

ATTACHMENT 9

Survey of Earned Doctorates Recent Methodological Research

Recent Methodological Research in the Survey of Earned Doctorates

Testing of the 2021 SED Field of Study and Educational History Data Collection Methods

Educational history data for research doctorate recipients is a rich and valuable source of information for federal agencies, policymakers, and researchers. However, collecting these data accurately and reliably is a complex undertaking and requires considerable amount of input from respondents. To improve data quality and reduce response burden, NCSES conducted a methodological study to test alternative methods for collecting respondents' field of study (FOS) and educational history data.

The study was conducted in two phases between December 2019 and January 2020, and included a quantitative analysis of missing and un-coded question items, response time, and an analysis of qualitative feedback from study participants. The study tested two versions of the field of study (FOS) questions and three versions of the educational history collection methods. The first phase test of the study consisted of 1,146 participants with graduate degrees recruited through Amazon's MTurk to complete a web survey online. The second phase consisted of 56 doctoral students and doctorate holders recruited from different colleges and universities to participate in cognitive interviews. All study participants were randomly assigned to one of the six possible combinations of the FOS and education history questions.

Results indicated that the current SED version of the FOS data collection worked better than the proposed alternative, and the participants who listed their education degrees on a single page answered the questions more quickly while reporting just as many and detailed postsecondary degree information as those in other versions. The FOS data collection based on the Classification of Instructional Program (CIP) code frame, and educational history list method were implemented in the 2021 SED web instrument.

Testing of the 2021 SED COVID-19 Pandemic Question Module

The novel coronavirus pandemic that began in the U.S. around March 2020 disrupted many aspects of higher education and began to present multiple challenges to the doctoral students in their path to doctorate completion. NCSES quickly developed a new question module for the 2021 SED to measure the impacts of the pandemic on doctoral students' graduate experience and post-graduation career plans. The module included seven filter questions about effects students may have experienced and follow up questions to collect more detailed information about their experience.

Cognitive interviews were conducted with 9 doctoral students in May 2020 to test whether the pandemic impact questions and response categories were understood by respondents and whether they captured the range of respondents' experiences. Test participants were recruited from doctoral students who had expressed an interest in participating in the 2020 SED FOS and Educational History Modules (described above).

The cognitive interviews revealed that participants were able to understand the seven filter questions about the impacts of pandemic on graduate experience and career plans. The only exception was the questions about changes in their career plans or goals, where participants were unclear in reference to the time frame. Based on the test results, the question about changes in career plans was modified to include the phrase "longer term," which clarified the reference period. These questions were included in the 2021 SED, which launched in June 2020.

Testing of the Sexual Orientation and Gender Identity Questions

NCSES has been evaluating sexual orientation and gender identity (SOGI) questions for possible inclusion on its surveys, including the SED, in response to Executive Order 13985 to expand the data available to the Federal Government to measure equity and capture the diversity of the American people¹ and more recently Executive Order 14075 to advance LGBTQI+ equality². Measuring and understanding the impact of sexual and gender minorities within the science and engineering enterprise is important and supports further research on the educational experiences and employment outcomes of these groups. However, the collection of such sensitive information on the SED poses a unique challenge as doctoral recipients' responses are often provided back to participating doctoral institutions at the individual level.

As part of NCSES's efforts to test SOGI items on its education and workforce surveys, NCSES SED contractor conducted cognitive interviews in the summer of 2022 with a convenience sample of 61 research doctoral students representing diverse sexual orientations, gender identities, and citizenship status. The sample included 24 gender minority participants, all of whom were also sexual minorities; 15 cisgender sexual minority participants; and 22 cisgender heterosexual participants. The first two groups combined are referred to as SOGI minorities and comprise all the sexual minorities in the sample. The latter group is referred to as non-SOGI minority participants. Because SED includes a high proportion of non-U.S. citizens with different cultural backgrounds and varying levels of proficiency with English, the cognitive interviews include 20 non-U.S. citizens.

There were three areas of inquiry in the interviews. First, the interviews assessed comprehension and the response process for each question tested. Three gender identity questions, two sexual orientation questions, and two biological sex questions were evaluated. The SOGI questions were selected from questions developed through prior research conducted by NCSES, the Census Bureau, and the National Center for Health Statistics. The question wording, the level of detail of the data collected, and the response format (i.e., select one or select all that apply) varied across questions.

Second, the interviews explored potential confidentiality sensitivities with respect to SOGI data. This was a topic of particular interest because the SED population is small by its nature which increases the potential for deducing the identity of individuals within small minority subgroups. Additionally, the SED data are shared with doctoral institutions because they play an integral part in disseminating the SED web survey, promoting the SED data collection, and use their graduates' data for institutional reporting and program review.

The third area of inquiry related to gender minority participants' preferences for presenting their responses in data tables produced for reports.

Key findings from the cognitive interviews are as follows:

While SOGI minority participants could and would answer most SOGI questions, they often worried about data confidentiality and the potential for professional consequences, including the loss of funding or employment opportunities if their gender identity or sexual orientation were disclosed. Others were concerned about maintaining their privacy as distinct from a fear of repercussions. As an example of these confidentiality concerns, almost half of gender minority participants mentioned

¹ <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>

² <https://www.whitehouse.gov/briefing-room/statements-releases/2022/06/15/fact-sheet-president-biden-to-sign-historic-executive-order-advancing-lgbtqi-equality-during-pride-month/>

that they would be fearful if their gender identity were revealed in data associated with their name or through deduction. From these findings, it is clear that many people share concerns about providing sex or SOGI data on surveys and that these concerns are heightened by the possibility that these data may be reported back to their doctoral institution.

In terms of question comprehension, in general, participants, including non-U.S. citizens, were able to understand all the test questions and choose the response option(s) that best described them. However, for some of the questions tested, the wording was so vague that the intent of the question was not apparent without referring to the response options. Also, while many participants were unsure how to define all the gender identities and sexual orientations presented as options in the most detailed questions, they were still able to easily select the response option(s) that describes them.

Several issues were identified that may affect data quality including:

1. Some participants stated that the terms “male” and “female” connote biological sex so the inclusion of these terms in questions about gender identity made the intent of the questions unclear to them. Replacing “male” and “female” with “man” and “woman” would make the response options align with the question wording. Gender minority participants seemed to be more attuned to this issue with a third (8 of 24) mentioning it with respect to the first gender identity question tested. Among cisgender participants, a fifth of sexual minorities (3 of 15) and 9% of non-SOGI minorities (2 of 22) mentioned this.
2. In the sexual orientation questions, the clauses defining straight as “not lesbian or gay” or “not gay, lesbian, or bisexual” were confusing to some participants because they were not inclusive of other sexual orientations such as queer or asexual. SOGI minority participants and non-SOGI minority participants were equally confused by this. Twenty-eight percent of SOGI minority participants (11 of 39), all of whom were sexual minorities by definition, and 28% of non-SOGI minority participants (6 of 22) expressed confusion about the clause defining straight in the first sexual orientation question tested. For this population, the term straight was clearer without the defining clauses.
3. Some gender minority participants (38%, 9 of 24) found it challenging to respond to gender identity questions that only allowed a single response option to be selected because more than one option applied to them. One-third of cisgender sexual minorities (5 of 15) and 9% of non-SOGI minorities (2 of 22) also noted that this would be difficult for some respondents.
4. Most SOGI minority and non-SOGI minority participants preferred the gender identity (87%, 53 of 61) and sexual orientation (80%, 48 of 60) questions that had the most response options and allowed for selection of multiple response options. Participants indicated that they liked these questions better due to greater inclusivity. SOGI minority participants also wanted representation of the LGBTQIA+ community in the survey data and these questions would influence their likelihood of responding to SOGI survey questions, yielding more accurate data.
5. Similarly, a majority of participants would like an additional response option beyond “male” and “female” in sex and sex at birth questions, often stating that it would make the question more inclusive. Eighty-two percent of participants (46 of 56) thought the addition of a third option would improve the biological sex question that has been included on the SED survey historically. However, there was no consensus on what the third option should be, and SOGI minority participants objected to response options they perceived as “othering” that defines them as outside the norm.

6. Over 90% of participants indicated that they would be likely to respond to gender identity and sexual orientation questions knowing that their de-identified answers would be shared with their doctoral institution and the NSF. If the data were linked with names, 80 percent of participants would still be likely to answer gender identity questions, but this percentage drops to about two-thirds for sexual orientation questions. SOGI minority participants had more concerns about sharing their data associated with their names than non-SOGI minorities. They were often worried about data confidentiality and the potential for professional consequences that may lead to losing funding or employment opportunities if their gender identity or sexual orientation was disclosed.
7. Many SOGI minority participants noted that knowing the reasons and relevance of the SOGI data collection would increase their likelihood of responding to these questions. For example, half of gender minority participants (10 of 20) said they would consider relevance when deciding whether to answer gender identity questions.
8. Although a majority of SOGI minority participants (81%, 26 of 32) indicated that they always answer questions about biological sex when they see them on surveys and forms, some noted that they did so because they assumed they could not skip these questions. Some SOGI minority participants (22%, 7 of 32) expressed annoyance or discomfort answering this question because they did not believe it was relevant information.
9. When asked how the gender minority participants would prefer their responses presented in data tables in reports, almost all (91%, 21 of 23) preferred the most detailed responses as opposed to showing data in aggregated form. These participants noted that this was the most accurate way of presenting the data and allowed gender minority participants to feel seen.

Based on the results of these cognitive tests and recently published OMB recommendations on best practices for the collection of SOGI data, NCSES will test several SOGI questions in the 2024 SED. Additional testing is needed for two primary reasons. First, although the cognitive test sample was diverse it was not representative of the entire SED population and cognitive test participants were responding to hypothetical scenarios for which they indicated their likelihood of responding to SOGI questions. Real world behavior across a broader set of doctorate recipients may differ. Second, our experiment will test whether contextual information about reason these data are being collected and whether the ability to opt out of sharing this data with the doctoral institution can be used to address and allay respondent concerns about privacy and confidentiality.

The cognitive tests suggest the following design considerations for the test question wording, response options, and presentation:

- Questions that offer a detailed set of response choices in a 'select all that apply' format may perform best. The findings showed that the questions with these characteristics were preferred by majority of participants regardless of sexual orientation or gender identity. The findings also suggest these types of questions would yield the best quality data for SOGI minority groups as the SOGI minority respondents may be more inclined to answer them.
- The interviews found a tension between SOGI minorities' desire for representation and concerns about being identified. Therefore, NCSES will also consider testing features to address respondents' concerns about confidentiality, privacy, and relevance of SOGI data. One way to do this would be to include "I prefer not to answer" response options to clearly signal to respondents that they may choose to skip these questions. Another approach would be to provide an explanation of how the data will be used and an assurance of data confidentiality on

screens where SOGI questions are asked. A third option to consider testing is the provision of a checkbox which allows respondents to answer SOGI questions but opt out of providing their SOGI data to their doctoral institution.

- Given some participants' negative reaction to questions about biological sex, NCSSES will consider presenting this question on the same screen as the question about gender identity. Doing so may forestall concerns SED respondents have increasingly expressed in recent years by simultaneously recognizing gender as an important construct.
- Given the desire of many participants to report these data in detail and OMB's guidance to collect only what can be reported, NCSSES will test questions that collect data at different levels of specificity.