# Gender Beliefs Measurement. How a Slightly Different Wording of the Same Question Changes the Story 

Andreea Constantin ${ }^{1 *}$<br>${ }^{1}$ University of Cologne, GK SOCLIFE, 2 Richard Strauss Str., Room 3.A01, 50931 Cologne, Germany

## Keywords

Attitudes towards gender roles Measurement Value surveys Wording


#### Abstract

Slightly different wordings are known to introduce important differences in the way people understand and answer survey questions and, moreover, in the quality of the items (Billiet 1991). This is what may happen also in the case of two wordings used to measure the attitude people express towards the effect of the women's job on their children. The aim of this study is to assess which of the two almost similar items, assumed to tap this kind of attitudes, produces a better measurement in terms of validity, reliability and overall quality. For this purpose I employ original data and use OLS models as well as SQP analysis. The findings reveal some surprising differences between the two items. This study starts with a short introduction, followed by the description of the method and the presentation of the findings. A short discussion concludes the text, focusing on implications for future studies.


## Introduction

In the last few decades the scientific interest on the attitudes toward gender roles has considerably increased in the academic world. Broadly, by attitudes towards gender roles we understand the normative beliefs about what gender relations in society should be like, or to the extent to which a person supports the norm of gender equality' (Bergh 2006: 6).

These attitudes have been measured in different cross-national studies like World Values Survey, International Social Survey Programme,

[^0]European Values Study, European Social Survey or Generations and Gender Survey. As the relatively high number of cross-national surveys studying this topic can show, these attitudes are highly relevant in a comparative perspective. Looking at the literature, one can observe that, in fact, this is the methodological approach used by most of the studies related with this set of attitudes. Among the themes linked with the attitudes towards gender roles are: changes in these attitudes across different cultures or social groups, their connection with some institutions or their link with other types of attitudes and values (Baxter and Kane 1995; Pfau-Effinger 2004; Sjöberg 2010; Geisler and Kreyenfeld 2011). This is why having cross-national and cross-time comparative measures is an important feature of these studies.

One of the possible biases found in the survey-based data comes from the different wording of questions assumed to measure the same concept. Among the important features of a question that can influence the answers given by respondents refer to how persuasive is the phrasing, how social desirable is the concept measured or the level of specificity implied by the item (Billiet 1991). When the items measured in different waves or surveys differ on one or more of these characteristics, important variation appears in the way people understand the questions.

Keeping this in mind, it is important to see how a slightly different wording of a survey question can give different results in order to prevent unwanted biases. This will make the objective of this research note. By using original data, collected through a Romanian convenience sample, the aim of this paper is to see how small changes in a gender belief item affect the quality of the collected information. The question discussed here - 'When a mother works for pay, the children suffer' vs. 'Pre-school child probably suffers when the mother is working' - has been addressed in both forms and considered equivalent. But is this assumption valid? Do the two forms give similar results? In the following, the present paper addresses this problem, trying to see which of the two wordings of the question is more suitable to be addressed in surveys. More explicitly, I will look at the quality (validity and reliability) of the two questions assessed in SQP 2.0, which allows survey researchers to create questions and obtain a quality prediction, based on information given by experiments done on more than 3000 questions. Moreover, I will look at the external validity of the two forms, by running a
regression model where the independent variables' effects are well known, and compare the results with the theoretically hypothesized ones. The data employed in this study are data collected through a short questionnaire in the summer of 2012, on a convenience sample of Romanian respondents. The short questionnaire was applied in order to test the country-specific variables for the Romanian 2012 WVS wave. Among the items we have included two different wordings of the same question measuring the attitude people express towards the effect of the women's job on their children. These two items will be compared here in order to see which of the two has a better quality and it's preferable to be included in the survey questionnaire.

## Theoretical and methodological specifications

In order to see which of the wordings leads to a better measurement, a first step is to look at the validity of the measurement. One estimation of the validity included in this study is the one given by SQL. Moreover, the evaluation of the two phrasings will also be based on the results of the external validity of the items. By external validity, I refer to whether the debated variables measure what we expect them to measure (Bollen 1989: 184), or, more specifically, whether the item accurately taps the theoretical concept meant to be measured.

