ALLBUS-Preis* sollen herausragende wissenschaftliche Veröffentlichungen in Zeitschriften und Sammelbänden prämiert werden, in denen ALLBUS-Daten eine zentrale Rolle spielen. Über die Verleihung des Preises, der mit 1.000 € dotiert ist, entscheidet eine Jury, die aus den Mitgliedern des wissenschaftlichen Beirats des ALLBUS besteht. Die Originalität und Bedeutung der in der Arbeit behandelten – methodischen oder inhaltlichen – Fragestellung sowie das Niveau der Analyse sind maßgebliche Kriterien der Bewertung.

Eingereicht werden können alle in Deutsch oder Englisch verfassten sozialwissenschaftlichen Arbeiten mit ALLBUS-Daten, die in den letzten drei Jahren in Zeitschriften oder Sammelbänden publiziert wurden. Sowohl Eigenbewerbungen als auch Vorschläge durch Dritte werden berücksichtigt.

Einzureichen sind:

- die Publikation, in Papierform sowie als pdf-Datei
- eine maximal zweiseitige Kurzfassung der Arbeit
- ein Lebenslauf des Autors/der Autoren (mit Verzeichnis der – maximal 10 – wichtigsten Publikationen)

Einsendungen bitte bis spätestens zum **31. März 2012** an:

GESIS – Leibniz-Institut für Sozialwissenschaften
ALLBUS-Preis
Postfach 12 21 55
68072 Mannheim

GOR 12
International GOR Conference
GENERAL ONLINE RESEARCH

Duale Hochschule Baden-Württemberg
Mannheim
5. – 7. März 2012

GOR 12
International
GOR Conference
GENERAL ONLINE
RESEARCH

Vom 5. bis 7. März 2012 findet die General Online Research Konferenz (GOR 12) an der Dualen Hochschule Baden-Württemberg in Mannheim statt. Die dreitägige Veranstaltung besteht aus einem Vortrags- und Workshop-Programm. Zum einen wird der Status Quo der Online-Forschung diskutiert: Welche aktuellen Ergebnisse gibt es zu Themen der Methodik oder der angewandten Marktforschung? Einen weiteren Schwerpunkt der Konferenz bilden neue Entwicklungen und Forschungsmethoden in den Feldern Social Media, Soziale Netzwerke und Zivilgesellschaft. Neben klassischen Vorträgen, und Posterpräsentationen wird es auch zum ersten Mal ein sogenanntes GORcamp geben, in dem im Anschluss an Kurzvorträge aktuelle Themen der Online-Forschung diskutiert werden. Erwartet werden rund 300 internationale Forscher aus Hochschulen, Instituten und Unternehmen. Die Konferenzsprache ist Englisch. Veranstaltungsort ist die Duale Hochschule Baden-Württemberg in Mannheim.

Weitere Informationen zur Konferenz sind der folgenden Website zu entnehmen: <http://www.gor.de>.

Autorinnen und Autoren Jahrgang 5 (2011)

- Dorothee Behr, Mannheim
- Jan Pablo Burgard, Trier
- Katrin Drasch, Nürnberg
- Andrea Dürnberger, Bamberg
- Siegfried Gabler, Mannheim
- Matthias Ganninger, Mannheim
- Britta Gauckler, Wiesbaden
- Jan-Philipp Kolb, Trier
- Thomas Körner, Wiesbaden
- Lars Eric Kroll, Berlin
- Britta Matthes, Nürnberg
- Natalja Menold, Mannheim
- Gerrit Müller, Nürnberg
- Ralf Münnich, Trier
- Andreas Pöge, Bielefeld
- Evi Scholz, Mannheim
- Cornelia Züll, Mannheim

Rezensentinnen und Rezensenten Jahrgang 5 (2011)

- Christian Deindl, Köln
- Thomas Gautschi, Mannheim
- Peter Kriwy, Erlangen-Nürnberg
- Ulrich Rosar, Düsseldorf

Gutachterinnen und Gutachter Jahrgang 5 (2011)

