


Are You...? Asking Questions on Sex with a Third Category in Germany

Field Methods
2022, Vol. 34(2) 91–107
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1525822X211072326
journals.sagepub.com/home/fmx


Patricia Hadler¹, Cornelia E. Neuert¹,
Verena Ortmanns¹, and Angelika Stiegler^{1*}

Abstract

A question asking for respondents' sex is one of the standard sociodemographic characteristics collected in a survey. Until now, it typically consisted of a simple question (e.g., "Are you...?") with two answer categories ("male" and "female"). In 2019, Germany implemented the additional sex designation *divers* for intersex people. In survey methodology, this has led to an ongoing discussion how to include a third category in questionnaires. We investigate respondents' understanding of the third category, and whether introducing it affects data quality. Moreover, we investigate the understanding of the German term *Geschlecht* for sex and gender. To answer our research questions, we implemented different question wordings asking for sex/gender in a non-probability-based online panel in Germany and combined them with open-ended questions. Findings and implications for surveys are discussed.

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

*Retired.

Corresponding Author:

Patricia Hadler, GESIS–Leibniz Institute for the Social Sciences, P.O. Box 12 21 55, 68072 Mannheim, Germany.

Email: patricia.hadler@gesis.org

Introduction

Respondents' sex/gender is one of the standard sociodemographic characteristics collected in a survey. The main reasons for collecting information on sex/gender are describing the sample composition using sociodemographic characteristics, using sex/gender as a filter variable so that some questions are answered only by male or female respondents, controlling for sex/gender in the analyses, and as a variable for hypothesis testing (Döring 2013). Up to now, the question asking for sex/gender traditionally consists of a short and simple question (e.g., "Are you...?") with two answer categories ("male" and "female"). In several social survey programs, sex and gender are conflated and used interchangeably (Bauer et al. 2017; Bittner and Goodyear-Grant 2017; Westbrook and Saperstein 2015). *Sex* is commonly based on genital anatomy (Tate et al. 2013) and assessed at birth (Broussard et al. 2018), while *gender* refers to a "psychological sense of self," which is also based on expectations about gender roles (Tate et al. 2013:767). This conflation of concepts is linguistically reinforced in languages like German that use the same term (*Geschlecht*) to refer to both gender and sex.¹

In interviewer-administered surveys, the question asking for sex/gender is typically not answered by respondents themselves but determined by the interviewer based on appearance or voice. Therefore, the question intends to capture the alleged biological sex, or the sex assigned at birth (Döring 2013) (although it more correctly measures gender expression). The biologically visible sex is then often used as proxy for gender in survey research (Bittner and Goodyear-Grant 2017). However, asking a binary question only and classifying people based on their physical appearance equates sex and gender and ignores or overlooks non-binary, such as intersex or transgender people. *Intersexuality* refers to people who cannot be clearly identified as male or female due to variations in sex characteristics (which must not necessarily be physically apparent at all; Döring 2013). *Transgenderism* means that the sex assigned at birth differs from the current gender identity (APA 2009).

In Germany, an additional sex designation was introduced in January 2019 for intersex people with the "Act Amending Information to be Entered into the Register of Births" (§20 , 3 PStG [Civil Status Act]). Besides leaving that section blank on birth certificates, intersex people now have the option to select a third category *divers*.² *Divers* can mean different or several (Duden 2020) and was selected due to its positive meaning. Although to date only few countries allow more than two legal sexes (Lindqvist et al. 2018), this might change in the future. In survey methodology, the change in legislation has led to discussions how to best include a third category in questionnaires.

As a survey also represents a type of communication between the researcher and the respondents, the question wording as well as the presented answer categories may impact questionnaire evaluation and the response

process. To ensure high data quality in surveys, questions and terms need to be consistently understood by all respondents (Fowler 1995). So far, it is not clear how the presentation of the sex category divers is perceived in a general population survey. On the one hand, respondents who feel excluded or discriminated by the way the question(s) is/are formulated may abandon the survey or systematically modify their response behavior. On the other hand, including a third category can be perceived as irritation or provocation. Previous research shows that there are objections to adding a non-binary category due to gender-binary beliefs and heteronormativity (Broussard et al. 2018).

