Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

www.diw.de

SOEP The German Socio-Economic Panel Study

SOEPpapers

on Multidisciplinary Panel Data Research

238

Carsten Sauer et al.

A Factorial Survey on the Justice of Earnings within the SOEP-Pretest 2008

Berlin, November 2009

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

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

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

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

The SOEPpapers are available at
http://www.diw.de/soeppapers

ISSN: 1864-6689 (online)

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

Contact: Uta Rahmann | urahmann@diw.de

A Factorial Survey on the Justice of Earnings within the SOEP-Pretest 2008

Carsten Sauer, Stefan Liebig, Katrin Auspurg, Thomas Hinz, Andy Donaubauer, and Jürgen Schupp

November 2009

The research project 'The Factorial Survey as a Method for Measuring Attitudes in Population Surveys' is funded by Deutsche Forschungsgemeinschaft (DFG).

Prof. Dr. Thomas Hinz

Dipl. Soz. Katrin Auspurg

Professur für empirische Sozialforschung mit Schwerpunkt Demoskopie
Universität Konstanz
Geisteswissenschaftliche Sektion
Postfach D-40
78457 Konstanz
Tel.: 07531 88-2349/-2167
Fax: 07531 88-4085
katrin.auspurg@uni-kostanz.de

Prof. Dr. Jürgen Schupp

DIW Berlin und FU Berlin

Prof. Dr. Stefan Liebig

Carsten Sauer, M.A.

Andy Donaubauer

Professur Soziale Ungleichheit und Sozialstrukturanalyse
Universität Bielefeld
Fakultät für Soziologie
Postfach 10 01 31
33501 Bielefeld
Tel.: 0521 106-4616/-6948
Fax: 0521 106-6479
carsten.sauer@uni-bielefeld.de

Homepages:

http://www.uni-konstanz.de/hinz/?cont=faktorieller_survey&lang=de

http://www.uni-bielefeld.de/soz/arbeitsbereiche/sozialstrukturanalyse/faktsurvey.html

A Factorial Survey on the Justice of Earnings within the SOEP-Pretest 2008

Abstract

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

JEL: C81, D63, J31

1 Introduction

Over the last decade there has been a marked increase of studies in academic and non-academic attitude and decision research which use a comparatively new method: the *factorial survey design*.[1] The factorial survey is an experimental method which confronts respondents with hypothetical descriptions of objects or situations (vignettes). In these descriptions some attributes (dimensions) are experimentally varied. The respondents' task is to normatively evaluate each of these descriptions or to indicate what they would recommend or how they would act in the presented situations. The aim is to identify those dimensions which affect the evaluation or the decision and to assess their relative impact. The issues addressed in various studies resemble attitudes towards the justice of income and wages (Alves and Rossi 1978; Hermkens and Boerman 1989; Jann 2003; Jasso 1994; Jasso and Meyersson Milgrom 2008; Jasso and Rossi 1977; Jasso and Webster 1997, 1999), on just taxation (Liebig and Mau 2005) and just punishment (Berk and Rossi 1977; Miller et al. 1986). There are also studies on the measurement of norms and values (Beck and Opp 2001; Jasso and Opp 1997; Mäs et al. 2005) and the degree of life satisfaction (Kapteyn et al. 2008). Other studies simulate bargaining situations (Auspurg and Abraham 2007; Auspurg et al. 2009b) or deal with trust (Barrera and Buskens 2007). The joint endeavor of this kind of studies is to measure the evaluation of certain outcomes e.g. income, grades, satisfaction, penalties, or certain decisions-making processes which strongly depend on the particular situation and the social context. The use of the factorial survey method is driven by the promise that it allows for a more differentiated measurement compared to classical item based approaches in attitudinal research. The main advantages of the factorial survey design in comparison to item-based measurement are: (1) The vignettes describe a situation more realistically – in everyday life people judge, decide or evaluate on a bundle of information and this is what factorial designs consider in their multidimensional descriptions; and (2) the experimental approach of the design, where respondents rate vignettes in which the dimensions vary independently from each other.

Despite the growing applications in attitudinal research there is little empirical knowledge on the methodological implications and effects of the factorial survey design. This is especially true for the use of factorial surveys in population surveys. Most of the studies are using homogenous respondent populations, most often students, and are carried out in the lab or comparable settings (e.g. classroom). As the research

[1] The factorial survey was established in the social sciences by Peter H. Rossi in his dissertation in 1951. It was used for the measurement of social status and prestige of households (Alves/Rossi 1978; Rossi 1979; Rossi/Nock 1982). Rossi's central goal was the development of a method of measurement that distinguishes between the relative relevance of several factors for social attitudes (Rossi/Anderson 1982:15 et seq..; Rossi/Nock 1982).

design and the respondents' rating task are usually very complex a number of methodological effects may occur and, as a consequence, may cause methodological artifacts. Against this background, we implemented a factorial survey module in the SOEP-pretest 2008 to find out which practical and methodological problems are associated with this technique, especially when it is used in large scale population surveys. The main focus is hereby on the acceptance of and comprehensibility for respondents and interviewers.

The subject matter of the implemented factorial survey module consists in the justice of wages. Respondents were confronted with 25 descriptions of fictitious earners who differed in certain characteristics such as gender, age, education, occupation or level of individual effort. In each of the cases respondents had to evaluate whether the presented gross income was just or unjust.

The paper is organized as follows: first, we give an overview on the construction of factorial survey designs (Chapter 2). Second, we describe the implementation of the instrument in the SOEP-Pretest 2008 and the respondent and vignette sample (Chapter 3). Third, we investigate the capability of the factorial survey design. We analyze the direct feedbacks of respondents and interviewers as well as the respondent behavior using data on the response times and consistency of responses (Chapter 4). Fourth, we present some results with regard to the perceived justice of earnings (Chapter 5) and in the last chapter we summarize our findings and stress the main methodological implications of this study.

2 The Factorial Survey Approach

Constructing the vignettes describing persons, situations or objects is the most important step in designing factorial surveys. At first those characteristics or dimensions of persons or objects have to be identified which hypothetically effect the response behavior. This step should be based on theoretical considerations (Alves 1982; Jasso 2006) and be carried out very carefully as seemingly marginal specifications (such as the definition of the number of levels used) have a great impact on the conceptual design and analysis of factorial surveys. The main task in defining the dimensions (i.e. the characteristics of the fictitious earners) is to find those that are relevant for the evaluations (of just earnings).

We intend to construct vignettes which describe persons who work full time and earn a certain amount of gross income. The rating task is to evaluate whether the income of the described person is just or unjust. Qualitative dimensions (such as the sex of the person) have a naturally limited number of levels (male and female). In contrast to this, the range and number of level of continuous dimensions (such as age) have to be defined. Age for example could be restricted to a range from 30 to 60 years with four (30, 40, 50

and 60 years) or seven different (30, 35, 40, 45, 50, 55, 60) levels. It is important to note that the number of parameters which have to be estimated in the data analyses increases exponentially with the number of dimensions and levels (Alves 1982; Jasso 2006).

After specifying the dimensions and levels the *vignette universe* of all possible vignettes is generated by multiplying all attribute levels with each other (Cartesian product). In the case of three dimensions with five levels each the universe consists of 5*5*5 = 125 vignettes.[2] Usually the complete vignette population cannot be rated by single respondents because of their vast extensiveness. The solution is to work with samples only (similar to matrix-sampling, for detail: Thomas et al. 2006). One may draw a unique sample for each respondent or a few samples rated by a number of respondents (so called decks) in order to obtain multiple ratings on each vignette (Jasso 2007).

The vignette sample can be obtained by using a *random* (Jasso 2006) or a *quota design* (Dülmer 2007; Kuhfeld 2005; Kuhfeld et al. 1994; Steiner and Atzmüller 2006). In both cases, the aim is to keep correlations low between different attributes. The dimensions stand orthogonal to each other in a full factorial design (vignette universe) and so all main and interaction effects can be estimated. This assumption gets relaxed in a reduced sample because some effects will be confounded. As recent studies suggest quota designs are more efficient compared to random samples due to their higher orthogonality and balance (that is: maximum variance of attribute levels). This is especially the case within small samples (Dülmer 2007; Steiner and Atzmüller 2006).[3]

For the evaluation task of each vignette different scales can be used. The main criterion hereby is: Is it necessary to employ a metric scale or is an ordinal scale appropriate? In most vignette studies rating scales with up to 15 categories are used (Dülmer 2001; Mäs et al. 2005; Schulte 2002; Thurman et al. 1988) but there are also a lot of applications using magnitude scaling (cf. Wallander 2009).

---

[2] In this full factorial design the correlation between the dimensions is zero. Some combinations lead to unrealistic scenarios (for example a medical doctor without a university degree) and were excluded from the vignette universe. This is why the correlation of dimensions in the resulting vignette universe is unequal to zero.

[3] These quota designs systematically draw vignettes out of the universe with the overall goal to have all level combinations uncorrelated. This can be done by statistical software. In regular cases the algorithm detects the maximum efficient design. Besides low correlation, efficiency also means a maximum variance of dimensions. Alongside the fractional factorial designs, which only maximize orthogonality, D-efficient designs are available. In D-efficient designs orthogonality looses on ground because maximum variance of attributes gets the main target criterion. The D-efficient design should be the preferred approach especially for vignette populations where implausible combinations have been deleted.

3 A Factorial Survey in the SOEP Pretest 2008

The program of the annual SOEP questionnaire for the following wave is pretested in each summer of the preceding year. The objective of this pretest is to test new modules and modifications of questions. Since a couple of years the SOEP-Pretest goes far beyond the standard format of a pretest. Since 2002 the sample size is around 1,000 respondents and considered representative for the German resident population of 16 years and older (Siegel et al. 2009).

Within the SOEP there are two main differences between the pretest and the main survey. First, all interviews in the SOEP-Pretest are programmed as CAPI versions (in contrast, in the main survey most of the interviews are based on paper and pencil questionnaires), that is why this SOEP-Pretest is useful to test experimental designs.[4] Second, whereas in the main survey all members of a household from the age 16 on are interviewed, the SOEP-Pretest is arranged in a much simpler fashion. There is *one* questionnaire to be filled out by *one* member of a household. The pretest sample is not related to the main survey, meaning that these respondents are not part of the panel study. The interviews of the SOEP-Pretest 2008 were conducted in the period from 1st to 31st August in 2008. The duration of the whole questionnaire was planed for 45 minutes which is matching with the realized median. In sum 1,066 interviews were conducted.

