

Testing the Validity of Gender Ideology Items by Implementing Probing Questions in Web Surveys

Behr, Dorothee; Braun, Michael; Kaczmirek, Lars; Bandilla, Wolfgang

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Behr, D., Braun, M., Kaczmirek, L., & Bandilla, W. (2013). Testing the Validity of Gender Ideology Items by Implementing Probing Questions in Web Surveys. *Field Methods*, 25(2), 124-141. <https://doi.org/10.1177/1525822X12462525>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by-nc/3.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more information see:
<https://creativecommons.org/licenses/by-nc/3.0>

Testing the Validity of Gender Ideology Items by Implementing Probing Questions in Web Surveys

Dorothee Behr¹, Michael Braun¹,
Lars Kaczmirek¹, and Wolfgang Bandilla¹

Field Methods

25(2) 124-141

© The Author(s) 2012

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/1525822X12462525

fm.sagepub.com



Abstract

This article examines the use of probing techniques in web surveys to identify validity problems of items. Conventional cognitive interviewing is usually based on small sample sizes and thus precludes quantifying the findings in a meaningful way or testing small or special subpopulations characterized by their response behavior. This article investigates probing in web surveys as a supplementary way to look at item validity. Data come from a web survey in which respondents were asked to give reasons for selecting a response category for a closed question. The web study was conducted in Germany, with respondents drawn from online panels ($n = 1,023$). The usefulness of the proposed approach is shown by revealing validity problems with a gender ideology item.

¹GESIS–Leibniz Institute for the Social Sciences, Mannheim, Germany

Corresponding Author:

Dorothee Behr, GESIS–Leibniz Institute for the Social Sciences, P. O. Box 122155, Mannheim 68072, Germany.

Email: dorothee.behr@gesis.org

Keywords

probing, web survey, gender ideology, validity

Introduction

In this article, we explore a new method to implement cognitive interviewing techniques, namely probing in web surveys with respondents drawn from online panels, to assess item validity. We focus on testing the applicability of the method by addressing hypotheses on the functioning of gender ideology items. Although we concentrate on the validity assessment of existing items, the method could equally be implemented at the pretesting stage.

Theoretical Background

Establishing validity of indicators is a necessary prerequisite of any substantive analysis. Otherwise, methodological artifacts might be interpreted as substantive results. To solve these problems, several data analytic approaches have been proposed for assessing measurement quality, such as correspondence analysis, confirmatory factor analysis, and multitrait-multimethod studies (Blasius and Thiessen 2006; Saris and Gallhofer 2007; Vandenberg and Lance 2000). Although the application of data analytic procedures is often an appropriate means for detecting problems in items and item batteries, they lack the power to explain the causes of these problems. Knowledge of these causes, however, could be used to improve questions for future use and to support substantive data analyses with existing data. With the dramatic increase of secondary analyses in the last decades, background information on interpretation patterns or interpretation differences for existing data is especially needed in the social sciences.

A possible solution for detecting methodological artifacts and their causes is to use cognitive interviewing techniques. These techniques are used to reveal cognitive processes in survey responding as well as unintended item interpretation. There are two major cognitive interviewing techniques used in survey research, namely the *think-aloud* technique, in which respondents verbalize their thoughts while they answer survey questions, and the *probing* technique, in which interviewers ask follow-up questions to obtain more specific information relevant to a specific item (Beatty and Willis 2007).

Among the various probing types, “category-selection probing” (Prüfer and Rexroth 2005) is a particularly appropriate means to assess validity.

In category-selection probing, respondents are asked why they selected a certain answer category for a closed question. Category-selection probing can be used to analyze different interpretation patterns among respondents. In particular, “silent misinterpretations” (DeMaio and Rothgeb 1996) can be detected, that is, when respondents seemingly do not have problems with the interpretation of an item but actually misinterpret its meaning in an unintended way. On the negative side, it has to be acknowledged that respondents may have problems answering “why” questions appropriately if, for example, the basis for their attitudes is not accessible to them (Willis 2005; Wilson et al. 1996).

Cognitive interviewing techniques are typically used in cognitive interviews that are part of the pretesting process prior to an actual survey but they can also be applied within or after a survey. In the following, we describe the conventional implementation of cognitive interviewing techniques. Based on this, we propose a supplemental approach to implement cognitive interviewing techniques.

The Conventional Implementation of Cognitive Interviewing Techniques

Cognitive interviewing techniques are mainly used in cognitive interviews (see reports at National Center for Health Statistics 2011). Despite the uncontested value of cognitive interviews, there are some limitations regarding their implementation.