Given the ambiguous gender- and sex-related terminology in the German language, and in light of recent changes in legislation leading to the introduction of a third sex category, this article tries to shed some light on the following research questions:

1. Does question wording and the presentation of a third response category impact response behavior and data quality in general population surveys?
2. How is the term “Geschlecht” understood in comparison to more specific terms such as “officially registered sex” and “gender identity”?
3. How is the third category divers understood?

Methods

To answer these questions, we experimentally varied the question wording, either showing the ambiguous German term *Geschlecht* or more distinct terminologies referring to officially registered sex or gender identity. Further, we varied whether respondents were shown the third category. Understanding of the respective terms was examined by implementing closed- and open-ended probes (Web Probing; see Behr et al. 2017).

Respondents

We conducted a study with respondents in Germany who were drawn from the opt-in panel of the Respondi AG. The experiment was placed at the end of a survey on a range of topics. Field time was from March 15–25, 2019. In total, $n = 1,135$ completed the survey (participation rate of 20.3%; AAPOR 2015). Recruitment was based on equal quotas for age, education, and sex. The sex-quota was based on a non-specific, binary question at the beginning of the questionnaire (“What sex/gender [*Geschlecht*] are you?”). This quota was necessary to compare the traditional binary question with the question variations used within this experiment. While this is disputable from a gender studies perspective and may cause non-binary respondents to break off the

survey, the purpose of our study was to examine the understanding of respondents in general population surveys. Average completion time for the entire survey was 10.2 min; the reported experiment took 2.1 min.

Experimental Set-Up and Conditions

We began the experiment by stating that in future, new questions on sex/gender will be implemented into surveys, and we therefore asked respondents to answer the questions, even if they had answered an identical or similar question earlier in the survey. Afterward, respondents were shown one of five versions of the question on sex/gender.

Experimental condition (A) “non-specific binary” was the only group in which respondents were not shown a new question but were reminded of the introductory quota question and their answer (“male” or “female”). Condition (B) “non-specific third category” used the same non-specific question wording (Geschlecht) as (A) but included the third response category (“divers”) and an explicit non-response (“I don’t want to say”). The remaining experimental conditions used the same answer categories as in (B) with varying question stems. Condition (C) “registered sex” specifically asked about the “(officially) registered sex.” Condition (D) “gender identity” asked about the respondents’ gender identity. Condition (E) “registered sex and gender identity” asked two questions on one screen: the question on registered sex as in (C) and on gender identity as in (D) (see [Online Supplement Figure 1](#) for an overview of the experimental design as well as of the question wording and answer categories across conditions).

Respondents were first asked to indicate whether they were familiar with the term divers in the context of sex/gender. They were then asked to give a definition of divers in an open-ended question.

Additionally, respondents were asked two questions to examine the underlying construct they were referring to when answering the experimental question. An open-ended probe asked about their *understanding* of the term used for sex or gender identity in the question as it was presented to them. In condition (E), respondents were presented two open-ended probes on the same screen, one for their registered sex and one for gender identity. A closed-ended probe asked about respondents’ assignment of their sex/gender question to the constructs of biological sex or gender identity depending on question wording (What were you thinking of when you answered this question “[Question text]?” with the answer categories “Mainly about my biological sex”; “Mainly about my gender identity”; “Both equally”). As respondents received two questions covering the two constructs, this question was not presented in condition (E).

Analytical Strategy

To answer the first research question, three types of respondent behavior indicative of irritation and subsequent low data quality were assessed. First, respondents may choose to discontinue the survey; thus, survey break-off during the experiment was examined. Second, respondents may give inconsistent answers to the initial binary question at the beginning of the survey and the non-binary question(s), for instance, by once indicating they are male, and later that they are female. Cohen's kappa for reliability of the response was calculated. Finally, the quality of the responses to the open-ended probes were examined by assessing whether probes were left unanswered or consisted of meaningless or even rude comments, such as judgmental or gender-binary content, indicating respondent irritation (Smyth et al. 2009).

To answer the second and third research question on the understanding of *Geschlecht* and divers, the responses to the open-ended questions were coded using a bottom-up approach. We developed two separate coding plans for the understanding of the terms referring to sex or gender identity and for divers. Responses could include more than one substantive code. All codes were developed inductively by one of the authors and refined by the other authors. All answers were coded by one of the authors and a student assistant. Discrepancies were discussed, the coding plan revised, and answers recoded independently. For sex/gender, Cohen's kappa was between 0.36³ and 0.93. For divers, Cohen's kappa was between 0.86 and 1.00. Remaining discrepancies were discussed, and final codes assigned together. The Coding Plans are available in the [Online Supplement](#).

