COLUMBIA

Jonathan B. Freeman
Columbia University
Department of Psychology
368 Schermerhorn Hall
New York, NY 10027

March 3, 2023

Sethuraman Panchanathan, Director
National Science Foundation
2415 Eisenhower Avenue
Alexandria, VA 22314

CC: Charles Barber, Chief Diversity and Inclusion Officer, National Science Foundation
Emilda Rivers, Director, NSF National Center for Science & Engineering Statistics
John Finamore, Chief Statistician, NSF National Center for Science & Engineering Statistics
Rhonda Davis, Head, NSF Office of Equity & Civil Rights
Amanda Greenwell, Head, NSF Office of Legislative & Public Affairs
Margo Schwab, Desk Officer for NSF, U.S. Office of Management & Budget
Christopher Marcum, Co-Chair, NSTC Scientific Integrity Task Force
Anne Ricciuti, Co-Chair, NSTC Scientific Integrity Task Force
Gretchen Goldman, Co-Chair, NSTC Scientific Integrity Task Force
Craig Robinson, Co-Chair, NSTC Scientific Integrity Task Force
Francesca Grifo, Co-Chair, NSTC Scientific Integrity Task Force
The Honorable Tammy Baldwin, U.S. Senate
The Honorable Dianne Feinstein, U.S. Senate

Dear Director Panchanathan,

Thank you for your letter in response to the concerns that 1,700 members of the scientific community and I raised with you on January 13, 2023. I share your goal of working together to realize a diverse and equitable scientific workforce. It is disappointing that the National Science Foundation's (NSF) National Center for Science & Engineering Statistics (NCSES) is abiding by its initial decision to exclude a sexual orientation question from its 2023 national workforce surveys. The decision is concerning not because there is a desire for NCSES' pilot results to "be ignored for the sake of expediency", as your letter seemed to suggest. It is concerning because NCSES' decision currently lacks scientific merit and is based on analyses with unaddressed threats to their internal validity.

I am grateful that NCSES has agreed to publicly release its pilot data and explain how the data informed its decision. In advance of the release, I offer suggestions regarding what specific data would be most helpful in alleviating the concerns that have been raised. If NCSES does not wish to conduct these additional analyses, it could provide the raw data as part of its public release so that scientists outside the government can ascertain the merit of its findings and provide NCSES with feedback to facilitate its future sexual orientation and gender identity (SOGI) measurement efforts. As NCSES conducts further testing, the analyses recommended here may also have the potential to prevent a waste of government resources and steer NCSES away from what appear to be red herrings.

While the concerns are scientific in nature, it should be noted that NSF still does not seem to recognize the urgency of the problem. While I can appreciate your letter's comparison of NCSES' drawn out process with SOGI efforts to a similarly drawn out process with non-degree certificate data, the costs are quite different. Each year that passes, NCSES' omission of SOGI questions has concrete and immediate consequences for the LGBTQI+ STEM community – a community that faces increasingly documented disparities, career barriers, underrepresentation, retention failure, and some of the highest levels of harassment in STEM. NCSES should not cut corners, but it also should not delay these efforts as it did for three years, or conduct unrigorous science that leads to running in circles, as it currently risks doing. A federal statistical expert and the co-chair of the 2022 National Academies consensus

COLUMBIA

Jonathan B. Freeman
Columbia University
Department of Psychology
368 Schermerhorn Hall
New York, NY 10027

report on SOGI measurement told *Nature* that it is "perplexing [NCSES] would leave the sexual orientation question off, given that it's been asked successfully [on other federal surveys] for ten years now". NCSES' stance on SOGI data is also at odds with the White House's own view. For instance, the National Science and Technology Council's (NSTC) recent *Federal Evidence Agenda on LGBTQI+ Equity* report clearly states that data collection "must start immediately" (p. 7):