Within the SOEP-Pretest the factorial survey module focuses on the justice evaluation of the wages of fictitious full time employees (40 hours per week) who are described by ten dimensions. The respondents had to rate in sum 25 vignettes, where the last vignette consisted of two additional dimensions on the nationality of the earner and his or her duration of stay in Germany. In the following we will concentrate on the results for the 24 vignettes with ten dimensions.[5]

3.1 Vignette Dimensions and Levels

The ten dimensions presented on the vignettes were based on the evidences of earlier vignette studies on the justice of earnings (Alves 1982; Alves and Rossi 1978; Jann 2003; Jasso 1978; Jasso and Rossi 1977; Jasso and Webster 1997, 1999). These studies show that the dimensions age, gender, number of children, occupation and education

---

[4] Further topics in the SOEP-Pretest are: (1) daily moods: self assessment of the respondents in regard to moods in a typical week, (2) Questions referring to *strength of character*: a German translation to the *Values in Action (VIA) – Classification of Strengths* concept, (3) new questions to measure (chronic) diseases.

[5] All respondents had to rate a blind vignette with the help of the interviewer at the beginning. The content of this vignette was: "A 35 year old single man with vocational training works as a hair dresser in a small company which achieves substantial gains. His performance on the job is outstanding and he earns a gross income of 350 Euro per month. Is the gross income of this employee in your opinion just or unjust?".

have a significant influence on justice evaluations. Further dimensions that are commonly known as relevant from justice research and related fields were added. These are the performance on the job and the marital status (Liebig and Schupp 2005, 2008a,b; Struck et al. 2006). As the size and economic situation of the company (Abraham and Hinz 2005a,b) are important for the actual income, we assume that these two dimensions are also relevant for the subjective justice evaluations. Table 1 gives an overview of the dimensions and the levels used to describe the fictitious earners.

Table 1: Dimensions and their levels

Dimension	Levels
Age	25/ 35/ 45/ 55 years
Sex	Male/ female
Marital status	Single earner, married/ double earner, married/ single
Vocational training	With/ without vocational training/ with university degree
Occupation	Manufacturing laborer/ door keeper/ locomotive engine driver/ administrative associate professional/ hairdresser/ social work professional/ computer programmer/ electrical engineer/ general manager/ medical doctor
Gross income[6]	500€/ 950€/ 1200€/ 1500€/ 2500€/ 3800€/ 5400€/ 6800€/ 10000€/ 15000€
Children	No child/ 1 child/ 2 children/ 3 children/ 4 children
Performance on the job	Below/ above average/ average
Economic situation of the company	High profit/ economical solid/ threatened by bankruptcy
Company size	Small/ medium/ large

3.2 Vignette Universe and Illogical Cases

The vignette universe is the combination of all attribute levels with each other. In the present study this combination of all dimensions and their levels sums up to 980,000 cases. Some combinations were excluded from the vignette universe as they describe cases which can definitely not be found in the *real world* and are therefore illogical. This is true for certain combinations of income and occupation:

- Gross income of more than 3,800 Euro for manufacturing workers
- Gross income of more than 5,400 Euro for doorkeepers and engine drivers
- Gross income of more than 6,800 Euro for administrative associate professional, hair dressers and social workers
- Gross income below 1,200 Euro for electrical engineers
- Gross income below 2,500 Euro for general managers or medical doctors

[6] The categories are related to the percentiles of the income distribution of fulltime employees 2007 in Germany (data source: SOEP 2007). The highest and lowest categories are added to have extreme cases.

There are also combinations of vocational training and occupation which are definitely unrealistic:

- Electrical engineers without vocational training
- Physicians without a university degree

We drew the vignette sample with a quota design (D-efficient design) under exclusion of the mentioned illogical cases (Kuhfeld 2005; Kuhfeld et al. 1994). Firstly, we drew 240 vignettes with a D-efficiency of over 90 and secondly we fractionalized them on ten decks with 24 vignettes[7] each.

3.3 Rating Task and Presentation of Vignettes

The rating task was a three step procedure: first the respondents had to evaluate whether the gross income of the person described on the vignette was just or unjust. The respondent continued with the next vignette if he/she had judged the income as just. If the respondent evaluated the income as unjust he/she had to reconcile in a second step whether the income was too high or too low. In the third step the respondent had to express the amount of felt injustice using a metric scale from 1, some injustice, to 100, extreme injustice. A disadvantage of this procedure could be that respondents are more familiar with rating scales which means that it may be more difficult to use this kind of scale. The advantage of a metric scale is that respondents have the opportunity to differentiate their judgments in a finer way compared to, for example, a five-point rating scale. Figure 1 presents a vignette with the rating steps.

The complete questionnaire within the SOEP-Pretest 2008 was designed as a computer assisted personal interview (CAPI)[8] and the interviewer read the questions to the respondent. In the vignette module, however, the vignettes were presented to the respondent on a computer screen. The interviewer was sitting next to the respondent to answer any questions that occurred during the evaluation task. In an introduction screen the respondent additionally got information about what to do and how to use the scale. Afterwards the respondent judged an example of a vignette and was able to ask the interviewer for help if there were any ambiguities. After this blind vignette the respondents were randomly assigned to one of the ten decks with 24 vignettes. The vignettes were programmed in a fixed order which means that respondents could not skip to the next vignette without a rating.[9] Therefore, the respondents were forced to rate every vignette.

---

[7] The maximum D-efficiency in a symmetric design is 100. Often the best achievable efficiency is less than 100 so one has to choose the best out from some alternative designs. A D-efficiency above 90 is deemed to be good.

[8] We thank Andreas Stocker, TNS Infratest Sozialforschung, Munich, who bestowed great care on the implementation and programming of the vignette module in the computer assisted questionnaire.

[9] This procedure is somewhat uncommon with regard to measuring the acceptance of a new module. But it is possible to reconstruct refusals by very short response durations (see Chapter 4.2.1).

Figure 1: Vignette with Ten Dimensions and Rating Task

A 45-year old woman, married, with two children,
and a husband who does not have own income,
she has vocational training and
works as a hairdresser in a large company, which is threatened of bankruptcy,
Her performance on the job is below the average,

She earns **1200 Euro** gross income per month before taxes.

Your rating:

F 1:
From your point of view, is the gross income for this person just or unjust?

- □ Gross income is just (→ carry on with the next person description)
- □ Gross income is unjust (→ carry on with F 2)

F 2:
Is the gross income unjustly too high or too low?

- □ unjustly too high (→ carry on with F 3)
- □ unjustly too low (→ carry on with F 3)

F 3:
With regard to your personal feeling, which number between 1 and 100 describes most adequately the amount of injustice?

3.4 Respondents and Vignette Sample (SOEP Pretest 2008)

The realized sample of the SOEP-Pretest relies on a three step probability sampling procedure according to the ADM-Design. The response rate reported by TNS Infratest Sozialforschung is about 50 percent (Siegel et al. 2009). The realized sample (N = 1,066) was weighted in regard to regional and demographic distribution. It is warranted that the weighted sample is representative for the German population, even though only unweighted data are used in the report at hand. Table A1 in the Appendix gives an overview of the realized sample.

Respondents were assigned to one of the ten vignette decks randomly. The distribution of respondents to each deck is reported in Table 2. The number of realized respondents by deck ranges between 96 (decks 2 and 7) and 127 (deck 9).[10] The correlations

[10] The range does not differ significantly from chance (Chi-Square-Test = 9.0; df = 9; p = 0.436).

between the dimensions in the whole vignette sample as well as in the single decks are very low (see Appendix A2), which means that the design is efficient in a statistical sense.

Table 2: Deck Frequencies

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

4 Methodological Results

The main research question focuses on methodological effects resulting from the higher-than-average complexity of a factorial survey and its application in population surveys. This is investigated by using three sources of information: (1) respondents feedback, (2) interviewer impressions, and (3) response behavior. After a short depiction of openly asked respondent feedbacks, the more profound analyses of the latter aspects are presented. We attempt to analyze in detail the differences in respondent behavior. As factorial surveys require much more attention and concentration from respondents, age and educational effects are likely to occur. Therefore we categorize respondents in the following analyses in three age groups (between 16 and 39 years, 40 to 65 years, and over 65 years)[11] and three educational groups (general educational level: lower (Hauptschule), middle (Realschule) and higher secondary school certificate (Abitur)).

4.1 Respondents Feedback and Interviewer Impressions

The questionnaire provided the opportunity to criticize and comment the vignette module in an open question. A total of 191 respondents made a comment. It is not traceable whether the other respondents had no critique at all or did not want to answer the open question. Table 3 shows the mostly mentioned comments.

[11] The distribution is almost equal between the groups. The intervals are similar with approximately 25 years each. The outer categories have the same amount of participants with 303 respective 325, whereas the inner category is over proportioned by 438.

Table 3: Commentaries

Content of commentaries	Percentage by mentioning	Percentage by the whole sample
Formulation of vignettes unrealistic	36.2	6.9
Set of questions too long	34.7	6.1
Comprehension problem	11.7	2.0
Assignment to categories just/ unjust	9.2	1.6
Miscellaneous	8.2	1.4
N (Amount of given comments)	191	1,066

In 36 percent of the cases (that is seven percent of the whole respondent sample) respondents declared the descriptions to be unrealistic in some cases. 35 percent (six percent of all respondents) of those who made a comment perceived the vignette part as too long. Only twelve percent (two percent of the whole sample) had problems with the comprehension of the rating procedure. Nine percent of the 191 respondents who made comments had difficulties to assign the income as just or unjust.

Based on the interviewers' assessment we are able take closer look on the respondents' comprehension and willingness to participate in the vignette module.[12] As shown in Table 4 more than 80 percent of the respondents understood the vignette part well (categories: very good and good). In comparison with the vignette module, the whole questionnaire has more than 90 percent in this category. This difference of ten percentage points indicates that the vignette module is more complex than other parts of the interview but it can still be considered similar to other complex modules in the SOEP-Pretest 2008.