Both coding plans contained two exclusive main categories, those being gender-binary beliefs and non-substantive answers. Gender-binary beliefs are judgmental statements about non-binary people or statements that deny the existence of non-binary sex/gender (e.g., "God made only male and female"). They depict gender as exclusive, based on biological aspects, congenital, and invariant; further, sex and gender are argued to be congruent and the binary genders as complementary (Debus and Laumann 2018). Non-substantive responses included non-response (e.g., "-"; "..."; "no comment"), random characters, don't know responses ("no idea"; "unsure"), and other uncodeable content.

Quantitative analyses were carried out using SPSS 24. We report Pearson's chi-square tests of independence. When expected frequencies are below five, Fisher's exact test is reported. Pairwise comparisons between experimental conditions were carried out when the overall test of independence was significant, and Bonferroni correction for multiple comparisons was applied.

Coding Plan Sex/Gender

Six substantive codes were grouped into two types of answer categories. The first referred to one or more of the underlying constructs (biological sex, gender identity, and officially registered sex). *Biological sex* was coded when a response clearly referred to sex characteristics and/or the anatomy of an individual's reproductive system. *Gender identity* contained all definitions that referred to either social roles defined by the persons' sex or definitions of personal identification of one's own gender. *Official documents* were assigned to all answers that refer to official documents, such as birth certificates or identity cards.

The remaining three responses referred to labels to categorize people which are summarized as operational labels. Frequently, respondents just named their own gender or gave responses like "male and female." The category *cisgender* was assigned when a respondent answered that his or her gender identity matches the sex assigned at birth or if one's personal answer (i.e., "I am male") was repeated. *Binary definition* included all entries that are limited to two sex categories. *Non-binary definition* was assigned to responses beyond male and female.

Coding Plan Divers

The coding plan divers consisted of 22 substantive codes in its most granular level, which were condensed to two main categories: Concrete conceptualizations defined the term along one (or more) of the four components of sexual identity, while responses remaining unclear whether they were referring to biological sex or gender identity were summarized as vague conceptualizations.

Concrete conceptualizations included four sub-categories that could clearly be assigned to one (or more) of the four dimensions of sexual identity (Jourian 2015). The category *intersexuality* was coded when a response gave a definition that clearly referred to biological *intersexuality* (and thus, biological sex) in the sense intended by German legislation. The category *transgender* referred to any form of incongruence between sex and gender identity. *Sexual orientation* was coded when responses described divers as including homo-, bi-, multi-, and/or asexuality. *Gender expression* was assigned to responses referring to transvestitism and cross-dressing.

Vague conceptualizations were used to code definitions that remained unclear as to whether they referred to sex or gender identity. Many of these responses defined divers in relation to the two binary categories. For instance, the code *both male and female* defined divers as including both male and female elements. The code *neither male nor female* was exclusion-based with divers belonging to neither binary category. The code *between male and female* defined divers as a middle category, implying that sex/gender is a

continuum of which male and female constitute the extremities. The code *unclear assignment* was used when divers was defined as not clearly, entirely, or decidedly male or female. The *third category* or the third sex/gender (“das dritte Geschlecht”) was a common phrasing in German-speaking media around the time of implementation. The codes *non-binary category* and *additional category* defined divers as a category beyond the traditional two categories, such as another or different sex/gender. The code *gender neutral* described a complete lack of sex/gender. Finally, the code *explicit non-response option* defined divers as an alternative response when respondents did not wish to report their sex/gender.

Results

Response Behavior and Data Quality

The first research question asked whether question wording and the presentation of a third response category impacts response behavior and data quality, which is analyzed by survey break-off, discrepancy of answers between the binary question on sex at the beginning of the survey and the later versions with a third category, and low-quality answers to the open-ended probes.

Survey Break-off. In total, only six respondents abandoned the survey during the reported experiment. Of these, four were male and two were female. All four men were assigned to condition (D) with the question on gender identity. Of the two females who broke off the survey, one was assigned to condition (A) and the other to condition (E). None of the dropouts showed visible signs of reactance, such as profanity in their answers to open-ended questions before dropping out.

