

October 3, 2022

NSTC Subcommittee on Equitable Data
White House Office of Science and Technology Policy
Eisenhower Executive Office Building
1650 Pennsylvania Ave. NW
Washington, DC 20504

CC: Elaine Slaugh, OMB Analyst for National Science Foundation
Yi Pei, OMB Analyst for National Science Foundation
Emilda Rivers, Division Director, National Center for Science & Engineering Statistics

Dear Members of the NSTC Subcommittee on Equitable Data:

On behalf of the American Association for the Advancement of Science (AAAS), the American Educational Research Association (AERA), the Council of Professional Associations on Federal Statistics (COPAFS), the Federation of Associations in Behavioral & Brain Sciences (FABBS), and the Consortium of Social Science Associations (COSSA), we are pleased to have the opportunity to comment on the development of the Federal Evidence Agenda on LGBTQI+ Equity. This is an initiative arising from President Biden's Executive Order (EO) 14075 that aims to improve the federal government's ability to make data-informed policy decisions toward improving the equity of LGBTQI+ populations. We commend the Biden administration in taking this historic step in advancing LGBTQI+ equity in American society.

For four years, we have engaged the National Science Foundation (NSF) and Office of Management & Budget (OMB) on the importance and feasibility of including sexual orientation and gender identity (SOGI) questions on surveys of the U.S. science, technology, engineering, and mathematics (STEM) workforce. These surveys, administered by NSF's National Center for Science & Engineering Statistics (NCSES), include the National Survey of College Graduates (NSCG), Survey of Doctorate Recipients (SDR), and Survey of Earned Doctorates (SED). The data and associated reports are used widely by researchers and policymakers to understand and address educational and career barriers in STEM; to inform national policies related to STEM and higher education; and to determine underrepresented groups' eligibility for funding and federal resources. Though key progress has been made through NCSES' pilot research on SOGI measurement, the administration has an opportunity to implement a more comprehensive approach to gathering this critical information.

Our comments address (I) the alarming disparities faced by LGBTQI+ people in U.S. STEM fields, and the need for NCSES to begin collecting SOGI data by 2023; (II) the feasibility of asking SOGI questions on NCSES surveys; (III) the privacy and confidentiality protections governing SOGI data and how to minimize identifiability risks; and (IV) the need to allocate NSF a budget starting in FY 2024 for developing a SOGI data infrastructure, tracking LGBTQI+ inequities in STEM, and addressing those inequities via NSF's programs and policies.

I. Alarming LGBTQI+ Disparities in U.S. STEM Fields: The Time is Now For NCSES To Measure, Track, and Report SOGI Data in the U.S. STEM Population

LGBTQI+ individuals are facing alarming disparities in U.S. STEM fields, which not only raise issues of equal opportunity but represent a waste of STEM talent. Our world faces complex and urgent scientific challenges, and all individuals wishing to contribute to science must be enabled to pursue their scientific potential. When groups of people are hindered from participating in STEM, we all lose as a society, and the competitiveness of the American science and technology enterprise is diminished. The welfare of LGBTQI+ scientists and engineers, who may go on to discover life-saving treatments or develop groundbreaking technologies, is not only a moral imperative but also in our national interest. As the U.S. continues to face urgent STEM talent gaps, Congress has long recognized that “underrepresented populations are the largest untapped STEM talent pools” and that “the United States should encourage full participation of individuals from underrepresented populations in STEM fields” (42 U.S.C. § 1862). Yet, the federal government currently lacks the necessary demographic data to inform policies that can address LGBTQI+ inequities and facilitate LGBTQI+ participation in STEM.

While a lack of SOGI data in NCSES surveys is preventing researchers’ and policymakers’ comprehensive understanding of the inequities faced by LGBTQI+ scientists, what data do exist point to serious issues. LGBTQI+ people are estimated to be approximately 20% less represented in STEM fields than statistically expected, and they are less likely than non-LGBTQI+ people to major in STEM, persist in STEM, earn STEM degrees, and be in STEM occupations.¹ Harmful biases and unsupportive STEM environments appear to be partly at fault. LGBTQI+ scientists experience more career barriers and workplace harassment than non-LGBTQI+ scientists, even when controlling for other demographic and career-related factors.² From a prevalence standpoint, such career barriers can have an enormous impact on American science. LGBTQI+ people comprise an estimated 7.1% of the U.S. population, and this number rises precipitously for younger generations who represent the future of American scientists, with 10.9% of Millennials and 22.8% of Gen-Z individuals identifying as LGBTQI+.³

After delaying the piloting of SOGI questions for three years, citing limited time and resources, NCSES began piloting on college graduates in April 2021. The results showed that SOGI questions did not raise methodological issues (virtually 0% breakoff and non-response rates), and respondents were very comfortable answering SOGI questions.⁴ NCSES has additionally conducted a field test with college graduates and cognitive interviews with doctoral recipients, and the results are expected to be finalized by November 2022. We urge NCSES, once the full set of results are in, to immediately adopt SOGI measures in the NSCG, SDR, and SED. Across its pilot studies, NCSES has been testing two variants of SOGI question wording and response options. In selecting which variant to adopt, we suggest that NCSES consider both methodological accuracy as well as the inclusiveness of the response options made available.

