

March 23, 2023

OMB Desk Officer for National Science Foundation
Office of Information and Regulatory Affairs
White House Office of Management and Budget
725 17th Street, NW
Washington, DC 20503

Suzanne H. Plimpton
Reports Clearance Officer
National Science Foundation
2415 Eisenhower Ave.
Alexandria, VA 22314

CC: Charles Barber, Chief Diversity and Inclusion Officer, National Science Foundation
The Honorable Tammy Baldwin, U.S. Senate
The Honorable Dianne Feinstein, U.S. Senate

Via reginfo.gov and email

RE: National Science Foundation; Notice of Submission for OMB Review; 2024-2025 Survey of Earned Doctorates (**Federal Register Doc. 2023-05642**)

Dear OMB and NSF Officials:

Thank you for the opportunity to comment on the National Science Foundation's (NSF) information collection request related to the 2024 and 2025 Survey of Earned Doctorates (SED). See 88 Fed. Reg. 16672 (March 20, 2023). My comments are concerned with the 2024 SED's Sexual Orientation and Gender Identity (SOGI) Experiment Plan ([Attachment 10](#)). I appreciate that NSF has provided sufficient details regarding its proposed SOGI pilot methodology prior to OMB approval, as advocated in the March 3, 2023 [letter](#) to the NSF Director and as is intended by [5 CFR 1320.8\(d\)](#). It is hoped that this improved public engagement can continue with all of NSF's future SOGI data efforts.

I applaud NSF's National Center for Science & Engineering Statistics (NCSES) for taking this crucial and long-awaited step toward implementing SOGI questions on its national surveys of the scientific workforce. As you know, SOGI data are critical to addressing the alarming [disparities](#) LGBTQI+ people are facing in STEM, including increasingly documented [career barriers](#), [underrepresentation](#), [retention failure](#), and [some of the highest levels of harassment in STEM](#). As our January 13, 2023 [letter](#) to the NSF Director pointed out, which was affirmed by a [subsequent letter](#) sent from Senators Tammy Baldwin, Dianne Feinstein, and 16 other U.S. senators, SOGI data in NCSES surveys are important for implementing federal policy, such as diversity-related funding provided by the CHIPS and Science Act, and for fulfilling the directives of [Executive Order 14075](#) to advance the use of SOGI data to address LGBTQI+ disparities, and [Executive Order 13985](#) to improve the equity of underserved communities.

The proposed large-scale, comprehensive pilot involving 20 unique combinations of SOGI questions across the full SED population of new Ph.D. recipients in the U.S. for the 2023-2024 academic year ($n \approx 57,000$) is a welcome next step. However, there are significant methodological concerns that NCSES should address prior to launching the SED. Below I offer suggestions both to improve the quality of the SOGI data that NCSES will collect and the inclusiveness of NCSES' approach.

A More Detailed Sexual Orientation Item Should Be Tested for Improved Data Disaggregation; It Should Replace the Flawed Item Referencing “Sexual Experience” (SO3-SEXIDNSCG-REV)

In all of the proposed sexual orientation versions, NCSES collapses response categories into only straight vs. sexual minority, premised on the idea that sexual minority respondents may feel more comfortable responding to a sexual orientation question when less detailed response options are available. However, NCSES’ own SED cognitive interviews have shown the opposite: Respondents prefer when more response options are available. Moreover, gay, lesbian, bisexual, and other sexual minority individuals are [likely to differ](#) in their STEM education and workforce experiences. [OMB’s Best Practices for the Collection of SOGI Data](#) recommends the reporting of “straight”, “gay or lesbian”, “bisexual”, and “another orientation” categories wherever possible, recognizing that aggregation can be used to minimize disclosure risks post hoc. NCSES should use at least one experiment path to test a sexual orientation question that can provide the OMB-recommended reporting categories. The extensive question context NCSES will provide on why it is collecting sexual orientation data and how it will protect the data, and giving respondents the ability to opt out of sending the data back to the doctoral institution, are all valuable approaches that should minimize privacy concerns. While straight vs. an aggregated sexual minority response category is one reasonable approach, NCSES should include at least one other approach entailing more detailed response options that permit further data disaggregation, which will increase the utility of the data and better align with OMB’s recommendations.