Response to New Questions on Sex/Gender. In conditions (B) through (E), the experiment began with respondents filling out one (or two) new question(s) on their sex/gender including the new category divers.

In total, 1% of respondents ($n = 17$) gave an answer that differed from the answer given to the binary question at the beginning of the survey. This amounts to Cohen’s kappa of 0.97. Inconsistent answers were dispersed among all experimental conditions (B: $n = 2$; C: $n = 5$, D: $n = 4$; E: $n = 6$). Two respondents used the option divers, both in answer to the question on gender identity in condition (E). Four respondents chose the explicit non-response option. The remaining 11 respondents gave one of the binary responses in answer to the question at the beginning of the survey and the other binary option in answer to the question in the reported experiment. In none of the cases do the responses to the open-ended questions indicate that the

respondent is intersexual or transgender, nor do the responses show signs of reactance to the topic.

Signs of Irritation to Open-ended Questions. Non-substantive answers were given by 11% of respondents in answer to the open-ended question on sex/gender, and 15% of respondents in answer to the open-ended question on the third category. This is comparable to other web probing studies (Lenzner and Neuert 2017). Gender-binary beliefs were expressed by 2% of respondents in answer to the open-ended question on sex/gender, and by 4% of respondents in answer to the open-ended question on divers. Neither the level of non-substantive answers (question on sex/gender: $\chi^2 = 4.951$; $df = 4$; $p = 0.292$; divers: $\chi^2 = 4.628$; $df = 4$; $p = 0.328$) nor gender-binary beliefs differed by question version (question on sex/gender: $\chi^2 = 2.786$; $df = 4$; $p = 0.594$; divers; $\chi^2 = 1.527$; $df = 4$; $p = 0.822$).

Men were more likely than women to give non-substantive answers to both the open-ended question on sex/gender (16% vs. 7%; $\chi^2 = 20.383$; $df = 1$; $p < 0.001$) and divers (18% vs. 12%; $\chi^2 = 6.746$; $df = 1$; $p = 0.009$). Moreover, they were more likely to voice gender-binary beliefs in answer to the open-ended question on divers (5% vs. 2%; $\chi^2 = 6.663$; $df = 1$; $p = 0.010$).

Understanding of Terms Asking about Sex/Gender

Associations of Sex/Gender Depending on Question Wording. Respondents in conditions (A) to (D) were asked whether they associated their sex/gender

Table 1. Association with Construct Depending on Term Used for Sex/Gender.

Answer categories	Experimental condition in % (n)				Total
	(A) non-spec. binary	(B) non-spec. third category	(C) registered sex	(D) gender identity	
Mainly my biological sex	55.2 ^d (117)	52.6 ^d (113)	54.6 ^d (118)	36.4 ^{a,b,c} (78)	49.7 (426)
Mainly my gender identity	4.2 ^d (9)	6.5 ^d (14)	1.9 ^d (4)	14.5 ^{a,b,c} (31)	6.8 (58)
Both equally	40.6 (86)	40.9 (88)	43.5 (94)	49.1 (105)	43.5 (373)
Total	100 (212)	100 (215)	100 (216)	100 (214)	100 (1135)

Note: Overall: $\chi^2 = 41.09$; $df = 6$; $p < 0.001$.

^{a/b/c/d} $p < .01$ pairwise Fisher's exact tests between conditions with superscripts indicating significant differences to the specified condition.

question more with their biological sex, gender identity, or with both equally. Except for condition (D), most respondents thought about their biological sex only and almost all others thought about both their sex and gender regardless of whether the question asked about sex/gender or registered sex (see Table 1). Only in condition (D) in which the survey question specifically asked about gender identity, significantly more respondents stated that they thought about gender identity (15% compared to 2%–7% in conditions A to C) and significantly less about their biological sex (36% compared to 53%–55% in conditions A to C).

Understanding of Sex/Gender. The related open-ended probe asked respondents to explain how they understood either Geschlecht, officially registered sex, or gender identity. In condition (E), two probes were asked for the terms officially registered sex and gender identity. The associated underlying constructs differed across conditions depending on which construct the question wording addressed (see Table 2). When asking about gender identity, reference is made to it (D: 71% and E₂: 76%; significantly different to conditions A, B, and C). When asking about officially registered sex, reference is made to official documents (C: 55%; E₁: 54%, significantly different to conditions A, B, and D) while official documents are not or seldom mentioned in the other conditions (A, B, D, and E₂). The two traditionally worded versions A and B showed no statistically significant differences.