While we understand NCSES’ desire to take a measured approach, we also note that demographic questions are never perfect. Federal-wide standards on race and ethnicity questions

¹ Freeman, J. B. (2020). Measuring and resolving LGBTQ disparities in STEM. *Policy Insights from the Behavioral and Brain Sciences*, 7, 141-148. <https://journals.sagepub.com/doi/abs/10.1177/2372732220943232>

² Cech, E. A., & Waidzun, T. J. (2021). Systemic inequalities for LGBTQ professionals in STEM. *Science Advances*, 7, eabe0933. <https://advances.sciencemag.org/content/7/3/eabe0933>

³ Gallup (2022). LGBT Identification in U.S. Ticks Up to 7.1%. <https://news.gallup.com/poll/389792/lgbt-identification-ticks-up.aspx>

⁴ Freeman, J. B. (2021). STEM disparities we must measure. *Science*, 374, 1333-1334. <https://www.science.org/doi/full/10.1126/science.abn1103>

have been revised several times and are currently being reviewed again. In our view, the urgent need to understand and address LGBTQI+ disparities in STEM by adopting SOGI measures now outweighs any desire NCSES may have to further exhaustively test SOGI measures, particularly when SOGI measures have been included in other federal population surveys for at least six years (see Section II). NCSES can and should always work to improve how it collects demographic information, but the time is now to begin collecting SOGI data in NCSES surveys.

II. Including SOGI Questions in National STEM Workforce Surveys Is Feasible, Does Not Raise Sensitivity Concerns, And Respondents Are Comfortable Providing SOGI Data

SOGI questions are highly feasible and have already been implemented in major federal population surveys. In 2015-2016, the Census Bureau conducted debriefing questionnaires, focus groups, and targeted interviews, and found that respondents reacted favorably to SOGI items, did not have any difficulty understanding them, and non-response and breakoff rates were extremely low.⁵ Numerous federal population surveys, including education- and employment-related surveys similar to NCSES surveys, have included SOGI questions for years, including the Baccalaureate & Beyond Longitudinal Study and High School Longitudinal Study (Department of Education), Current Population Survey (Department of Labor), National Health Interview Survey (Centers for Disease Control & Prevention), National Crime Victimization Survey (Department of Justice), and most recently the Household Pulse Survey (Census Bureau).

Since early testing by the Census Bureau, the SOGI questions implemented across various federal population surveys have consistently been shown to not raise sensitivity issues such as excessive breakoff or non-response rates, and to behave on par with – if not better than – other common demographic questions, such as income or disability. For instance, in the Department of Education’s 2016 High School Longitudinal Study, SOGI questions triggered fewer breakoffs and item non-responses than income and disability questions.⁶ In NCSES’ own recent testing of SOGI questions with college graduates, breakoff and non-response rates were virtually 0%, and LGBTQI+ and non-LGBTQI+ respondents alike overwhelmingly reported feeling comfortable providing SOGI data to a federal agency like NCSES.⁴

III. SOGI Data Are Covered Under Federal Privacy and Confidentiality Protections; Levels of LGBTQI+ Data Disaggregation Can Be Calibrated To Minimize Risk

The privacy and confidentiality of any personally identifiable data in NCSES surveys, such as potential SOGI data, are already protected by federal laws to which NCSES strictly adheres, including 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. NCSES also uses methods to avoid intentional or unintentional disclosure of identifiable information. It removes names and all identifying information, and out of an abundance of caution uses suppression techniques to protect confidentiality. For example, if a data cell has too few respondents such that an individual might possibly be identified (e.g., when cross-tabulated with other demographics or identifiers), NCSES suppresses the data cell. In cases where NCSES provides identifiable data to outside researchers or deans of graduate schools, NCSES’ data sharing is governed by strict agreements that require the data to remain confidential, be

⁵ Ellis, R., Virgile, M., Holzberg, J., Nelson, D.V., Edgar, J., Phipps, P., & Kaplan, R. (2017). Assessing the feasibility of asking about SOGI in the Current Population Survey: Results from cognitive interviews. <https://www.bls.gov/osmr/research-papers/2017/html/st170210.htm>

⁶ Christopher, E.M. & Burns, L. (April 2021), *Federal CASIC Workshops*. https://www.census.gov/fedcasic/fc2021/pdf/4B_Christopher.pdf

exclusively used for statistical purposes, and never be disclosed to other parties. Breaches to such agreements or unauthorized use of NCSES data are associated with severe penalties.

