Jonathan B. Freeman, Ph.D.
Associate Professor of Psychology

July 10, 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; 2023 Survey of Doctorate Recipients (**Federal Register Doc. 2023-13279**)

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 2023 Survey of Doctorate Recipients (SDR). *See* 88 Fed. Reg. 40870 (June 22, 2023). I am writing to supplement my previous comments submitted on June 26, 2023 in this same docket. My additional remarks follow new information provided in a July 7, 2023 response letter from NSF's National Center for Science & Engineering Statistics (NCSES), as well as the disclosure of NCSES' recently presented data at the annual meeting of the American Association for Public Opinion Research (AAPOR). I am grateful to NCSES for its detailed responses to my previous public comments and its transparency in making the data presented at AAPOR available on request.

The Proposed Design Confounds Three Independent Effects: A Corrected Design (e.g., Appendix) Is Needed So That NCSES Can Accurately Evaluate the Causes of Potential Data Quality Issues

NCSES' response letter explaining its revisions to the 2024 Survey of Earned Doctorates' (SED) SOGI Experiment Plan in light of submitted comments helps connect the dots regarding some of the methodological choices NCSES now proposes for the 2023 SDR's SOGI Experiment Plan (Attachment D5). NCSES' additional details would seem to clarify why it now in the SDR proposes to vary the response option for heterosexual respondents ("Straight, that is not gay or lesbian" vs. "Straight"), as well as vary the opt-out option provided ("Prefer not to answer" vs. "I don't know" vs. neither). I now assume that these were not arbitrary choices and that NCSES wishes to test these effects directly.

Data from the 2021 National Survey of College Graduates' (NSCG) bridge panel recently presented at AAPOR highlights how critically important the issue of this opt-out response option is for NCSES' sexual orientation question. Besides the 2.1% item nonresponse rate reported for this question, an additional 7.4% of respondents selected the "Prefer not to answer" option, effectively totaling a 9.5% nonresponse rate. This is not especially surprising given that the sexual orientation question in the 2021 NSCG suffered

COLUMBIA UNIVERSITY DEPARTMENT OF PSYCHOLOGY

Schermerhorn Hall   Office: 368   Lab: 352   1190 Amsterdam Avenue New York, NY 10027   Phone: 212-853-4317   Email: jon.freeman@columbia.edu

from several flaws: The question stem included the unusual "sexual experience" reference, which likely confused respondents and/or made them uncomfortable (I thank NCSES for clarifying that its goal with this stem was to be more inclusive). Moreover, unlike NCSES' current approaches with SOGI, it did not provide any question context for why it seeks to collect SOGI information or how data will be protected.

However, it also stands to reason that offering "Prefer not to answer" may have harmed, rather than helped, in the 2021 NSCG. Giving respondents an explicit option to opt out may have made the possibility of nonresponse unduly salient and effectively encouraged nonresponse, as compared to "I don't know" or a lack of an explicit opt-out option. The AAPOR presentation disclosed by NCSES also discussed SOGI testing in the Department of Education's 2022 National Teacher and Principal Survey (NTPS), which did include question context on why SOGI data were being collected and, critically, did not provide respondents with any opt-out option such as "Prefer not to answer" or "I don't know". The 2022 NTPS achieved a nonresponse rate of 3.0% for sexual orientation, which aligns with other Department of Education surveys – a far better result than the effective nonresponse rate of 9.5% in the 2021 NSCG.

It would seem critical for NCSES to know whether including "Prefer not to answer" encourages more nonresponse than warranted, especially for sexual orientation. However, a major problem is that the proposed design cannot unequivocally measure the opt-out option's effect on nonresponse rates or other metrics, as it does not independently manipulate this variable. Instead, manipulation of the opt-out option is confounded with other variables of interest. Currently, the use of "I don't know" is confounded with use of "Straight, that is not gay or lesbian" (Battery 1); use of "Prefer not to answer" is confounded with the binary Yes/No design (Battery 2); and use of neither opt-out option is confounded with more expansive response options (e.g., "Pansexual", "Asexual") (Battery 3). In other piloting of sexual orientation, NCSES is using "Prefer not to answer" in all 2024 SED designs yet using "I don't know" in all 2023 NSCG designs; thus, any observed differences are confounded with different privacy concerns in the SED vs. NSCG. Similar confounds occur with the opt-out option for gender identity. Furthermore, use of "Male"/"Female" (Battery 1) vs. "Man"/"Woman" (Battery 3) is confounded with use of more limited response options (Battery 1) vs. more expansive ones (e.g., "Non-binary", "Gender non-conforming") (Battery 3).