A regression model will be estimated, where the dependent variables are the two different phrasings of the question, while the independent variables are theoretical predictors of the attitudes towards gender roles. Given the limitations imposed by the dataset used, the models will include a variable measuring the age, income, education level, gender and a control item measuring attitudes towards gender role, which is expected to be highly correlated with the dependent variable.

First, the age of the respondent is expected to have a strong relationship with the attitudinal items, older people usually expressing less egalitarian views towards gender roles. Higher educated people and, in consequence, as well those with larger incomes are expected to express more egalitarian gender beliefs. When looking at gender, the literature specifies that women express more egalitarian attitudes towards gender roles
compared to men (Baxter and Kane 1995; Kane 1998; Andrews et al. 2004; Becker 1981; Haas 2005; Edlund 2007). An important aspect that might influence the answers given by respondents to the two different worded questions is the specification of the age of children that might suffer when their mother is working. While the first wording talks about children in general, the second version specifies that the question refers to pre-school children. Previous research indicates different attitudes towards gender roles according to how young are the children in the family, saying that the presence of small children reshapes partners' division of roles toward a more traditional version even in the cases of a previous fairly equal division of labor (Geisler and Kreyenfeld 2011).

A second step in assessing the differences between the two wordings of the question discussed here is to look at the reliability of the questions. Reliability is defined as the consistency or the repeatability of a measure, or more specifically, the consistency of the responses individuals give to the same question over trials or across individuals. We can have reliable measures which are not valid (Bollen 1989). The reliability of the two wordings is assessed by using the SQP application. SQP 2.0 allows researchers to create questions and obtain a quality prediction. It also provides information about Multitrait-Multimethod Matrix ${ }^{1}$ (MTMM) experiments done on more than 3000 questions. For each survey item scholars provide information about several characteristics of the questions. Then, a quality prediction for the question is estimated based on these characteristics using MTMM data and analysis previously done by The Research and Expertise Centre for Survey Methodology (RECSM) for the European Social Survey (ESS) and for other surveys in many different countries ${ }^{2}$.

## Results

As the two phrasings are expected to measure the same belief towards the effect of the women's job on their children, a correlation coefficient has been estimated. Surprisingly, the correlation is very weak ( -0.149 , $\operatorname{sig}<0.05$ ),

[^1]indicating that the two variables do not measure the same concept. Moreover, I looked at the number of missing values which is very low (0 missing values for Q1; $0.5 \%$ for Q2) for both variables, so this cannot be used as a criterion in selecting the preferred wording.

SQL computes one validity and one reliability coefficient. It also provides an overall quality coefficient. These coefficients are based on the characteristics of the two wordings defined by the user which are compared to the characteristics of the over 3000 survey questions available in the database. Among these characteristics are to be found the response options, introductions, information about the show-cards, about the lights of the questionnaire or the data collection method. Moreover, using all this information the validity and reliability coefficients as well as the quality one are computed using Breiman's (2001) random forests of regression trees ${ }^{3}$. These coefficients take values from 0 to 1 . Looking at the validity coefficient, the results show a small difference between the two (see Table 1 and 2), indicating a slightly better validity for Q2. This would imply that the second phrasing measures better than the first one what people believe about the effect of the women's job on their children.

Table 1. Q1. When a mother works for pay, the children suffer

| Prediction | Value |
| :--- | :--- |
| Reliability | 0.689 |
| Validity | 0.975 |
| Quality | 0.672 |

Table 2. Q2. Pre-school child probably suffers when the mother is working

| Prediction | Value |
| :--- | :---: |
| Reliability | 0.700 |
| Validity | 0.978 |
| Quality | 0.685 |

[^2]Moreover, when we focus on the external validity, the differences become clearer. Although the effects of the theoretical determinants of the gender beliefs are not statistically significant in the models estimated in order to assess the external validity of the items, looking at the direction of the effects some observations can be made. First, the effects of the traditional predictors are in the opposite direction than expected for Q1 (see Table 3). More specifically, the OLS regression shows that people who think children suffer when their mother works are also those who disagree with the idea of women being equally fulfilled as housewives as when having a paid job. This relationship implies the association of two contradicting attitudes, one indicating the preference for women to stay home, while expressing the belief that women would be less fulfilled by doing so. This model (see Table 3) also shows that women are less egalitarian than men, as well as people with higher incomes and higher education. All these results indicate a very low external validity for this version of wording, as none of these effects are usually found in the previous research and are not in line with the theoretical expectations. The positive relationship between this variable and the age is the expected one, older people tending to express less egalitarian gender views.