> *Data collection must start immediately. Especially in the context of federal surveys, recommendations for SOGI data measures and successful data collection methods exist, are well-documented, and have been successful in federal data collection efforts. Just as continued research is necessary to improve the quality of race, ethnicity, disability, and other demographic data collection, the Federal Statistical System should continuously strive to improve its SOGI data collection methods, but in most circumstances, the Federal Government has enough knowledge to responsibly begin collecting SOGI data now.*

As our letter 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 NSF's surveys are critical to implementing federal policy, such as diversity-related funding provided by the CHIPS and Science Act, and to fulfilling the directives of Executive Order 14075 to advance SOGI data practices and address LGBTQI+ disparities, and Executive Order 13985 to improve the equity of underserved communities.

Over the years, my colleagues and I have consistently heard from NCSES that the resources it has to devote to SOGI efforts are severely constrained. Both Executive Order 14075 and the White House's FY 2024 R&D priorities urge agencies to request specific funding to advance their SOGI data practices. As we previously advocated, if funds are an issue, NSF should request the necessary funding to pursue whatever remaining studies it believes are necessary to accomplish this work as soon as possible.

Moving forward, I am hopeful that NCSES' SOGI efforts can take a path of greater openness, public engagement, and improved science. Below I outline (a) suggestions for fixing significant transparency issues that continue to hinder the quality of NCSES' science; and (b) a recommended set of results that NCSES should provide as part of its forthcoming release to alleviate the concerns raised, fulfill its scientific integrity commitments per the NSTC's *Framework for Federal Scientific Integrity Policy and Practice*, and potentially save NCSES time and resources by identifying spurious effects.

NSF's Transparency with SOGI: Improving Public Comment in Line with 5 CFR 1320.8(d)

Thank you for affirming the importance of NSF's transparency and its intention to "continue… listening to the community and learning where [it] can do more." A major roadblock is that NCSES still does not provide any mechanism for experts or stakeholders to make informed comments about its planned SOGI question designs or methodology. For the past five years, when NCSES is asked, it has been willing to disclose information about its SOGI plans once they have been finalized and approved by OMB. However, it does not solicit feedback from experts or stakeholders to help guide the making of those plans or permit the public to relay concerns about potentially flawed study designs. This has had significant consequences for the quality of NCSES' science and resulted in a waste of NSF resources.

As discussed in our letter, NCSES initially indicated that it did not intend to pilot sexual orientation on the 2021 National Survey of College Graduates' (NSCG) non-production bridge panel, per its September 2020 filing with the U.S. Office of Management & Budget (OMB). Following reporting in *Science* and NCSES' discussions with the Federal Committee on Statistical Methodology's (FCSM) SOGI Research Group, it ultimately did include a sexual orientation item for the bridge panel study. However, the SOGI question wording and response options that NCSES used in the 2021 NSCG bridge panel are still not publicly available in any OMB records. NCSES did not provide this information

COLUMBIA

Jonathan B. Freeman
Columbia University
Department of Psychology
368 Schermerhorn Hall
New York, NY 10027

to members of the public in any 60-day comment period or even a 30-day final comment period, as is intended by 5 CFR 1320.8(d). When asked for specific SOGI question designs, following OMB approval, NCSES did disclose them in a direct correspondence (personal communication with NCSES; April 9, 2021). However, neither I nor any outside expert or stakeholder had an opportunity to provide comments or concerns to NCSES or OMB regarding the specific SOGI question designs or methodologies that NCSES implemented on 5,000 individuals in the U.S. population.

