Memorandum

Date: February 11, 2021

To: Margo Schwab, Desk Officer
             Office of Management and Budget

From: Emilda B. Rivers, Director
             National Center for Science and Engineering Statistics
             National Science Foundation

Via: Suzanne Plimpton, Reports Clearance Officer
             National Science Foundation

Subject: Request for approval of an online data collection recruited from Amazon's Mechanical
             Turk (MTurk) for experimentation with the National Survey of College Graduates

---

The purpose of this memorandum is to inform you that the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) plans to conduct exploratory quantitative testing under the generic clearance for improving survey projects (OMB control number 3145-0174). This project is a follow-up to previous research using crowdsourcing platforms to recruit subjects for an online survey. The previous research (outlined in a memorandum dated August 25, 2017) examined the effectiveness and utility of one specific crowdsourcing platform, Amazon's Mechanical Turk (MTurk), and an online survey to pretest questionnaire items. This proof of concept research effort was successful: it provided evidence of the feasibility of using an online convenience sample as a cost-effective and time-efficient approach to obtain useful feedback on the salary question format and confidentiality pledge wording used in NCSES surveys.

The project discussed in this memorandum continues examining the effectiveness and utility of online convenience samples by examining two research questions with one sample:

1. Whether a non-probability sample may be used to derive estimates of the college-educated population in the U.S., and

2. Whether different ways of asking about sexual identity and gender identity in the college-educated population affect estimates and response distributions

Background

In an effort to control costs for pretesting questionnaires and collecting survey data, agencies are exploring alternate methods to cognitive interviews and full-scale national surveys. One such method involves using online convenience samples where participants are sometimes recruited using crowdsourcing platforms that pay people to perform small tasks called human intelligence tasks (HITs). NCSES and other statistical agencies, notably the Bureau of Labor Statistics (BLS) and the National Institutes of Health-National Cancer Institute (NIH-NCI), have used these crowdsourcing platforms to recruit participants to complete online surveys.

NCSES recently used Amazon's Mechanical Turk (MTurk) to recruit participants for an online survey that assessed how participants understand and react to the confidentiality pledge wording and interpret questionnaire items on basic annual earned income (Chandler, 2018). Before using MTurk to recruit sample cases for an online survey, NCSES conducted an assessment of online convenience sample sources such as crowdsourcing platforms and online survey sample providers (Chandler et al., 2017). From this assessment, MTurk was identified as one of most promising crowdsourcing platforms because it has a larger available sample than other crowdsourcing platforms. It also has the most extensive features, allowing requesters control over who may participate in their surveys. In addition, BLS and NIH-NCI have used MTurk in their past efforts to recruit participants to complete online surveys. NCSES proposes to continue exploring the use of these types of convenience samples by conducting a study to explore the feasibility of deriving useful estimates for selected questions from NCSES's National Survey of College Graduates (NSCG).

In addition, a convenience sample of college-educated individuals, a population of interest to NCSES, allows for examining the potential measurement error associated with asking questions related to sexual identity and gender identity. The questionnaire for the online survey (Attachment A) includes selected questionnaire items from the NSCG, questions on sexual identity and gender identity (not currently asked on the NSCG), and follow-up questions on sexual identity and gender identity. The follow-up questions are modeled on probes that would typically be asked in a face-to-face cognitive interview; asking them in this setting gives NCSES some quantitative information to assess question performance.

Purpose of research

This submission builds on NCSES's previous research on convenience samples in two ways: 1) by determining if an NSCG-like population (i.e., U.S. residing college graduates) can be recruited through the use of a non-probability based sample and 2) using this convenience sample to further develop innovative questionnaire items.