Wir danken folgenden Kolleginnen und Kollegen, die für den Jahrgang 5 (2011) der Methoden – Daten – Analysen Manuskripte begutachtet haben:

- Tobias Bachteler, Duisburg
- Nico Dragano, Düsseldorf
- Sabine Häder, Mannheim
- Rüdiger Jacob, Trier
- Thomas Klein, Heidelberg
- Brian Kleiner, Lausanne
- Tobia Lakes, Berlin
- Oliver Lipps, Lausanne
- Henning Lohmann, Berlin
- Florian Meinfelder, Bamberg
- Gerrit Müller, Nürnberg
- Alexandra Nonnenmacher, Hannover
- Klaus Pforr, Mannheim
- Jürgen Schiener, Mainz
- Bernhard Schimpl-Neimanns, Mannheim
- Barry Schouten, Den Haag
- Martin Spieß, Hamburg
- Uwe Warner, Differdange
- Cornelia Weins, Bochum
- Heike Wirth, Mannheim

Hinweise für unsere Autorinnen und Autoren

Methoden – Daten – Analysen (MDA) veröffentlicht Beiträge aus dem Bereich der Empirischen Sozialforschung, insbesondere aus dem Bereich der Umfragemethodik. Im Vordergrund stehen Artikel, welche die methodischen und/oder statistischen Kenntnisse der Profession erweitern, sowie Beiträge, die sich mit der Anwendung der Methoden der Empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder solche, in denen ein statistisches Verfahren exemplarisch angewandt wird. Obwohl der Schwerpunkt auf Umfragemethoden liegt, sind Beiträge zu anderen methodischen Bereichen willkommen. Die Artikel sollen für eine breite Leserschaft von Wissenschaftlern und Praktikern im Bereich der Empirischen Sozialforschung verständlich sein.

Manuskripte, die bereits an anderer Stelle veröffentlicht sind oder gleichzeitig anderen Publikationsorganen zur Veröffentlichung angeboten worden sind, werden grundsätzlich nicht berücksichtigt. Eine spätere Veröffentlichung eines in der MDA erschienenen Beitrages ist möglich, sofern an exponierter Stelle auf die Ersterscheinung des Beitrages in der MDA hingewiesen wird.

Jeder Beitrag, der zur Veröffentlichung in MDA eingereicht wird, wird zunächst von den Herausgebern danach bewertet, ob er für eine Veröffentlichung grundsätzlich in Frage kommt.

Falls die Herausgeber einer Veröffentlichung grundsätzlich ablehnend gegenüber stehen, werden die Autoren unter Angabe von Gründen für diese Entscheidung informiert.

Falls die Herausgeber zur Ansicht gelangen, dass der Beitrag grundsätzlich zur Veröffentlichung in Frage kommt, wird er anonymisiert an mindestens zwei unabhängige Gutachter verschickt, die um eine Stellungnahme gebeten werden. Im Zweifelsfalle wird ein drittes Gutachten eingeholt.

Wird ein Beitrag nach Beschluss der Herausgeber in das Begutachtungsverfahren gegeben, erfolgt die abschließende Entscheidung über ein Manuskript auf der Basis der Gutachten durch die Herausgeber. Im Falle einer Ablehnung erhalten die Autoren eine ausführliche Begründung für die Ablehnung. Wird eine Überarbeitung eines Beitrages für erforderlich gehalten, erhalten die Autoren detaillierte Überarbeitungshinweise.

Unabhängig vom Ergebnis des Begutachtungsverfahrens werden die Autoren von der Entscheidung durch die Redaktion per E-Mail informiert.

Die folgenden Regeln sind bei der Abfassung von Manuskripten zu beachten:

Manuskripte müssen per E-Mail (mda@gesis.org) eingereicht werden. Der Umfang der Manuskripte soll inklusive Leerzeichen alles in allem nicht mehr als 70.000 Zeichen betragen.

Den Beiträgen sind Abstracts in Deutsch und Englisch (jeweils ca. 15 Zeilen) voranzustellen. Auch der Titel des Beitrages ist in Deutsch und Englisch einzureichen.

