NATIONAL CENTER FOR HEALTH STATISTICS

Collaborating Center for Questionnaire Design and Evaluation Research

November 2021

An Initial Cognitive Evaluation of a 2-step Gender Identity Measure

By: Kristen Miller, PhD, Stephanie Willson, PhD, Valerie Ryan, PhD

INTRODUCTION

This study, conducted by the Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER) at the National Center for Health Statistics, investigates the performance of a gender identity measure for federal surveys. Specifically, the study examines a 2-step measure, whereby respondents are first asked sex assigned at birth followed by a question on current gender identity. The primary focus of the study was two-fold: 1) to determine whether respondents perceived the two questions as asking about distinct constructs, as opposed to seeing them as repetitive, and 2) to identify the ways in which respondents defined or conceptualized those constructs when formulating answers. As such, the study sought to investigate construct validity as well as to provide insight into potential response error.

Suggested citation:
Miller, K., Willson, S., Ryan, V. (2021). An Initial Cognitive Evaluation of a 2-Step Gender Identity Measure. National Center for Health Statistics - CCQDER. Hyattsville, MD
Available from: https://wwwn.cdc.gov/qbank/report.aspx?1219

Despite the small sample size, a range of interpretations was identified—though not necessarily problematic. Many respondents saw clear, but varied distinctions between the two questions; a few respondents were either confused or did not perceive a difference. While no cases of response error were identified, some evidence suggests a possible theory for false positive responses among gender non-minority respondents who do not discern a difference between the two questions. Future cognitive interviewing studies with larger, more diverse samples may confirm and/or expand upon these interpretive findings. Quantitative studies, specifically methodological surveys with embedded construct and error probes, would provide more understanding about the extent to which identified interpretative and error patterns would occur in large-scale surveys.

BACKGROUND

To better understand disparities, the federal government requires improvement of data collection to adequately capture experiences of transgender and other gender minority persons. This interest is also compelled by the recognition of a growing transgender population and the fact that notions of sex and gender, conventionally understood and asked as s single question (what sex are you?), and increasingly conceptualized by younger generations as two distinct ideas.

Although there is little empirical research to support the design, recent efforts to incorporate a gender identity measure have adopted a '2-step approach,' one that distinguishes 'sex' from 'gender' by first asking sex assigned at birth followed by gender identity. While there is variation involving specific wording, ordering

[1] For questions related to this report, contact Kristen Miller, Ph.D. at ksmiller@cdc.gov. Also, see Q-Bank: Question Evaluation for Surveys (cdc.gov) and CCQDER - Collaborative Center for Questionnaire Design and Evaluation Research Homepage (cdc.gov).

of the two questions, and category labels, a commonly used measure (and the specific set examined in this study) is as follows:

1) What sex were you assigned at birth, on your original birth certificate?
   Male
   Female
2) How do you describe yourself? You can select all that apply.
   Male
   Female
   Transgender
   Something Else

With this approach, gender minority respondents are identified as those having discordant responses to the two questions (answering 'male' and 'female,' or 'female' and 'male') as well as those selecting 'transgender.' Currently, it is unknown whether the category 'something else' captures gender minorities, non-minorities, or some combination, and so this category cannot be collapsed into one or the other of those groups.[1] This design feature, alone, is a noted flaw, particularly since the percentage of those reporting 'something else' has shown to be relatively large compared to those identified as transgender.[2] If this group is sizeable, the exclusion of gender minority respondents selecting 'something else' would result in a meaningful underestimation of the population.

Another concern pertaining to the 2-step design is the potential for false-positive cases, specifically, those respondents who erroneously select discordant categories in the sex and gender questions. Because the percentage of gender minorities is very small, the effects of this type of misclassification are likely to be substantial, biasing estimated differences toward the null and erroneously diluting apparent disparities. There is limited study on false positive rates

using the 2-step approach. However, a Pew study found that 24–56% of individuals who selected opposing sex and gender categories were in error.[3]

While there has been speculation as to the cause of false positive cases,[4] there is little empirical research into the actual reason or, more generally, how these measures perform on federal surveys—measures that distinguish a difference between 'sex' and 'gender' as well as allow for multiple, non-binary gender response options. The purpose of this study is to afford some understanding of how this specific measure operates with an eye toward development of an optimal gender identity measure, one that produces the least error for population-based, largescale surveys.

