



National Center for Science and
Engineering Statistics

Title: 2024 Survey of Earned Doctorates SOGI Data Collection Experiment Summary

Date: May 2024

Contract Awardee: RTI International
Contract Number: GS-00F-354CA 49100422F0033

Disclaimer: The National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) supports research opportunities to inform its strategic objectives. This report summarizes research conducted under contract with NCSES and is being shared to inform interested parties of ongoing activities and to encourage further discussion. Any opinions, findings, conclusions, or recommendations expressed in this report do not necessarily reflect the views of NCSES or NSF. Please send questions to ncsesweb@nsf.gov.

May 2024

2024 Survey of Earned Doctorates Sexual Orientation and Gender Identity Data Collection Experiment

Preliminary Findings Summary Report

Prepared for

National Center for Science and Engineering Statistics

2415 Eisenhower Avenue, Suite W14200

Alexandria, VA 22314

Phone: (703) 292-8780

Prepared by

Y. Patrick Hsieh, Laura Burns Fritch, Peter Einaudi

RTI International

3040 E. Cornwallis Road

Research Triangle Park, NC 27709

RTI Project Number 0218508

Abstract

The National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) has been evaluating sexual orientation and gender identity (SOGI) questions for possible inclusion on its surveys, including the Survey of Earned Doctorate (SED). This report summarizes the findings of a research experiment in the 2024 SED data collection to compare several different versions of SOGI questions. The results of this experiment combine with the qualitative findings from previous SED SOGI research to inform the feasibility of asking SED respondents to report both their gender identity and sexual orientation for the future data cycles beginning with the 2025 SED. Based on these results, considerations for the 2025 SED data collection cycle include adding separate questions for sexual orientation and for gender identity with detailed sets of response options and open-ended write-ins as well as assuring respondents that the individual-level data will not be shared with their doctorate institutions to address concerns about confidentiality and privacy raised by SOGI minorities.

Background

The National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) has been evaluating sexual orientation and gender identity (SOGI) questions for possible inclusion on its surveys, including the Survey of Earned Doctorate (SED), in response to Executive Order 13985—which seeks to expand the data available to the federal government to measure equity and capture the diversity of the American people—and, more recently, Executive Order 14075—which aims to advance equality for lesbian, gay, bisexual, transgender, queer, and intersex (LGBTQI+) individuals.¹

As part of this larger effort, in summer 2022, NCSES conducted cognitive interviews with a convenience sample of 61 research doctoral students representing diverse sexual orientations, gender identities, and citizenship status. The results of this work can be found in attachment 9 of the 2024–25 SED Information Collection Request, *Survey of Earned Doctorates Recent Methodological Research* (https://www.reginfo.gov/public/do/PRAViewDocument?ref_nbr=202305-3145-005).

Following the qualitative cognitive interviews, NCSES is now conducting a research experiment in the 2024 SED data collection to compare several different versions of SOGI questions. A module of the SOGI experimental questions was embedded in the demographic information section at the end of the 2024 SED Web instrument.

Based on the best practices for the SOGI data collection recommended by the Office of Management and Budget (OMB) and the 2022 cognitive interview participants' preferences, the 2024 SED SOGI experiment includes eight test questions in the module:² (1) a sex assigned at birth (SAB) question, (2) a combined SOGI question that asks if the respondent considers themselves to be LGBTQ+, (3) three gender identity (GI) questions, and (4) three sexual orientation (SO) questions. The experiment uses compound random assignment to allocate respondents across the 10 experimental paths described in Table 1. Respondents in all 10 paths encounter the SAB question first, followed by either the combined SOGI question (Figure 1) or one of the three GI test questions on the same Web screen (Figure 2). For those in paths 2–10, the SO test questions were presented next on the following screen (Figure 3). For the rest of this document, the questions are referenced according to the question name noted in Figures 1–3.

Table 1.
2024 SED SOGI experiment design

Concept	Path 1	Paths 2–10
Sex assigned at birth (SAB)	SAB	
Gender identity	SOGI1 (25%)	G1 (25%) G2 (25%) G3 (25%)
Sexual orientation		SO1 (25%) SO2 (25%) SO3 (25%)
Sample distribution	25%	75%

SED = Survey of Earned Doctorates; SOGI = sexual orientation and gender identity.

Figure 1.
Survey screen showing the sex assigned at birth and combined SOGI questions

Sex Assigned at Birth

(SAB) What sex were you assigned at birth, on your original birth certificate?

Male
 Female

Combined SOGI (1 Test Question)

(SOGI1) Do you consider yourself to be LGBTQ+?