NCSES should use the question wording of SO1-OMB-SO-REV (“Which of the following best represents how you think of yourself?”) when testing an item with more detailed response categories. This more detailed item should replace the poorly worded SO3-SEXIDNSCG-REV item (“Regardless of your sexual experience, what is your sexual orientation or identity?”). There is no justification for referencing “sexual experience”. This unusual reference to “sexual experience” is likely the cause of the ~2% breakoff rate and other quality concerns in the 2021 National Survey of College Graduates’ (NSCG) non-production bridge panel, as described in our [January 13, 2023](#) and [March 3, 2023](#) letters to the NSF Director. Perhaps NCSES is interested in testing this wording to determine whether similar quality concerns arise in the 2024 SED as did in the 2021 NSCG bridge panel, thereby implicating the flawed question wording and helping guide SOGI decision-making for the NSCG. However, these instances are not directly comparable, as the 2024 SED is providing significant context on why NCSES is collecting SOGI data, whereas the 2021 NSCG bridge panel did not; this context may attenuate respondents’ negative reactions to the flawed wording. It would be far more valuable for NCSES to test a more detailed sexual orientation item able to disaggregate sexual minority subgroups, rather than continue to test the flawed item referencing “sexual experience” that lacks any federal precedent.

The Collapsed LGBTQ+ Status Question (SOGI1) Is Not an Appropriate Demographic Question; Sexual and Gender Minorities Differ in their STEM Education and Workforce Experiences

NCSES proposes to dedicate a sizable portion of the SED population to piloting a collapsed SOGI question asking about LGBTQ+ identity status. This question has no federal precedent and is not recommended by OMB (the [OMB-recommended](#) collapsed SOGI measure entails a check-all-that-apply list of specific SOGI terms, not LGBTQ+ vs. non-LGBTQ+). NCSES indicates that this proposed LGBTQ+ status measure “may be preferable [sic] unless distinction between sexual and gender minorities and distinguishing subgroups within these minority populations are necessary and possible”. Yet, existing data have [already shown](#) clear differences in STEM experiences between sexual and gender minorities; distinguishing them is therefore necessary. If NCSES believes SOGI minority respondents would be more willing to respond to SOGI items when they are in this collapsed form due to privacy concerns, then the opt-out question not to have data sent to the doctoral institution can be used to prevent those concerns for a more disaggregated item. OMB is [clear](#) in that “[r]esponses can...be aggregated to different levels (e.g., [sexual or gender minorities]) as needed for

compliance with agency disclosure prevention protocols” post hoc. SOGI questions should be demographic questions, not a question about LGBTQ+ community affiliation. Nevertheless, if NCSES does test this measure, it should certainly not ask about assigned sex at birth (see below).

OMB and the National Academies Are Clear: Assigned Sex at Birth Should Never Be Asked Without Also Asking About Gender (Experiment Paths 1-2 Should Be Revised)

In experiment paths 1-2, NCSES proposes to ask respondents about their assigned sex at birth followed by a collapsed LGBTQ+ status question (SOGI1). Both the [2022 National Academies' consensus report](#) on SOGI measurement and [OMB's Best Practices for the Collection of SOGI Data](#) are clear that assigned sex at birth should never be asked without also asking gender. This may be offensive to some gender minorities and signal that NSF cares more about biological sex than respondents' self-identified gender. The [National Academies' consensus report](#) indicates that the “collection of data on sex as a biological variable should be accompanied by collection of data on gender” (Recommendation #1). OMB's recommendations state that “respondents should not be asked to provide their sex assigned at birth unless they are also given the opportunity to provide their current gender identity” (p. 4). Asking about LGBTQ+ identity status (SOGI1) does not mitigate this concern.