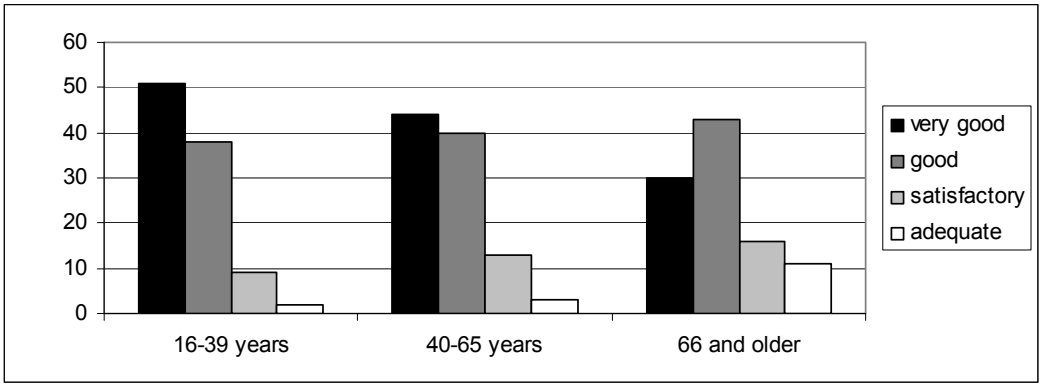
Table 4: Comprehension (in Percent)

Evaluation	Vignettes	Total questionnaire
Very good	41.74	51.13
Good	40.34	40.15
Satisfying	12.95	6.75
Adequate	3.19	1.41
Inadequate	.75	.28
Deficient	1.03	.28
Total	100	100
N	1,066	1,066

[12] The question given to the interviewer was: 'Please state precisely for the last question, or group of questions, in regard to the topic 'Income Justice', how you would evaluate the respondent's performance with respect to comprehension and willingness to reply.' (closed question, categories: very good, good, satisfying, adequate, inadequate, deficient). The assessments of the interviewers are obviously subjective and should be interpreted with caution. Still, these impressions provide some valuable insights on the interview situation itself.

Figure 2 shows that some differences between the age groups occurred. More than 50 percent of the youngest respondents had a very good understanding of the vignettes. In comparison, only 30 percent of the oldest interviewees are in this group. 40 percent of the latter group comprehended the task well (category: good). Only ten percent of respondents over 65 years understood the vignettes worse than satisfying.

Figure 2: Comprehension by Age Groups (in Percent)

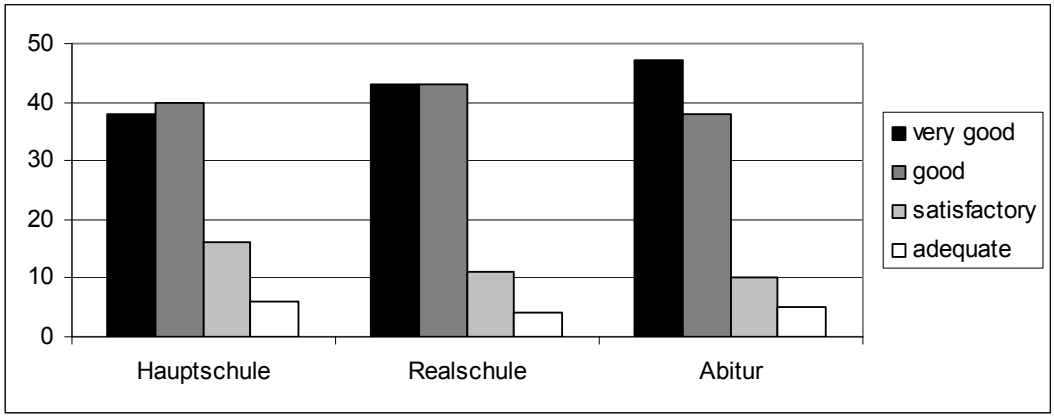


Note: The categories adequate, inadequate and deficient are combined.

As shown in Figure 3, there are only few differences between educational groups in regard to comprehension. In 50 percent of the cases respondents with the highest education level had a very good comprehension of the vignette module. In 40 percent of the cases the comprehension was still good. The middle group performed almost as well with a total of 90 percent who understood the task at least well. From those who have a lower secondary school certificate (Hauptschule) almost 80 percent understood the task well. The differences between educational levels can be considered smaller than the differences between age groups. Notable is the fact that the differences between age and educational groups are similar in the whole questionnaire (not displayed analyses). This means that no vignette specific comprehensive problems occurred.[13]

[13] This is remarkable because effects of age and education have a high chance of leading to misinterpretations in regard to content (undiscovered effects of age and education could be interpreted as subject matters, see Schwarz and Knäuper 2006).

Figure 3: Comprehension by Education (in Percent)



Note: The categories fair, poor and unsatisfactory are combined.

The respondents' willingness to answer – as the interviewer perceived it – is presented in Table 5. In over 80 percent of the cases the willingness to participate in the vignette module was good or very good in comparison to almost 90 percent for the whole questionnaire.

Table 5: Respondent Willingness (in Percent)

Evaluation	Vignettes	Total questionnaire
Very good	44.18	51.69
Good	37.52	37.43
Satisfying	11.35	7.41
Adequate	5.35	2.53
Inadequate	1.22	.66
Deficient	.38	.28
Total	1,066	1,066

As Table 6 shows the willingness to answer differed between age groups. The youngest group performed significantly better than the oldest group. The middle group performed the task almost similar in comparison to the youngest group.

Table 6: Willingness in Vignette Part by Age (in Percent)

Evaluation	Age group (Years)		
	16-39	40-65	66-
Very good	52.48	46.58	33.23
Good	37.29	36.76	38.77
Satisfying	6.60	11.19	16.00
Adequate/ fair	3.30	4.34	8.62
Inadequate/ poor	.00	.68	3.08
Deficient/ unsatisfactory	.33	.46	.31
Total	303	438	325

Table 7 reports only marginal differences between educational groups. In the group of respondents with a lower secondary school certificate (Hauptschule) 78 percent of the interviewers classify their willingness to answer at least good, in comparison to 83 percent and 87 percent in the other groups.

Table 7: Willingness in Vignette Part by Education (in Percent)

Evaluation	Education		
	Secondary School Certificate		
	Low (Hauptschule)	Middle (Realschule)	High (Abitur)
Very good	40.98	45.32	48.98
Good	36.68	41.69	33.47
Satisfying	13.32	9.06	10.61
Adequate/ fair	6.76	3.93	4.49
Inadequate/ poor	1.84	00	1.63
Deficient/ unsatisfactory	.41	.00	.82
Total	488	331	245

In sum, interviewers' impressions do not show dramatic differences between age or educational groups. This can be interpreted as a first hint that vignettes are applicable in public surveys.

4.2 Respondent Behavior

Respondent behavior provides valuable insight on the rating situation and allows for drawing conclusions from the evaluation with regard to the capability of the vignette tool. In the following, a closer look on response time, the use of the rating scale and the consistency of the judgments is taken.

4.2.1 Response Time

The response time is only available for the whole vignette module. The analysis of this kind of process produced data is problematic because of the fact that important context information like interruptions during interviews is often neglected. Nevertheless the gathered data provide useful information – for instance in respect to factual refusal. The CAPI programming excluded the possibility of *refusing* or *drop outs* during the module (compare part 4). A measured response time of 20 seconds for the complete module can be interpreted as a factual refusal. Approximately five percent needed less than three and a half minutes to complete this part of the questionnaire which is an average of eight seconds per vignette. Two interviews build the counterpart with 137 and 139 minutes process time (on the average five minutes per vignette). This distortion is an indicator for unmeasured breaks. Besides these outliers the data seems analyzable. The respondents needed an average of thirteen and a half minutes for a completion of the vignettes (24 plus example and vignette with two further dimensions, see Footnote 5). The median is twelve (12.4) minutes. Table 8 informs about important data points in

regard to process time. Mentionable is also that the respondents started the vignette module on average after 25 minutes of questioning.

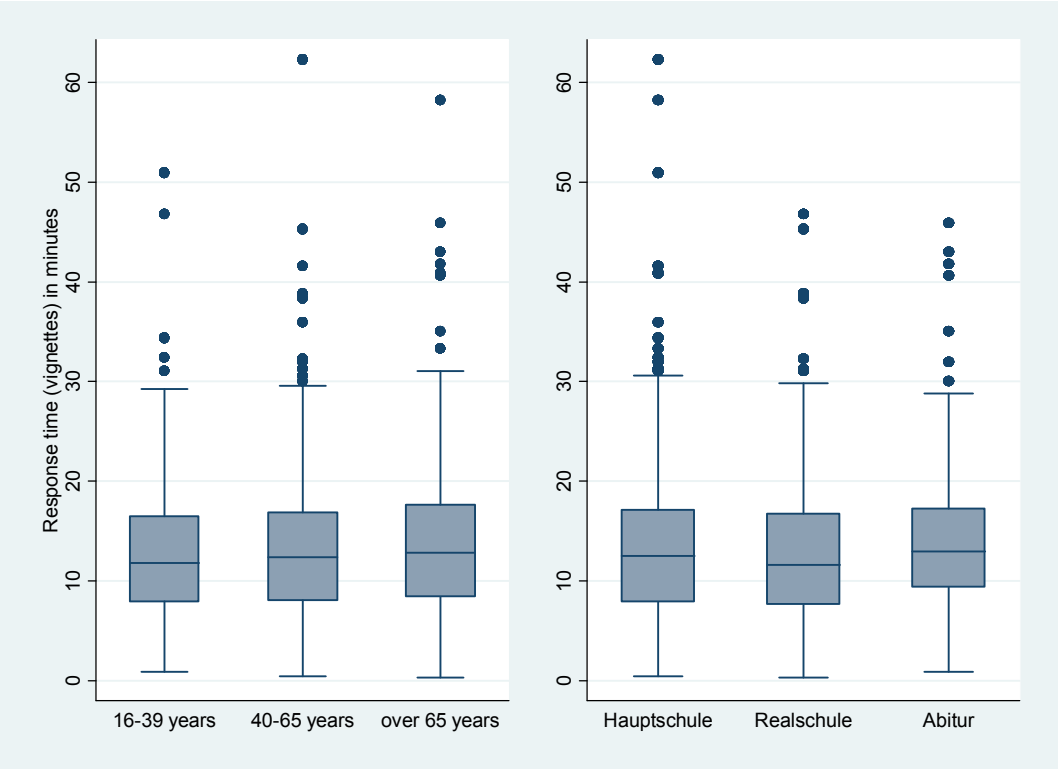
Table 8: Interview Time: Mean and 5-Point Statistic (in Minutes)

	N	Mean	S.D.	Min	.25	.5	.75	Max
Vignette module	1,066	13.52	9.26	.30	8.10	12.38	17.02	138.97
Questionnaire	1,063	50.65	24.69	18.97	36.68	45.27	57.92	341.22

Figure 4 shows the box plots of the process time for respondents' age (left box plots) and education (right box plots). There are no dramatic differences between the groups, respondents with higher education level and older respondents were in need of slightly more time to fulfill the questionnaire (median for older aged being one minute more).