In its reports and data releases, NCSES has long included aggregate statistics on racial and ethnic groups that have a far smaller prevalence in the U.S. population than that of LGBTQI+ people (7.1%)³, such as Asians (6.1%), American Indians or Alaska Natives (1.3%), and Native Hawaiian or other Pacific Islanders (0.3%).⁷ Thus, NCSES will not encounter issues in providing statistics on the LGBTQI+ population in the aggregate as well. When data are cross-tabulated, in a situation where a data cell lacks sufficient sample, NCSES can always use suppression techniques to avoid identifiability risk. That said, the utility of SOGI data can be increased even further with disaggregation into LGBTQI+ subgroups. While comprehensive data are lacking, differences in STEM disparities between sexual-minority men vs. sexual-minority women or between sexual minorities vs. gender minorities have already been observed, highlighting the importance of parsing diversity within the LGBTQI+ umbrella.¹⁻²

We suggest that NCSES prioritize implementing the full range of response options in reports and releases of SOGI data so that diversity within the LGBTQI+ population can be understood. In specific data tabulations where a full range of response options might pose identifiability risk (e.g., when cross-tabulated with other variables), NCSES could suppress such data cells and provide only an aggregate LGBTQI+ statistic. In other tabulations where the range of response options does not pose risk, no suppression will be needed and NCSES can provide disaggregated data so that variability by LGBTQI+ subgroups can be parsed. In short, even if fully disaggregated LGBTQI+ data may not be possible in certain contexts or tabulations due to identifiability risk, in no way does that warrant the wholesale exclusion of SOGI data altogether.

IV. NSF’s FY 2024 R&D Budget Should Include Funding for Developing a SOGI Data Infrastructure, and Tracking and Addressing LGBTQI+ Inequities in STEM

Adding SOGI questions and incorporating SOGI data into NCSES’ reports and data releases (e.g., the *Women, Minorities, and Persons with Disabilities in Science & Engineering* report) should be only the first, albeit long-awaited, step in NSF’s responsibility to ensure the equity of LGBTQI+ people in U.S. STEM fields. EO 14075 specifically asks that agencies include in their budget submission to OMB “a request for any necessary funding increases to support improved SOGI data practices”, and the OMB Director’s July 22, 2022 memorandum regarding the White House’s FY 2024 R&D priorities urges agencies to request funding for creating more equitable data infrastructures, including in relation to SOGI data in particular.⁸

Where NCSES identifies LGBTQI+ disparities, including how they may intersect with other marginalized characteristics, NSF should exchange relevant data with other agencies and stakeholders and use the data to inform policies and protocols that can broaden the participation of LGBTQI+ people in U.S. STEM fields. For instance, disparities identified in NCSES surveys are used to set National Institutes of Health (NIH) diversity policies, such as eligibility for funding and trainee fellowships. NSF has a large portfolio of grants in STEM education and workforce research that could be directed at developing and testing solutions for the LGBTQI+ disparities identified via NCSES’ analyses. NSF currently asks for demographics in relation to its own programs and opportunities, such as fellowships or research grants; such data collection

⁷ U.S. Census Bureau, Population Estimates (July 1, 2021). <https://www.census.gov/quickfacts/fact/table/US/PST045221>

⁸ White House Memorandum (July 22, 2022). <https://www.whitehouse.gov/wp-content/uploads/2022/07/M-22-15.pdf>

should be expanded to include SOGI measures, and NSF should use these data to track LGBTQI+ equity in its own programs and opportunities. Creating new data pipelines or revising existing ones, exchanging SOGI statistical data across agencies, and ongoing policy analyses may be needed to track and address potential LGBTQI+ disparities identified by NSF.

Beginning in FY 2024, OMB should allocate funds in NSF's R&D budget not only for continued research to improve SOGI and other demographic measures, but to develop a broader infrastructure for SOGI data. This would equip NSF with tools and procedures necessary to translate any LGBTQI+ underrepresentation, opportunity gaps, retention failure, or other disparities into practices and programs that can help address those disparities. This would include NSF sharing relevant data with other agencies and stakeholders so that they, too, can help address those disparities. This infrastructure should be closely linked to NSF's processes for tracking and addressing other disparities related to gender, race, ethnicity, disability, economic background, and first-generation status, given their likely intersectional nature.

Thank you for your consideration. Please direct any correspondence to jon.freeman@columbia.edu.

Sincerely,



Jonathan B. Freeman, PhD
Associate Professor
Columbia University



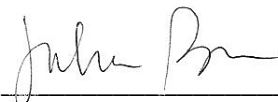
Felice J. Levine, PhD
Executive Director
American Educational Research Association
(AERA)



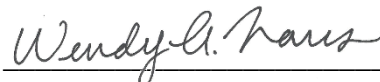
Sudip S. Parikh, PhD
Chief Executive Officer
American Association for the Advancement
of Science (AAAS)



Paul B. Schroeder, MA
Executive Director
Council of Professional Associations on
Federal Statistics (COPAFS)



Juliane Baron, MPAff
Executive Director
Federation of Associations in Behavioral &
Brain Sciences (FABBS)



Wendy A. Naus
Executive Director
Consortium of Social Science Associations
(COSSA)