

META-ANALYSIS OF DCMRB STUDIES FROM FY21-FY23

Data Collection and Methodology Research Branch (DCMRB)
Economic Statistical Methods Division (ESMD)
U.S. Census Bureau

Amy Anderson Riemer – Supervisory Survey Statistician
Hillary Steinberg – Survey Statistician

coding it forward >



JACOB DEPINET
Data Analytics
The Ohio State University

*Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate disclosure avoidance protection of the confidential¹ source data (Disclosure Review Board (DRB) approval number: CBDRB-FY23-ESMD001-016).

CONTEXT

- What is the Data Collection and Methodology Research Branch (DCRMB)?
- Issues of Interest
 - Unclear questions
 - Unclear functionalities
 - Areas of high burden
- Results of DCMRB studies are summarized in reports and/or presentations

PROBLEM

- Issues found across studies aren't highlighted – there is little interaction between DCMRB studies
- Illuminating these themes can optimize future research for the DCMRB and improve survey quality for survey designers in the Economic Directorate
 - Saves agency time and resources
 - Improves data and survey quality
- Two avenues to overcome this: NVivo & two AI models

PROCESS (NVIVO)

- What is NVivo?
- Review of previous DCMRB reports
 - Informal open coding
- Consulted with branch members to discuss problems they noticed or are interested in
- Through consultation, open coding previous studies, and iterative development, an NVivo codebook was developed to perform an initial coding of documents by hand

BUILDING THE CODEBOOK

- Four main aspects
 - Codes: Themes that are of interest to the DCMRB
 - Question Bank: Written questions in appendices of reports
 - File Attributes: Traits for the reports and presentations that are being analyzed
 - Cases: Units of observation – Surveys

⊕ Name	↔ Files	References
⊕ ○ Methods	6	43
⊕ ○ Procedural Information	4	21
⊕ ○ Tools	3	18
⊖ ○ Improvable Areas	6	92
⊕ ○ Layout	1	1
⊕ ○ Instrucional Clarity	1	1
⊕ ○ Question Design	2	3
⊕ ○ Visual Clarity	0	0
⊕ ○ Functionality	1	2
⊕ ○ Burden	4	7
⊕ ○ Sensitivity	0	0
⊖ ○ Scope	4	13
○ Current collection is t	1	1
○ Current collection is t	0	0
○ Issues with current le	3	6
○ Support for current s	3	6

NVIVO CODEBOOK & ACTIVE CODING

Name of Code	Description of Code	Options	Example	Natural Language Question
Methods				
Sampling Frame	Who are we interested in reaching out to in this study	<ol style="list-style-type: none"> Completed a specific Census survey in the past Company meets certain size criteria Company meets certain complexity 	We are interested in what establishments were reached out to within these reports for their respective studies and why. Find and summarize who was reached out to based on	

In-Scope Businesses and Respondents

Eligibility Criteria:

- Sampled in at least two in-scope surveys
- In at least two industrial sectors
- More than one establishment

	Phase 1	Phase 2
Number of industries*		
Three or fewer	16	25
Four or more	5	5
Number of establishments*		
30 or fewer	9	19
31 or more	12	11

*Numbers may not sum to total interviews because of missing data.

Shape your future START HERE >

2020CENSUS.GOV

Codes

Name	Files	References
Methods	6	43
Sampling Frame	5	14
Completed a specific	1	1
Company meets cert	4	8
Company meets cert	1	1
Company has other c	2	4
Other reason for incl	0	0

<Files\\ICES Presentation updated 5-11-21> - § 1 reference coded [0.08% Coverage]