First, cognitive interviews are mainly used to detect bad items and improve a questionnaire. That is, they are mainly used as a pretesting device and not as part of a post-survey assessment. Second, cognitive interviews are often conducted in a lab. This leads to questioning whether results in the lab transfer to the field (Willis and Schechter 1997). Third, cognitive interviewing is conducted by an interviewer. However, the more interviewers are supposed to play an active role in the sense of proactively investigating hidden comprehension problems, the lower the comparability of the results obtained by different interviewers might be (Conrad and Blair 2004, 2009). Fourth, cognitive interviewing is traditionally based on small quota samples of 5–15 interviews (Willis 2005), a fact that is challenged, for example, by Blair et al. (2006). Although even few interviews can help detect major problems with items (Beatty and Willis 2007), low case numbers do not allow quantifying the findings in a meaningful way, assessing the prevalence of problems, or unraveling interpretation patterns of special

subpopulations characterized by their response behavior. Small sample size is possibly the major limitation of traditional cognitive interviewing.

To obtain more generalizable results or information on rare cases, respondent debriefing is occasionally used as a supplemental testing method. This includes follow-up probes with all or a sample of respondents after completion of a pilot survey interview (DeMaio and Rothgeb 1996; Hess and Singer 1995; Nichols and Hunter Childs 2009). In a similar vein, random probes have been asked as part of the actual interview to allow assessment of item validity in the actual survey (Schuman 1966; Smith 1989).

Supplemental Implementation of Cognitive Interviewing Techniques: Web Surveys

Methods for analyzing cognitive processes do not have to be restricted to conventional cognitive interviewing or to respondent debriefing and random probes with (pilot) survey respondents. On the contrary, the methods could usefully be extended to probing in web surveys with respondents from online panels.

Web surveys allow us to cost effectively survey a high number of cases. Thus, they pave the way for meaningful quantification of results and for tackling special or rare response combinations. They also guarantee standardization of probing and hence prevent potential interviewer effects. Research on open-ended questions on the web has started only recently, but first results are encouraging. Narrative open-ended questions in web surveys have been found to fare as well as or better than open-ended questions in paper-and-pencil self-administered surveys (Denscombe 2008; Holland and Christian 2009; Smyth et al. 2009). Admittedly, open-ended questions on the web can also cause drop-out or item nonresponse (Galesic 2006). In addition, the answer quality of these open-ended questions is affected by education, age, sex, or respondents' interest in the topic (Denscombe 2008; Holland and Christian 2009; Oudejans and Christian 2010).

With appropriate design and wording, as well as proper use of interactive features, however, the chances of obtaining meaningful answers can be enhanced (Dillman et al. 2008). Behr et al. (2012) have demonstrated this, particularly with regard to category-selection probing. Furthermore, web surveys provide respondents with time to answer, the possibility to elaborate or modify their statements, and anonymity of answers. The latter, of course, hinges on the level of trust that respondents have with surveying agencies or general data protection procedures. If respondents can be motivated to answer probing questions in the first place, probing in web surveys

seems promising overall. Last but not least, web surveys are a suitable means to bring probing into the field context. Category-selection probing especially can fit perfectly into the normal process of responding, under the condition that respondents are not required to reiterate the same justification several times (i.e., the probed items should not be too similar). At the same time, the number of probes should remain restricted to prevent artificiality and reactivity (Oksenberg et al. 1991) and to keep response burden low. Web surveys could run during the development stage of a questionnaire to inform questionnaire design but equally alongside or after regular surveys to assess measurement error with actual survey items.

Nowadays, online panels (i.e., pools of registered persons who have consented to regularly participate in web surveys) offer a convenient way to sample respondents for a web survey from a wider segment of the population. However, since almost all of these panels take a nonprobability approach in recruiting respondents, they should not be used to estimate general population values (Baker et al. 2010). In Germany, where this study was carried out, a probability-based panel is not yet available, which explains the focus on nonprobability panels in this article. If over- or underrepresentation of certain subgroups and related bias is adequately taken into account in the analyses of online panel survey data, academics and other researchers can still profit from using nonprobability online panels, especially with regard to exploratory studies and experiments.