Similar to the previous study, participants for this study will be recruited from MTurk. The goal is to determine whether it is possible to select a college-educated non-probability sample recruited through MTurk that is similar to the NSCG's target population on key demographic characteristic and whether this sample can produce estimates that align with the NSCG estimates other than employment or related and the employment estimates from other federal surveys (e.g., the Current Population Survey (CPS)). Through this evaluation, we hope to obtain an increased understanding of the measurement error associated with estimates of the college-educated population collected through a non-probability sample. Previous research has provided an indication that the college graduates recruited from MTurk differ from the overall college-educated population on some demographic characteristics such as age and citizenship status, but are similar on other characteristics like race and ethnicity. In this evaluation exploring the measurement error of estimates from non-probability samples, we will collect data using a sample allocation approach that aligns the distribution of the MTurk and NSCG samples across selected demographic characteristics.

In response to the coronavirus pandemic and its potential impact on the comparability of estimates derived from this study, NCSES has added response options to the 'reasons for not working' question (see Attachment A, item A3) and added questionnaire items to evaluate

employment status earlier in the year prior to the pandemic (see Attachment A, item G1). These modifications are designed to allow NCSES to assess the changes in the response distribution arising from the pandemic and will aid in the comparison of the non-probability sample data to the other federal data sources.

This submission also includes NCSES's first effort to examine the collection of sexual identity and gender identity data. Building on the work of the Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity (SOGI) in Federal Surveys, NCSES will include four questions not currently asked on the NSCG: sexual identity, sex at birth, current gender identity, and pronouns used by the respondent.

Sexual identity is one dimension of sexual orientation. The sexual identity question posed in this study is currently used in the National Crime Victimization Survey and National Health Interview Survey. It had undergone rigorous cognitive pretesting and successfully been fielded in in-person production surveys. This study will allow us to compare results from interviewer-administered surveys to a self-administered survey. This study also includes an experiment regarding sexual identity. Each respondent will receive one of four sexual identity questions. The question stem is the same for each of the four treatments, though the response options vary. Running this experiment will demonstrate possible measurement error effects in the presentation of response options.

The questionnaire will use a two-step approach to collect gender identity information. The two-step method of measuring gender identity is based on empirical research that indicates a one-question method results in significant undercounting of transgender populations. The two-step method classifies the transgender population by asking sex assigned at birth and current gender identity. This allows both individuals whose sex assigned at birth does not comport with current gender identity *and* those who identify with the term "transgender" to be classified as transgender. This study includes an experiment for gender identity. All respondents will be asked both questions, but the experiment varies the order in which the two questions are presented. This experiment will demonstrate possible context effects for the gender identity questions.

The instrument also contains a series of follow-up probes to the SOGI and pronoun questions. They address topics such as question comprehension, accuracy of response options, and sensitivity, and are the sort of things that an interviewer would ask in the context of an in-person cognitive interview.

Methodology

The survey will be conducted via web, and participants will be recruited from MTurk. As MTurk participants are self-selecting, it is likely the resulting estimates will not be statistically representative. As a result, NCSES will not use the data collected in this study to make inferences or produce statistical estimates. The results from the survey will be compared with the NSCG estimates and estimates from other federal surveys to provide insight on whether a convenience sample from MTurk can shed light on the college-educated population. In addition, the survey results will further our understanding of potential measurement errors associated with collecting data on sexual identity and gender identity.

Samples obtained from MTurk will be studied to examine the internal validity and the demographics of the MTurk participants. All results, especially those exploring wording options for the sexual identity and gender identity questions, will be interpreted with caution given the sample was pulled from MTurk rather than a probability-based sample. Although the MTurk sample may not be statistically representative, studies using MTurk samples obtain similar results to surveys using probability-based samples (e.g., Mullinix et al., 2015).

A draft of the MTurk post is provided in Attachment A. The ad will be shown only to MTurk workers who reside in the United States (including Puerto Rico). Once participants are recruited, they will be given a link to the online survey instrument, which will be hosted by NCSES's contractor for this research. The data collected as part of this survey will be stored on the contractor's servers.