The remaining codes, which were grouped as “operational labels,” revealed that for most respondents in our sample, sex and gender are strongly intertwined (see Table 3). Asking respondents to describe how they understand sex/gender [Geschlecht]/registered sex/gender identity often led to answers like “male and female” or just naming their own sex/gender. By far, the most common codes were cisgender (overall: 60%) and a binary definition (overall: 27%). However, distributions across conditions differed significantly ($\chi^2 = 111.55$; $df = 24$; $p < 0.001$). Cisgender was most often mentioned in conditions asking for the understanding of officially registered sex (C: 79%; E: 80%; statistically significant different from A, B, and D). Responses coded as binary definition were most represented in the conditions asking for “Geschlecht” (A: 44%; B: 42%; statistically significant differences to C to E).

Understanding of the Response Option Divers

Familiarity of Divers. Respondents were asked whether they were familiar with the term divers in the context of questions on sex/gender. Most respondents (84%) claimed to be familiar with the term; 4% were not certain, and 12% did not know the term. There were no significant differences across conditions ($\chi^2 = 2.928$; $df = 8$; $p = 0.939$).

Understanding of Divers. Approximately one-third of respondents (32%) included a concrete conceptualization of the term divers in their response;

Table 2. Frequencies of Codes Related to Sex/Gender Constructs Across Conditions, % (n).

Condition Codes	(A) non-spec; binary	(B) non-spec; third category	(C) registered sex	(D) gender identity	(E ₁) registered sex and gender identity	(E ₂) registered sex and gender identity
Biological sex	61.5 ^{c,d,e1,e2} (48)	55.2 ^{c,d,e1,e2} (48)	28.8 ^{a,b,e2} (32)	15.7 ^{a,b,e1,e2} (14)	33.3 ^{a,b,d} (53)	11.5 ^{a,b,c,d} (19)
Gender identity	12.8 ^{d,e1,e2} (10)	17.2 ^{c,d,e1,e2} (15)	4.5 ^{b,d,e2} (5)	70.8 ^{a,b,c,e1} (63)	1.9 ^{a,b,d} (4)	75.8 ^{a,b,c} (125)
Official documents	1.3 ^{e1} (1)	0 ^{c,e1,e2} (0)	55.0 ^{a,b,d,e1} (61)	0 ^{c,e1,e2} (0)	54.1 ^{a,b,d} (86)	9.1 ^{b,c,d} (15)
Biol. sex and gender identity	24.4 ^{c,e1,e2} (19)	27.6 ^{c,e1,e2} (24)	4.5 ^{a,b} (5)	13.5 ^{e1,e2} (12)	2.5 ^{a,b,d} (4)	2.4 ^{a,b,d} (4)
No. of responses	100 (78)	100 (87)	100 (103)	100 (89)	100 (146)	100 (163)

Note: In condition (E), we probed for the understanding of registered sex as in (C) and gender identity as in (D); the probed term is shown in bold. The results of these codings are presented separately in columns E₁ and E₂. Therefore, we do not test differences between condition E₁ and E₂. ^{a,b,c,d,e1,e2} $p < .0035$; pairwise Pearson's χ^2 -tests or Fisher's exact test between conditions with superscripts indicating significant differences using Bonferroni correction for multiple comparisons to the specified condition.

Table 3. Frequencies of Codes Related to Operational Understanding of Sex/Gender Across Conditions, % (n).

Condition Codes	(A) non-spec; binary	(B) non-spec; third category	(C) registered sex	(D) registered identity	(E ₁) registered sex and gender identity	(E ₂) registered sex and gender identity
Cisgender	44.3 ^{c,e1,e2} (54)	41.9 ^{c,e1,e2} (57)	79.1 ^{a,b,d} (68)	53.3 ^{c,e1,e2} (64)	80.0 ^{a,b,d} (84)	74.1 ^{a,b,d} (83)
Binary definition	41.8 ^{c,d,e1,e2} (51)	44.9 ^{c,d,e1,e2} (61)	14.0 ^{a,b} (12)	20.8 ^{ab} (25)	14.3 ^{a,b} (15)	18.8 ^{a,b} (21)
Non-binary definition	9.8 ^{e1} (12)	9.6 ^{e1} (13)	3.5 ^d (3)	17.5 ^{c,e1,e2} (21)	0 ^{a,b,d} (0)	2.7 ^d (3)
Other	4.1 (5)	3.7 (5)	3.5 (3)	8.3 (10)	5.7 (6)	4.5 (5)
No. of responses	100 (122)	100 (136)	100 (86)	100 (120)	100 (105)	100 (112)