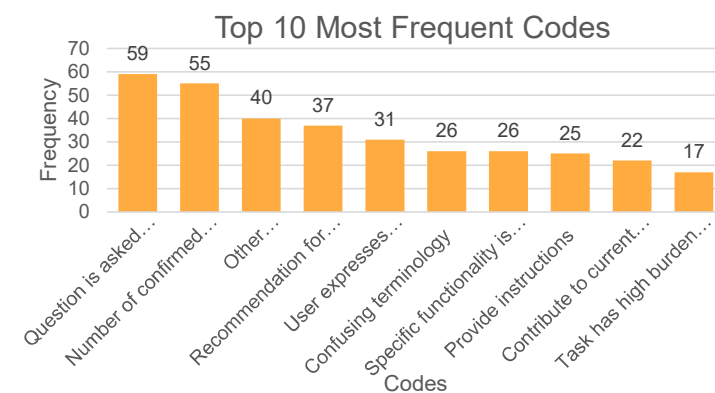
Reference 1 - 0.08% Coverage

Sampled in at least two in-scope surveys

NVIVO RESULTS: A GLANCE

CODES	REFERENCES	PROPORTION
Improvable Areas		
Instructional Clarity	43	10%
Question Design	113	27%
Data Accessibility	121	29%
Functionality	55	13%
Burden	21	5%
Layout	22	5%
Scope	15	4%
Unit Problem	15	4%
Sensitivity	13	3%
Visual Clarity	4	1%
Sub Total	422	100%
Recommendations		
Instructional Recommendations	64	21%
Question design recommendations	120	39%
Functionality Recommendations	69	23%
Data Accessibility Recommendations	17	6%
Layout Recommendations	17	6%
Burden Recommendations	7	2%
Scope Recommendations	5	2%
Visual Design Recommendations	2	1%
Sensitivity Recommendations	2	1%
Unit Problem Recommendations	1	0%
Sub Total	304	100%

CODE	FREQUENCY
Question is unclear in what is asking	23
Question is asked appropriately and is understood by the user	59
Other mentions of question design	12
Question options are unclear	9
Question options are non-exhaustive	9
Question is in inappropriate place	1
Question is double-barreled	0
Question contains bias	0
Question is loaded	0
Question options overlap	0
CODE	FREQUENCY
User expresses difficulty in accessing certain data	31
User mentions having to collaborate with others to obtain data	21
Mentions of disparities in Census' collection of data and how unit keeps record of that data	19
User mentions ease in accessing certain data	8
Mentions of misalignment of understanding of data between Census and respondent	14
Mentions of misunderstanding of how the Census thinks a unit is run vs. how the unit is actually run	14
other mentions of data (in)accessibility	10
Mentions of Census having information for a portion of a unit and respondents are unsure to access data just for that portion or for the larger unit	4
Data must be cleared or approved before being released to the Census	0



CODE	FREQUENCY
Question is asked appropriately and is understood by the user	59
Number of confirmed participants	55
Other recommendations in designing questions	40
Recommendation for integrating new functionalities	37
User expresses difficulty in accessing certain data	31
Confusing terminology	26
Specific functionality is unclear to the user	26
Provide instructions	25
Contribute to current instructions	22
Task has high burden or increases burden for users	17



Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau.

NVIVO RESULTS: CRITICAL FINDINGS

- Accessing data is difficult due to the number of people and departments required to get the data
 - “The question isn’t difficult, it’s difficult to find the right person who has the right answer”
- Disparities of how the unit stores data vs. how the Census collects it
- Integrating new functionalities, such as downloadable PDFs, search engines, and navigational tools helps reduce burden – but only if they’re clear in what they do

NVIVO RESULTS: CRITICAL FINDINGS

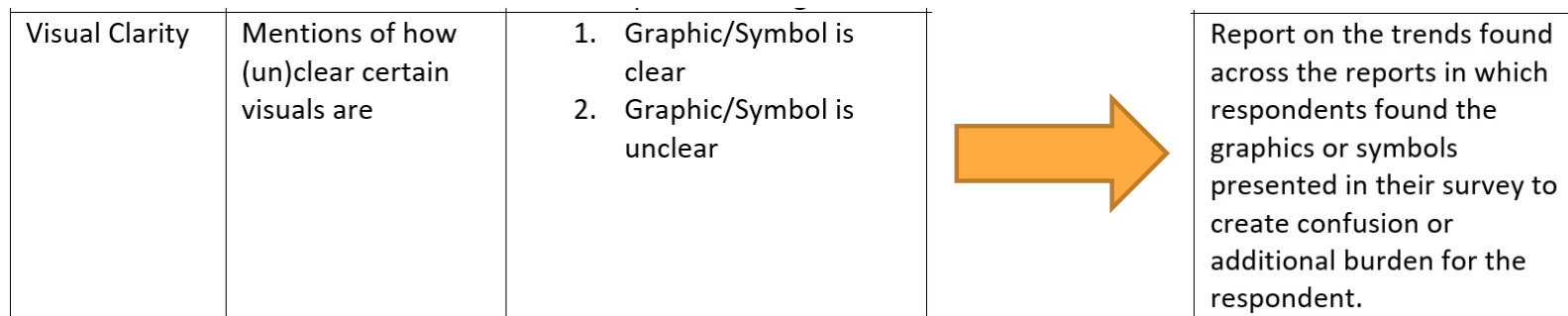
- Having additional yet concise instructions or definitions is recommended much more frequently than removing instructions or definitions
 - This includes instructions on buttons
- Like questions should be put together, with the most burdensome at the end
- Many functionalities still need to be improved
- A 5-page document summarizing all codes is available at the path provided at the end