Analysis Plan
For the two research questions outlined above, NCSES intends to engage in several analyses, outlined below:
- Comparing estimates between the MTurk sample and federal surveys (e.g., NSCG, CPS) on key questions, such as:
    - Employment
    - Level of the highest degree
    - Community college enrollment
    - The amount of money borrowed for education
- Comparing response distributions among the four different sexual identity questions (X1.1a-d), to see if a) the order of response options results in different estimates, and b) the presence of an option for "Something else – specify" changes the response distributions or affects item nonresponse.
- In the two sexual identity questions that include a "Something else – specify" option (questions X1.1a and X1.1b), analyzing the content and nature of the verbatim responses.
- In the gender identity question experiment, identifying whether estimates of non-binary individuals are affected by the order of the questions.
- Analyzing response distributions to the follow-up questions to the sexual identity, gender identity, and pronoun usage items, by treatment condition, to identify which questions might reasonably be asked on the NSCG into the future.
- For the sexual identity, gender identity, and pronouns questions, breakoff rates, item nonresponse rates, and other metrics will be used to identify which of these questions might reasonably be asked on the NSCG in the future.

Using the information gathered through the follow-up probing for the sexual identity and gender identity questions. assess equity in question interpretation and performance with respect to race, ethnicity, disability, and other demographic characteristics.

Participants

NCSES will use MTurk to recruit 2,800 participants who must have a bachelor's degree or higher and live in the United States (including Puerto Rico). The proposed sample size takes into account the anticipated loss of information due to break-offs, incomplete responses, and participants who do not follow the task instructions. Similar sample sizes have been used for studies of this nature (e.g., Shapiro et al., 2013; Edgar, 2016).

Burden Hours

NCSES estimates that it will take approximately 15 minutes to read the announcement on MTurk, click the link to the survey, and complete the survey. With 2,800 participants, a total of 700 burden hours (2,800*15 minutes) are requested for this study.

Payment to Participants

All participants will receive an amount that is typical for similar survey-taking tasks on MTurk. NCSES estimates that participants will receive up to $4.00. The final rate will be determined following consultations with experts.

Informed Consent

At the beginning of the survey, participants will be informed of the OMB control number, the expected survey completion time, and the voluntary nature of the study. In addition, participants will be informed that the data they provide in this study will reside on a server outside of the NCSES domain and that NCSES cannot guarantee the protection of survey responses.

Survey Schedule

The tentative schedule for the survey is as follows:

Proposed Date	Activity/Deliverable
February 12, 2021	OMB submission for approval
February 26, 2021	OMB clearance
April 1, 2021	Launch survey
May 1, 2021	Survey due date
July 18, 2021	Final report

Contact Person

Flora Lan
Project Officer
Human Resources Statistics Program
National Center for Science and Engineering Statistics
National Science Foundation
flan@nsf.gov
703-292-4758

Attachment A: Announcement and Questionnaire

References

Edgar, J (2016) "OMB clearance 1220-0141 'Submission of Materials for Testing of Revised CIPSEA Pledge'."

Chandler, J., Poznyak, D., Sinclair, M., and Hudson, M. (2017) "Use of Online Crowdsourcing and Online Survey Sample Providers for Survey Operations." Mathematica Policy Research Report for the National Center for Science and Engineering Statistics.

Chandler, J. (2018) " A Feasibility Study of Using Mechanical Turk to Test Survey Questions" Mathematica Policy Research Report for the National Center for Science and Engineering Statistics.

Kaplan, R (2017) "OMB clearance 1220-0141 'Cognitive and Psychological Research'"

Mullinix, K.J., Leeper, T.J., Druckman, J.N. and Freese, J. (2015) 'The Generalizability of Survey Experiments', *Journal of Experimental Political Science*, 2(2), pp. 109–138. doi: 10.1017/XPS.2015.19.

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science*, *1*(2), 213-220.

Attachment A: Announcement and Questionnaire

MTurk Announcement

Posts on MTurk are typically brief. What follows may be modified.

Heading:

Take a survey about education and employment

Description:

This task involves taking a 15-minute survey on several topics, including education and employment. Payment: [TBD, up to $4.00].

Language: English

Mode: web only

Survey distribution: open survey (MTurk)

Record meta-data: workerID, IP address, user agent string

Back button: Display through E22r

Question checks: Soft-check on questions throughout by default

Notes on question numbering:

All questions that begin with the letters A through E are identical to items in the NSCG with the same question number.