This lack of proper public engagement resulted in NCSES testing a flawed sexual orientation question that contains a highly unusual reference to "sexual experience" and has no federal precedent, as was discussed in our January 13, 2023 letter. Had experts or stakeholders been given the opportunity to provide public comment on the question design prior to it being finalized, NCSES could have addressed these concerns. In the *Nature* coverage described earlier, a SOGI expert at the Williams Institute converged on our letter's assessment that the bridge panel's sexual orientation question was flawed and noted it was a "fail on the part of NSF not to include more traditional measures of sexual orientation". The federal statistical expert and National Academies consensus report co-chair mentioned earlier also told *Nature* that the unusual reference to "sexual experience" likely confused respondents. Even the current FCSM SOGI Research Group co-chair stated that "[i]n hindsight, it might have been a better methodological approach [for NCSES] to employ a split panel design to test the augmented version of sexual orientation that was tested [i.e., the one now widely criticized] with [more traditional] version[s]" (personal communication with FCSM SOGI Research Group co-chairs; January 6, 2023). Thus, much wasted time and resources could have been avoided had NCSES properly allowed public comment on the question designs and methodology it decided to implement.

Now, two years later, history is poised to repeat itself. If NCSES intends to use the 2023 NSCG's bridge panel for testing a revised sexual orientation item or engage in further piloting of SOGI questions, members of the public should be allowed to provide comments on specific SOGI question designs or testing plans. Even if NCSES still wishes to circumvent the inclusion of specific SOGI methodological details in its formal comment periods, then outside the information collection clearance process NCSES should engage experts and stakeholders on these designs and plans before they are finalized. This should also apply to SOGI research approved as part of NCSES' Generic Clearance mechanism (OMB No. 3145-0174), such as the Mechanical Turk (MTurk) non-probability sample discussed in our letter and in your response. These changes would better align NSF's SOGI efforts with the NSTC's *Framework for Federal Scientific Integrity Policy and Practice*. If NSF is committed to transparency and engaging the community, as your letter stated, then I hope you agree that allowing sufficient public comment on its specific SOGI plans should be a bare minimum.

RECOMMENDATION: Most immediately, if NCSES intends to pilot SOGI questions on the 2023 NSCG non-production bridge panel, it should provide a mechanism for the public to give feedback on specific question designs and methodology as soon as possible (and for future bridge panels, via the 60-day and 30-day comment periods), as is intended by 5 CFR 1320.8(d). If NCSES seeks approval for other SOGI pilot studies via its Generic Clearance mechanism (OMB No. 3145-0174), such as new MTurk non-probability samples, it should allow experts and stakeholders to provide specific feedback as part of the OMB approval process for the pilot activity or some alternate method outside this process.

Significant Scientific Concerns Remain: What NCSES' Forthcoming Release Should Address

Below I provide suggestions as to what NCSES should release in order to address the concerns raised and help improve the methods it employs for future SOGI measurement research.

COLUMBIA

Jonathan B. Freeman
Columbia University
Department of Psychology
368 Schermerhorn Hall
New York, NY 10027

The Response Order Effect Is Spuriously Produced By a Well-Documented Primacy Bias Among Inattentive MTurk Respondents; It Can Be Corrected by Established Data Screening Methods

Your letter reaffirmed a likely flawed finding and did not appear to address the relevant concerns raised on pp. 45-49 of our letter. Your letter noted that "...[NSF's] research showed that simply reversing the order in which sexual orientation options were presented – gay or lesbian, bisexual, straight, or another orientation – resulted in the size of the sexual minority STEM population changing from 2.9% to 4.6%... [T]hat difference raises a significant concern about accuracy".

By "size of the sexual minority STEM population" I believe your letter meant to indicate that the estimated size of the gay and lesbian population changed between 2.9% and 4.6% depending on the two orderings. (The size of the sexual minority population would include those who identified as "bisexual" and "another orientation", which was 12.3% and 13.6% depending on the two orderings; the study also did not distinguish STEM vs. non-STEM). But as noted in our letter, NCSES also found that the size of the straight population changed from 87.7% to 86.4% as well. This additional element helps make clear NCSES' true finding: Respondents were 1.7% more likely to select the gay/lesbian option when listed first, and were 1.3% more likely to select the straight option when listed first.