METHODOLOGY

[1] While some surveys provide open-ended text boxes for respondents to provide their preferred label, this data is rarely analyzed and recoded for use.

[2] In week 34, 2021 of the Census PULSE, 1.16% selected 'none of these,' while just 0.3% selected 'transgender.' Those with discordant sex/gender categories (M/F and F/M) represented 0.2% of the sample.

[3] Amaya A, et al. Does Adding a Third Gender Response Option Really Improve Measurement? AAPOR annual conference presentation. May 2021.

[4] Smith, T. Transgender and Alternative Gender Measurement on the 2018 General Social Survey. GSS Methodology Report No. 129. NORC, University of Chicago. July 2019. [9] https://www.cdc.gov/nchs/nsfg/index.htm

The methodological design for this study was developed to identify constructs captured by the two questions as well as definite cases or potential causes of response error. In-depth, personal cognitive interviews were conducted in two stages. First, interviewers read both questions to respondents and recorded their answers. Second, interviewers asked retrospective follow-up questions to understand respondents' interpretations and their processes for formulating answers. While interviews were video-recorded, interviewers also noted rationales and any confusion or response error. This study took place within the context of a larger study to examine questions for the National Survey of Family Growth (NSFG)[9] and was not a targeted study of only gender identity questions. The gender identity questions, however, occurred first in the questionnaire and were not impacted by subsequent context effects. Interviews were a maximum of 60 minutes in length and took place virtually using the Zoom Internet meeting platform.

Sampling and Respondent Demographics: As a qualitative study, cognitive interviewing methodology employs a purposive non-random sample. Rather than aiming for statistical representation, individuals are chosen for characteristics relevant to the questions and topics under investigation. Because the original intent of this study was to examine more general NSFG questions, the sampling criteria was based on requirements for those questions. Had the initial and sole purpose of the study been to examine gender identity measures, the sample would have differed, particularly by increasing the number of gender minority respondents.

Table 1 summarizes the sample composition. A total of 31 English-speaking, adult male and female respondents were interviewed. Twenty-four respondents were gender non-minorities (their sex assigned at birth matched their gender identification); seven were gender minorities. Other areas of demographic diversity included age, race, and educational attainment.

Table 1: Sample Composition

	Total Number (n=31)
Age	
Under 30	12
30 – 39	8
40 and over	11
Education	
High School or less	10
2-year college degree	3
4-year college degree	13
Graduate degree	5
Race/Ethnicity	
Non-Hispanic	28
White	17
Black	9
Native Hawaiian/Pacific Islander	1
Multiracial	1
Hispanic	3

Gender Identity	
Cisgender	24
Female	12
Male	12
Non-cisgender	7

Analysis: Analysis of interview data included a multi-stage process similar to the constant comparative method first developed by Glaser and Strauss (1967)[5] and adapted to cognitive interviewing studies by Miller et al. (2014).[6] Data analysis was assisted by the use of Q-Notes, a CCQDER-developed software application specifically designed for cognitive interview studies.[7]

Analysis first occurred within each interview as respondents were asked to explain their answers, revealing how they made sense of and went about answering the questions. If not previously apparent, respondents were typically asked whether they saw the two questions as being similar or different, and in what ways they might have differed. Careful attention was

paid to ensure that respondents' explanations pertained to the interpretive processes within their question response experience and not hypothetical or academic speculation. Transcripts and detailed notes were then produced from the recorded interviews.

After interviews were conducted and transcripts compiled, comparisons across interviews were made to identify interpretive themes and patterns of potential error—the basis for investigating construct validity. Comparisons were then made across the sex and gender questions to discern the ways in which, and the extent to which, the two questions captured distinct constructs. Finally, comparisons were made across subgroups, particularly, across gender minority and non-minority respondents, to examine comparability and issues of measurement equity.