Note: In condition (E), we probed for the understanding of registered sex as in (C) and gender identity as in (D). The results of these codings are presented separately in columns E₁ and E₂.
^{a,b,c/d,e1/e2} $p < .0045$; pairwise Pearson's χ^2 -tests between conditions with superscripts indicating significant differences using Bonferroni correction for multiple comparisons to the specified condition.

Table 4. Frequencies of Concrete and Vague Conceptualizations of Divers.

Concrete conceptualization	No. of responses	% of concrete categorization	Vague conceptualization	No. of responses	% of vague categorization
Intersexuality (only)	67	18.3	Both male and female	67	12.0
Transgender (only)	233	63.7	Between male and female	18	3.2
Sexual orientation/gender expression (only)	30	8.2	Neither male nor female	217	38.8
Intersexuality and Transgender	21	5.7	Unclear assignment	192	34.3
Intersexuality and Sexual orientation/gender expression	4	1.1	The third category	40	7.1
Transgender and Sexual orientation/gender expression	10	2.7	Additional category	66	11.8
Total	365	100.0	Non-binary category	8	1.4
			Neutral gender	8	1.4
			Explicit non-response option	4	0.7
			Total	620	100.0

Note: Multiple answers possible; for vague conceptualizations, $n = 620$ refers to the total number of answers.

however, vague conceptualizations were more common, making up almost half of the responses (49%). Only 4% of respondents voiced gender-binary beliefs. The amount of non-substantive responses was 15%. Question wording had no influence on the understanding of divers ($\chi^2 = 10.380$; $df = 12$; $p = 0.583$). However, respondents familiar with the term were more likely to offer both concrete and vague conceptualizations, and less likely to give non-substantive responses ($\chi^2 = 191.966$; $df = 3$; $p < 0.001$).

Table 4 shows all categories and their combinations for respondents who gave either a concrete or a vague conceptualization of divers. As the code for gender expression was only named by few respondents, the categories sexual orientation and gender expression were merged for analysis.

Within the group of concrete conceptualizations, the most common understanding of divers was not intersexuality, as intended by German legislation. Instead, divers was most commonly understood as a category to describe transgenderism. The most frequent overlaps were respondents naming both intersexuality and transgender (6%) in their understanding of divers and a combination of transgender and sexual orientation or gender expression (3%).

About half of the sample gave a vague conceptualization of the term divers. Within the group of vague conceptualizations there were two dominant entries. The first was an exclusion-based definition of divers as being “neither male nor female,” which was coded for 39%. The second common definition emphasized that divers indicates that sex/gender cannot be clearly or completely assigned to one of the other two categories. This accounted for 34%. The inclusion-based category both male and female and the definition of divers as an additional category were each coded for 12%. Divers as the third category, commonly used in media at the time of implementation, was coded for another 7%. All other codes accounted for less than 5%.

Discussion and Conclusion

One concern when it comes to including a non-binary response option for questions on sex/gender is a decrease in data quality. In our study, we encountered only few break-offs and few discrepant answers between the un-specific, binary question on sex/gender at the beginning of the survey and the later answer to non-binary question variations. This indicates that adding a non-binary sex category does not lead to respondents giving less reliable answers.

However, other findings point to irritation particularly among male respondents. The majority of non-substantive answers to probes and judgmental, gender-binary beliefs were made by male respondents, in line with previous research on higher aggressive heteronormativity in cisgender males (i.e., Nagoshi et al. 2008). Thus, we cannot rule out that including a third

category may lead to systematic bias on the part of cisgender males in larger samples, perhaps contributing to a higher share of extreme responding for attitude questions.

We varied question wording to investigate whether using different terms for sex or gender triggered different associations when responding. The wording had hardly any influence on whether respondents thought of their sex, gender identity, or both equally when answering. For most respondents, the underlying concepts were identical and they did not distinguish between them for themselves.

Regarding the understanding of divers, about half of our sample gave vague definitions that remained unclear whether they pertained to sex or gender. These definitions were mostly based on exclusion from or unclear assignment to a binary category. The most common concrete conceptualization was transgenderism and not intersexuality. It seems that respondents are more familiar with the concept of transgenderism than with intersexuality. There is no official statistic on the number of registrations of divers, but analyses by advocacy groups indicate that changes in registered sex carried out based on the new legislation for intersex people are mainly being used by transgender people (Queer 2019). Also, transgender people are fighting in court to open divers to this group (Dritte Option 2020). The German government is currently reforming legislation for transgender people (Tagesspiegel 2019). Thus, the association of divers as transgender may be in line with German legislation in the future. Few respondents associated divers with sexual orientation or gender expression. However, of those who did, many also mentioned transgenderism and/or intersexuality. Possibly, these respondents conflate different aspects of sexual identity.

The study has some limitations. First, the results are not representative for and cannot be generalized to the general population of Germany. Data collection was based on a non-probability sample using quotas, which does not ensure that each respondent has the same chance of being selected and does not allow to estimate sampling bias. Second, the sample probably consists mainly of cisgender respondents. Third, the experiment was placed toward the end of the survey, possibly explaining the low break-off rate (Rossmann et al. 2015).

There are also unsolved practical issues. Although it is relatively simple to add the category divers as a sex designation in self-administered surveys, introducing this category into face-to-face interviews might be more challenging. Most interviewer-administered surveys do not ask about respondents' sex/gender, but the interviewers themselves answer this question. Explicitly asking respondents on their sex/gender might make interviewers uncomfortable and for some respondents this might be irritating (Westbrook and Saperstein 2015). Hence, new strategies of asking this question in face-to-face and telephone interviews must be found.

Data management, data publication, and analysis also face challenges dealing with respondents who select the category divers. Due to de-anonymization concerns, these respondents may need to be deleted from the published data file or (randomly) allocated to a binary category. Data analysts must decide how to treat these probably small number of respondents when weighting data or in statistical analysis. Also, respondents might fear threats to their privacy or confidentiality when selecting a rarely selected response category. Finally, due to differences in legislation, the operationalization of non-binary categories poses challenges in cross-cultural research. In the United States, for instance, the category “X” includes transgender people but this does not exist in all states.

Developing survey questions that capture both the complexity of non-binary sex and gender and can easily be implemented in general population surveys remains challenging (Ryan 2019). Further research is needed to provide general recommendations for sociodemographic questions that allow surveying the whole diversity of respondent groups.

Acknowledgments

We are thankful to Friederike Quint for her help with coding the open-ended answers. We would also like to thank the editor and the three anonymous reviewers for their remarks, which helped improve this article.

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) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. It is beyond our work to evaluate the discussion on whether gender is socially constructed or natural. In line with Bauer et al., we see it as given that the two constructs of gender and sex are distinct (2017:1022).
2. According to legislation and in contrast to the definition of intersexuality given above, divers can only be selected by intersex people who cannot be medically assigned to one sex according to their physical (primary or secondary) sex

characteristics. This has been heavily criticized by many inter* and trans* associations (Federal Anti-Discrimination Agency 2019).

3. The rather fair value (Landis and Koch 1977) occurred for the code gender-binary beliefs, as some responses were ambiguous in tone. For the final coding, only responses that clearly expressed rejection were coded as gender-binary beliefs.

References

- American Association for Public Opinion Research (AAPOR). 2015. *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. 8th ed. Alexandria, VA: AAPOR.
- American Psychological Association (APA). 2009. *Task force on gender identity and gender variance*. Report of the task force on gender identity and gender variance. Washington, DC: American Psychological Association. <https://www.apa.org/pubs/info/reports/gender-identity> (accessed April 29, 2020).
- Bauer, G. R., J. Braimoh, A. I. Scheim, and C. Dharma. 2017. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PLoS One* 12:e0178043.
- Behr, D., K. Meitinger, M. Braun, and L. Kaczmirek. 2017. Web probing—Implementing probing techniques from cognitive interviewing in web surveys with the goal to assess the validity of survey questions. *GESIS—Survey Guidelines*. Mannheim: GESIS-Leibniz Institute for the Social Sciences.
- Bittner, A., and E. Goodyear-Grant. 2017. Sex isn't gender: Reforming concepts and measurements in the study of public opinion. *Political Behavior* 39:1019–41.
- Broussard, K. A., R. H. Warner, and A. R. Pope. 2018. Too many boxes, or not enough? Preferences for how we ask about gender in cisgender, LGB, and gender-diverse samples. *Sex Roles* 78:606–24.
- Debus, K., and V. Laumann, eds. 2018. *Pädagogik geschlechtlicher, amouröser und sexueller Vielfalt. Zwischen Sensibilisierung und Empowerment*. Berlin: Dissens-Institut für Bildung und Forschung. <https://interventionen.dissens.de/materialien/handreicherung> (accessed March 28, 2020).
- Döring, N. 2013. Zur Operationalisierung von Geschlecht im Fragebogen: Probleme und Lösungsansätze aus Sicht von Mess-, Umfrage-, Gender- und Queer-Theorie. *GENDER—Zeitschrift für Geschlecht, Kultur und Gesellschaft* 5:94–113.
- Dritte Option. 2020. *Pressemitteilung: Erfolg für 'Dritte Option' diesmal bereits in der 1. Instanz*. January 23. [dritte-option.de/pm-erfolg-vor-dem-amtsgericht-muenster/#more-1092](https://www.dritte-option.de/pm-erfolg-vor-dem-amtsgericht-muenster/#more-1092) (accessed February 25, 2020).
- Duden. 2020. *Divers*. <https://www.duden.de/rechtschreibung/divers> (accessed October 20, 2020).
- Federal Anti-Discrimination Agency. 2019. *Mann–Frau–Divers: Die 'Dritte Option' und das Allgemeine Gleichbehandlungsgesetz*. https://www.antidiskriminierungsstelle.de/DE/ThemenUndForschung/Geschlecht/Dritte_Option/Dritte_Option_node.html (accessed October 29, 2020).

- Fowler, F. J. 1995. *Improving survey questions: Design and evaluation. Applied Social research methods*. Vol. 38. Thousand Oaks, CA: Sage.
- Jourian, T. J. 2015. Evolving nature of sexual orientation and gender identity. In *Gender and sexual diversity in U.S. higher education: Contexts and opportunities for LGBTQ college students*, edited by D.-L. Stewart, K. A. Renn, and G. B. Brazelton, 11–23. New Directions for Student Services. San Francisco: Jossey-Bass.
- Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–74.
- Lenzner, T., and C. E. Neuert. 2017. Pretesting survey questions via web probing - Does it produce similar results to face-to-face cognitive interviewing? *Survey Practice* 10:1–11.
- Lindqvist, A., E. A. Bäck, H. Bäck, and M. G. June. 2018. Measuring gender in surveys. Social psychological perspectives. Paper presented at the Gender Diversity in Survey Research Workshop, June 11–12, Gothenburg, Sweden.
- Nagoshi, J. L., K. A. Adams, H. K. Terrell, E. D. Hill, S. Brzuzy, and C. T. Nagoshi. 2008. Gender differences in correlates of homophobia and transphobia. *Sex Roles* 59:521–31.
- Queer. 2019. *Erste Zahlen der Bundesregierung: Rund 250 trans Menschen beantragten Änderung des Geschlechtseintrags*. https://www.queer.de/detail.php?article_id=33473 (accessed February 25, 2020).
- Rossmann, J., J. E. Blumenstiel, and M. Steinbrecher. 2015. Why do respondents break off web surveys and does it matter? Results from four follow-up surveys. *International Journal of Public Opinion Research* 27:289–302.
- Ryan, J. M. 2019. The problematics of assessing trans identity in survey research: A modest proposal for improving question design. *Societies* 9:1–13.
- 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.
- Tagesspiegel. 2019. *Gesetz für Transsexuelle soll reformiert werden*. <https://www.tagesspiegel.de/gesellschaft/queerspiegel/geschlechtseintrag-gesetz-fuer-transsexuelle-soll-reformiert-werden/24322112.html> (accessed February 25, 2020).
- Tate, C. C., J. N. Ledbetter, and C. P. Youssef. 2013. A two-question method for assessing gender categories in the social and medical sciences. *Journal of Sex Research* 50:767–76.
- Westbrook, L., and A. Saperstein. 2015. New categories are not enough: Rethinking the measurement of sex and gender in social surveys. *Gender & Society* 29: 534–560.