Yes
 No
 I prefer not to answer

LGBTQI+ = lesbian, gay, bisexual, transgender, queer, and intersex; SAB = sex assigned at birth; SOGI = sexual orientation and gender identity.

Figure 2.
Survey screen showing the sex assigned at birth and gender identity (GI) test questions

Sex Assigned at Birth

(SAB) What sex were you assigned at birth, on your original birth certificate?

Male
 Female

Gender Identity (1 of 3 Test Questions)

(G1) How do you currently describe yourself?
Select all that apply.

Male
 Female
 Transgender
 I use a different term: _____

(G2) Do you consider yourself to be a gender minority (e.g., transgender, non-binary)?

Yes
 No
 I prefer not to answer

(G3) How do you currently describe yourself?
Select all that apply.

Man
 Woman
 Transgender
 Non-binary, Gender non-conforming, Genderfluid, Genderqueer
 I use a different term: _____
 I prefer not to answer

SAB = sex assigned at birth.

Figure 3.
Survey screen showing the Sexual Orientation (SO) test questions

Sexual Orientation (1 of 3 Test Questions)		
<p>(SO1) How you think of yourself?</p> <p><input type="radio"/> Straight or heterosexual</p> <p><input type="radio"/> Gay, lesbian, bisexual, queer or another orientation</p> <p><input type="radio"/> I prefer not to answer</p>	<p>(SO2) Do you consider yourself to be a sexual minority (e.g., gay, lesbian, bisexual)?</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> I prefer not to answer</p>	<p>(SO3) Which of the following best represents how you think of yourself?</p> <p><input type="radio"/> Straight or heterosexual</p> <p><input type="radio"/> Gay or lesbian</p> <p><input type="radio"/> Bisexual</p> <p><input type="radio"/> I use a different term: _____</p> <p><input type="radio"/> I prefer not to answer</p>

From the time the 2024 SED data collection began on 31 May 2023 to 31 March 2024,³ a total of 30,745 respondents had completed the Web survey. For the preliminary SOGI experiment data analysis, the question performance was measured by item nonresponse rates, breakoff rates, construct validity of the estimates of SOGI minorities, and respondents’ ratings of the ease of understanding and answering the SOGI test questions. Additional evaluation metrics included respondents’ concerns about confidentiality and privacy of their SOGI data and their perceived representation of SOGI response options because such concerns or perceptions may be associated with respondents’ willingness to answer SOGI questions outside of the experimental research context. (See the appendix for the details of these evaluation questions.)

Methods

The following criteria were used to assess the preliminary SOGI data:

- Item nonresponse rates of all test questions in the SOGI experiment were compared with the historical item nonresponse rates of SED questions and were considered acceptable if they were under 10%, a benchmark for the item nonresponse rates for other similarly sensitive questions, such as questions about functional limitations. When examining the item nonresponse rates for the sets of GI or SO test questions, the item nonresponse rates were compared only among the questions measuring the same construct.
- Breakoff rates for the SOGI module were assessed by counting the number of respondent breakoffs occurring on a given SOGI question screen divided by the total number of breakoffs in the survey. The breakoff rates in the SOGI module were then compared against the average breakoff rate across all survey screens (1.7%, excluding screens with zero breakoffs from the calculation). Breakoff rates in the SOGI module that were less than the average breakoff rate were classified as acceptable.
- Construct validity was assessed by comparing responses to different test questions and assessing if the percentage of respondents identified as SOGI minorities was similar, meaning that the percentage difference between estimates was less than 20%; for example: $\text{absolute value}(\text{estimate1} - \text{estimate2}) / \text{estimate2}$. The next section summarizes the key findings of this preliminary data analysis and recommendations.

Key Findings

SOGI test questions in the SED yielded acceptable item nonresponse and survey breakoff rates while achieving construct validity. Item nonresponse rates presented here were calculated based on the sum of survey respondents who did not answer and those who selected the “I prefer not to answer.” The item nonresponse rates for the gender identity questions were 1.8% for G1, 3.0% for G3, and 4.6% for G2. For the sexual orientation questions, item nonresponse rates were 5.6% for SO2, 6.7% for SO1, and 6.9% for SO3. The item nonresponse rate for SOGI1 was 6.5%. Between 2020 and 2022, the questions about respondents’ functional limitations and their parents’ education have exhibited approximately 10% and 13% item nonresponse rates,

respectively, while the item nonresponse rate for other questions—such as race/ethnicity, citizenship, place of birth, and month and year of birth—range between 5% and 7%.⁴ Therefore, the SOGI test questions yielded comparable or better data quality than these questions as measured by item nonresponse. Among the respondents who started the 2024 SED Web instrument, 667 respondents had broken off without finishing the survey as of 31 March 2024. Of these, 10 respondents broke off on a screen requesting SOGI information; 8 respondents broke off on the first screen (i.e., Figure 1 or Figure 2), and 2 respondents broke off on the screen that ask for confirmation of their gender identity responses. These 10 breakoffs accounted for approximately 1.5% of all survey breakoffs, which is slightly lower than 1.7%, the average breakoff rate across all survey screens with breakoffs.