FINDINGS

Table 2 presents respondents' answers to both sex and gender questions. For the sex question, 19 respondents answered 'male' and 12 answered 'female.' For the gender question, 12 respondents answered 'male,' 13 answered 'female,' and 6 answered 'transgender.' Because the question allows for more than one answer, 3 respondents answered 'female' in addition to their 'transgender' response, and 2 answered 'something else' along with 'transgender.'

[5] Glaser, B. and Strauss, A. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Hawthorne, NY: Aldine de Gruyter.

[6] Miller, K., Willson, S., Chepp, V., & Padilla, J.-L. (2014) *Cognitive Interviewing Methodology: A Sociological Approach for Survey Question Evaluation.* Hoboken, NJ: Wiley.

[7] For information, see: https://www.cdc.gov/nchs/ccqder/products/qnotes.htm

Table 2: Responses to 'sex at birth' and 'current gender'

	[SEX] What sex were you assigned at birth, on your original birth certificate?		
[GENDER] How do you describe yourself? You can select all that apply.		Male	Female	
	Male	12	0	12
	Female	1	12	13
	Transgender	1	0	1
	Transgender and Female	3	0	3
	Transgender and Something Else	2	0	2
Total Number = 31		19	12	31

Through their responses, twenty-four respondents were identified as gender non-minority, that is, respondents' answers to both sex and gender questions matched. Seven respondents were identified as gender minority, with one identified through discordant sex and gender responses, and the others through their selection of the 'transgender' category. All but one of those respondents selecting 'transgender' picked more than one gender category—though that one respondent explained she would have chosen 'female' along with her initial 'transgender' response but missed the instruction to 'select all that apply.' Two respondents selected the category 'transgender' as well as 'something else.' In their specification of 'something else,' both chose to use the label 'non-binary.' No cases of definite error were identified, though a few respondents expressed confusion, and some evidence emerged suggesting a possible theory for the false positive responses previously demonstrated in survey data, which is discussed further toward the end of this report.

Seeing difference between sex and gender

A few respondents were unable to see a difference between the two questions, saying essentially that they were repetitive. When asked, for example, one respondent simply stated, "Same thing. They're just asking me who I am. Like, what are you?" Similarly, another respondent stated, "I guess I just never... I just don't see them as different. I don't know. I guess I've just always been female and that's just who I am." For these respondents, the questions were very basic and clear-cut; they knew little of a distinction between 'sex' and 'gender' so had little to offer. When asked to expand, for example, one respondent explained:

I'm not thinking about anything in particular…. I'm a stereotypical male. I'm sorry. I'm thinking that I have secondary sex characteristics. I'm not really thinking about [much] … or… that's how I know what I am.

The vast majority of respondents did, however, perceive the questions as asking about two distinct concepts, though varying in their ability to articulate the difference: some respondents were especially fluent, while others relayed more rudimentary explanations.

Respondents who identified as transgender were the most knowledgeable and articulate in describing the difference since the subject is relevant to their everyday life. For all seven respondents, their unique position not only afforded them the ability to describe the difference but also to comment on the ways in which the

constructs are even more multifaceted than what the questions portray. Some suggested adding an 'intersex' category for the sex assigned at birth question as well as 'transfeminine' and 'transmasculine' for the gender question. This is not to say, however, that the questions were seen as unsatisfactory because they were overly simplified. One transgender woman with suggestions, for example, explained that she appreciated the design of the gender question, particularly within the context of a survey questionnaire:

I wasn't going to make you write down some crazy other thing that's like 500 words. 'Well, let me tell you my whole life's story,' it's like, you're just trying to click a box. So, I clicked those two. [female and transgender].

Other than transgender respondents, those most able to articulate concise explanations described having a friend or acquaintance or, as in one case, a child, that identifies as transgender. To maintain these types of close relationships, an understanding of 'sex' and 'gender' as one-in-the-same becomes problematic. For these respondents, their explanation of the distinction came with self-reflection as to their own social location. For example:

One's what was assigned to you at birth, and one is as you grew up, you came into an understanding of what you are. It's not as simple as what the doctor sees in that first moment. In a biological and in a social sense that's the way that I relate and understand myself. Biological in that genitals, chromosomes, and secondary sex characteristics, like a beard. And I just think like societal roles—I mean, I consider myself a father because I have two kids. And that's a male-identified role.