Table 3. Linear regression. Dependent variable: 'Q1. When a mother works for pay, the children suffer'

|  | Unstandardized <br> Coefficients | Standardized <br> Coefficients |
| :--- | :--- | :--- |
| (Constant) | 2.37 |  |
| Being a housewife is just as fulfilling as working | -0.04 | -0.04 |
| for pay (4-strongly agree). | 0.10 | 0.05 |
| Female | 0.01 | 0.04 |
| Income respondent | 0.01 | 0.05 |
| Education | $0.01^{+}$ | 0.22 |
| Age |  |  |

+ p <0.1; $\mathrm{R}^{2}$ adjusted $=0.025$
$1=$ strongly disagree; $2=$ disagree; $3=$ agree; $4=$ strongly agree

Second, the direction of the effects observed in the regression model estimated for the second wording - Q2 - is the expected one (see Table 4). Here, individuals who think that pre-school children would probably suffer
when their mother works also agree more with the idea of women being equally fulfilled when she stays home as when she has a paid job. Also, more egalitarian individuals are females, highly educated and earn a higher income. These effects are the theoretically expected ones, previous research emphasizing similar results. Still, in this model age is not in line with what we would expect, as it has a positive effect, while it should have a negative one, more egalitarian people being usually younger, not older as the analysis indicates (see Table 4).

Table 4. Linear regression. Dependent variable: 'Q1. Pre-school child probably suffers when the mother is working'

|  | Unstandardized <br> Coefficients | Standardized <br> Coefficients |
| :--- | :--- | :--- |
| (Constant) | 1,82 |  |
| Being a housewife is just as fulfilling as working <br> for pay (4-strongly agree). | $-0,10$ | $-0,09$ |
| Female | 0,12 | 0,06 |
| Income respondent | 0,01 | 0,03 |
| Education | $0,04^{+}$ | 0,15 |
| Age | 0,01 | 0,09 |

+ $\mathrm{p}<0.1$; R 2 adjusted $=0.006$
$1=$ strongly agree; $2=$ agree; $3=$ disagree; $4=$ strongly disagree

When looking at the reliability assessment made by the SQP application, it can be seen that the results are congruent with what the previous steps showed. More exactly, the reliability coefficient shows a slightly higher reliability for Q2 (see Table 1 and 2). This indicates that the second wording gives more consistent measures for the concept discussed here, or more specifically, that the responses individuals give to this question are more similar/consistent over trials or across individuals. Also, the measure of quality estimated in SQP, based on experimental data, indicates that Q2 is a better survey question than Q1, which would point out the conclusion that overall, Q2 is the preferred wording from the two discussed here.

A question that might appear here is what can be done to improve the phrasing that we choose to work with or where specifically come the
problems from. An important feature of the SQP analysis is the potential improvements suggested for each of the questions assessed. Based on the characteristics defined by the author, the software compares the two items of interest with some other similar ones found in its database. This comparison permits the software to indicate what characteristics of the questions may be changed in order to improve the reliability, validity and the overall quality coefficients of the items. Among the suggestions made by SQL we find some that could help us improve the quality of the question. One of them is to have a slightly longer question. This could bring additional information to the respondent and make clearer what the individual has to think about. However, a longer scale would induce difficulties in applying it, particularly when collecting data by using long questionnaires like the World Values Survey one.

Another suggestion refers to the number of categories of the response scale. Here, the experimental data shows that, for the question to be better, we should offer the respondents at least 10 categories form which to choose. Still, this would imply a change in the item's scale and would bring comparability problems across survey waves. Moreover, a slightly better quality of the question would be reached by developing a self-applied questionnaire, not an interviewer applied one. Here, it is important to be mentioned that some of the differences observed between the two wordings quality might be due to the interviewer influence.

From what the pretesting report shows, in case the questionnaire will be applied by an interviewer, it is absolutely necessary to explain better to the interviewers the meaning of the questions and to make it clear that they should not add explanations that are not written in the questions as: mother's work place is in Romania or abroad. It is probable that some of the inconsistencies encountered in the items analyzed here to be due to the fact that respondents answered not only according to how they understood the question, but also to how the interviewer addressed it.