The percentage of respondents identifying themselves as SOGI minorities was similar across the SOGI test questions, suggesting construct validity among the presented question wording options. When analysis focuses on reporting of gender identity, the percentage of respondents who identified as gender minorities ranged between 2.5% (G1) and 2.9% (G2), with a 15.3 percent difference between the highest and lowest estimates. Similarly, the percentage of respondents who identified as sexual minorities ranged from 13.0% (SO1) to 14.7% (SO3), with a 13.4 percent difference between the highest and lowest estimates. The percentage of respondents who identified as LGBTQ+ in SOGI1 (path 1) (13.4%) fell within the range of 13.1% and 16.4% of respondents identified as SOGI minorities from the GI and SO test questions (paths 2–10).⁵

Taken together, despite the difference in the test question design and the specificity of the response options, the SOGI test questions exhibited acceptable item nonresponse rates and a similar ability to collect and identify SOGI minority status from respondents.

The majority of SOGI minority and nonminority respondents indicated that reporting their SOGI information in a federal government survey was comfortable and easy for them to do. Across all seven SOGI test questions, between 77.5% (SOGI1) and 89.2% (G3) of SOGI nonminorities were comfortable providing their SOGI information. The majority of SOGI minorities, to a lesser extent, were also comfortable providing SOGI information in response to six of the seven test questions (G2, 62.7%; G3, 58.3%; SO1, 69.3%; SO2, 71.3%; SO3, 72.2%; and SOGI1, 60.4%), with the exception of G1 (45.3%). Similarly, over 90% of SOGI nonminorities rated the SOGI test questions as being easy to answer, while more than two-thirds of SOGI minorities gave the same rating for six of the seven test questions (G2, 74.2%; G3, 72.4%; SO1, 83.5%; SO2, 82.9%; SO3, 82.5%; and SOGI1, 73.3%), with the exception of G1 (50.5%).

Among those 1.8%–6.9% of respondents who chose not to answer the SOGI questions, between 12.0% (G1) and 27.6% (SO2) were somewhat or very comfortable providing their SOGI info, and between 38.1% (G2) and 46.8% (SO3) rated the SOGI questions very or somewhat easy to answer. Taken together, with the exception of those who answered the GI test question G1, the majority of both SOGI minorities and nonminorities were comfortable with answering the SOGI questions and said they were easy to answer in a federal government survey.

The proportion of respondents expressing concerns about privacy and confidentiality varied by SOGI status. SOGI minorities were about twice as likely as SOGI nonminorities to be concerned about how their SOGI data would be used (from 39.7% [SO2] to 61.0% [G1]). Respondents who declined to provide their SOGI information also exhibited similar levels of concerns (from 46.4% [SO2] to 63.2% [G1]) as their SOGI minority counterparts. SOGI nonminorities were least concerned about the use of their SOGI information (from 18.8% [SO1] to 25.2% [SOGI1]). With regard to privacy concerns, about a third of SOGI minorities (from 31.1% [G3] to 37.7% [SOGI1]) and between half and two-thirds of respondents who did not provide their SOGI status (from 52.6% [G1] to 67.8% [SO1]) considered SOGI information to be their own private information. However, less than 25% of SOGI nonminorities (from 14.6% [G3] to 23.4% [SOGI1]) exhibited the privacy concerns. More than three-fourths of

SOGI minorities (from 75.4% [SO2] to 92.0% [G1]) indicated they had concerns about sharing SOGI information, whereas about half of nonminorities did so (from 45.1% [SO1] to 54.8% [SOG1]).

Besides the overall patterns of how respondents answered and perceived the SOGI test questions and how such response patterns varied by each respondent's SOGI status, several notable findings regarding data quality of specific test questions also emerged from the data analysis.