To provide only a few examples of how these confounds may create issues in NCSES' interpretation of effects: Using "Straight, that is not gay or lesbian" could lead to more "I don't know" responses due to respondents' confusion about the phrasing. The forcing of respondents to dichotomize their minority status (Yes/No) rather than allowing them to self-identify could create discomfort and lead to more "Prefer not to answer" responses. Or, observed differences in the response distributions of binary gender options could be driven by more inclusive vs. restrictive response options, rather than sex vs. gender terminology. Thus, NCSES will be unable to tease apart the causes of differences observed between the question designs. NCSES should fix these confounds by independently manipulating the three factors of interest (opt-out response option; heterosexual response option; binary gender response options). A schematic example of a corrected design that would allow NCSES to adequately estimate the three independent effects is provided in the Appendix. Keeping with the sample size per cell NCSES has proposed, a corrected design would only require an *n* of 24,000-36,000 (rather than 12,000).

The benefit of accurately understanding the causes of potential data quality issues in NCSES' question designs is far greater than the cost of a slight increase in SOGI pilot sample. The current approach tries to do too many things at once, resulting in unnecessary confounds out of a fear of survey burden. It is of significant public interest for NCSES to generate satisfactory solutions for SOGI questions across its workforce surveys with this current round of piloting, so it may begin officially collecting SOGI data immediately thereafter. Expanding the SOGI pilot in the 2023 SDR from *n* of 12,000 to *n* of 24,000-36,000 using a corrected design (e.g., Appendix) is highly feasible. By comparison, NCSES proposes to use the entire SDR sample (*n* = 130,000) to pilot new race and ethnicity questions that OMB only began considering months ago. While the effort to improve race and ethnicity questions is urgent and important, NCSES has been considering SOGI questions now for five years with relatively little to show. An accurate and complete understanding of the three effects of interest is worth the minor increase in survey burden.

COLUMBIA UNIVERSITY DEPARTMENT OF PSYCHOLOGY

Schermerhorn Hall   Office: 368   Lab: 352   1190 Amsterdam Avenue New York, NY 10027   Phone: 212-853-4317   Email: jon.freeman@columbia.edu

COLUMBIA

Jonathan B. Freeman, Ph.D.
Associate Professor of Psychology

NCSES Should Provide an Accurate Level of Data Exposition in Its Forthcoming Public Releases

NCSES' goal should be to optimize SOGI questions, while recognizing that they are like any other sensitive demographic items. Thus, it may not be possible to achieve perfectly 0% item nonresponse and breakoff rates. Consistent with the approaches of NCSES' peer statistical agencies, quality metrics for SOGI questions should be compared to items such as salary, earned income, disability, race, and ethnicity. OMB's *Best Practices for the Collection of SOGI Data* follows the same logic, indicating, for example, that "respondents are unlikely to skip SOGI questions (low item nonresponse), especially compared to other sensitive data items" (p. 3). Although NCSES has not provided comparisons with other sensitive items in the past, in its recent AAPOR presentation it valuably drew comparisons in nonresponse rates between SOGI and salary and earned income items in the 2021 NSCG. Confirming the predictions of our January 13, 2023 letter, salary (5.5%) and earned income (7.2%) elicited far higher item nonresponse rates than sexual orientation (2.1%), assigned birth sex (0.4%), or gender identity (0.6%). Still, NCSES did not provide a comparison with disability, race, or ethnicity, and it did not compare breakoff rates between SOGI questions and other demographic questions.