A good online panel with a sound quality assurance system excludes panelists who continually provide questionable data (Baker et al. 2010). Also, it provides information on data protection procedures and laws during the recruitment stage, which leads to a bond of trust between the panel provider and respondents. Despite this, uncertainties remain as to what extent respondents from a panel predominately dedicated to market research—the nonprobability panels usually belong to this segment—are willing to usefully answer social science items and, in particular, probes about these items. The panelists might satiate (Krosnick 1999) by giving less elaborate answers, which eventually may be useless to the researcher, or by not answering at all. While Behr et al. (2012) demonstrate that online panelists are indeed willing to answer category-selection probes on social science items (roughly between 70% and 80% of panelists provided basic or more elaborate substantive answers across three category-selection probes), no assessment has yet been made as to whether the substantive answers given are sufficiently elaborate in order to answer research questions.

In summary, probing in online panel web surveys seems to be a promising approach to assess item validity, especially when quantification of

results or the tackling of specific response combinations is sought. However, uncertainties remain, particularly with regard to answer quality. This article, therefore, focuses on hypotheses on the functioning of gender ideology items and thereby puts the probe answers to the test.

Validity Problems: The Case of Gender Ideology

Gender ideology, that is, attitudes regarding the proper roles of men and women in family and working life, is a regularly investigated topic in social research. Frequently, it is measured with traditionally slanted items, that is, items that focus on traditional perspectives and that posit, for instance, that the primary responsibility of the woman should be the home and that of the man to earn a living. Although these items permit respondents to reject a traditional stance, they do not allow them to explicitly express an egalitarian view. This limited perspective has been criticized by social scientists and respondents alike.

Against this backdrop, Braun (2008) explored the use of egalitarian slanted items (i.e., those depicting a particular nontraditional role model) and investigated the difficulties involved in using these items as opposed to those with a traditional slant. Based on a multimode probing study, including conventional cognitive interviewing, probing in telephone surveys, and probing in web surveys mainly based on family-related discussion lists, he found that less traditional respondents, as measured through a traditional benchmark item, do not exhibit particularly strong agreement with egalitarian items that lay down specific egalitarian stances.

Gender egalitarianism is obviously not simply the reverse of gender traditionalism. Instead, it includes very different stances, such as reaching gender equality or facilitating individual solutions for each couple. These different positions are connected with different responses to egalitarian slanted items such that the answers of nontraditional respondents are spread across the entire range of the respective answer scale. In addition, some traditional respondents have been found to agree with egalitarian items. For example, they simply ignore parts of an egalitarian item and focus their answers instead on what is compatible with their traditional view.

Goals and Hypotheses

We aim at replicating substantive findings from Braun's multimode probing study (2008) in our web survey. A successful replication of results would speak in favor of using probes in web surveys. Our analysis focuses on

respondents that have a particular—contradictory—response combination with two gender ideology items. For these respondents, we examine the answers they provide to a related category-selection probe. For specific response combinations, Braun (2008) identified several answer patterns among probe answers that were not intended by the researcher. We expect to replicate these patterns, under the condition that substantive answers given by the panelists are sufficiently elaborate and certain subgroups (such as traditional respondents or respondents with new emerging egalitarian stances) are sufficiently covered in online panels. The answer patterns we intend to replicate are:

Error pattern: Agreement with a traditionally slanted benchmark item combined with agreement to an egalitarian item runs counter to measurement goals. We posit that this contradiction can be explained by misunderstandings of (at least one of) the items (e.g., by a particular idiosyncratic reinterpretation of the egalitarian item by traditional respondents). Such responses can be categorized as being “wrong,” given the measurement intentions of the researchers.

Individual solution pattern: Disagreement with a traditionally slanted benchmark item combined with disagreement (or neutral stances) with an item that depicts a specific egalitarian stance is equally troubling at first sight. However, we suggest that this pattern can be explained by the emerging preference for individual solutions for each couple. Such responses cannot be regarded as being “wrong.” On the contrary, they might reflect a well-considered personal position: rejection of the traditional role model without requiring one specific egalitarian model as binding for all.

Middle response pattern: Respondents who select middle values for both the traditional and the egalitarian item may belong to two entirely different types: the no-opinions and the strong supporters of an individual solution model for whom even a strict rejection of the traditional role model is incompatible with their views.

Data and Methods

Data Source

The data in this article come from two identical web surveys conducted in Germany in June/July 2010. Respondents for these surveys were drawn from two different online panels (around 500 cases were targeted in each

panel). Quotas for the samples were based on region (eastern vs. western Germany), sex, age (18–30 years, 31–50 years, 51–70 years), and education (less than university entrance requirement vs. university entrance requirement). The commissioning of two different panels was part of a panel experiment (Behr et al. 2012), but the experiment does not play a decisive role in the substantive analysis presented in this article. The data from the two web surveys were merged for the purposes of the analyses.

