Optimizing Response Rate in Customer Satisfaction Survey

Datasci 241 - Final Project

Connor Ethan Yen, Matthew Dodd, Leena Bhai, and Heather Rodney

2023-04-05

Abstract

Internal IT often seeks to improve customer experience and satisfaction. However, targeting satisfaction through customer satisfaction surveys (CSS) is complicated by the trade off between survey specificity and response rates. Very simple CSS achieve high response rates but provide teams with little actionable direction, while highly detailed CSS risk statistical bias and uncertainty due to low response rates. In this study, we focus on internal IT at the midsize pharmaceutical company and experiment with different types of CSS—ranging from simple thumbs-up/thumbs-down to multiple targeted ordinal questions—to find the type of survey that provides enough actionable and statistically representative information for IT-teams to perform productive work. We find that as surveys increase in length from a baseline of 1 question to 3 and 5 questions, their response rates fell by 17% and 34.6% respectively. We further find that there is no statistically significant effect on response rate in using a binary or likert response scale.

Introduction

The tradeoff between CSS specificity and response rates is well documented in the literature. Factors such as question length/amount (e.g., Burchell & Marsh 1992, Galesic & Bosnjak 2009, Revilla & Ochoa 2017), survey phrasing (e.g., Ograjenšek & Gal 2011, Nicolini & Valle 2011), and user demographics (e.g., Revilla & Höhne 2020) have all been considered. Nevertheless, little work has been done to quantitatively characterize the tradeoff between CSS specificity and response rates from an experimental point of view.

This study partners with an unspecified midsize pharmaceutical company to experiment on the CSS complexity versus response rate tradeoff. IT teams at this company are interested in improving the quality of service delivered and hope to do so by using customer satisfaction surveys. While complex surveys allow IT teams to parse out individual pain-points, the literature suggests that these surveys also lead to low response rates. On the other hand, simple surveys may achieve high response rates, but lack the specificity required for teams to identify where to improve. This study characterizes survey complexity across two dimensions: number of questions and response scale. We experiment with how response rate is effected by the number of questions asked and the response scale deployed.

Research Questions and Hypothesis

1. **How does increasing the number of questions on a survey effect response rate?** We hypothesize that increasing the number of questions on a survey decreases survey response rate.
2. **Does changing the response scale from likert to binary increase response rate?** We hypothesize that changing the response scale from likert to binary increases response rate.

Experimental Design and Mesurement

When a user raises an issue to the IT help desk, their issue is raised as a "ticket." These tickets document details relevant to the user's case. With the ticket raised, the IT help desk assigns a lead "agent" to work on the issue. Users can check the status of their tickets through a web portal. Once the ticket has been resolved, the user can close the ticket through the portal.

One of the features provided by the service management platform is collection of satisfaction surveys. If enabled, users are directed to a survey page immediately after clicking the "close" button on their ticket. We leveraged this feature to perform our experiment. Our team had the advantage of circumventing prior user bias due to former experiences with the survey process since the survey feature had been disabled in early 2020. As the company nearly doubled in size between 2020 and 2023, most users of the ticket platform had never been exposed to internal help desk surveys before this experiment was run.

The service management platform has a centralized data repository for tickets with rows indexed by unique ticket IDs. Columns for if the survey was sent (True/False), question-1 response (1-5 or +/-), question-2 response (1-5 or +/-), question-3 response (1-5 or +/-), question-4 response (1-5 or +/-), and question-5 response (1-5 or +/-) were added. Thus, survey results populate in the database as field values. We define response to be True if at least one of the values in the "question-1 response" through "question-5 response" columns is not empty. In the case of a positive response, the user successfully clicked "submit" on the survey page. Note that all questions on the survey were required so users who successfully submitted a survey responded to all the survey questions asked. We define response to be False if the "question-1 response" through "question-5 response" columns are all empty. A derived column corresponding to survey response (True/False) was added for ease of analysis.

The column tracking if the survey was sent automatically populates with "False." If upon closure of a ticket, the service management platform directs the users browser to the survey page, the "survey sent" column switches to True. In cases where the user did not respond to the survey (i.e., the survey response column is False), the user either closed out of the survey page or had a network timeout. In cases where the survey redirect was blocked by the user's browser, the "survey sent" column does not always record a consistent value. However, browsers are configured at the enterprise level with default privileges enabling the survey redirect. While it is possible for individual users to have configurations that block the survey redirect, this issue should only exist for a small minority of users.