This MTurk SOGI study is only NCSES' second time ever using the MTurk platform, and these results speak more to NCSES' insufficient screening and quality assurance procedures rather than represent a novel finding about SOGI questions. It is highly unlikely that these respondents forgot their sexual orientation or were confused by the question. Instead, this pattern is more consistent with a subset of inattentive MTurk respondents who showed primacy bias, i.e, they selected the first-listed option to rush through the survey and receive payment. Such low-effort, careless, and rushed responding among MTurk respondents is extremely well characterized. A 2021 review states:

> *MTurkers often complete HITs [studies] in distracting environments and at rapid speed to maximize monetary returns, which translates into about 15% of MTurkers failing attention and compliance checks. MTurkers are less likely to pay attention to study instructions or manipulations, and more likely to engage in insufficient effort or careless responding, as compared with [other] samples.*

NCSES' use of MTurk samples is commendable, but it should do so responsibly. Researchers are expected to abide by established guidelines to avoid bias and conduct proper data screening. If not properly screened, "bad" respondents will rush, take shortcuts, and disproportionately use the first-listed option in a vertically oriented question such as NCSES' sexual orientation question. This primacy bias is well studied by survey researchers and has been documented among MTurk respondents around the world. When "bad" respondents are left in the data, a 1.7% or 1.3% greater likelihood to select the first-listed option in NCSES' sexual orientation question is fully expected. It is commonplace and even the very first tip that some online survey platforms warn their researchers of.

Established guidelines include: (a) only recruiting MTurk participants with high-quality track records (≥95% HIT approval rate or 'Master Workers'); (b) using attention checks (i.e., instructional manipulation checks) throughout the survey to detect participants engaged in low-effort responding; (c) excluding inattentive participants who are rushing through the survey, either at the level of the question item (e.g., thresholds such as <1 or <2 seconds per item), or at the level of the survey (e.g., thresholds such as <300 milliseconds × total survey word count); and/or (d) removing participants with low response variability and extreme response patterns via well-described statistical methods.

NCSES excluded only 4.6% of the data, which appears to be due entirely to ineligibility and suspicious IP addresses rather than low-effort screening. It did not require MTurk respondents to have a minimum

COLUMBIA

Jonathan B. Freeman
Columbia University
Department of Psychology
368 Schermerhorn Hall
New York, NY 10027

approval rating, and it did not use any systematic screening criteria to remove low-effort, rushed responders. The research community largely expects that MTurk studies discard at least 15% of the data due to standard low-effort screening procedures, including for studies using only high-quality MTurk users. Thus, it is virtually guaranteed that NCSES' final MTurk sample still contains a subset of inattentive respondents who rushed through the survey and may have exhibited primacy bias – a bias producing what NCSES now believes is a real change in how sexual orientation was reported.

In fact, NCSES' only publicly available MTurk study has already shown a similar pattern of bias due to inattentive respondents. In that study, NCSES discovered an anomalous effect and explicitly interpreted it as likely produced by inattentive respondents (see p. 16). Unlike the MTurk SOGI study, this other MTurk study did restrict participation appropriately (≥95% approval rating); however, similar to the SOGI study, it did not include attention checks or use systematic screening criteria.

Unfortunately, NCSES cannot retroactively implement attention checks or restrict the SOGI study to high-quality MTurk participants. However, assuming they were recorded, it can use item-level completion times; if these were not recorded, it should have timestamp information and therefore have at least measured completion times for the entire survey. Prior research has shown that primacy effects on MTurk, such as a 1.7% or 1.3% greater likelihood to select a first-listed option on an item, are eliminated when data from rushed responders are removed (e.g., excluding participants who completed the survey in less than 300 milliseconds × total survey word count).