- **Among SOGI minority and nonminority respondents, SOGI1 yielded higher item nonresponse and greater privacy concerns than other SOGI test questions.** As noted earlier, SOGI1 is one of the two test questions with higher item nonresponse, at approximately 6.5%, than other test questions. Among SOGI minority respondents, the proportion of respondents who considered SOGI status as their own private information (37.7%) was the highest for those who answered the combined question SOGI1 in path 1 as compared to their counterparts answering different combinations of SOGI test questions in other paths (from 29.1% [G2] to 32.9% [SO1]). As one of the higher percentages, the 83.8% of SOGI minorities who answered SOGI1 had concerns sharing their SOGI information, and another 49.5% were concerned with how SOGI data may be used. Although not the lowest percentage, 60.5% of SOGI minorities who answered SOGI1 felt somewhat or very comfortable providing SOGI information to the federal government.

Similarly, among the nonminority respondents, the proportion of respondents who considered SOGI status as their own private information (23.4%) and the proportion of those who were concerned about sharing their SOGI information (54.8%) and how SOGI information may be used (25.2%) all were the highest for those who answered SOGI1. Furthermore, the proportion of nonminority respondents feeling very or somewhat comfortable providing SOGI information to the federal government was the lowest for those who answered SOGI1 (77.5%).

- **G2 yielded higher item nonresponse and request for clarification than other GI test questions.** Among the three GI test questions, G2 had the highest item nonresponse rate, at 4.6%, with G1 and G3 at 1.8% and 3.0%, respectively. The question also seems to invoke some comprehension issues, given that some respondents provided comments in the feedback section at the end of the survey instrument stating that they interpreted "gender minority" to include cisgender women.

Among gender nonminority respondents, the proportions of respondents who expressed concerns about G2 were the highest for more than five tested dimensions. For example, the percentage of gender nonminority respondents who did not understand G2 (2.3%) was about twice the percentage of those who did not understand G1 (1.1%) or G3 (1.2%). Those who were concerned about how GI information might be used were highest for the respondents who answered G2 (22.9%) when compared with those who answered G1 (21.5%) or G3 (20.7%). Similarly, the proportion of gender nonminority respondents who felt somewhat or very comfortable providing their SOGI information in a federal government survey was the lowest for those who answered G2 (85.4%) when compared with those answering G1 (89.0%) or G3 (89.2%).

- **Among gender minorities, G1 appeared to be more difficult to understand and drew more privacy concerns than other GI test questions.** The proportion of gender minorities who considered G1 somewhat or very difficult to answer was the highest (49.5%) when compared with their counterparts who answered G2 or G3 (25.8% or 27.6%, respectively). The proportions of gender minorities who were concerned about sharing their GI information and of those concerned about how GI information might be used, were both the highest for the gender minorities who answered G1 (92.0% and 61.0%, respectively) when compared with those answering G2 (79.1% and 46.6%, respectively) or G3 (86.4% and 58.3%, respectively). In contrast, the proportion of gender minorities who feel somewhat or very comfortable providing GI information in a federal government survey was the lowest for those who answered G1 (45.3%) when compared with those who answered G2 (62.7%) and G3 (58.3%). The proportion of gender minorities indicating that the response options did reflect how they think of themselves is also higher for those who answered G1 (39.0%) than those who answered G3 (8.25%).⁶

- **The percentages of sexual minorities were generally similar across the three SO test questions; however, due to its more detailed response options, SO3 provides the highest potential analytic utility.** As noted earlier, all three SO test questions yielded similar level of estimates of sexual minorities. Among the sexual minorities, the proportion who indicated that the response options did not reflect how they think of themselves ranged from 11.6% for respondents who answered SO3 to 6.1% for those who answered SO1. However, among respondents who chose not to answer the SO test questions, the opposite pattern appeared: 18.3% indicated that SO1 did not reflect how they think of themselves, compared with 9.2% for SO3. This contrast in respondents' engagement with SO1 and SO3 may be associated with the inclusion of the term "queer" in the response options of SO1 and its exclusion in SO3; sexual minorities who consider themselves queer would have seen themselves in SO1 but not in SO3. This conclusion is supported by the fact that "queer" is the most common term provided in the open-ended "I use a different term" text for SO3. SO3 also provides more clearly defined analytic subgroups and an open-ended fill-in option that could be used to improve the response options over time as the SOGI terminologies continue to evolve, such as exploring the addition of "queer" as a response option in future iterations.
- **The majority of SOGI minorities were comfortable sharing their SOGI responses to NSF and their doctorate institutions anonymously, but only about half were comfortable sharing these data when associated with their names.** About 80% of SOGI minorities indicated that they were comfortable sharing their SOGI answers anonymously with NSF (82.0%) and their doctorate institutions (81.1%), but only about half were comfortable sharing their information associated with their names (49.1% for NSF and 52.5% for academic institutions).