For this study we deployed a 2 by 3 design with dimensions of response granularity and number of questions. Response granularity had two levels: binary (a simple thumbs-up and thumbs-down) and likert (one to five scale). The surveys further had either one, three, or five survey questions. The following questions were used:

1. How satisfied are you with your overall experience of the service desk?
2. How satisfied are you with the resolution speed of your ticket?
3. How satisfied are you with the degree of communication provided during the resolution process?
4. How satisfied are you with the usability of the platform?
5. How satisfied are you with the resolution provided by the service desk?

Surveys with one question asked question 1, surveys with three questions asked questions 1 through 3, and surveys with five questions asked questions 1 through 5. Users who received the binary scale option had a green thumbs-up icon and a red thumbs-down icon under each question. Users who received the likert scale option had a 1 through 5 scale under each question with 1 corresponding to "not satisfied" and 5 corresponding to "very satisfied." The syntactic question format of "how satisfied are you with _____?" was chosen intentionally to keep the complexity of each question constant.

Three covariates are included in the analysis: department, salary grade, and resolution time. Department and salary grade are anonymized in this report due to privacy concerns. For the department covariate, the company was divided into three units: manufacturing and research, business, and IT. For the salary grade, salaries were grouped according to salary quantiles provided by HR corresponding to 0-33%, 33-66%, and 66-100%.

Randomization and Enrollment

Surveys were only enabled for each user once. That is, for each user (defined by Active Directory profiles), only the first ticket "closed" during the experimental timeframe received a survey redirect. Randomization was conducted at ticket closure. The process flow is as follows:

1. The user clicks "close" on their ticket.
2. The platform database collates the user ID on tickets where a survey was sent (i.e. "sent survey" == True).
3. If the user ID associated with the ticket is NOT in the collated list from step-2, the user is randomly redirected to one of the six surveys (i.e., treatments).
4. If the user ID associated with the ticket is in the collated list from step-1, the user is redirected to their profile home page.

This method ensures that each "participant" of the study is unique and avoids dependence between data points. Note that because even populations exist for department and salary grade levels, stratified or quota sampling was not deemed necessary.

The power analysis for this study demonstrated that to observe a large true effect, our sample size should be over 500, and to observe a small true effect, our sample size should be over 2500. Our study involved 672 individuals providing us the power to observe a large true effect.

Experimental Timeline

The experiment was conducted over a 3 week period from February 13, 2023 to March 3, 2023 (PST). This period was chosen to not correspond with any major business processes (e.g., financial quarter ends) that may have uniformly influenced the urgency of work and subsequent inclination to engage in non-critical tasks.

The confined time frame also reduced spillage between treatment groups since as the experimental window increases, discussion between coworkers might cause individuals to realize that different satisfaction surveys were being conducted. Nevertheless, because survey response is not a metric reported at an executive level through which survey response would be heavily encouraged, we do not expect spillover effects to be significant.

Observations and Outcome Measures

We monitored various outcome measures, including the respondents' engagement with the survey, which is determined by their participation or lack thereof. Given the internal nature of the employee surveys, we recorded several key attributes: the employee's department, their salary grade, and the duration required to address the issue outlined in the ticket. Additionally, we took into consideration two treatment factors - the quantity of survey questions and the response scale utilized. The following table provides a detailed breakdown of the outcomes in relation to the two treatment aspects.

Response Scale	Number of Questions	Total Count
binary	One	107
binary	Three	112
binary	Five	113
likert	One	116
likert	Three	121
likert	Five	103

Table 1: Total count of responses by number of questions and response scale

## Time Taken to Complete a Request
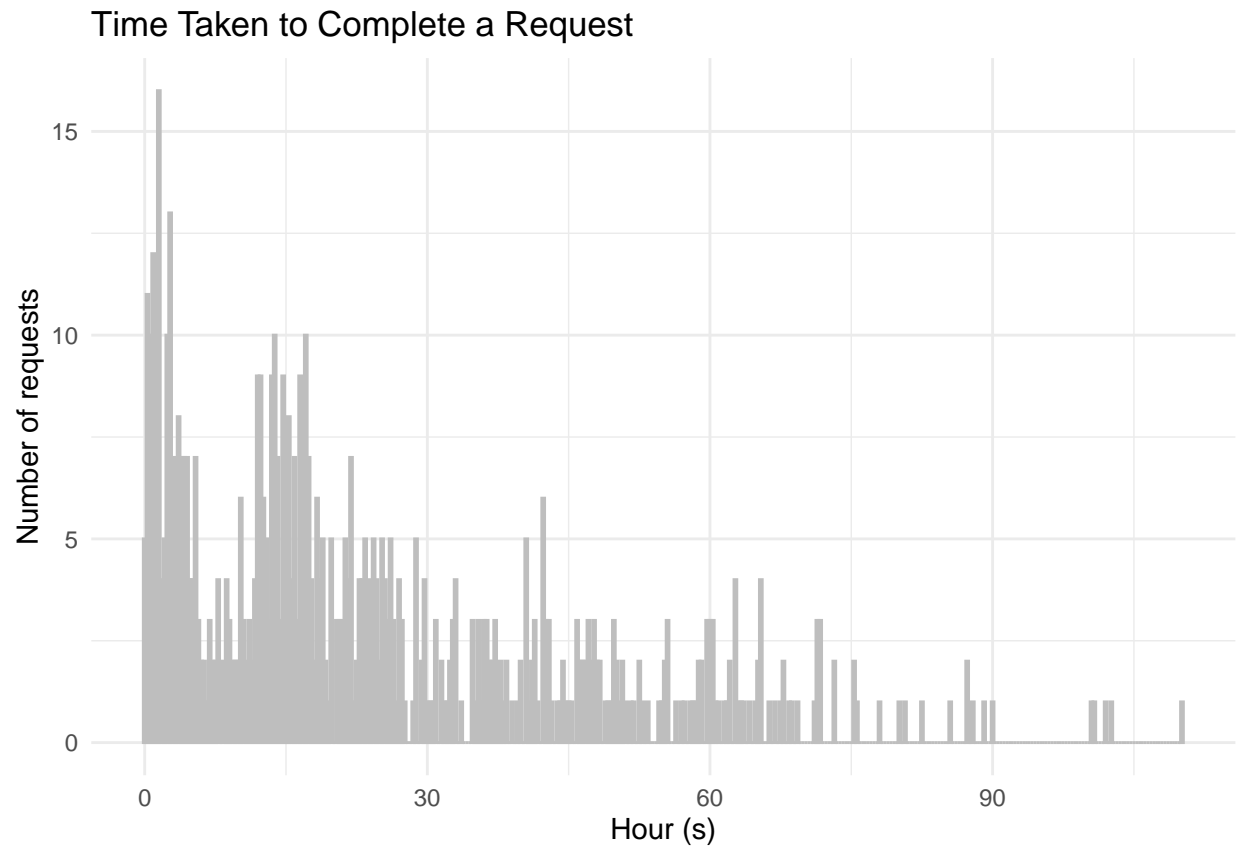
The histogram exhibits a right-skewed distribution, with most tickets being resolved within a day, while some tickets taking about three days to resolve . Therefore, it is recommended to apply a logarithmic transformation to the data for more accurate analysis.

Data Completeness

Results

```
## 'summarise()' has grouped output by 'survey_response_scale',
## 'survey_number_of_questions'. You can override using the '.groups' argument.
## Joining, by = c("survey_response_scale", "survey_number_of_questions")
```

Percentage of Responses by Length and Type of Questions
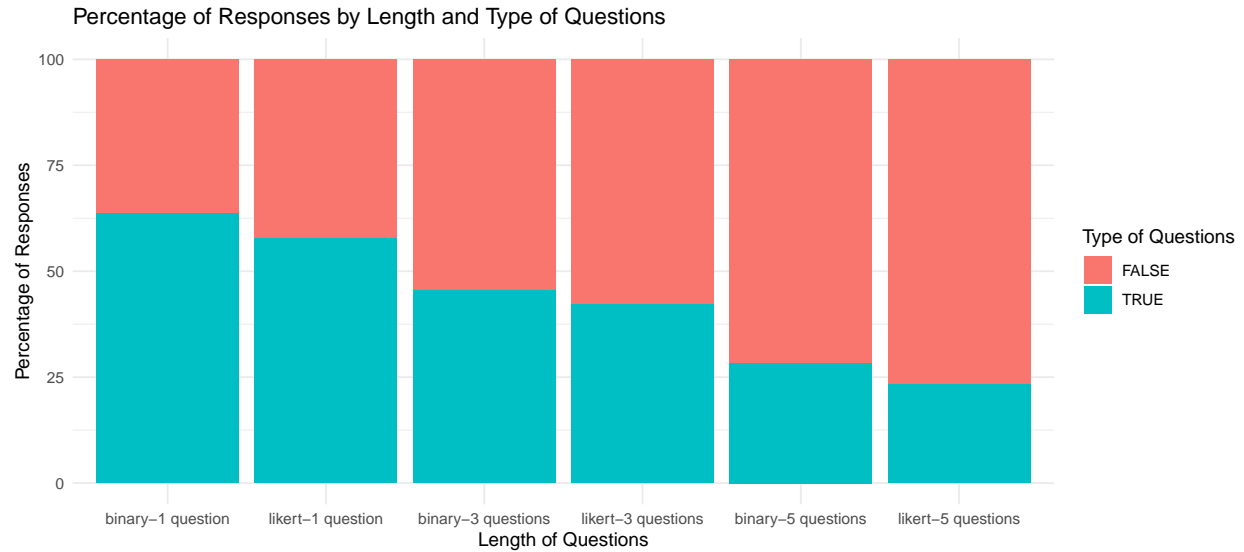
Our analysis reveals a negative correlation between survey complexity and response rate. The simplest survey type, featuring a single binary question, yields the highest response rate at 63.5%. In contrast, the most intricate survey, consisting of a 5-item Likert scale questionnaire, demonstrates the lowest response rate of 23.3%. This suggests that as the survey's complexity increases, participant engagement tends to decrease.

```
## 'summarise()' has grouped output by 'department'. You can override using the
## '.groups' argument.
## 'summarise()' has grouped output by 'salary_grade'. You can override using the
## '.groups' argument.
```

Response Rate by Days to Complete Requests
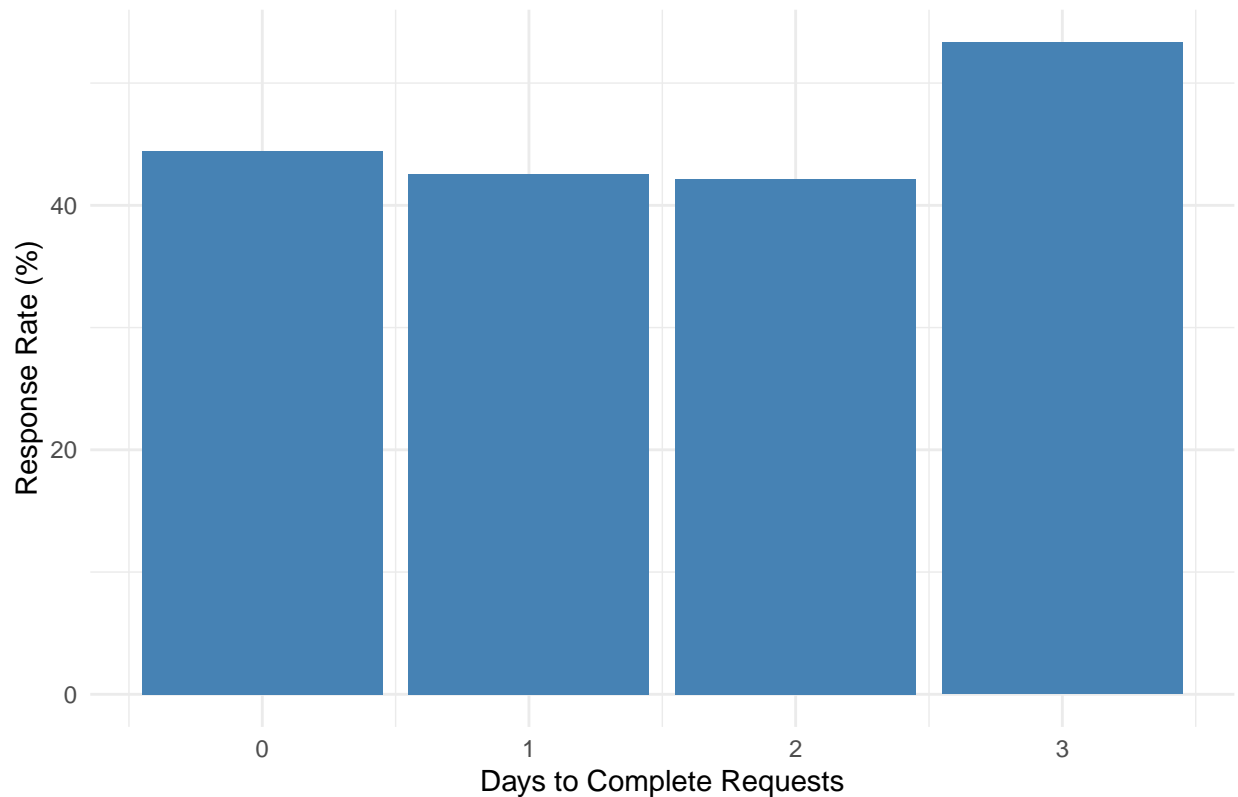
Upon analyzing the relationship between ticket resolution time and survey response rates, we observed an inconsistent pattern. Response rates are 44% for same-day resolutions, slightly lower at 42.5% for one-day resolutions, and 42% for two-day resolutions. Interestingly, the response rate significantly increases to 53% for three-day resolutions. This suggests that complex factors influence the relationship, warranting further investigation.