NCSES should provide this greater level of data exposition for members of the public in the forthcoming releases it has committed to. This contrasts with a different approach NCSES has taken in other contexts. For instance, NCSES recently revised its justification to OMB in its Supporting Statement A as to why it excluded sexual orientation from the 2023 NSCG. This revision followed concerns raised in our January 13, 2023 letter regarding NCSES' lack of appropriate comparisons in assessing the 2021 NSCG's sexual orientation breakoff rate. NCSES' revised justification states that "[i]n fact, the sexual orientation question had a breakoff rate that exceeded the rate for over 90% of the questions collected on the NSCG production survey" yet provided no further delineation of the results (p. 5). This is an incomplete and inaccurate characterization, as the vast majority of the NSCG's 109 questions are non-sensitive and innocuous and therefore not appropriate comparisons. The ~10 items that elicited more breakoffs than sexual orientation are very likely the most comparable: demographic questions and other sensitive items (e.g., earned income, salary, disability). NCSES should compare SOGI items' performance to its qualitatively similar questions, not the full set of standard innocuous items. Any sensitive demographics would be expected to fall in the upper tail of this distribution on a federal survey.

If this kind of data exposition foreshadows what is to come in NCSES' future public releases, I urge NCSES instead to follow and expand on its recent precedent with the AAPOR presentation and provide more detailed and clearer analyses. In forthcoming releases, NCSES should provide direct comparisons with other comparable demographic items and provide a full description of results on any metrics that contribute to its decision-making regarding SOGI questions. This transparency will aid researchers outside the government so that they can accurately understand the nature of NCSES' effects and how SOGI questions function in the broader context of NCSES surveys. In turn, this will improve external scientific feedback that aims to facilitate NCSES' SOGI measurement efforts and will better align with the NSTC's *Framework for Federal Scientific Integrity Policy and Practice*.

I thank NCSES for its responses to my public comments and for making its presented data available on request, and I applaud NSF on taking this important step toward advancing LGBTQI+ equity in STEM.

Respectfully Submitted,

Jon Freeman
Associate Professor, Columbia University

COLUMBIA UNIVERSITY DEPARTMENT OF PSYCHOLOGY

Schermerhorn Hall   Office: 368   Lab: 352   1190 Amsterdam Avenue New York, NY 10027   Phone: 212-853-4317   Email: jon.freeman@columbia.edu

APPENDIX

Battery 1 (OMB Recommended)			Battery 2 (SOGI Y/N)			Battery 3 (Higher-Level Specificity)		
GI Version I *n=4,000* "Prefer not to answer"	**GI Version II** *n=4,000* "I don't know"	**GI Version III** *n=4,000* No explicit opt-out	**GI Version I** *n=4,000* "Prefer not to answer"	**GI Version II** *n=4,000* "I don't know"	**GI Version III** *n=4,000* No explicit opt-out	**GI Version I** *n=6,000* "Man" / "Woman"	**GI Version II** *n=6,000* "Male" / "Female"	
[counterbalance versions for GI and SO]			[counterbalance versions for GI and SO]			[counterbalance versions for GI and SO]		
SO Version I *n=4,000* "Straight, that is not gay or lesbian"	**SO Version II** *n=4,000* "Straight or heterosex."	**SO Version III** *n=4,000* "Straight"	**SO Version I** *n=4,000* "Prefer not to answer"	**SO Version II** *n=4,000* "I don't know"	**SO Version III** *n=4,000* No explicit opt-out	**SO Version I** *n=4,000* "Prefer not to answer"	**SO Version II** *n=4,000* "I don't know"	**SO Version III** *n=4,000* No explicit opt-out

GENDER IDENTITY (GI)

Battery 1 | Version I
How do you currently describe yourself?
Select all that apply.
1. Male
2. Female
3. Transgender
4. I use a different term – Specify: [free-text]
5. Prefer not to answer

Battery 1 | Version II
How do you currently describe yourself?
Select all that apply.
1. Male
2. Female
3. Transgender
4. I use a different term – Specify: [free-text]
5. I don't know

Battery 1 | Version III
How do you currently describe yourself?
Select all that apply.
1. Male
2. Female
3. Transgender
4. I use a different term – Specify: [free-text]
[Neither "I don't know" or "Prefer not to answer"]

Battery 2 | Version I
Do you consider yourself to be a gender minority (e.g., transgender, non-binary)?
1. Yes
2. No
3. Prefer not to answer

SEXUAL ORIENTATION (SO)

Battery 1 | Version I
Which of the following best represents how you think of yourself?
1. Gay or lesbian
2. Straight, that is not gay or lesbian [2023 NSCG]
3. Bisexual
4. I use a different term – Specify: [free-text]
5. I don't know

Battery 1 | Version II
Which of the following best represents how you think of yourself?
1. Gay or lesbian
2. Straight or heterosexual [2024 SED]
3. Bisexual
4. I use a different term – Specify: [free-text]
5. I don't know

Battery 1 | Version III
Which of the following best represents how you think of yourself?
1. Gay or lesbian
2. Straight [Proposed 2023 SDR Battery 3]
3. Bisexual
4. I use a different term – Specify: [free-text]
5. I don't know

Battery 2 | Version I
Do you consider yourself to be a sexual minority (e.g., gay, lesbian, bisexual)?
1. Yes
2. No
3. Prefer not to answer

COLUMBIA UNIVERSITY DEPARTMENT OF PSYCHOLOGY

Schermerhorn Hall   Office: 368   Lab: 352   1190 Amsterdam Avenue New York, NY 10027   Phone: 212-853-4317   Email: jon.freeman@columbia.edu

Battery 2 | Version II
Do you consider yourself to be a gender minority (e.g., transgender, non-binary)?
    1. Yes
    2. No
    3. I don't know

Battery 2 | Version III
Do you consider yourself to be a gender minority (e.g., transgender, non-binary)?
    1. Yes
    2. No
    [Neither "I don't know" or "Prefer not to answer"]

Battery 2 | Version II
Do you consider yourself to be a sexual minority (e.g., gay, lesbian, bisexual)?
    1. Yes
    2. No
    3. I don't know

Battery 2 | Version III
Do you consider yourself to be a sexual minority (e.g., gay, lesbian, bisexual)?
    1. Yes
    2. No
    [Neither "I don't know" or "Prefer not to answer"]

Battery 3 | Version I
How do you currently describe yourself?
Select all that apply.
    1. Man
    2. Woman
    3. Transgender
    4. Gender non-conforming
    5. Non-binary
    6. Genderfluid
    7. Genderqueer
    8. I use a different term - Specify: [free-text]
    [Neither "I don't know" or "Prefer not to answer"]

Battery 3 | Version II
How do you currently describe yourself?
Select all that apply.
    1. Male
    2. Female
    3. Transgender
    4. Gender non-conforming
    5. Non-binary
    6. Genderfluid
    7. Genderqueer
    8. I use a different term - Specify: [free-text]
    [Neither "I don't know" or "Prefer not to answer"]

Battery 3 | Version I
Which of the following best represents how you think of yourself?
Select all that apply.
    1. Gay or lesbian
    2. Straight or heterosexual **[2024 SED]**
    3. Bisexual
    4. Asexual
    5. Pansexual
    6. Fluid
    7. Queer
    8. I use a different term - Specify: [free-text]
    9. Prefer not to answer

Battery 3 | Version II
Which of the following best represents how you think of yourself?
Select all that apply.
    1. Gay or lesbian
    2. Straight or heterosexual **[2024 SED]**
    3. Bisexual
    4. Asexual
    5. Pansexual
    6. Fluid
    7. Queer
    8. I use a different term - Specify: [free-text]
    9. I don't know

Battery 3 | Version III
Which of the following best represents how you think of yourself?
Select all that apply.
    1. Gay or lesbian
    2. Straight or heterosexual **[2024 SED]**
    3. Bisexual
    4. Asexual
    5. Pansexual
    6. Fluid
    7. Queer
    8. I use a different term - Specify: [free-text]
    [Neither "I don't know" or "Prefer not to answer"]

COLUMBIA UNIVERSITY DEPARTMENT OF PSYCHOLOGY

Schermerhorn Hall   Office: 368   Lab: 352   1190 Amsterdam Avenue New York, NY 10027   Phone: 212-853-4317   Email: jon.freeman@columbia.edu