PROCESS (AI)

- Economy Wide Statistics Division (EWD) and Reveal Consulting Team introduction to AI
 - Internal Census model: 30 Billion Parameters
 - Open AI model: Over 1 Trillion Parameters
- Natural Language questions were developed to operationalize themes found and to test the effectiveness of the AI models
 - Natural Language Questions: Human designed questions using colloquial terminology

FROM CODES TO NL QUESTIONS

- Operationalizing these codes depends on the complexity of the topic and the subcodes to follow
 - These needed to be iteratively developed with the AI to ensure all the appropriate themes were found with the appropriate level of specificity
- Example of one of the simpler operationalizations:



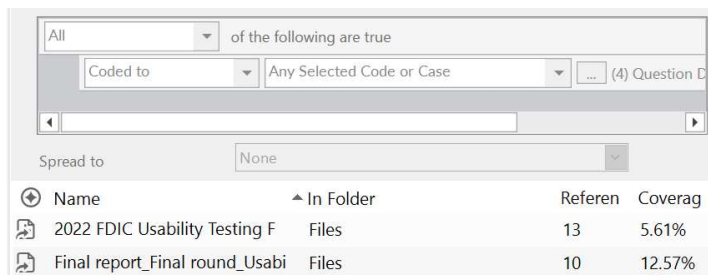
TESTING THE NL QUESTIONS EXAMPLE

Question: Within these reports, what are the some of the common themes found when examining questions? What types of questions were found to be poorly designed for users and why?

Documents: *Findings and Recommendations for Usability Testing for the 2022 SBLS and Usability Evaluation for the 2022 Economic Census*

NVIVO FINDINGS

Can I answer this in NVivo?



Name	In Folder	Referen	Coverag
2022 FDIC Usability Testing F	Files	13	5.61%
Final report_Final round_Usabi	Files	10	12.57%

- Query was run on both documents searching for “Instructional Clarity” and “Question Design”
 - Found confusion with gross annual revenue and other vocabulary like “Lending for business purposes”
 - Instructional limitations on using tools to answer questions (such as screener questions or using search tools)
 - Screener placement position wasn’t optimal to reduce burden

AI RESULTS (CENSUS MODEL)

- Results are broad
 - “Participants highlighted problematic areas such as instructions/features that were difficult to understand”
 - “Common themes include navigation, layout, terminology, and confusing steps”
- Hallucinations and Falsehoods
 - “The usability study conducted with 11 bankers...”
- Less Practical Citing
 - “The top-level categories were readily recognized, as these steps were most often cited.”

AI RESULTS (OPEN MODEL)

- New formatting:
 - **“Ambiguity and Misinterpretation of Questions:** Some questions were not clear or straightforward, leading to misinterpretation. For instance, in the SBLs survey, participants misinterpreted questions regarding gross annual revenue (GAR), causing confusion about the data to be reported.”
- Still citing studies, just less direct
- Much more complex and synthesized answers
- Building upon previous questions and work

CENSUS AI VS. OPEN AI

Census AI

- Can handle sensitive documentation
- Provides citations from original documents
- Slower
- Less likely to synthesize non-explicit information
- Less complex results

Open AI

- Cannot handle sensitive documentation
- Does not provide citations
- Faster
- Able to synthesize and produce information
- Complex and thoughtful responses