Importantly, while these respondents understood a difference between 'sex' and 'gender,' they also understood themselves as being part of a majority group, with a couple referring to themselves as being 'cisgender.'

Those respondents who did not have a personal relationship but were aware of the phenomena, also tended to see a distinction, but were less articulate. Some described learning the distinction of 'sex' and 'gender' in school and attempted—with varying accuracy—to relay a textbook definition. Others described knowing only through the media or in passing conversations and, therefore, had very little to say. For example:

One is on your birth certificate—what you were assigned. And the second is more your preference, I guess. Some people change their genders and so on. For me, yes, [it was straightforward]. Maybe I'm old fashioned. Maybe my generation is a little different, who knows.

In a couple cases, respondents expressed a bit of confusion and concern over not quite fully understanding *the* difference, outside of just knowing that there *is* a difference. For example, one respondent expressed her concern:

I had to pause for a minute and think about what was being asked [laughs]. I am female, so... For starters, those questions are becoming a lot more prevalent today. And in fact, I had one of my [college] professors ask us our pronouns and stuff. So, like, a lot of times the questions, if they're worded strangely, I just get tripped up. And I don't identify as anything other than female. Like, my pronouns are what I would expect them to be. But I know some people are very otherwise. I always get tripped up by questions like those because I'm trying to—I want to make sure I'm not saying the wrong thing.

Given the structure of the interview, it is not clear whether these respondents with no direct connection to gender minorities would identify as 'cisgender,' or even see themselves as being part of a majority group. However, it was clear that they did not use this language when describing their understanding of the questions.

Regardless of the varying levels of awareness and ability to articulate difference, there was a clear sense among most respondents that notions of 'sex' and 'gender' are evolving. Many respondents remarked how "gender

is more fluid now," and that "people can change genders now." When explaining the difference between the two questions, another respondent described the ways in which she has witnessed changes in the 'sex/gender' construct, evolving from a predisposed fact to an idea of agency and self-determination:

Very different questions. One is kind of ingrained. You don't have a chance when you're born. As soon as you're born, they give you something and you run with it until you learn it's something that you can even question. At least at my age it wasn't even something we thought we could question. You just are this. And now the studies and people expressing themselves, he/him pronouns, and such. So, they're very different questions because you could choose to identify by either or as both. And that can be based on your sexual preference or your personal experiences. And it can be based on a sense of rebelliousness from a system that forced you to pick something or be something that may be limiting.

Constructs captured

Similar to the varied experiences and degrees of recognizing difference, there was a range in the ways respondents understood what the questions were asking, that is, the specific constructs that the questions captured. All respondents who saw the questions as being identical understood them to be asking about a biological phenomenon, though they typically blended this with social and personal characterizations of gender. For example, in explaining his answer to both questions, one respondent stated:

> *I am biologically male. I am genetically male. I look like a male. I feel like a male. I am everything male personified. So, I describe myself as a male.*

For this respondent, these various components, that is, biology, genetic makeup, others' perceptions, personal feelings and identity, are entirely interrelated and essentially the same. Conceptualizing 'sex/gender' as a single, homogeneous construct was not only not problematic, it made sense for him.

Indeed, this was a common theme for those respondents who saw the questions as being repetitive. For example, this respondent began her reasoning with a biological understanding to explain why she selected 'female' to the two questions:

> *What I can offer to the human race from a biological standpoint is female associated. So that's why I would say I identify as a female. I can't provide semen to the human race. I'll never be able to provide semen—well, I don't know—but from a biological standpoint, I'll never be able to contribute anything other than a certain set of things. Breastmilk, eggs. That's what I can contribute. From that perspective I consider myself female.*

As she continued her explanation, she linked her biological explanation to a more social understanding:

> *There are certain things that happen biologically, and I can share those common experiences with other people who were assigned female at birth. I see it similar to an [electrical] outlet. The female part is the wall, and the male part is the plug. Meaning that I feel like there's a yin to each yang. I feel like I fit into the description of the yang. I identify with the role. I kind of like the role that society has created for women. Yes, there's room for improvement... and more equality, but I identify with the role as well.*

She continued, making an explicit link between the biological and social components of being 'female:'

> *It's different from home and work. Because at work I'm more inclined to not stand out as a female—unless it's a topic of diversity and inclusion. Outside of that I don't want people to stop and say 'this is a female.' [But] at home I like the idea that, as a female, it's important to keep a clean house, to make sure that there's food available, to bring sensitivity to situations, to react differently than what is expected of a male. To bring a different perspective and sensitivity to a situation. And providing compassion. Which looks different for male and female. Men are more black-and-white, and women can see all the color in between, associating with our emotions, connecting to our emotions. I don't know if it's because we have PMS or things like that send us into spaces where we have to learn to manage our emotions more than a male would have to worry about.*

Other than these few respondents who hold notions of 'sex/gender' as a single construct, most did see a difference, and did not conflate the two ideas. Many, including some transgender respondents, saw the primary difference as 'sex' being biological and 'gender' being social. For example, explaining why she answered 'male' to the sex question, she said, "The first one's pretty standard. I mean, it's just true logically—biologically." When answering the gender question, she chose 'female' as opposed to 'transgender' because, "I am transgender by the definition, but I don't identify as that."

A corollary to understanding 'sex' as a biological phenomenon is to also see it as being fixed, categorical and lacking agency or self-determination. For example, one respondent described these multifaceted differences:

> *Sex is what you're assigned at birth. Sex is binary, and gender is fluid. So, I think your gender can change as you grow into yourself. But when you're born you have no say. They stamp you with something.*

Gender, on the other hand, as a social construct, is seen as a non-categorical, fluid form of selfexpression. As one respondent explained: "I know there's a spectrum of gender and only a couple of sexes."

A few respondents went beyond thinking of 'sex' as a fixed and concrete concept, noting that people "can get a sex change operation" and "some people who are intersex… have biological parts of both male and female." And, as this respondent notes, for intersex babies, "You had doctors actually assign the person a gender. So that was all going into my head." In a similar view, another respondent explained:

> *Assigned at birth is clearly based on your parents and doctor, religious norms. I'm sure the country you're born in. Whereas the second question, how do you see yourself, is really up to your own experiences in life.*

From this perspective, both 'sex' and 'gender' are seen as social constructs, though the first is based on others, that is, a social identity, and the second, a personal identity.[8]

Again, because of their unique position and multifaceted understandings of 'sex' and 'gender,' transgender respondents were able to provide insight into question design features that would allow for this complexity. For example:

> *I thought of my birth certificate. Which has a sex label on it. Although, you can get those amended. So, I think that you used the word 'original' is probably pretty important there.*

And, because of the evolving belief in agency and self-determination, another respondent suggested:

[8] Goffman, E. 1959. *The Presentation of Self in Everyday Life*. Garden City, N.Y.: Doubleday.

> *Well, the question implies—and there's a lot of truth in it—that a lot of trans people don't necessarily want to talk about or admit their previous gender identities, and this is one way to clarify their assigned sex. At birth. Their 'true sex' quote, unquote.*

Finally, it is important to note that, while almost all of these respondents had relatively clear— although varied notions—there were a few who were confused and had difficulty selecting a response. Some of this confusion, as described earlier, pertained to respondents' lack of certainty about the full distinction between 'sex' and 'gender.' However, in a couple cases, the confusion related to respondents conflating sexual identity with notions of 'sex' and 'gender.' For example, one respondent explained the trouble:

> *When I'm doing a survey or when I'm at a doctor's office or filling out paperwork… when it comes to personal conversations, I know how to answer gender. 'What is your gender?' My gender is male. But if somebody asks me what my sex is, man that's a whole rabbit hole of a conversation. Which I subscribe to being heterosexual, straight, however you want to call it.*

Although not possible to conclude, this confusion may point to a potential cause of false positive responses, particularly for those not seeing a difference between 'sex' and 'gender.' Since they may believe that they have already answered the question, they might assume that the second question is asking about 'sexuality', thus, indicating the 'sex' of the partner with whom they would choose a sexual relationship.

DISCUSSION

Several implications pertaining to the design of an optimal gender identity measure can be drawn from this study. Specifically, the study helps to identify particular criteria for an ideal question as well as points to design elements that could be used in the specific wording of such a question.

Most apparent, the optimal question is one that appropriately classifies all respondents, while producing the least error. When forced to choose between error types, the decision should be based on the purpose of data collection: False negative error is preferred over false positive error, when the purpose is to characterize the population. False positive error is preferred when the purpose is to estimate size. Regardless, understanding and then documenting that error (the type as well as the extent it exists) is necessary for appropriate interpretation of resulting gender identity data.

Additionally, to reach the optimal design, the question must make sense—and be consistent with—a range of conceptualizations pertaining to 'sex' and 'gender' (and likely 'sexuality'). This study alone, with its small sample, illustrates a broad range in understandings. A larger study with a more diverse population, including a wide range of age, would help to understand how these constructs play out on a national scale.

Furthermore, the optimal design must incorporate as much detail as possible to characterize the complexity of 'sex' and 'gender' without causing false positive error. It appears, for example, that 'sex assigned at birth' and the 'select all' response option allow for a layer of complexity that do not generate false positives. This study also suggests that distinguishing the two constructs with separate questions (the 2-step approach) allows for a multifaceted characterization. However, more research would help to determine wording or question design strategies to limit the relatively large amount of false positive error that was described in the introduction of this report. Some surveys, including the Census PULSE, for example, use a follow-up confirmation question after the two questions, asking respondents to confirm their answers to both the sex and gender questions. Specifically, it asks:

> Just to confirm, you were assigned {FILL} at birth and now you describe yourself as {FILL}. Is that correct?

The addition of this third question increases the chances that the misclassification will be identified and immediately corrected by the respondent. Reliance on this question, however, is not optimal since surveys administrators may not include this additional question because they do not appreciate its importance and/or do not want to add respondent burden. It also assumes that respondents will recognize their mistake and make the correction. It is more advantageous to develop a measure that does not produce this type of error in the first place.

In terms of response categories, the ideal design would use categories that reflect the ways in which people describe their own selves. The term 'cisgender,' for example, could be seen as a viable response category unless non-minority respondents do not see themselves as being part of a majority group using that particular identity. This was a similar problem in the development of the National Health Interview Survey (NHIS) sexual identity question whereby sexual non-minorities did not see themselves as being part of a majority group and did not necessarily identify as heterosexual.[9] In fact, a meaningful number of these respondents were unfamiliar with the term 'heterosexual,' which caused respondents to answer 'none of these' or 'bisexual.' The question was, therefore, revised to replace the word 'heterosexual' with 'not gay'—the actual way in which these respondents understood themselves in terms of sexual identity. This wording change almost entirely eliminated the false positive error. For this study, interviewers did not specifically ask respondents about the cisgender label, and so it is not possible to address this issue here; it will be an important issue to examine in future studies. Likely for gender, most non-minority people simply identify as being 'woman' or 'man.'

Finally, there are several methodological takeaways from this study. Achieving the optimal question requires interpretive study as well as a quantification of those interpretative patterns. Cognitive interview, focus group and pile sort studies provide insight into the breadth of understandings related to 'sex' and 'gender' as well as the various ways in which the US population categorizes themselves. A comprehensive study, particularly across age ranges, would illustrate the degree and rate to which the constructs have evolved, providing insight into future development. Quantitative studies, specifically methodological surveys with embedded construct and error probes, would provide more understanding of the extent to which identified interpretative and error patterns would occur in actual surveys. By conducting split-sample experiments, it will be possible to identify specific wording options with the least error. Thus, a mixed-method, data-driven approach to question design would be useful in determining the optimal gender identity measure for population-based surveys.

[9] See Miller, K. and Ryan, M. 2011. Design, Development and Testing of the NHIS Sexual Identity Question. https://wwwn.cdc.gov/qbank/Report.aspx?1087.