These results indicate that vignettes in a population survey can be evaluated in a tolerable amount of time. The median shows an average of 30 seconds per vignette. The differences between education and age are narrow. All respondents are able to process the vignette module in a similar span of time.

Figure 4: Distribution of Time of the Vignette Part by Education and Age Groups (in Minutes)



Note: Two outliers (137 and 139 minutes) are not displayed in this figure.

4.2.2 Use of the Scale

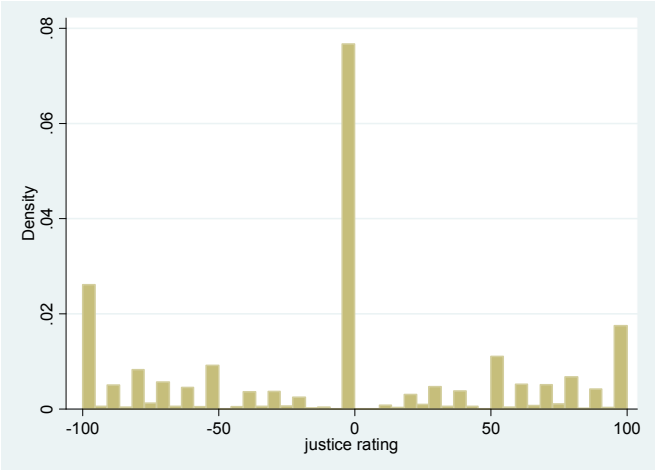
How do respondents use the response scale? The range of the scale reaches from -100 to -1 describing the view that the fictitious earner on the vignette is underrewarded. The zero point of the scale marks a just income and positive values from +1 to +100 reflect a situation where the presented fictitious earner is overrewarded. Table 9 shows the frequencies of the categorized responses distinguishing between underrewarded, just income and overrewarded. About 9,000 vignettes were rated as just, slightly fewer as underrewarded and about 8,000 vignettes as overrewarded. At the first glance the ratings show a dominance of the "just" category.

Table 9: Distribution of the Variable "Justice Evaluation"

Distribution	Judgments (N)
Underrewarded	8.759
Just income	8.897
Overrewarded	7.928
Total	25.584

Figure 5 displays the distribution of the evaluations using not the categorized responses but the scale values. As shown in the graph the category "zero" respectively "just" extremely dominates the other scale values. The agglomerate at the borders of the distribution shows a ceiling effect, especially in the negative number range. In addition, some often mentioned values stand out (-100, -50, 0, 50, 100). The respondents did not fully use the metric scale. For the following analyses this result is taken into consideration by using the categorical variable with three values (see Table 9).

Figure 5: Distribution of the Judgments



To determine in more detail how the respondents used the response scale we concentrate on two aspects. First of all, the clustering of the category "just" is remarkable. It might indicate that respondents wanted to accelerate the procedure by overleaping the second and third part of the evaluation task (see part 1). Second, we

analyze in more detail, how many different values the respondents really use to make their judgments (maximal 24) and what kind of scaling they apply.

From Table 10 it can be seen that the respondents rated on average 8.3 vignettes (out of 24 – more than one third) as just with marginal differences between educational groups.

Table 10: Frequency of Category "just" by Education

R's general educational level	Mean	S.D.	N
Lower secondary school certificate (Hauptschule)	8.26	4.53	488
Middle secondary school certificate (Realschule)	8.29	4.30	331
Higher secondary school certificate (Abitur)	8.66	4.24	245
Total	8.33	4.36	1,064

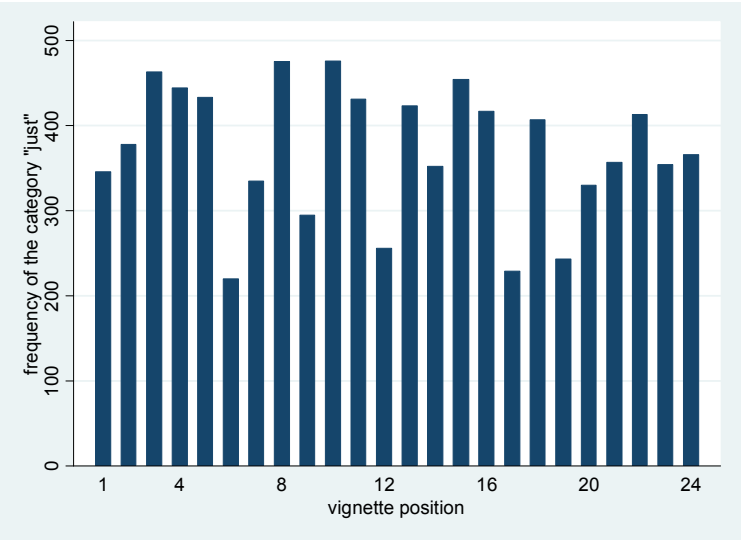
Table 11 shows a significant difference ($p<.01$) between age groups. The group of the 16 to 39 year old respondents uses the "just"-category more than the group of 40 to 65 year old respondents. Respondents of 66 years and older lie between the other two groups with an average of 8.5 vignettes rated as just.

Table 11: Frequency of Category "just" by Age

Age group	Mean	S.D.	N
16-39	8.87[a]	4.27	303
40-65	7.86[a]	4.14	438
66-	8.50	4.69	325
Total	8.35	4.37	1,066

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

Figure 6: Frequency of Category "just" by Vignette Position



One could assume that choosing the "just"-category reflects the wish to speed up evaluation task and goes hand in hand with fatigue effects as respondents have to fulfill only one rating step instead of three. Both would imply a structure where more just ratings can be found at the end of the module. However, the correlation between the position of the vignette within the module and the rating is low (see Figure 6). This means that there are no hints for more "just"-ratings in later positions. The use of the category "just" might not reflect fatigue effects or the desire to speed up evaluation task.

Table 12: Frequency of Different Magnitude Ratings by Education

R's general educational level	Mean	S.D.	N
Lower secondary school certificate (Hauptschule)	8.33[a]	3.22	488
Middle secondary school certificate (Realschule)	8.52	2.98	331
Higher secondary school certificate (Abitur)	8.96[a]	3.06	245
Total	8.53	3.12	1,064
a: group comparing sig. (p < 0,05)			

Do the respondents really use the full 100 point scale to differentiate their judgments? The average use of different values is 8.53, the median is 8. As seen in Table 12, there are differences between educational groups. Respondents with a lower secondary school certificate (Hauptschule) use significantly fewer values ($p<.05$) for the income rating compared to respondents with a higher secondary school certificate (Abitur). This could indicate that people with higher education are using the scale in a more fine-grained fashion. Similar results exist with regard to magnitude scales in methodological studies within Conjoint-Analysis (Steenkamp and Wittink 1994; Teas 1987). The age groups do not differ significantly from each other (Table 13).

Table 13: Frequency of Different Magnitude Ratings by Age

Age group	Mean	S.D.	N
16-39	8.66	3.15	303
40-65	8.69	3.08	438
66-	8.19	3.11	325
Total	8.53	3.11	1,066

This analysis provides only a first hint about how the respondents used the scale. In a next step we focus on the range of values and the distances between them. Table 14 shows which numbers were used. Nearly eight percent of the respondents used numbers with a distance of 25 in each of the 24 vignettes (25, 50, 75, 100). For two third a ten-point scale would have been appropriate as they only used decimal steps. Together with persons who additionally used finer five-point steps (which is covered by a 20-point scale), 90 percent of the respondents are detected. Only ten percent use additional

numbers and thus only this small group would be narrowed in their ratings by a 20-point scale instead of a 100-point scale.

Table 14: Used Scale

Scale (Ratings)	Percent	N
4-point Scale (25, 50, 75, 100)	7.69	82
10-point Scale (10, 20, ..., 90, 100)	65.01	693
20-point Scale(5, 10, 15, …, 95, 100)	90.34	963

4.2.3 Consistency of Judgments

There are two recognized strategies to check for consistency of responses in factorial surveys. Both strategies rely on the results of individual regression models where the rating is the dependent variable and the independent variables resemble the dimensions. The first strategy is to take a look at the model fit during different response sequences (in OLS-models the model fit would be specified by the R^2). The questions are: Is there an adaptation phase in the first judgments and is that sequence therefore not comparable to latter ones? Is there a phase in which the respondents judge most consistently? Is there evidence for fatigue effects at the end of the vignette module? To answer these questions we compare the consistencies in terms of R^2 within different phases of the vignette module.

The second strategy is to investigate how the consistencies depend on respondent attributes such as age or education level.[14] In fact the design of vignettes is more complex in comparison to item batteries (in order to complete the rating it can take up to three steps). Therefore we may question if old or young as well as high or low educated respondents are to the same degree able to deliver consistent responses? For an answer we compare the model fits of each age and education group.