NVIVO VS. AI

NVivo

- Slower
- Relies on human precision
- Human dependent synthesis

AI

- Faster
- Can miss or pick up information that humans do not
- Potential bias
- Researchers must know the studies either way
 - Hallucinations

FUTURE WORK

- More studies need to be coded and integrated into the NVivo project before FY21 and those without DAO Approval
- Analysis should be done yearly to incorporate new findings
- References should be made to NVivo files when designing new studies
- The internal Census model needs further development and testing

PROBLEM

- Based on this research, question clarity and developing more accessible data collection efforts are two of the top areas for improvement
- NVivo project can be expanded to meet new needs
- Prime opportunity for integrating AI into Census Work
 - Neither model is perfect, but both offer insights to their potential

QUESTIONS?

Contact Information

Census Email: jacob.a.depignet@census.gov

Personal Email: depignet.23@osu.edu

LinkedIn: <https://www.linkedin.com/in/jacobdepignet>

AI RESULTS (CENSUS MODEL)

"The usability study conducted with 11 bankers evaluated the web instrument's performance in terms of efficiency, accuracy, and user satisfaction. Participants highlighted problematic areas such as instructions/features that were difficult to understand.

Common themes found in question may include issues with navigation or layout, difficulty understanding certain terms used in the survey, and confusion regarding certain steps in the survey process.

Direct quotes:

The findings from the usability evaluation were based on session observations across all participants. The findings cited are those that occurred most frequently or could have potential impacts on the user experience or data quality.

the types of questions respondents could expect on subsequent screens (for SU Main Mailing cases). Although most participants seemingly perused through the content, they did highlight the steps for reporting to the survey and the options for reporting.

The top-level categories were readily recognized, as these steps were most often cited.

For single unit participants, whose screen outlined that they would report to organization/classification questions, they did comment on the types of questions they would be reporting, based on the sub-listing provided under Step 1-Report.

The 'Suggested Personnel' screen was also the only screen with a link to download the PDF worksheet.

Most the participants that evaluated this screen did not note the PDF availability. However, when later prompted, many participants felt that having a PDF of the entire survey would be helpful, and some of those participants noted that having smaller PDFs for each individual module would also be helpful.

It was not obvious to all participants that the screeners were only a small part of a larger survey. The nature of the screener questions also caused some participants to anticipate a more quantitative data heavy survey. The write-in box for 'Other' answer choices are located outside of the grid.

Some participants did not realize that answers to the screener section were not changeable once submitted."



Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau.

AI RESULTS (OPEN MODEL)

“Here are some common themes found when examining the questions within the two reports:

Ambiguity and Misinterpretation of Questions: Some questions were not clear or straightforward, leading to misinterpretation. For instance, in the SBLS survey, participants misinterpreted questions regarding gross annual revenue (GAR), causing confusion about the data to be reported.

Complex or Technical Terminology: Both reports highlighted instances where users struggled with complex or technical terms. In the EC2022 survey, not all users understood the 9-digit NAICS codes. Such terms can cause confusion or errors in response.

Assumptions about User Knowledge or Abilities: Some questions seemed to assume a certain level of user knowledge or ability that may not have been present. For example, in the EC2022 survey, participants were expected to navigate and utilize the machine learning search tool, but not all found it intuitive.

Lack of Clear Instructions or Guidance: In both surveys, users struggled with questions where instructions or guidance were unclear or lacking. In the EC2022 survey, users had difficulty recognizing the function of the NAICS summary display, and in the SBLS survey, participants were uncertain about the immutability of their responses once submitted.

In terms of poorly designed questions, a few types stand out:

Questions with Ambiguous Wording: For instance, in the SBLS survey, the use of the term “for any GAR” instead of “by any GAR” led to confusion.

Questions without Adequate Explanation or Context: Questions that did not provide sufficient context or explanation were challenging for users. For instance, in the EC2022 survey, participants found it non-intuitive to return to the pre-list screen after making a selection.

Questions Requiring Technical Knowledge: Questions that required an understanding of specific technical terms or codes (like the 9-digit NAICS codes in the EC2022 survey) proved difficult for some users.

Questions with Complex Input Mechanisms: Questions that required users to interact with complex elements (like the machine learning search tool in the EC2022 survey) were challenging for some participants.”



Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau.