A User-Driven Method for Using Research Products to Empirically Assess Item Importance in National Surveys

Ai Rene Ong; Robert Schultz; Sofi Sinozich; Jennifer J. Sinibaldi; Brady T. West; James Wagner; John Finamore

ESRA 2021
Acknowledgement

This research was supported by the National Center for Science and Engineering Statistics (NCSES) through a Broad Agency Announcement (BAA).

Disclaimer: content on these slides do not reflect official NCSES policies.
Background

• Large-scale, nationally representative surveys provide essential statistics for the general population
  • E.g., unemployment statistics from the U.S. Bureau of Labor Statistics

• However, these surveys are often long and burdensome, and survey non-response has increased over the years\(^1\)

• Decisions about how to redesign surveys to reduce burden have to take into account
  1. The importance of each question to the stakeholders (e.g., congress, academics)
  2. The burden of each question to the respondents

Proposed method

• To measure the **importance** of the individual survey items for the stakeholders, we coded documents that use the survey data and other research products for their **item use**

• To measure **respondent burden**, we looked at **time spent** on each page in a web survey and **breakoff rates** to questions

• We then use these measures of importance and burden to propose some ideas for **survey redesign**
Survey of Doctorate Recipients

• Design: NCSES-sponsored longitudinal survey with a fixed panel design

• Sample:
  • Individuals under the age of 76 who have obtained a Science, Engineering, or Health doctorate from the United States
  • Sampled from Survey of Earned Doctorates (SED)

• Modes: Web, mail, CATI
  • This case study will only be focused on web mode
    • 93% of SDR respondents completed in web
    • Sample size in web mode: 70,770\(^1\)

• No. of items: 202
  • Each option in “select all that apply” questions was considered a separate item

1. Number is restricted to respondents who completed the whole instrument
Other data sources

• To measure item importance:
  • Bibliography of NCSES publications, peer-reviewed journal articles, and other documents that have used SDR variables (information provided by NCSES)
  • Data table download statistics from NCSES’s Online Data Tool that provides data users access to SDR variables (information provided by NCSES)

• To measure item burden:
  • Page-level timing and break-off paradata from web respondents to the 2019 SDR cycle
Method

• Bibliography was coded for their use of SDR variables and number of citations to calculate overall item use and average influence
  • \( \text{Average influence} = \frac{\text{No.of times item is used}}{\text{Total citations}} \)

• The data table download statistics for each variable were summed up

• Items that are displayed on the same page have page-timing averaged

• As it is unclear where the break-off happened, the same page break-off rate is used for all the items on the same page
Analysis

• Descriptive analysis of item usage, influence, and burden
• We will present:
  • Unused items across the bibliography coded, data table download statistics
  • Most burdensome items according to the web timing and break-off paradata in SDR 2019
A selection of unused items

<table>
<thead>
<tr>
<th>Item</th>
<th>Question topic</th>
<th>Time (sec)</th>
<th>Break-off (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A33.2</td>
<td>No. of people supervised indirectly</td>
<td>18.2</td>
<td>0</td>
</tr>
<tr>
<td>E14</td>
<td>Earliest age of difficulty</td>
<td>16.8</td>
<td>0.2</td>
</tr>
<tr>
<td>A29.1</td>
<td>Most important reason for working in an area outside the field of your first U.S. doctoral degree</td>
<td>14.1</td>
<td>0.1</td>
</tr>
<tr>
<td>A40A</td>
<td>Year retired</td>
<td>13.2</td>
<td>0</td>
</tr>
<tr>
<td>C01</td>
<td>Attend any work-related training</td>
<td>11.7</td>
<td>1</td>
</tr>
<tr>
<td>A29.2</td>
<td>Second most important reason for working in an area outside the field of your first U.S. doctoral degree</td>
<td>9.7</td>
<td>0.2</td>
</tr>
<tr>
<td>C03</td>
<td>Most important reason for taking training</td>
<td>9.5</td>
<td>0.2</td>
</tr>
<tr>
<td>D12</td>
<td>School-related costs paid for by an employer</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td>D08</td>
<td>Enrolled full-time/part-time/not enrolled but taking courses</td>
<td>6.8</td>
<td>0</td>
</tr>
<tr>
<td>D07</td>
<td>Enrolled in a college/taking courses on the week of Feb 1, 2019</td>
<td>5.6</td>
<td>0.3</td>
</tr>
<tr>
<td>E03.3</td>
<td>Partner's duties on job require technical expertise of a bachelor's degree and above</td>
<td>4.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Recommendations (1)