NCSES Should Experiment With Omitting Assigned Sex at Birth When Using the Gender Minority Status Question (G2-Y/N-OPTION) to Increase Some Gender Minorities' Comfortability

The two-step method of asking assigned sex at birth and gender identity has been used in most federal surveys, which is based on the idea that a fuller spectrum of gender minority respondents (i.e., any respondent whose gender identity differs from their assigned birth sex) can be captured. It has long been recommended by the [Williams Institute](#) and more recently by the [National Academies](#). However, the [2022 National Academies consensus report](#) also recognized that for federal data collections that do not pertain to health the two-step method may not be appropriate: “Collection of data on sex as a biological variable should be limited to circumstances where information about sex traits is relevant, as in the provision of clinical preventive screenings or for research investigations of specific genetic, anatomical, or physiological processes and their connections to patterns of health and disease” (Recommendation #1). Indeed, some gender minority respondents may feel that asking about assigned sex at birth in NCSES surveys is unnecessarily intrusive, inappropriately health-focused, or irrelevant to their STEM education and workforce experiences.

The Williams Institute has long [recommended](#) a secondary alternative, where gender is assessed in addition to a gender minority status question similar to G2-Y/N-OPTION – but without asking assigned sex at birth. Some major federal surveys, including the CDC's Behavioral Risk Factor Surveillance System Survey (BRFSS), have [long used](#) this method of collecting gender and gender minority status to capture gender minority respondents. While the two-step method involving assigned sex at birth is intended to capture the broadest spectrum of gender minority respondents, if a Yes/No gender minority status item can serve this purpose, then asking about assigned sex at birth may not be necessary. NCSES should test at least one experiment path where respondents are asked about their gender and their gender minority status (G2-Y/N-OPTION), without asking about assigned sex at birth. This has the potential to improve some gender minority respondents' comfortability and reduce survey burden.

Added Privacy Features Need Not Be Limited to Sexual Orientation; Ability to Opt Out of Sending Sensitive Data to Doctoral Institution Could Be Explored for Other Items like Disability

NCSES' sexual orientation data is afforded the same privacy and confidentiality protections as any other identifiable data in NCSES surveys, namely those guaranteed by the National Science Foundation Act of 1950, the Privacy Act of 1974, and the 2018 reauthorization of the Confidential

Information Protection and Statistical Efficiency Act. Providing question context for respondents on why NCSES is collecting SOGI data and how privacy will be protected is a valuable approach that should mitigate privacy concerns.

Among NCSES surveys, the SED is unique in that potentially identifiable data are sent to doctoral institutions, which is governed by strict data-sharing agreements between those institutions and NCSES. For respondents who initially indicate they do not wish to complete the sexual orientation question, NCSES has proposed a follow-up question that would allow respondents to still provide sexual orientation to NCSES but opt out of the data being sent to their doctoral institution. Federal statisticians have long regarded sexual orientation as a potentially sensitive question, similar to race, disability, income, salary, and other items. Heightened concerns of identifiability and data use are not unique to sexual orientation but would also apply to similarly sensitive questions (e.g., disability). NCSES should consider exploring similar follow-up, opt-out questions for other sensitive questions, which may improve the quality and item response rate of other sensitive data collected on the SED.

NCSES' singling-out of sexual orientation with the opt-out question may unnecessarily increase the perception that sexual orientation is overly and uniquely sensitive information. Unless NCSES has tested comfortability with disability, income, race, and other items and made comparisons with sexual orientation, there is no empirical or theoretical basis to suggest this. Other items, especially disability, would benefit from the same enhanced privacy features related to the SED, and NCSES should explore this. This issue does not pertain to NCSES' other surveys, such as the NSCG or the Survey of Doctorate Recipients (SDR), as for those surveys data are not provided to respondents' institutions.

I congratulate NSF on taking this important step toward advancing LGBTQI+ equity in STEM.

Respectfully Submitted,



Jon Freeman
Associate Professor, Columbia University