Questionnaire

The questionnaire covered the topics of gender, family, and immigrants. In total, it comprised 33 closed-ended questions and six probes per respondent. Among the closed items, this article focuses on 2 items from the gender and family block, namely *egalitarian division* (A man and a woman should share housekeeping chores and taking care of the children equally, so that both can combine work and family life) and *role segregation* (A man's job is to earn money; a woman's job is to look after the home and family). The latter is a traditionally slanted item from the International Social Survey Program (International Social Survey Programme [ISSP] 2002), which, according to MacInnes (1998:243) "is not only a classic statement of male breadwinner ideology, but captures one of the essentials of a patriarchal sexual division of labour: that men are naturally suited to public activity and women to private nurturance." As such, it can be regarded as a benchmark item.

The role segregation item is widely used in the literature to represent gender role attitudes. The item egalitarian division avoids the traditional slant by presenting a nontraditional division of labor that nontraditional respondents do not have to reject to express their egalitarian stance. However, while on the surface this item might be a perfect operationalization of an egalitarian stance, it contains multiple stimuli that are likely to cause difficulties in interpretation, as will be seen below. Both items are measured on a 5-point scale (1 = *strongly agree*, 2 = *agree*, 3 = *neither agree nor disagree*, 4 = *disagree*, 5 = *strongly disagree*).

With regard to the probes, this article focuses on the *category-selection probe* following the *egalitarian division* item. Respondents were asked, for example: "Please explain why you have chosen [answer value inserted]." Behr et al. (2012) explain in detail different wording and design experiments that were implemented for the category-selection probe and their impact on answer behavior.

Coding Procedure

The answers to the probing of the egalitarian division item were coded. The coding schema differentiated between nonsubstantive answers (such as “?”, “no answer,” “ccc,” “why not,” or “it simply is like that”), three substantive codes, and an “other” code. The substantive codes are as follows: (1) positive consequences for the children/joint responsibility in child-raising (e.g., “children need both parents”), (2) equality arguments (e.g., the catchword “equality”), and (3) the necessity of finding individual solutions (e.g., “each couple must decide for themselves”).

The restriction to these three substantive codes was motivated by the focus on the hypotheses. They helped us explain discrepancies between the answers to the two closed-ended items role segregation and egalitarian division. The code “other” included mixed arguments that would not have been incompatible with the answers given to the closed items as well as arguments not covered by the three substantive codes. The answers within the category “other” are definitively not useless but could become the main source of data for further research questions. Independent coding by two coders of 10% of the answers resulted in an agreement of 0.87, an acceptable value given that the answers can be regarded as data with medium complexity.

Data Analysis

The existing time series for the ISSP item role segregation from 1988 until 2002 will serve as a benchmark to gauge the plausibility of the traditionality level obtained for the web surveys compared to the general population (the next relevant ISSP module will be fielded in 2012). An accord between the data sources would indicate that the web survey results in terms of traditionality are realistic to some extent. Directly corresponding to the hypotheses formulated for nontraditional and traditional respondents, we will then analyze patterns of the responses to the probing question. This will be done quantitatively and illustrated with citations from the probing answers.

Results

In total, 1,023 respondents completed our two web surveys. The drop-out rate was at 7.1%, and the median response time amounted to 10:39 minutes. The probe to egalitarian division was answered by 82% of respondents on a

Table 1. Means of Role Segregation in Western and Eastern Germany.

| | ISSP 1988 | ISSP 1994 | ISSP 2002 | Web Surveys 2010 |
|-----------------|-----------|-----------|-----------|------------------|
| Western Germany | 2.9 | 3.2 | 3.5 | 3.8 |
| Eastern Germany | — | 4.0 | 3.9 | 4.0 |

Note: Role segregation measured by the item “A man’s job is to earn money; a woman’s job is to look after the home and family” on a scale from 1 = *strongly agree* to 5 = *strongly disagree*.

(basic) substantive level. The remaining 18% of answers were nonsubstantive. Behr et al. (2012) address in detail design, panel, and individual characteristics that influenced the chances of providing (non)substantive answers.

Table 1 shows the means of role segregation in western and eastern Germany for the ISSP studies 1988, 1994, and 2002 as well as for our merged web surveys. For the benchmark item for gender ideology, western Germany scores 3.8 in the web. Given the ISSP time series, which shows a strong nontraditional trend, the web sample is in line with this trend. For eastern Germany, there is hardly any difference between the web sample and the ISSP time series, which does not show any trend. Thus, our web surveys provide us with a good approximation of the plausible traditionality level in Germany. Nevertheless, we are unable to establish representativeness of our data: The comparability of the traditionality levels does not preclude that the web sample might still differ with regard to other variables.

Table 2 shows the number of observations in the different combinations of the traditionally slanted benchmark and the egalitarian item (don’t knows [DKs] excluded). The distributions for both items are skewed. However, as shown above, the distribution of role segregation is very likely similar to what we can expect to find in the general population today. The distribution of egalitarian division is dramatically more skewed. Unfortunately, we cannot compare it with scores of the general population since the item has not been used in the ISSP. However, given the responses to role segregation, this distribution of responses to egalitarian division would not be expected.

First, how can respondents who clearly prefer the traditional role model, characterized by role segregation, at the same time be in favor of an equal division of tasks? Second, why is the agreement of the nontraditional respondents with an egalitarian division not even stronger? The answers given to the probing question provide us with insights about what is happening here.

Table 2. Number of Observations in Combinations of Role Segregation and Egalitarian Division.

| Egalitarian division | Role Segregation | | | | | Total |
|----------------------|------------------|-------|-------------------------|-----------------------------|-------------------|-------|
| | Strongly Agree | Agree | Neither/ Nor | Disagree | Strongly Disagree | |
| Strongly agree | 12 | 24 | 74 | 135 | 323 | 568 |
| Agree | 7 | 37 | 95 | 123 | 68 | 330 |
| Neither/nor | 8 | 12 | 39 | 9 | 14 | 82 |
| Disagree | 4 | 10 | 3 | 3 | 2 | 22 |
| Strongly disagree | 1 | 0 | 1 | 0 | 1 | 3 |
| Total | 32 | 83 | 212 | 270 | 408 | |
| | Error pattern | | Middle response pattern | Individual solution pattern | | |

Note: $n = 1,005$. The gray-shaded boxes indicate the collapsed categories: error pattern (left), middle response pattern (center), and individual solution pattern (right).

Given the skewed distributions of both items and very small case numbers in some categories, we collapsed some categories within the error, individual solutions, and middle response patterns for further analyses (see gray-shaded boxes in Table 2). Having few case numbers is a disadvantage; at the same time, this show the merit of using web probing compared to conventional probing. Quotas for conventional cognitive interviewing are normally not based on the combinations of two closed-ended items, so it is unclear whether conventional cognitive interviewing would have allowed us to examine specific response combinations at all.

With regard to role segregation, we do not differentiate between those who strongly agree and agree (and for symmetry reasons, we also do not differentiate between those who disagree and strongly disagree). With regard to egalitarian division, due to the extremely skewed distribution and the different meanings the response categories might have compared to the role segregation item, we keep the distinction between those who agree and those who strongly agree (large enough case numbers in each cell). However, we collapse the remaining three categories, which are neither/nor, disagree, and strongly disagree.

The focus will now be on the three answer patterns: error, individual solution, and middle response in line with our research hypotheses. Table 3 displays the respondents' argumentation strands for these three answer patterns as revealed by the probe answers.

Table 3. Answer Patterns in Percent (and Absolute Numbers) as Revealed by the Probe in Collapsed Combinations of Role Segregation and Egalitarian Division.

| Egalitarian Division | | Role Segregation | | |
|---------------------------------|-----------------|----------------------|--------------------------------|------------------------------------|
| | | (Strongly) Agree | Neither/Nor | (Strongly) Disagree |
| Strongly agree | Not substantive | 31% (11) | 20% (15) | 11% (52) |
| | Children | 6% (2) | 9% (7) | 5% (22) |
| | Equality | 28% (10) | 26% (19) | 25% (115) |
| | Ind. solutions | 0% (0) | 0% (0) | 2% (7) |
| Agree | Mixed/other | 36% (13) | 45% (33) | 57% (262) |
| | Not substantive | 30% (13) | 17% (16) | 14% (26) |
| | Children | 14% (6) | 7% (7) | 8% (15) |
| | Equality | 16% (7) | 18% (17) | 24% (46) |
| Neither/nor – strongly disagree | Ind. solutions | 5% (2) | 9% (9) | 4% (8) |
| | Mixed/other | 36% (16) | 48% (46) | 50% (96) |
| | Not substantive | 26% (9) | 51% (20) | 24% (7) |
| | Children | 26% (9) | 5% (2) | 7% (2) |
| | Equality | 0% (0) | 3% (1) | 0% (0) |
| | Ind. solutions | 14% (5) | 38% (15) | 45% (13) |
| | Mixed/other | 34% (12) | 3% (1) | 24% (7) |
| | | <i>Error pattern</i> | <i>Middle response pattern</i> | <i>Individual solution pattern</i> |

Note: The gray-shaded boxes indicate the response patterns of interest: error pattern (left), middle response pattern (center), and individual solution pattern (right). The four respondents (strongly) disagreeing with the egalitarian division are excluded from the middle response pattern.

Error Pattern

The error pattern subsumes those respondents that (strongly) agree to both closed-ended items. When looking into the answers of these respondents for the probe egalitarian division (left gray-shaded cells in Table 3), the following argumentation strands are revealed: Some respondents (6% and 14%, respectively) exclusively refer to positive consequences for the children or the joint responsibility in child-raising. They offer comments such as “Children need both parents” or “because child-raising is the task of both parents.” This is compatible with a preference for traditional gender roles but shows a narrow interpretation of the item egalitarian division that does not match the measurement goals of the researchers any more. A nonnegligible part of respondents who combine (strong) agreement with traditional gender roles with (strong) agreement to an equal division of labor refers to equality arguments (e.g., “gender equality” or “gender equality should

prevail nowadays” [28% and 16%, respectively]). Although such reasons perfectly fit to the scale value respondents have selected for egalitarian division, they are entirely incompatible with the answer these respondents have selected for role segregation. Respondents in this combination should simply not mention equality reasons. For these respondents, we cannot deduce from the two closed-ended items alone what their gender ideology is or whether they have a consistent concept at all.

Individual Solution Pattern

On the opposite end (right gray-shaded cell in Table 3), respondents reject traditional gender roles without agreeing to an equal division of labor. Almost half of these respondents (45%) establish a reference to individuality or individual solutions. Virtually no one refers to equality arguments—which are the single most frequently given justification by nontraditional respondents who are in favor of an equal division of labor.

It is worth noting that respondents have a variety of individual solutions in mind, *depending on the job* (If one of them has no or only a less well-paid job, the other one [irrespective whether this is the husband or the wife] should be the main breadwinner.), *personal preferences* (This is the ideal solution; however, only if both want this. Maybe one of them likes to be househusband/-wife?), or *external restrictions* (This entirely depends on the particular case, e.g., the job someone has. Sometimes it is not possible today to freely decide who goes out to work and who cares for the children, because not everyone who wants to work can get a job and not everyone who wants to take care of the children is able to get a part-time job or time off from the job.)

Another reason pertaining to the individual solution pattern is the *preference for completely unrestricted freedom of decision making* (It does not matter who stays at home and who goes out to work, irrespective of whether this is the man or the woman [. . .]). No one has to do both things or share anything. Everyone should be free to decide what he or she wants to do, as long as this can be properly organized.).

Some of these respondents—though clearly opposing traditional roles—are in favor of an asymmetric role assignment. These respondents mention, for instance, the *advantage of role specialization* (Specialization is better than if everyone does everything.) or the *constancy of one reference person for the children* (I think it is important that one parent can stay at home, such that the children always have a contact person.). Since these respondents do

not prescribe a specific role for a specific gender, their answers can be regarded as supporting individual solutions.

These argumentation patterns illustrate a possible fallacy for researchers: The more respondents favor individual solutions, the weaker the support for an egalitarian division of tasks becomes. That is, respondents would increasingly disagree with an item depicting such an egalitarian way of life. Researchers would then conclude a traditional trend if they knew nothing of the respondents' reasoning. This, however, would not be in line with "reality."

Middle Response Pattern

Finally, we look at those respondents who offer a "neither–nor" response for both closed questions (gray-shaded cell in the middle column, Table 3). This combination displays the highest percentage of nonsubstantive answers. Half of these respondents, therefore, seem to have no opinion on this issue—or are not willing or motivated to voice it (51%). A closer investigation of these respondents shows that 85% of them answer none in three category-selection probes in the survey on a substantive level. Also, 50% of them belong to the 10% of respondents who finish the survey the quickest. The second most important argumentation pattern is for respondents mentioning individual solutions (38%). Thus, choosing the middle response for both items is, in this instance, mainly a mixture of no-opinion/no-motivation and individualism.

Again, respondents preferring individual solutions mention a variety of reasons that are very similar to the reasons of the individual solution response pattern and that can also explain the choice of the middle response category of the role segregation item: *time availability* (The decision should be based on who has more time [contingent on the job].), *resource dependency* (This is always a case-by-case decision—often based on financial considerations: As the woman [unfortunately] often earns less than the man, it is easier for the family if the woman cares for the children.) or *completely unrestricted freedom of decision making* (Everyone should decide on his or her own and find the best variant for herself or himself without pressure from society.).

Discussion

We were able to demonstrate that category-selection probing can usefully be implemented in web surveys with respondents drawn from online panels. A majority of respondents answered the category-selection probe in a

substantive way rather than just clicking through the survey or giving non-sense answers.

By focusing on specific combinations of response categories selected for a traditionally slanted benchmark item and an egalitarian item, we were able to reproduce findings by Braun (2008) and thus to confirm our hypotheses. Agreement with a traditionally slanted item combined with agreement to an egalitarian item can be explained by misunderstandings of (at least one of) the items (*error pattern*). Disagreement with a traditionally slanted benchmark item combined with disagreement (or neutral stances) with an egalitarian item, which presents a specific egalitarian model, can be explained by the emerging preference for individual solutions (*individual solution pattern*). This preference reflects a well-considered personal position that combines a rejection of the traditional role model with a rejection to make one specific egalitarian model binding for all. Finally, we demonstrated that respondents selecting middle values for both the traditional and the egalitarian items mainly belong to two types: the no-opinions/those who were not motivated to write text and strong supporters for an individual solution model. The no-opinions and unmotivated respondents especially require further investigation in future studies. The successful replication of Braun's findings (2008), which were based on a multimode probing study, backs the feasibility and usability of our web-probing method. At the same time, the probing results emphasize that the egalitarian item warrants improvements for future surveys.

If the web-probing method is implemented alongside or after major (population) surveys, the information gathered can be used to evaluate actual survey data. It can then guard against drawing wrong conclusions. If the method is implemented as part of a pretest and validity problems are uncovered, items can still be rephrased and improved to increase validity. The open answers can then serve as a pool of what is relevant and important for respondents and what might be worth being explicitly mentioned in items.

An important limitation of this study pertains to using a nonprobability sample that does not allow conclusions on the general population. However, compared to conventional cognitive interviewing, the use of online panels has the clear advantage of resulting in a markedly higher case number that can be used to clarify the meaning of (relatively) rare response combinations or assessing the prevalence of certain interpretation patterns. When probability-based online panels become available in countries other than the United States and the Netherlands, the representation problem might be mitigated in the future.

Another important limitation is the lack of interactivity in our study: Nonsense, insufficient, or incomprehensible answers were not followed up by additional probes. Here, in particular, we see areas for further research, for example, with regard to motivational texts, better instructions, or follow-up probes to initial probes.

Sample size and interactivity needs are certainly major determinants when choosing between conventional probing and web probing. We do not recommend replacing conventional cognitive interviewing with web probing. However, we understand web probing as a supplemental method when the investigation of response combinations or the prevalence of problems and argumentation patterns is needed and when in-depth information, which might only be obtained with intensive and repeated probing, is not necessarily sought. Furthermore, we see web probing as a possibility to assess item validity if cognitive labs or interviewers are not available.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This research was funded by the German Research Foundation (DFG) as part of the PPSM Priority Programme on Survey Methodology (SPP 1292) (project BR 908/3-1).

References

- Baker, R., S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, D. Dillman, M. R. Frankel, P. Garland, R. M. Groves, C. Kennedy, J. Krosnick, and P. J. Lavrakas. 2010. AAPOR report on online panels. *Public Opinion Quarterly* 74:711–81.
- Beatty, P. C., and G. B. Willis. 2007. Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly* 71:287–311.
- Behr, D., L. Kaczmirek, W. Bandilla, and M. Braun. 2012. Asking probing questions in web surveys: Which factors have an impact on the quality of responses? *Social Science Computer Review* 30:487–98.
- Blair, J., Conrad, F., Ackermann, A. C., and G. Claxton. 2006. The effect of sample size on cognitive interview findings. Paper presented at the AAPOR Conference, Montreal, Canada, May 18–21. http://www.abtassociates.com/presentations/aapor06_sample_size_cognitive_interviews.pdf (accessed August 9, 2011).

- Blasius, J., and V. Thiessen. 2006. Assessing data quality and construct comparability in cross-national surveys. *European Sociological Review* 22:229–42.
- Braun, M. 2008. Using egalitarian items to measure men's and women's family roles. *Sex Roles* 59:644–56.
- Conrad, F., and J. Blair. 2004. Aspects of data quality in cognitive interviews: The case of verbal reports. In *Methods for testing and evaluating survey questionnaires*, eds. S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, and E. Singer, 67–88. Hoboken, NJ: Wiley.
- . 2009. Sources of error in cognitive interviews. *Public Opinion Quarterly* 73:32–55.
- DeMaio, T. J., and J. M. Rothgeb. 1996. Cognitive interviewing techniques: In the lab and in the field. In *Answering questions: Methodology for determining cognitive and communicative processes in survey research*, eds. N. Schwarz and S. Sudman, 177–95. San Francisco: Jossey-Bass.
- Denscombe, M. 2008. The length of responses to open-ended questions: A comparison of online and paper questionnaires in terms of a mode effect. *Social Science Computer Review* 26:359–68.
- Dillman, D. A., J. D. Smyth, and L. M. Christian. 2008. *Internet, mail, and mixed-mode surveys: The tailored design method*. 3rd ed. Hoboken, NJ: Wiley.
- Galesic, M. 2006. Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics* 22:313–28.
- Hess, J., and E. Singer. 1995. *The role of respondent debriefing questions in questionnaire development*. Washington, DC: U.S. Census Bureau. <http://www.census.gov/srd/papers/pdf/sm9518.pdf> (accessed August 19, 2011).
- Holland, J. L., and L. M. Christian. 2009. The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Social Science Computer Review* 27:196–212.
- ISSP. 2002. International Social Survey Programme 2002: Family and Changing Gender Roles III (ISSP 2002). GESIS Data Archive, Cologne, Germany, ZA3880. Source Questionnaire.
- Krosnick, J. A. 1999. Survey research. *Annual Review of Psychology* 50:537–67.
- MacInnes, J. 1998. Analysing patriarchy capitalism and women's employment in Europe. *Innovation* 11:227–48.
- National Center for Health Statistics. 2011. Centers for Disease Control and Prevention. *Q-Bank. List of reports*. <http://wwwn.cdc.gov/QBANK/Reports.aspx> (accessed August 8, 2011).
- Nichols, E., and J. Hunter Childs. 2009. Respondent debriefings conducted by experts: A technique for questionnaire evaluation. *Field Methods* 21:115–32.
- Oksenberg, L., C. Cannell, and G. Kalton, 1991. New strategies for pretesting survey questions. *Journal of Official Statistics* 7:349–65.

- Oudejans, M., and L. M. Christian. 2010. Using interactive features to motivate and probe responses to open-ended questions. In *Social and behavioral research and the internet: Advances in applied methods and research strategies*, eds. M. Das, P. Ester, and L. Kaczmirek, 304–32. London: Routledge.
- Prüfer, P., and M. Rexroth. 2005. Kognitive Interviews [Cognitive Interviews]. ZUMA How-to-Reihe 15. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/How_to15PP_MR.pdf?download=true (accessed August 2, 2011).
- Saris, W. and I. Gallhofer. 2007. *Design, evaluation, and analysis of questionnaires for survey research*. Hoboken, NJ: Wiley.
- Schuman, H. 1966. The random probe: A technique for evaluating the validity of closed questions. *American Sociological Review* 31:218–22.
- Smith, T. W. 1989. Random probes of GSS questions. *International Journal of Public Opinion Research* 1:305–25.
- Smyth, J. D., D. A. Dillman, L. M. Christian, and M. McBride. 2009. Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly* 73:325–37.
- Vandenberg, R. J., and C. E. Lance. 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods* 3:4–69.
- Willis, G. B. 2005. *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.
- Willis, G. B., and S. Schechter. 1997. Evaluation of cognitive interviewing techniques: Do the results generalize to the field? *Bulletin de Méthodologie Sociologique* 1997:40–66.
- Wilson, T. D., S. J. LaFleur, and D. E. Anderson. 1996. The validity and consequences of verbal reports about attitudes. In *Answering questions: Methodology for determining cognitive and communicative processes in survey research*, eds. N. Schwarz and S. Sudman, 91–114. San Francisco: Jossey-Bass.