• For unused items, one can consider either dropping the items from the survey or asking these questions every other year.
• The average time to answer these questions is not very long (less than median page time).
• Break-off rates on these pages are low but added together may have an effect on lessening respondent burden.
## A selection of the most burdensome items

<table>
<thead>
<tr>
<th>Item</th>
<th>Question topic</th>
<th>Freq. of use</th>
<th>Avg. influence</th>
<th>Data table download</th>
<th>Time (sec)</th>
<th>Break-off (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A06</td>
<td>Duties and responsibilities in the last job</td>
<td>2</td>
<td>26.5</td>
<td>0</td>
<td>96.35</td>
<td>0.6</td>
</tr>
<tr>
<td>A20</td>
<td>Duties and responsibilities in the job</td>
<td>14</td>
<td>61.07</td>
<td>0</td>
<td>63.37</td>
<td>4.7</td>
</tr>
<tr>
<td>A09</td>
<td>Principal employer in the week of February 1, 2019</td>
<td>20</td>
<td>53.1</td>
<td>24</td>
<td>55.8</td>
<td>4.8</td>
</tr>
<tr>
<td>A13</td>
<td>Description of principal employer</td>
<td>4</td>
<td>70</td>
<td>0</td>
<td>30.81</td>
<td>2.5</td>
</tr>
<tr>
<td>A44</td>
<td>Total earned income in 2018</td>
<td>12</td>
<td>109.83</td>
<td>6</td>
<td>30.7</td>
<td>3.5</td>
</tr>
<tr>
<td>A04</td>
<td>Date last worked for pay or profit</td>
<td>3</td>
<td>12</td>
<td>3</td>
<td>30.37</td>
<td>0.5</td>
</tr>
<tr>
<td>A10</td>
<td>No. of people who worked for principal employer</td>
<td>6</td>
<td>72.17</td>
<td>0</td>
<td>30.26</td>
<td>1.9</td>
</tr>
<tr>
<td>A05</td>
<td>Title of last job held</td>
<td>3</td>
<td>17.67</td>
<td>0</td>
<td>29.4</td>
<td>0.6</td>
</tr>
<tr>
<td>A21</td>
<td>Job category code for current job</td>
<td>27</td>
<td>72.96</td>
<td>33</td>
<td>25</td>
<td>0.2</td>
</tr>
<tr>
<td>A07</td>
<td>Job category for last job</td>
<td>3</td>
<td>16</td>
<td>0</td>
<td>25</td>
<td>0.03</td>
</tr>
<tr>
<td>D04</td>
<td>Month and year degree awarded</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>20.74</td>
<td>0.1</td>
</tr>
</tbody>
</table>
### Recommendations (II)

<table>
<thead>
<tr>
<th>Item</th>
<th>Question topic</th>
<th>Freq. of use</th>
<th>Avg. influence</th>
<th>Data table download</th>
<th>Time (sec)</th>
<th>Break-off (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A06</td>
<td>Duties and responsibilities in the last job</td>
<td>2</td>
<td>26.5</td>
<td>0</td>
<td>96.35</td>
<td>0.6</td>
</tr>
<tr>
<td>A20</td>
<td>Duties and responsibilities in the job</td>
<td>14</td>
<td>61.07</td>
<td>0</td>
<td>63.37</td>
<td>4.7</td>
</tr>
</tbody>
</table>

- These are open-ended text questions, which explains why they take the longest
- They are also important questions for NCSES and survey data users and need to be kept in the survey
- Knowing that these are burdensome items:
  1. Consider including wording such as, “Your answer to this question is very important…” to motivate respondents
  2. Ask these questions only to respondents who have switched jobs/job titles
Summary

• Through a combination of examining survey item usage and web question timing and break-off paradata, we identified a few items that can be dropped.

• We also identified potentially burdensome items that are important that would benefit from a redesign.
Limitations and future directions

• This method is dependent on having good data on variable use
• Other measures of respondent burden need to be considered, and methods to combine different measures of burden should also be investigated
Thank you!

Ong, Ai Rene
aireneo@umich.edu
Average page timing

Median = 23.92 sec
Page break-off rate