RECOMMENDATION: NCSES should consult established guidelines and screen its final MTurk sample (n = 2,680) for inattentive participants. Using completion times either for the sexual orientation item or for the entire survey, NCSES should set a threshold for excessively fast responding (e.g., <1 or <2 seconds for the question item, or <300 milliseconds × total survey word count for the entire survey). Distributional inspection of completion times and checking threshold robustness could guide choices. NCSES should remove rushed responders falling below the threshold and reanalyze response distributions. NCSES should also conduct a 2 (rushed, non-rushed) × 2 (gay/lesbian-first condition, straight-first condition) × 2 (gay/lesbian responses, straight responses) chi-square test and analyze the conditional odds-ratios between rushed/non-rushed × ordering condition. This will likely reveal that the order effect is present/amplified for rushed responders, but absent/attenuated for non-rushed responders. For future MTurk studies, NCSES should restrict to users with ≥95% approval rating, implement attention checks, and use systematic screening criteria to detect inattentive participants. If NCSES continues to explore order effects, it should include a control, non-sexual-orientation item with the same manipulation to distinguish what effects are due to the item vs. a bias in the overall data.

A Sexual Orientation Item Poses No Unique Privacy Risks and Should Be Compared to Similarly Sensitive Questions Such As Race, Disability, Income, and Salary

Some of your letter's statements seem to conflate the perceived sensitivity of a sexual orientation question with the actual data protections afforded it. The letter states that NCSES' "studies consistently found that some individuals are hesitant to provide sexual orientation data in the context of their STEM education and workforce experiences" and that "some of the initial survey respondents told [NSF] they were hesitant to share personal information with the government, or had concerns that the survey response would identify them to their colleagues". It was then concluded that "[a]n individual's data privacy is non-negotiable". To be clear, 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.

COLUMBIA

Jonathan B. Freeman
Columbia University
Department of Psychology
368 Schermerhorn Hall
New York, NY 10027

Federal statisticians have long regarded sexual orientation as a potentially sensitive question, similar to race, disability, income, salary, and other items. NCSES should therefore start with the assumption that a small fraction of respondents may be hesitant to complete any of these questions or have heightened concerns about their data use. If NCSES were to probe respondents on their comfortability with race, disability, income, or salary, it would certainly observe a minority of respondents with highly similar perceptions. Respondents' perceptions that any of these questions are sensitive in no way changes the actual data protections afforded to them. Privacy is not being negotiated.

NCSES' MTurk study tested comfortability by asking to what extent respondents felt comfortable revealing SOGI data to the federal government, a friend, a family member, a doctor, or a police officer. NCSES found that respondents were overwhelmingly comfortable revealing their sexual orientation to the government, and critically, they were equally comfortable revealing their sexual orientation as they were their gender identity. Thus, given its own self-reported comfortability data, NCSES should be equally prepared to move forward with a sexual orientation item as it is with a gender identity item.

Your letter also reaffirmed the issue of the bridge panel's sexual orientation item's breakoffs. As argued in our letter on pp. 42-45, the breakoffs are highly likely to be due to NCSES' unusual reference to "sexual experience", discussed earlier. As prior federal surveys have shown, once using a revised item that avoids such issues, breakoffs should approach 0%. That being said, NCSES' benchmark should not be 0% (although 0% can certainly be strived for). Its benchmark for a sexual orientation item should be the breakoffs of other potentially sensitive items such as income and disability, as argued in our letter and consistent with the approach of NCSES' peer statistical agencies.

RECOMMENDATION: NCSES should directly compare the bridge panel's quality assessment metrics for the sexual orientation item – breakoff rates, item nonresponse rates, completion times, and number of changed responses – with those of its comparable items: race, income, salary, and disability. It should also provide breakoff and item nonresponse rates for SOGI questions in the MTurk study. In future SOGI testing, any exploratory question wording or response options that lack clear federal precedents should be incorporated into split-panel designs, so that tried-and-true SOGI measures (e.g., the OMB-recommended sexual orientation question used now for ten years) can serve as a baseline.

Thank you again for your letter. I hope the suggestions I have provided can inspire greater openness and engagement and improve the efficiency and rigor of NSF's SOGI data efforts. If I can be of any service in these efforts, please do not hesitate to let me know.

Sincerely,

Jon Freeman
Associate Professor, Columbia University