Methodological Research Concerning the Survey of Earned Doctorates

Note: Reports available upon request from Lynn Milan (lmilan@nsf.gov).

An Examination of the Institution Eligibility Criteria for the Survey of Earned Doctorates: This report was the culmination of work conducted during 2013 at the request of NSF that examined the eligibility criteria for institution inclusion in the SED against a broader national and international context. The report also made recommendations for an adjudication process for determining the eligibility of institutions and programs not currently in the SED but appearing to meet the criteria for inclusion. The report is currently under consideration by NSF. No changes have yet been implemented to the SED eligibility review process. (Kirby et al., 2014)

Survey of Earned Doctorates Confidentiality Report: This report presents the findings from the cognitive interviews and focus groups conducted in 2013 and 2014 with doctorate recipients, graduate deans, institution contacts and institution researchers concerning the confidentiality procedures employed by the SED. Additional topics covered in the interviews and focus groups included a revised consent statement and a proposed data linking and data sharing project among select institutions. The report included findings and recommendations concerning the data linking and data sharing project under consideration at NSF. No changes have yet been implemented based on the findings. (Welch et al., 2014)

Timeline Data Quality Improvements for the Survey of Earned Doctorates: This report presents findings from a study conducted in 2014 that analyzed the current approach the SED employs to collect, edit and report timeline data. The report provided recommendations for the increased utility and improved quality of the data through potential questionnaire changes, edit changes and changes to how the data are contained in the DRF and presented in reports. Findings were used to inform a number of process revisions, including: expansion of the auto-coding process for timeline variables; modification of rules used to flag nontraditional timeline sequences; addition of timeline variables to the Doctorate Records File (DRF) for use in further research; and revision of select imputation rules. In addition, revisions to timeline questions in the instrument are included in the planned future cognitive interview activities. (Bautista et al., 2014)

Enhancements in Auto-Coding in the Survey of Earned Doctorates: This report presents findings from a study conducted in 2014 on the feasibility of employing an automated coding application to additional SED variables that are currently coded manually by trained clerks. The report recommended additional coding rules that would allow for automated coding, reducing the need for manual coding which in turn would improve quality and lower labor costs. Findings from this study are being implemented in the SED 2015 round auto-coding activities. (Groenhout et al., 2014)

After the Breakoff Part 2: Converting Web Break-offs to Completes This presentation was discussed at the 2014 International Field Directors & Technologies Conference in 2014. The aim of the presentation was to identify predictors of survey breakoff and identify the most successful prompts for converting breakoffs to completions. Conclusions from the presentation included (1)

certain characteristics found in survey paradata are significant predictors of survey breakoff (2) tailored email prompts were much less successful than standard email prompts and (3) standard email prompts were more successful than mail prompts in converting breakoffs. Drawing on the third finding listed above, the results of this study informed the 2015 non-respondent prompting experiment, which is examining if sending email prompts before mail prompts result in a higher survey completion rate. Respondents eligible for the experiment are those for whom both email and mailing addresses are available. The control group will receive the standard protocol of mail prompts first, followed by email prompts, if needed. The treatment group will receive the prompts in the opposite order: that is, email prompts followed by mail prompts, if necessary. (Groenhout et al., 2014)

Disclosure Analyses of Tabular Data: In this report, it is recommended to enhance cell suppression in tabulation by estimating suppressed cells via log-linear modeling. The resulting complete table is expected to have high utility for users at large because model-based best prediction is used for suppressed cells from the available unsuppressed information. In particular, the paper illustrates that estimates can be used to check the underlying trend over different subgroups cross-sectionally and for a given subgroup longitudinally. The report found that the proposed method uses only information released under cell suppression in tabulation and, therefore, does not increase the disclosure risk. Also, it preserves all the unsuppressed cells and marginal counts. No changes have yet been implemented based on this report. (Cohen et al., 2013)

Department Coding Feasibility Study: Since 1969, the SED questionnaire has included a verbatim item asking respondents to name the department of the university that supervised their doctoral studies. Each year these data are captured in the annual SED data files, but they are not cleaned, coded, or stored in the DRF. This study was conducted using verbatim responses to this particular item from the 2011 SED round to examine the feasibility and costs for coding. According to the report, coding the SED department verbatim responses is feasible for respondents who reported that their field of study is in science and engineering. If put into production the entire process could be integrated into standard SED verbatim coding to further simplify the process and reduce setup and maintenance costs. Hence, the report recommends adding a coded department verbatim variable to the DRF. No changes have yet been implemented. (Groenhout and Toit, 2013)

Comparison between Screener/Follow-Up Item Format and Yes/No Item Format: In 2010, SED changed the format of two demographic items (ethnicity and disability) from screener/follow-up to yes/no format. This paper explored the impact of this change on the responses to these items. The report analyzed the four most recent rounds of SED data (2008-2011); two rounds with screener/follow-up format and two rounds with yes/no format. The authors did not find evidence of increased data quality in the new format, as they have observed more "other" responses in the yes/no format. No changes were implemented as a result of this analysis. (Hernandez, Arakelyan, and Welch, 2013)

Converting Web Breakoffs to Completes and the Effects on Data Quality: This presentation examines web survey breakoffs in terms of respondent propensity to breakoff, the survey administrator's efforts at converting breakoffs to completes, and the data quality from these

surveys. In the first section, we found that the same demographic groups with higher rates of unit nonresponse are more likely to break off after beginning a web survey. In the second section, we found that mail prompts were more effective than email prompts at converting web breakoffs to completed surveys, but recognized some limitations in this comparison and recommended a prompt mode experiment to evaluate whether email prompts could be more successful. In the final section, we compared rates of item nonresponse between web surveys completed in a single session and those that were breakoffs and later converted to completes. We found no significant difference, concluding that prompting efforts aimed at respondents who have begun but not completed a web survey are worthwhile as the breakoff has no impact on data quality as measured by item nonresponse. Findings from this study led to further research in 2014 (see "After the Breakoff Part 2" above), which had differing results. During the 2015 round, a non-respondent prompting experiment is being conducted in order to determine if sending email prompts before mail prompts results in a higher survey completion rate. (Groenhout, Bilgen, and Latter, 2013)

Mode Effects on Item Response Rates: This report examines item response rate by mode, controlling for time of completion. According to the findings, web item response rates (RRs) tend to be higher than paper and pencil item RRs. The presentation concluded that prompts may play an important role in increasing item RR on the web. Also, the authors found that prompts may convince early responders to provide very sensitive information (such as responses to Social Security number and birthday questions), but not late responders. Thus, the web prompts have remained in the web questionnaire. The web questionnaire redesign in 2015 will further improve the prompts both for clarity and user experience. The effectiveness of these redesigned prompts will also be included in the subsequent cognitive interview activities. (Webber, Welch, and Hernandez, 2013)

Metadata Standards and Technology Development for the NSF Survey of Earned Doctorates: This presentation was made at the Federal Committee on Statistical Methodology's annual conference in the fall of 2013. The presentation discussed the project by NCSES, NORC and Metadata Technologies North American (MTNA) to develop new technologies to capture comprehensive metadata, automate the production of essential documentation, and generate an archival package for the annual NSF Survey of Earned Doctorates (SED). The work on this project links to a larger project where NCSES is developing a unified data system to store and disseminate all NCSES survey data. SED materials have been successfully delivered to the NCSES data system using the automated tools discussed in this presentation. (Noonan et al., 2013)