## Discussion

Overall, it can be said that Q2: 'Pre-school child probably suffers when the mother is working' is a better way of measuring the attitude expressed by
people towards the effect of the women's job on their children, as it has a better reliability, a higher validity and an overall better quality.

This might be due to any of the small differences in the two wordings. First, the quality difference may rise from the fact that Q2 has a more specific subject, the pre-school children. Respondents might find it easier to answer the question when the subject is clearly identified compared to the situation where they have to answer questions about children of all ages.

Second, the probabilistic terms in which Q2 is formulated might be another reason why people understand and answer more coherently this question. People tend to elude expressing radical opinions, so that saying that all children will surely suffer when their mother is working might be a too strong sentence compared to the more sceptical and probabilistic formulation saying that there are chances for the very young children to suffer when their mother has a job.

These different results for the two similar wordings should call the attention of the researchers working with the gender beliefs items, especially when they tend to use them interchangeably. Although the present study is very limited by the sample size and data collection method, and the results should be taken with caution, the observations made here indicate that the two wordings are not equivalent and should not be used interchangeably.

## Acknowledgement

Working on this paper was supported by CNCS grant PN-II-ID-PCE-2011-3-0210.

## References:

Andrews, Arlene B., Irene Luckey, Errol Bolden, Judith Whiting-Fickling and Katherine A. Lind. 2004. Public Perceptions about Father Involvement: Results of a Statewide Household Survey. Journal of Family Issues 25 (5): 603-633.
Baxter, Janeen and Emily W. Kane. 1995. Dependence and Independence: A Cross-National Analysis of Gender Inequality and Gender Attitudes. Gender $\mathcal{E}$ Society 9 (2): 193-215.
Becker, Gary S. 1981. A Treatise on the Family. London and Cambridge, MA: Harvard University Press.
Bergh, Johannes. 2006. Gender Attitudes and Modernization Processes. International Journal of Public Opinion Research 19 (1): 5-23.

Billiet, Jacques B. 1991. Research on Question Wording Effects in Survey Interviews. Graduate Management Research 5 (4): 66-80.
Bollen, Kenneth A. 1989. Structural Equations with Latent Variables. New York: John Wiley \& Sons
Breiman, Leo. 2001. Random forests. Machine learning 45 (1): 5-32.
Campbell, Donald T. and Donald W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 56 (2): 81-105.
Edlund, Jonas. 2007. The Work Family Time Squeeze: Conflicting Demands of Paid and Unpaid Work among Working Couples in 29 Countries. International Journal of Comparative Sociology 48 (6): 451-480.
Geisler, Esther and Michaela Kreyenfeld. 2011. Against all odds: Fathers' use of parental leave in Germany. Journal of European Social Policy 21 (1): 88-99.
Haas, Barbara. 2005. The Work-Care Balance: Is it Possible to Identify Typologies for Cross-National Comparisons? Current Sociology 53 (3): 487-508.
Kane, Emily W. 1998. Men's and Women's Beliefs About Gender Inequality: Family Ties, Dependence, and Agreement. Sociological Forum 13 (4): 611-637.
Pfau-Effinger, Birgit. 2004. Development of Culture, Welfare States and Women's Employment in Europe. Aldershot: Ashgate.
Saris, Willem E., Daniel Oberski, Melanie Revilla, Diana Zavala, Laur Lilleoja, Irmtraud Gallhofer and Tom Gruner. 2011. Final report about the project JRA3 as part of ESS Infrastructure. RECSM Working No. 24. Barcelona: Research and Expertise Centre for Survey Methodology.
Sjöberg, Ola. 2010. Ambivalent Attitudes, Contradictory Institutions: Ambivalence in Gender-Role Attitudes in Comparative Perspective. International Journal of Comparative Sociology 51 (1-2): 33-57.


[^0]:    * Contact address: constantin@wiso.uni-koeln.de (A. Constantin)

[^1]:    ${ }^{1}$ See Campbell and Fiske (1959) for a description of the basic principles of the MTMM.
    ${ }^{2}$ More information available at http:/ / www.sqp.nl/

[^2]:    ${ }^{3}$ For more details about the coefficients' computation see Saris et al. (2011).