Regressions

```
##      department salary_grade ticket_res_time_hours survey_number_of_questions
##   1:          B      grade 3                  3.21                3 questions
##   2:          B      grade 2                 23.35                3 questions
##   3:          C      grade 3                 32.98                 1 question
##   4:          B      grade 3                  1.14                 1 question
##   5:          A      grade 3                 42.21                3 questions
##  ---
## 668:          B      grade 3                 62.81                 1 question
## 669:          C      grade 3                 66.86                5 questions
## 670:          C      grade 3                 25.38                5 questions
## 671:          A      grade 2                  2.50                 1 question
## 672:          C      grade 1                 16.65                5 questions
##      survey_response_scale responded_to_survey
##   1:                likert                TRUE
##   2:                likert                TRUE
##   3:                binary                TRUE
##   4:                binary               FALSE
```

```
##   5:              binary             FALSE
## ---
## 668:            likert              TRUE
## 669:            binary             FALSE
## 670:            binary              TRUE
## 671:            likert              TRUE
## 672:            likert             FALSE


##
## ===========================================================================
##                                                             Dependent variable:
##                                                       ----------------------------
##
##                                                          (1)       (2)       (3)
## -------------------------------------------------------------------------------
## survey_number_of_questions3 questions                -0.170*** -0.180*** -0.176***
##                                                        (0.045)   (0.064)   (0.065)
##
## survey_number_of_questions5 questions                -0.346*** -0.352*** -0.349***
##                                                        (0.046)   (0.064)   (0.065)
##
## survey_response_scalelikert                                     -0.058    -0.056
##                                                                 (0.064)   (0.064)
##
## salary_gradegrade 2                                                       -0.076
##                                                                          (0.086)
##
## salary_gradegrade 3                                                       -0.117
##                                                                          (0.078)
##
## departmentB                                                               0.067
##                                                                          (0.052)
##
## departmentC                                                               0.024
##                                                                          (0.049)
##
## log(ticket_res_time_hours)                                               -0.004
##                                                                          (0.014)
##
## survey_number_of_questions3 questions:survey_response_scalelikert  0.019    0.013
##                                                                 (0.090)   (0.090)
##
## survey_number_of_questions5 questions:survey_response_scalelikert  0.008    0.005
##                                                                 (0.091)   (0.092)
##
## Constant                                              0.605***  0.636***  0.713***
##                                                        (0.032)   (0.046)   (0.104)
##
## ===========================================================================
## ===========================================================================
## Note:                                                    *p<0.1; **p<0.05; ***p<0.01
```

In order to analyze the data we specified three different regression models. Our initial baseline model regresses our binary "responded to survey" variable on the number of questions contained in the survey

which took the value of 1, 3 or 5. We found that surveys with one question had a response rate of 60.5%. The regression showed that as survyes increased in length from 3 to 5 questions their response rates fell by -17 and -34.6% respectively, both statistically significant results.

Our second specification added an interaction term denoting the response scale of the survey, likert or binary. Interestingly, this specification shows that there is no statistically significant effect of response scale to response rates for any of the survey lengths.

Our final model adds to the interaction specification by adding controls for respondant department, salary grade and the log of a tickets response time. This log transform was used to the the right hand skew in our ticket resolution time series. The results again show that it only seems that survey length has a statistically significant effect on response rates, however the added controls do increase our point estimates for the decrease in response rates for 3 and 5 questions to -17.6 and -34.9% respectively

Conclusions

Limitations and Future Enhancements

Appendix A