The basis of the following analyses is a multinomial Logit-Model which conservatively considers the reached scale level.[15] We transform the dependent (metric) variable into one with three outcomes, -1 (unjustly underpaid), 0 (just) and 1 (unjustly overpaid). We measure the model fit in the case of categorical regression analysis by the Pseudo-R^2 by McFadden (Long 1997; Long and Freese 2006). The Pseudo-R^2 does not measure the *variance explanation* (unlike the R^2 in OLS-Models), but gives a hint for the goodness

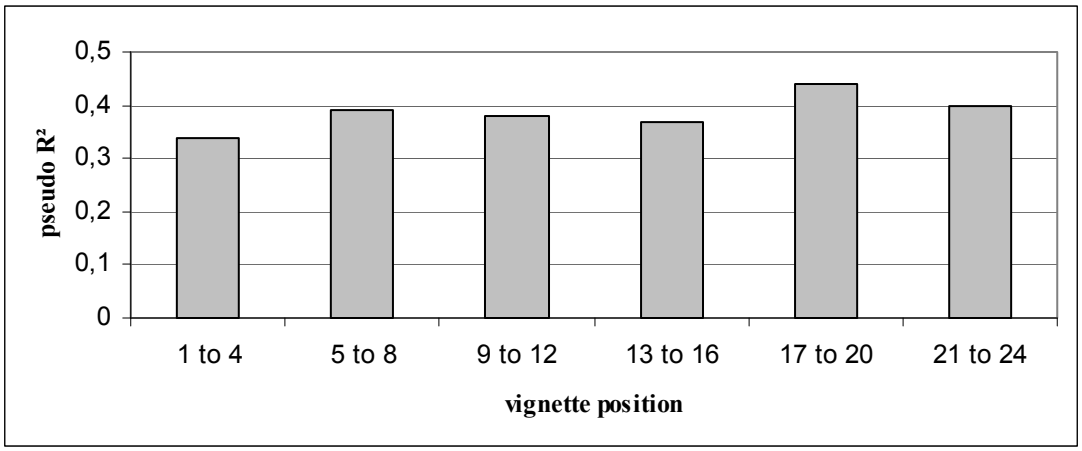
[14] However, most researchers use the consistency measured by the model fit to underline the fact that their dimensions are adequate. But this criterion is not sufficient (Auspurg et al. 2009a) because respondents may produce consistent judgments also when they fade out some dimensions in cases of over-burden or fatigue. Therefore, the effect sizes of different coefficients and their significance (standardized number of cases) is also important.

[15] Alternative censured data can be estimated by Tobit-regression models. In this case the zero leads to a modeling problem which is the reason for choosing a Logit-model.

of fit of the model and at the same time for the consistency of respondent behavior. The ten dimensions are included into the model as independent variables (see Table 1).

Figure 7 lists the Pseudo-R^2 values in six phases of the vignette module (multinomial Logit-Models for each sequence under consideration of the clustered data structure).[16] Every sequence includes all judgments of the respondents that is we have pooled regression results. The most consistent phase is the fifth (vignette 17 to 20) with a Pseudo-R^2 value of .4. In the first sequence the Pseudo-R^2 is far below .4 and also the lowest value of all parts. In the middle part the Pseudo-R^2 is slightly less than .4. There are only marginal differences between the phases of the vignette module indicated by the goodness of fit.

Figure 7: Model Fit in Six Phases of the Vignette Module

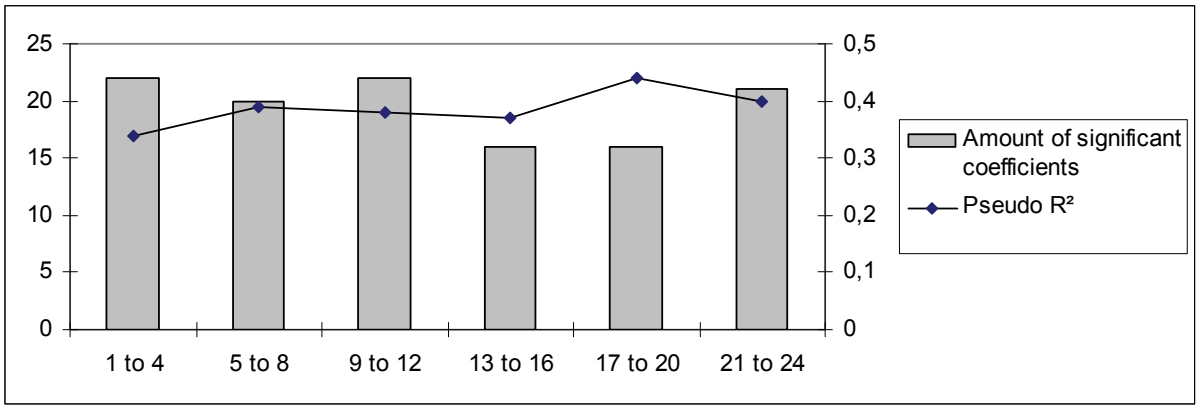


At the first glance these results imply the absence of fatigue effects in the vignette part. But the respondents could also have produced consistent ratings by fading out some dimensions. Therefore we investigated in a further research step the number of significant effects (significance at the .05 level) per sequence. A multinomial Logit-Model, which includes all ten vignette attributes, has 15 independent coefficients (because of the dummy-split for the variables marital status, vocational training, performance, firm size and economic situation of the firm). The dependent variable has three categories. Without the constant, a prediction of 15 x 2 = 30 coefficients is conducted in a multilogit-model. Obviously, the maximum number of significant coefficients is 30. Figure 8 displays the number of significant coefficients (grey bars, left scale) as well as the Pseudo-R^2 from Figure 7 (connected line, right scale) in one graph. The figure shows a result that contrasts strongly with the results from above. In the first, third and sixth phase of the vignette module more than 20 coefficients are

[16] The data are 'clustered' in so far as individual respondents rate a bundle of vignettes. The ratings are therefore not independent from each other, but a data structure likewise to panel data exists. This data structure is considered by estimating robust standard errors (Huber-White-Correcting).

significant. In the second phase 20 coefficients are significant but in the fourth and fifth sequence we find only 16 significant effects. Figure 8 clearly highlights the differences between the consistency measured by the Pseudo-R^2 and the number of significant coefficients. The highest Pseudo-R^2 and at the same time the lowest number of significant effects are found in the fifth sequence. The respondents are seemingly reaching a higher consistency by using a heuristic to simplify their ratings.

Figure 8: Number of Significant Coefficients and Pseudo-R^2 in the Six Phases of the Vignette Module



One could challenge these results with regard to the fact that they are the result of the specific split into six parts. That is why these findings have to be confirmed by analyzing smaller and wider splits. Nevertheless, the results remain stable also in alternative splits.[17]

In the further part of this analysis we take a look at the results of the second research strategy, the analysis of the differences between age and education groups. As we have seen in the previous analyses both measures have to be taken into account, on the one hand the Pseudo-R^2 and on the other hand the number of significant coefficients. We estimate pooled multilogit-models by age and education group and report the model fits in Table 15.[18] The results for the education groups show that respondents with Abitur achieved marginal higher Pseudo-R^2 values than the other two groups (.38 with respect to .37 (Realschule) and .35 (Hauptschule)). The number of significant coefficients

[17] Pseudo R^2-values for a split in two halves are: .35 and .37; the numbers of significant coefficients are: 20 and 18. Additional analyses with three groups also show that the number of significant dimensions decreases during the response process while the R^2-values increase. The different numbers of significant coefficients could also result from differently strong correlations between independent variables (due to different 'efficient' vignette samples, compare Chapter 2 and 3.2). However, the correlation tables for singular phases presented in the appendix show that this interpretation may be excluded: The correlations and variances of the vignette variables differ only marginal between the phases (compare Tables A2).

[18] One has to take into consideration that the number of observations differs in each education group. To avoid interferences, we drew ten random samples of a size of N=245 out of the respondents with the education level Haupt- and Realschule. We proceeded the same way in the case of the age groups. The table reports the respective means of these samples.

varied inversely to education levels. Regression models of respondents with Abitur lead to 16 significant coefficients. For the other groups 17.6 (Realschule) and 17.9 (Hauptschule) can be reported. Common to both measures of model fit is that the differences between the groups are marginal.

For the three age groups we find that respondents of the middle group achieved the highest Pseudo-R^2 (.37) in comparison to the others (.36, youngest group, and .35, oldest group). The differences are, however, marginal. There are no major differences in the number of significant coefficients. The regression for the youngest group shows 22 significant effects, for the middle group 18.4 and the oldest group 14.7. This could be a hint for a fatigue effect in the older group. Further analyses – under control of the six phases – indicate relatively constant respondent behavior regarding significant coefficients and model fit in this age group (not displayed). We find differences in the number of significant effects between age group but we can not conclude that this is a fatigue effect rather than a result of different justice evaluations.

Table 15: Model Fit and Significant Coefficients by Education and Age

Respondent group	Pseudo-R^2	Significant Coefficients	Number of Respondents
R's general educational level			
Lower secondary school certificate (Hauptschule)	.35	17.9	245*
Middle secondary school certificate (Realschule)	.37	17.6	245*
Higher secondary school certificate (Abitur)	.38	16.0	245
Age group			
16 to 39 years	.36	22.0	303
40 to 65 years	.37	18.4	303*
66 years and older	.35	14.7	303*

* Mean from ten samples (compare Footnote 18)

An analysis of the consistency of ratings shows that the response fit cannot be measured only by goodness of fit values such as Pseudo-R^2. Respondents who take fewer dimensions into account may also achieve high Pseudo-R^2 like those who do not use such heuristics. An examination of the significant coefficients points out that the best sequences (respectively broadest judges) are found between the ninth and twelfth vignette.

5 Justice Attitudes

To show the potential of factorial design for analyzing different research questions we will present in the following some content based results on the justice attitudes of the participants of the SOEP-Pretest 2008. These results stem from a multinomial logit-model as introduced in Chapter 4. The analyses are based on the 24 evaluations each of the 1066 respondents made within the vignette module, resulting in 25.584 justice judgments. The Logit- Model is estimated with robust standard errors correcting for the clustered data structure. The dependent variable is the categorical variable of justice evaluations (see Table 9). The independent variables are the vignette dimensions described in Table 1. In order to generate an interpretable table, marginal effects are reported instead of beta coefficients. For example, in Table 18, the likelihood that the income is perceived as too low (underrewarded) increases by 2.35 percent if the described person has a university degree instead of no vocational training. Without vocational training, the likelihood of being overrewarded decreases by 6.14 percent .

We take a look at the different variables step by step. The gender of the vignette person has a significant effect on the justice evaluation. The income of male earners is more often rated as too low or just than the income of females. Because the regression model controls for all covariates this effect goes reduces to the gender of a vignette person. This is an evidence for the just-gender-wage gap known from earlier studies (Jann 2003; Jasso und Webster 1997, 1999).

The age of the vignette person is also a relevant criterion for justice evaluation. Respondents award higher wages to older people which can be seen as an indication for the seniority principle.

The dimension “vocational training” is included as a dummy set into the model. The reference category is a vignette person without vocational training. Vocational training and the university degree are relevant for income evaluations independent from the actual occupation, in other words: more training should lead to a higher income.

The same effect can be observed for the occupational prestige operationalized by the magnitude prestige score (MPS): People in occupations with a higher prestige should earn more (than people with a job of lower prestige).

The gross income of the vignette person is the basis of the just evaluation of the respondent and has a strong effect. The higher the stated income the more likely it is perceived as just or even too high (for 1,000 Euro income more the likelihood for rated as overrewarded increases by 9.45 percent).

Table 18: Determinants of Justice Income (Multinomial Logit-Model), Marginal Effects

	Gross income		
	Too low	Just	Too high
Sex [1 = male]	.0084***	.0193**	-.0277***
	(.0018)	(.0072)	(.0073)
Age [10 years]	.0037***	.0150***	-.0188***
	(.0007)	(.0030)	(.0030)
Vocational training[1]	.0155***	.0217*	-.0373***
	(.0024)	(.0094)	(.0095)
University degree[1]	.0235***	.0379***	-.0614***
	(.0029)	(.0097)	(.0096)
Prestige [10 MPS-Scores]	.0057***	.0183***	-.0240***
	(.0006)	(.0013)	(.0012)
Gross income [1,000 Euro]	-.0554***	-.0391***	.0945***
	(.0047)	(.0047)	(.0025)
Performance: average[2]	.0220***	.0538***	-.0758***
	(.0032)	(.0089)	(.0086)
Above average[2]	.0329***	.104***	-.137***
	(.0040)	(.0094)	(.0090)
Comp. economically stable[3]	-.0047*	.0307***	-.0260**
	(.0019)	(.0091)	(.0090)
Threatened by bancruptcy[3]	-.0030	-.0301***	.0332***
	(.0018)	(.0090)	(.0090)
Medium company[4]	.0012	.0193*	-.0205*
	(.0019)	(.0083)	(.0084)
Large company[4]	.0054*	.0237*	-.0292**
	(.0021)	(.0097)	(.0098)
Double earner[5]	-.0131***	-.0239**	.0370***
	(.0024)	(.0088)	(.0089)
Single[5]	-.00684***	-.0244**	.0312***
	(.0019)	(.0087)	(.0090)
Number of children	.0014**	.0089***	-.0104***
	(.0005)	(.0025)	(.0026)
Pseudo-R^2		.360	
ll_0		-28074.3	
Ll		-17976.9	
N respondents		1066	
N vignettes		25584	

Beta coefficients; Standard errors in parentheses

* $p < .05$, ** $p < .01$, *** $p < .001$

[1] Ref.: without vocational training

[2] Ref.: below average

[3] Ref.: high profit

[4] Ref.: small company

[5] Ref.: single earner, married

Performance on the job (in the model as a dummy set with the reference category "below average") also has the expected effect: Higher performance should lead to higher payment. A person who performs below average instead of above average has a higher probability (13.7 percent) to be rated as overrewarded.

The economic situation of the firm has an impact too. In contrast to the reference category, high profit employees in companies are more often in the "just"-category. The vignette persons who work in companies which are threatened by bankruptcy are more likely to be in the overrewarded group. The company size does have a similar but smaller effect: In medium and large companies the likelihood to be in the overrewarded category is lower in comparison to small companies.

The last two dimensions are related to the family situation. Married and sole earners are the reference group. Singles and double earners are more often in the overrewarded category. From the respondents' viewpoint, single earners with a partner should earn more. The number of children is also a relevant predictor for the justice evaluation: The more children the vignette person has the more this person should earn.

But factorial surveys do not only provide the opportunity to analyze the reactions of respondents to the variation of certain dimension of the vignettes, they also allow studying differences in the response behavior between groups of respondents and the effects of the characteristics of the respondents on their evaluations. To give an example of this kind of analyzes we ask in a first step, if the occupation of a vignette person has a different effect on the justice evaluation depending on the age or the education of the respondent. Our representative sample is predestined for this kind of content based questions and has more potential for analyses than homogeneous student samples or other specific populations, which until now are the standard in factorial surveys.

We categorize our respondents in the same age and educational groups like in the preceding analyses. For each age or educational group we calculate the means of justice evaluations regarding the ten occupations of the vignette persons (Figure 10).[19] Interestingly, the curves of the three age groups and the three educational groups do not differ in their shape. There is a consensus – in the different age groups as well as in the education groups – what people in the ten occupations should earn and, more important, that occupations with a higher prestige should also earn more than occupations with a low prestige. It is noteworthy that social work professionals and locomotive engine

[19] Occupations with higher prestige scores were judged more often as overrewarded. The first impression is that it is a contradiction to the regression. But the reason is evident: The occupations correlate with income that is controlled in the regression.

drivers lightly fall out of this order: These two occupational groups should earn a bit more than the prestige scores pretend.

Figure 10: Justice Evaluation by Occupation of the Vignette Persons and Respondent Groups

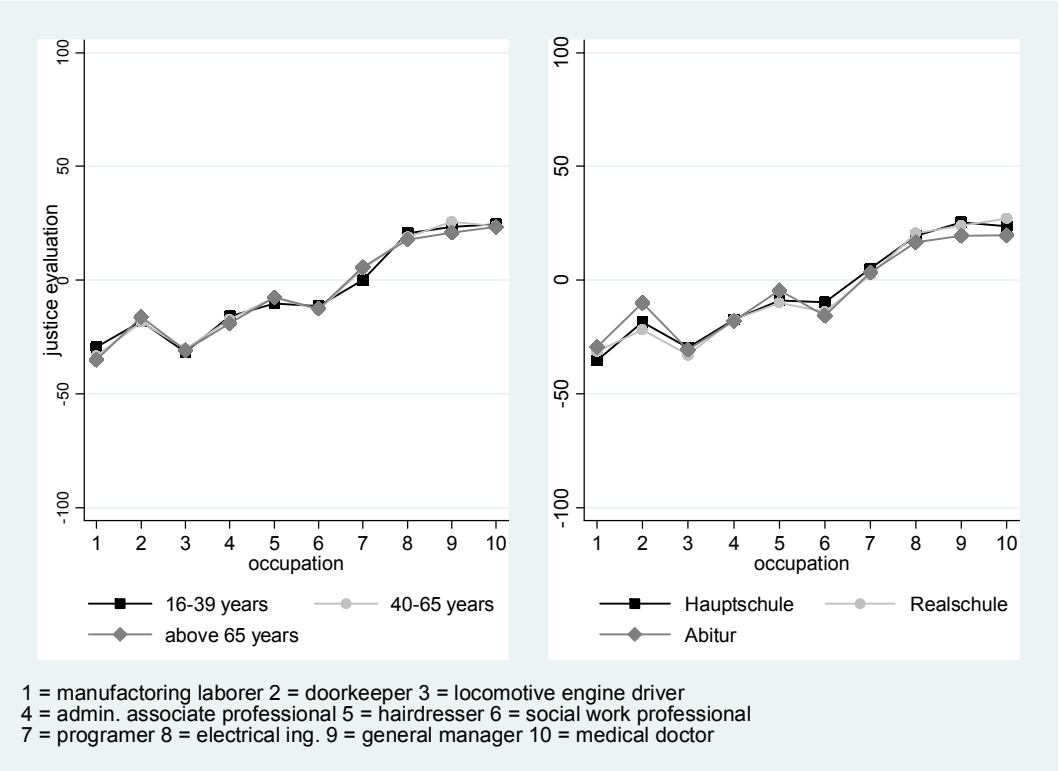
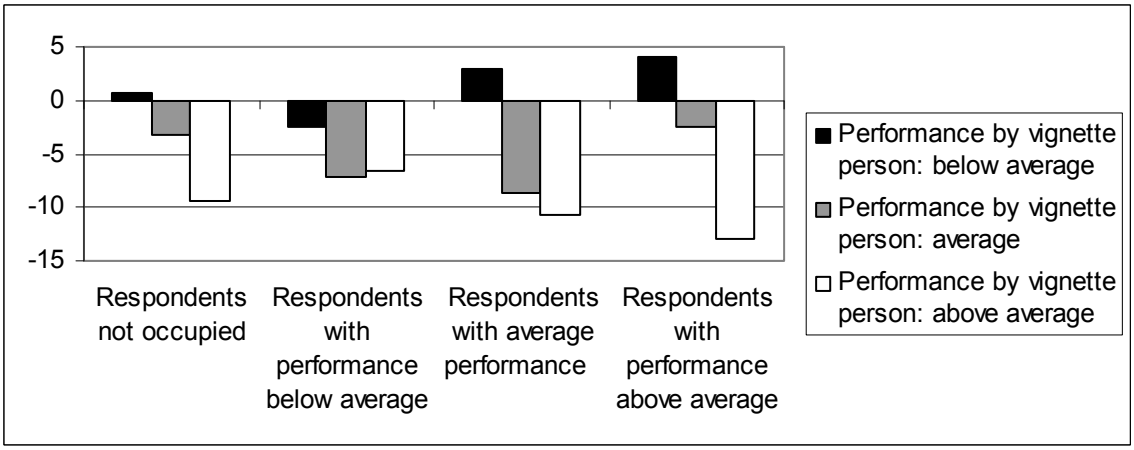


Figure 11: Justice Evaluation by Performance (Mean Differences)



Note: self assessment of the respondents

In order to indicate the further reaching potential of factorial surveys we present two more respondent specific analyses. In the regression model, the performance has a

significant impact on the justice evaluation. To find out if this is a consensual judgment, three groups beside the unemployed are considered. All employees in the SOEP-Pretest had to rate their own performance on the job – analog to the three categories of the vignette dimension "performance on the job". Figure 11 displays that these grouped respondents differ in their income evaluation. Respondents with a low performance on the job do not consider vignette performance in the same way than respondents who declare they perform well.[20]

Another dimension with significant effect in the model is the gender of the vignette person: Men and women assign men a higher just income. The difference of the effect size is similar as well.