Considerations for the 2025 SED

Combined with the qualitative findings from the previous SED SOGI cognitive interviews and the quantitative preliminary findings from the SOGI experiment in the 2024 SED, asking SED respondents to report both their gender identity and sexual orientation is viable for the future data cycles beginning with the 2025 SED.

Among the three GI questions tested in the experiment, G3 resulted in the least issues across the tested dimensions of item nonresponse rates, breakoff rates, ease of answering, and construct validity; also, it raised the least concerns about confidentiality and privacy among gender minorities. When the results are taken altogether, G3 yielded the highest data quality. This question also has more inclusive response options than G1 and G2, including "Non-binary, Gender Non-conforming, Gender fluid, or Genderqueer" as a response option. While the results obtained by the three SO test questions may be at a similar level, SO3 (with its greater number of response options) may be preferable and consistent with the SO question example provided in the OMB best practices. The design of G3 and SO3 questions, with more detailed sets of response options with an open-ended write-in, should be considered for the 2025 SED data collection cycle. Furthermore, given that "queer" is the most frequently reported open-ended response to SO3 in the experiment, an addition of the "queer" option to SO3, with the necessary format change to allow selecting multiple responses may also be considered to improve the SO question design in the future.

Given the concerns expressed by many SOGI minorities about privacy and confidentiality of SOGI data and the potential increase in the item nonresponse rates on the SOGI data collection outside of an experimental condition, NCSES may consider informing 2025 SED respondents that their individual responses to the SOGI questions will not be shared with their institutions and will only be reported in aggregated form to help mitigate the risk of higher item nonresponse.

Appendix: Survey of Earned Doctorates (SED) Sexual Orientation and Gender identity (SOGI) Experiment Evaluation Questions Analyzed in the Report

1. Assessment of ease of understanding and answering sexual orientation and gender identity (SOGI) questions

How easy or difficult was it for you to report your [FILL-1] information in this question?

- Very comfortable
- Somewhat comfortable
- Somewhat uncomfortable
- Very uncomfortable

[FILL-1 for G1, G2, G3]: gender identity

[FILL-1 for SO1, SO2, SO3]: sexual orientation

[FILL-1 for SOGI1]: sexual and gender identity

2. Assessment of respondents' concerns about privacy, confidentiality, and perceived representation of their SOGI information

Which of the following statements are true for you when thinking about this question?

Select all that apply.

- This information is important for ensuring equity for [FILL-2] minorities
- I consider my [FILL-3] to be my own private information
- I have questions about how this information would be useful to the SED
- I have concerns about how this information may be used
- The response options do not reflect how I think of myself (only for G1/G3/SO1/SO3)
- I did not understand the question
- I have other concerns. Please specify: _____
- I do not have any concerns about sharing this information

[FILL-2 for G1, G2, G3]: gender

[FILL-2 for SO1, SO2, SO3]: sexual

[FILL-2 for SOGI1]: sexual and gender

[FILL-3 for G1, G2, G3]: gender identity

[FILL-3 for SO1, SO2, SO3]: sexual orientation

[FILL-3 for SOGI1]: sexual and gender identity

Notes

¹ For Executive Order 13985, see <https://www.federalregister.gov/documents/2021/01/25/2021-01753/advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government>; for Executive Order 14075, see <https://www.federalregister.gov/documents/2022/06/21/2022-13391/advancing-equality-for-lesbian-gay-bisexual-transgender-queer-and-intersex-individuals>.

² See <https://www.whitehouse.gov/wp-content/uploads/2023/01/SOGI-Best-Practices.pdf>.

³ A nonresponse bias analysis comparing early and late responders in previous SED data collections indicates that in each of the past five survey cycles, a set of key characteristics of early respondents through 31 March was not significantly different from all respondents in the survey cycle. Based on this finding and on the stability of the preliminary 2024 SED SOGI experimental findings examined monthly, the use of survey respondents from 31 May 2023 through 31 March 2024

does not appear to produce any bias in determining whether to collect the SOGI data in the 2025 SED survey cycle before the full results are available after the data collection cycle ends. The 2024 SED survey data collection cycle is from June 2023 through December 2024; the 2025 cycle is from June 2024 through December 2025.

⁴ See <https://nces.nsf.gov/pubs/nsf24300/technical-notes#technical-tables>.

⁵ For paths 2–10, respondents who identified as a gender minority and a sexual minority were counted only once for the purposes of calculating the percentage who identified as a SOGI minority.

⁶ G2 presented “Yes”/“No”/“I prefer not to answer” options to the respondents. Therefore, respondents who answered G2 were not asked to assess whether the response options reflect how they think of themselves.