Um die Anonymität der Beiträge zu wahren, darf in einem Manuskript nur der Titel des Beitrages enthalten sein, nicht aber Namen oder Anschriften der Autoren; Name und Anschrift der Autoren müssen, gemeinsam mit dem Titel des Beitrages, auf einer separaten Seite eingereicht werden.

Beiträge sind mit dem Dezimalklassifikationssystem zu untergliedern (1 - 2 - 2.1 - 2.2 - 3 usw.). Die Gliederungstiefe geht dabei höchstens auf *eine* Stelle nach dem Punkt.

Tabellen enthalten Tabellenummer und Titel im Tabellenkopf, Abbildungen werden analog behandelt.

Grafiken sind mittels gängiger Grafiksoftware zu erstellen. Ist eine spezielle Grafiksoftware erforderlich, übernimmt der Autor/die Autorin die endgültige Formatierung der Grafiken in eigener Regie.

Bei der Erstellung von Tabellen und Grafiken ist zu berücksichtigen, dass der Satzspiegel 11,5 cm (Breite) x 18,5 cm (Höhe) beträgt. Die Grafiken sind als jpeg- oder tif-Dateien in *Graustufen (CMYK)* mit einer Auflösung von mindestens 300 dpi zu liefern.

Die Beiträge sind unter Wahrung der gültigen Rechtschreiberegeln (neue Rechtschreibung) zu erstellen.

Werden in einem Beitrag empirische Daten verwandt, muss die Möglichkeit der Replikation bestehen. Im Falle einer Veröffentlichung in der MDA erklären sich die Autoren daher schriftlich bereit, Dritten auf deren Anfrage hin die Daten und ProgrammROUTINEN zur Verfügung zu stellen.

Anmerkungen und Fußnoten sind mit der Fußnotenfunktion des Schreibprogrammes (im Normalfall Word) zu erstellen; bitte nicht gesondert formatieren. Fußnoten sind nur für inhaltliche Kommentare vorzusehen, nicht für bibliographische Hinweise.

Literaturhinweise im Text sind nach den folgenden Mustern aufzuführen: Müller (2002) – Schulze und Mayer (2003) – Müller, Mayer und Schulze (2004) – Müller et al. (2005) – Müller (2006: 75) – (vgl. Müller 2007: 75) – (Müller 2008; Mayer/Müller/Schulze 2009).

Das Literaturverzeichnis ist wie folgt zu gestalten:

Buchveröffentlichungen:

Strobl, R. und W. Kühnel, 2000: Dazugehörig und ausgegrenzt. Analysen zu Integrationschancen junger Aussiedler. Weinheim/München: Juventa.

Zeitschriftenbeiträge:

Becker, R., R. Imhof und G. Mehlkop, 2007: Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz. Empirische Befunde eines Methodenexperiments. *Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung* 1 (2): 131-159.

Beiträge in Büchern:

Braun, M. und I. Borg, 2004: Berufswerte im zeitlichen und im Ost-West-Vergleich. S. 179-199 in: R. Schmitt-Beck, M. Wasmer und A. Koch (Hg.): Sozialer und politischer Wandel in Deutschland. Analysen mit ALLBUS-Daten aus zwei Jahrzehnten. Wiesbaden: VS-Verlag für Sozialwissenschaften.

Internetquellen:

Stadtmüller, S. und R. Porst, 2005: Zum Einsatz von Incentives bei postalischen Befragungen. *GESIS How-to-Reihe*, Nr. 14. Mannheim: GESIS. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/how-to14rp.pdf (1.12.2008).

Esser, H. 1986: Über die Teilnahme an Befragungen. *ZUMA-Nachrichten* 18: 38-47. PID: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-210300>.

Datenfile:

Forschungsgruppe Wahlen, Mannheim: Zur politischen Lage in Niedersachsen im Januar 2008. *GESIS Köln, Deutschland ZA Studie* Nr. 4863; doi: 10.4232/1.4863.