Figure 12: Justice Evaluation by Occupation of the Vignette Person

Thus some possibilities for analyses are indicated. For multivariate models some other procedures are applicable which estimate the level of measurement of the dependent variable less conservatively. Furthermore more analyses (based on content problems) about subgroup specific judgment rules are conceivable, especially for interactions of vignette and respondent attributes.

6 Conclusions

This research note describes the factorial survey and its implementation in the SOEP-Pretest 2008. The main research objective of this study was to investigate the capability of factorial surveys in large population surveys. Therefore, we created a vignette module that was part of the CAPI-questionnaire with 25 descriptions of fulltime employees. Respondents expressed their ratings using a three step procedure. Afterwards, interviewers and respondents gave feedbacks about difficulty and comprehensiveness of the rating task. We analyzed these evaluations and the response

[20] Negative values mean vignette persons gets not enough and positive he/she gets too much.

behavior within the vignette module in order to get detailed insights about the usability of this method.

To sum up the most important methodological results:

(1) The factorial survey is a useful instrument for attitude measurement if researchers follow some ancillary conditions, such as the creation of realistic vignettes. Respondents of all age and education groups are capable of rating the vignettes. The higher complexity – regarding to general surveys – seems to be manageable by a vast majority of the respondents.
(2) The response time is about 30 seconds per vignette with only marginal differences between education and age groups. Vignette modules with a moderate number of vignettes and dimensions are capable of being integrated in large population surveys like the SOEP.
(3) The analysis of the use of the 100-point response scale applied in the study shows that respondents only use a few values – with an accumulation at integer values (50, 100). Therefore there is no need for applying large response scales exceeding the common range used in other SOEP item batteries (e.g. life satisfaction). With a rating scale that is common to SOEP respondents most ratings can be covered.
(4) The analysis of the consistency of response behavior shows that the average respondents can deal with the rating tasks of factorial surveys.

The second objective of this investigation was to learn more about respondents' attitudes towards income justice. The results exemplify that besides the occupation, the vocational training and performance – thus factors in direct reference to employment – familiar aspects like marital status, the occupational status of the partner and the number of children are relevant criteria for justice evaluations, too. The factorial survey features a various analysis potential, both, in respect of methodological research problems and also in regard to substantial research questions. The positive experience of the SOEP Pretest 2008 encourages the use of vignettes in the main survey.

7 Literature

Abraham, Martin, and Thomas Hinz (Eds.). 2005a. *Arbeitsmarktsoziologie. Probleme, Theorien, empirische Befunde*. Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH.

—. 2005b. "Theorien des Arbeitsmarktes: Ein Überblick." Pp. 17-68 in *Arbeitsmarktsoziologie. Probleme, Theorien, empirische Befunde*, edited by Martin Abraham and Thomas Hinz. Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH.

Addelman, Sidney. 1962a. "Symmetrical and asymmetrical fractional factorial plans." *Technometrics* 4:47-57.

—. 1962b. "Orthogonal Main-Effect Plans for asymmetrical factorial experiments." *Technometrics* 4:21-46.

Alexander, Cheryl S., and Henry Jay Becker. 1978. "The Use of Vignettes in Survey Research." *Public Opinion Quarterly* 42: 93-104.

Alves, Wayne M. 1982. "Modeling distributive justice judgments." in *Measuring social judgments. The factorial survey approach*, edited by Peter H. Rossi and Steven L. Nock. Beverly Hills: Sage.

Alves, Wayne M., and Peter Rossi. 1978. "Who Should Get What? Fairness Judgments of the Distribution of Earnings." *American Journal of Sociology* 84:541-564.

Auspurg, Katrin, and Martin Abraham. 2007. "Die Umzugsentscheidung von Paaren als Verhandlungsproblem." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59:271-293.

Auspurg, Katrin, Thomas Hinz, and Stefan Liebig. 2009a. "Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey." *Methoden - Daten - Analysen* 3 (1):59-96.

Auspurg, Katrin, Martin Abraham, and Thomas Hinz. 2009b. "Die Methodik des faktoriellen Surveys in einer Paarbefragung." Pp. 179-210 in *Klein aber fein! Quantitative empirische Sozialforschung mit kleinen Fallzahlen*, edited by Peter Kriwy and Christiane Gross. Wiesbaden: VS Verlag für Sozialwissenschaften.

Auspurg, Katrin, Thomas Hinz, Stefan Liebig, and Carsten Sauer. 2008. "Wer verdient welches Einkommen? Ergebnisse eines faktoriellen Surveys zur Einkommensgerechtigkeit in Deutschland." Mimeo. University of Konstanz.

Barrera, Davide, and Vincent Buskens. 2007. "Imitation and Learning under Uncertainty: A Vignette Experiment." *International Sociology* 22:367-396.

Beck, Michael, and Karl-Dieter Opp. 2001. "Der faktorielle Survey und die Messung von Normen." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53:283-306.

Berk, R. A., and P. H. Rossi. 1977. *Prison reform and state elites*. Cambridge, Mass.: Ballinger.

Dülmer, Hermann. 2001. "Bildung und der Einfluss von Argumenten auf das moralische Urteil. Eine empirische Analyse zur moralischen Entwicklungstheorie Kohlbergs." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53:1-27.

—. 2007. "Experimental Plans in Factorial Surveys: Random or Quota Design?" *Sociological Methods Research* 35:382-409.

Giesecke, Johannes, and Roland Verwiebe. 2008. "Die Lohnentwicklung in Deutschland zwischen 1998 und 2005 - Wachsende Ungleichheit." *WSI-Mitteilungen* 61(2):85-90.

Grabka, Markus M., and Joachim R. Frick. 2008. "Schrumpfende Mittelschicht in Deutschland - Anzeichen einer dauerhaften Polarisierung der verfügbaren Einkommen." *Wochenbericht des DIW Berlin* 75(10):101-108.

Hermkens, Piet L. J., and Frank A. Boerman. 1989. "Consensus With Respect to the Fairness of Incomes: Differences Between Social Groups." *Social Justice Research* 3:201-215.

Jann, Ben. 2003. "Lohngerechtigkeit und Geschlechterdiskriminierung: Experimentelle Evidenz." Mimeo. Swiss Federal Institute of Technology Zurich (ETH).

Jasso, Guillermina. 1978. "On the justice of earnings: a new specification of the Justice Evaluation Function." *American Journal of Sociology* 83: 1398-1419.

—. 1994. "Assessing Individual and Group Differences in the Sense of Justice: Framework and Application to Gender Differences in the Justice of Earnings." *Social Science Research* 23:368-406.

—. 2006. "Factorial Survey Methods for Studying Beliefs and Judgments." *Sociological Methods Research* 34:334-423.

—. 2007. "Studying Justice: Measurement, Estimation, and Analysis of the actual reward and the just reward. IZA DP 2592. Bonn.

Jasso, Guillermina, and Murray Jr. Webster. 1997. "Double Standards in Just Earnings for Male and Female Workers." *Social Psychology Quarterly* 60:66-78.

—. 1999. "Assessing the Gender Gap in Just Earnings and Its Underlying Mechanisms." *Social Psychology Quarterly* 62:367-380.

Jasso, Guillermina, and Eva M. Meyersson Milgrom. 2008. "Distributive Justice and CEO Compensation." *Acta Sociologica* 51:123-143.

Jasso, Guillermina, and Karl-Dieter Opp. 1997. "Probing the Character of Norms: A Factorial Survey Analysis of the Norms of Political Action." *American Sociological Review* 62:947.

Jasso, Guillermina, and Peter Rossi. 1977. "Distributive justice and earned income." *American Sociological Review* 42:639-651.

Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2008. "Are Americans Really Less Happy With Their Incomes?" RANDWorking Paper WR-591. Santa Monica (CA).

Kuhfeld, Warren F. 2005. "Experimental Design, Efficiency, Coding, and Choice Designs." Pp. 47-97 in *Marketing Research Methods in SAS: Experimental Design, Choice, Conjoint, and Graphical Techniques*, edited by W. F. Kuhfeld.

Kuhfeld, Warren F., Randell D. Tobias, and Mark Garrett. 1994. "Efficient Experimental Design with Marketing Research Applications." *Journal of Marketing Research* 31:545-557.

Liebig, Stefan, and Steffen Mau. 2002. "Einstellungen zur sozialen Mindestsicherung. Ein Vorschlag zur differenzierten Erfassung normativer Urteile." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54:109-134.

—. 2005. "Wann ist ein Steuersystem gerecht?" *Zeitschrift für Soziologie.* 34:468-491.

Liebig, Stefan, and Jürgen Schupp. 2005. "Empfinden die Erwerbstätigen in Deutschland ihre Einkommen als gerecht?" *Wochenbericht des DIW Berlin* 72(48):721-725.

—. 2008a. "Leistungs- oder Bedarfsgerechtigkeit? Über einen normativen Zielkonflikt des Wohlfahrtsstaats und seiner Bedeutung für die Bewertung des eigenen Erwerbseinkommens." *Soziale Welt* 59:7-30.

—. 2008b. "Immer mehr Erwerbstätige empfinden ihr Einkommen als ungerecht." *Wochenbericht des DIW Berlin* 75(31):434-440.

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.

Long, Scott J., and Jeremy Freese. 2006: *Regression Models for Categorical Dependent Variables Using Stata*. College Sation: Stata Press.

Mäs, Michael, Kurt Mühler, and Karl-Dieter Opp. 2005. "Wann ist man Deutsch? Empirische Ergebnisse eines faktoriellen Surveys." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 57:112-134.

Miller, J. L., P. H. Rossi, and J.E. Simpson. 1986. "Perceptions of Justice: Race and Gender Differences in Judgment of Appropriate Prison Sentences." *Law & Society Review* 20:313-334.

OECD. 2008. "Growing Unequal? Income Distribution and Poverty in OECD Countries."

Prüfer, Peter, Lisa Vazansky, and Darius Wystup. 2003. "Antwortskalen im ALLBUS und ISSP. Eine Sammlung." ZUMA-Methodenbericht 2003/11.

Schulte, Aileen. 2002. "Consensus versus Disagreement in Disease-Related Stigma: A Comparison of Reactions to AIDS and Cancer Patients." *Sociological Perspectives* 45:81.

Schwarz, Norbert, and Bärbel Knäuper. 2006: Kognitionspsychologie und Umfrageforschung: Altersabhängige Kontexteffekte. Pp. 203-216 in: *Methoden der Sozialforschung*. Sonderheft 44 der Kölner Zeitschrift für Soziologie und Sozialpsychologie. edited by Andreas Diekmann. Wiesbaden: VS Verlag für Sozialwissenschaften.

Schwarze, Johannes. 2007. "Gerechte Löhne? Eine empirische Analyse subjektiver Erwerbseinkommen." Pp. 80-107 in *Arbeitsmarkt- und Sozialpolitikforschung im Wandel - Festschrift für Christof Helberger zum 65. Geburtstag*, edited by Johannes Schwarze, Jutta Räbiger, and Reinhold Thiede. Hamburg: Dr. Kovac.

Siegel, Nico A., Andreas Stocker und Sebastian Warnholz. 2009. SOEP Testerhebung 2008: Persönlichkeit, Gerechtigkeitsempfinden und Alltagsstimmung. Methodenbericht. Munich: TNS Infratest Sozialforschung.

Smith, Tom W. 1986. "A Study of Non-Response and Negative Values on the Factorial Vignettes on Welfare." GSS Methodological Report No. 44. Chicago. NORC.

Steenkamp, Jan-Benedict, und Dick R. Wittink. 1994. "The Metric Quality of Full-Profile Judgments and the Number-of-Attribute-Levels-Effect in Conjoint Analysis." *International Journal of Research in Marketing* 11:275-286.

Steiner, Peter M., and Christiane Atzmüller. 2006. "Experimentelle Vignettendesigns in Faktoriellen Surveys." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58:117-146.

Struck, Olaf, Gesine Stephan, Christoph Köhler, Alexandra Klause, Christian Pfeifer, and Tatjana Sohr (Eds.). 2006. *Arbeit und Gerechtigkeit Entlassungen und Lohnkürzungen im Urteil der Bevölkerung*. Wiesbaden: VS Verlag für Sozialwissenschaften

Teas, R.K. 1987. "Magnitude scaling of the dependent variable in decompositional multiattributive preference models." *Journal of the Academy of Marketing Science* 15:64-73.

Thomas, Neal, Trivellore E. Raghunathan, Nathaniel Schenker, Myron J. Katzoff, and Clifford L. Johnson. 2006. "An Evaluation of Matrix Sampling Methods Using Data from the National Health and Nutrition Examination Survey." *Survey Methodology* 32(2):217-231.

Thurman, Quint C., Julie A. Lam, and Peter H. Rossi. 1988. "Sorting out the Cuckoo's Nest: A Factorial Survey Approach to the Study of Popular Conceptions of Mental Illness." *The Sociological Quarterly* 29:565.

Winkler, Niels, Martin Kroh and Martin Spiess. 2006. "Entwicklung einer deutschen Kurzskala zur zweidimensionalen Messung von sozialer Erwünschtheit." DIW Discussion Papers 579. Berlin. DIW Berlin (German Institute for Economic Research).

Appendix

Table A1: Realized Sample in the SOEP-Pretest 2008

Sample Size	N	1066
Gender	Male	47.3%
	Female	52.7%
Age (in years)	Mean	51.7
	Median	53
Education	Lower secondary school certificate (Hauptschule)	43%
	Middle secondary school certificate (Realschule)	31%
	Higher secondary school certificate (Abitur)	25%
Employment status	Not employed	56%
	Full-time	29%
	Part-time	10%

Table A2: Correlations of the Vignette Dimensions

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Sex	1.000									
(2) Age	.008	1.000								
(3) Vocational training	.005	-.030	1.000							
(4) MPS	.028	.020	.251	1.000						
(5) Gross income.	-.007	.014	.128	.586	1.000					
(6) Performance	.001	.003	.008	-.008	-.009	1.000				
(7) Econ. Sit. Company	.018	.014	.020	.035	.044	.008	1.000			
(8) Company Size	.017	-.010	-.010	-.005	-.022	-.043	.003	1.000		
(9) Marital status	-.000	.021	-.037	-.036	-.021	.024	.004	.012	1.000	
(10) children	.011	.008	-.011	.018	-.005	-.056	-.021	.017	.000	1.000

Table A3a: Correlations of the Vignette Dimensions for the Vignettes 1 to 4

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Sex	1.000									
(2) Age	.339	1.000								
(3) Vocational training	-.198	.136	1.000							
(4) MPS	.084	-.139	.317	1.000						
(5) Gross income.	.045	-.211	.054	.651	1.000					
(6) Performance	.174	-.018	-.249	-.153	-.095	1.000				
(7) Econ. sit. Company	.091	-.003	-.140	-.148	-.043	.223	1.000			
(8) Company Size	-.170	.103	-.238	-.392	-.379	.207	-.034	1.000		
(9) Marital status	-.091	.160	-.147	.025	.100	.074	-.061	.106	1.000	
(10) children	-.009	.055	.145	.223	-.039	-.322	-.066	-.031	.025	1.000

Table A3b: Correlations of the Vignette Dimensions for the Vignettes 5 to 8

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Sex	1.000									
(2) Age	-.010	1.000								
(3) Vocational training	-.149	-.059	1.000							
(4) MPS	-.042	.223	.335	1.000						
(5) Gross income.	-.122	.341	.020	.509	1.000					
(6) Performance	-.033	.002	.086	.291	.391	1.000				
(7) Econ. Sit. Company	.083	.064	-.123	.124	.052	.151	1.000			
(8) Company Size	.007	-.036	.161	-.113	-.286	-.269	-.111	1.000		
(9) Marital status	-.155	.242	-.084	-.016	-.193	-.084	.238	.055	1.000	
(10) children	.091	-.112	.070	.073	-.108	-.084	.015	.287	-.007	1.000

Table A3c: Correlations of the Vignette Dimensions for the Vignettes 9 to 12

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Sex	1.000									
(2) Age	-.185	1.000								
(3) Vocational training	.382	-.114	1.000							
(4) MPS	.217	-.174	.375	1.000						
(5) Gross income.	.195	-.102	.003	.642	1.000					
(6) Performance	-.109	.187	.073	-.194	-.257	1.000				
(7) Econ. Sit. Company	-.205	-.138	.060	.057	.070	.065	1.000			
(8) Company Size	.355	-.206	.033	.295	.324	-.106	-.040	1.000		
(9) Marital status	.156	.045	-.038	.188	.027	.084	-.163	.208	1.000	
(10) children	.049	.100	-.127	-.033	.113	.140	.107	-.085	.089	1.000

Table A3d: Correlations of the Vignette Dimensions for the Vignettes 13 to 16

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Sex	1.000									
(2) Age	-.185	1.000								
(3) Vocational training	.382	-.114	1.000							
(4) MPS	.217	-.174	.375	1.000						
(5) Gross income.	.195	-.102	.003	.642	1.000					
(6) Performance	-.109	.187	.073	-.194	-.257	1.000				
(7) Econ. Sit. Company	-.205	-.138	.060	.057	.070	.065	1.000			
(8) Company Size	.355	-.206	.033	.295	.324	-.106	-.040	1.000		
(9) Marital status	.156	.045	-.038	.188	.027	.084	-.163	.208	1.000	
(10) children	.049	.100	-.127	-.033	.113	.140	.107	-.085	.089	1.000

Table A3e Correlations of the Vignette Dimensions for the Vignettes 17 to 20

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Sex	1.000									
(2) Age	-.489	1.000								
(3) Vocational training	.184	-.055	1.000							
(4) MPS	.076	.036	.217	1.000						
(5) Gross income.	.057	-.002	.142	.617	1.000					
(6) Performance	.031	.056	.061	.106	.134	1.000				
(7) Econ. Sit. Company	-.022	-.001	.233	.068	.120	-.170	1.000			
(8) Company Size	-.019	.026	-.300	-.046	-.128	-.113	.116	1.000		
(9) Marital status	.006	-.168	-.153	-.301	-.097	-.201	-.020	.015	1.000	
(10) children	-.189	.047	.045	.031	.035	.113	-.038	.069	-.027	1.000

Table A3f: Correlations of the Vignette Dimensions for the Vignettes 21 to 24

Dimension	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Sex	1.000									
(2) Age	.182	1.000								
(3) Vocational training	-.157	-.070	1.000							
(4) MPS	-.191	-.033	.069	1.000						
(5) Gross income.	-.230	-.087	.497	.558	1.000					
(6) Performance	-.052	-.096	.127	-.209	.040	1.000				
(7) Econ. Sit. Company	.070	.059	-.128	-.034	-.058	-.318	1.000			
(8) Company Size	.029	.137	.172	.295	.414	.132	.116	1.000		
(9) Marital status	-.045	.011	.205	-.230	.137	-.030	-.063	-.203	1.000	
(10) children	.167	.000	-.121	-.066	-.123	.115	-.206	-.140	.069	1.000

Table A3g: Variances of Vignette Dimensions for the Singular Vignette Sections

	Varianz der Ausprägungen der Dimensionen									
Vignette Section	Sex	Age	Vocational training	MPS	Gross income	Performance	Economical situation	Company size	Marital status	Children
1 bis 4	.2441	123.0	.6561	2215	2.41e+07	.6406	.5464	.6982	.7974	2.057
5 bis 8	.2491	145.4	.6661	2346	1.81e+07	.5217	.6912	.6753	.6546	2.045
9 bis 12	.2444	123.9	.7256	2158	2.39e+07	.6773	.8281	.7166	.6848	2.495
13 bis 16	.2452	12.6	.6081	1764	1.38e+07	.6745	.5847	.4947	.6463	1.956
17 bis 20	.2345	122.0	.6647	1536	1.99e+07	.6397	.6786	.7284	.7068	2.281
21 bis 24	.2483	104.1	.6912	1659	1.71e+07	.7466	.5844	.6735	.5312	2.088