Cooperative Congressional Election Study

Guide for
CCES Page Timing Data

prepared by *Ivelisse Cuevas-Molina, Research Assistant*
University of Massachusetts Amherst

Stephen Ansolabehere, PI
Harvard University

Brian Schaffner, Co-PI
University of Massachusetts Amherst

May 9, 2017

1 Introduction

The Cooperative Congressional Election Study (CCES) has put forth page timing data from the 2012 and 2014 CCES for public use. This guide provides users with a description of this data and general practices for its use in conjunction with the CCES Common Content from the the aforementioned survey years. The CCES is an online biennial large sample survey administered by YouGov with a Pre Election and Post Election wave; further details about the sampling methodology and how the CCES is conducted can be found in the guides corresponding to the 2012 and 2014 CCES Common Content data.

2 Background

A response latency is the time it takes a respondent to answer a question. Response latencies (sometimes referred to as response times) were first measured in computer-based and computer-assisted surveys, commonly measured as the time span between the moment after a question is asked by an interviewer and the moment a respondent starts to provide an answer. Technically, response latencies measure the reaction time for a survey participant's response to a question. More specifically, response latencies are the operationalization of the mental process that occurs during response formation (Mayerl, 2013).

However, self-administered online surveys like the CCES measure response latencies differently because there is no interviewer involved and the medium through which the questionnaire is completed has its own constraints. Response latencies in the CCES are measured in the form of page timings because each question or group of questions is presented to respondents on an individual page of a web browser. Page timings measure the time during which a question or group of questions becomes visible to the respondent, the respondent reads the question(s), formulates a response, selects or writes a response, and then moves on to the next page in the questionnaire. This time measurement is unobtrusive because CCES participants are unaware that YouGov's online questionnaire platform is counting how long it takes for the respondent to go through each page of the questionnaire.

Response latencies have been used to measure cognitive effort, information processing and attitude accessibility in relation to social and political phenomena (Mulligan et al., 2003; Schaffner & Roche, 2016). Shorter response latencies can indicate less cognitive effort and high attitude accessibility or strong attitudes while longer response latencies indicate more cognitive effort and less attitude accessibility or weaker attitudes.

3 Complications in Using Page Timings

Cognitive effort is not the only factor that can affect the length of response latencies. Indeed, page timings may be affected by (1) factors unique to the question, (2) factors unique to the respondent, and (3) idiosyncratic interventions during the interview process. Questionnaire construction, mode of administration and individual factors can affect the length of response latencies.

In the case of questionnaire construction, longer wording, complex questions, and questions with more response alternatives will result in lengthier page timings. The more options the respondent has to consider the more time she will spend making a judgment on how to respond to the question presented. Thus, one cannot simply compare response times across questions without accounting for such differences.

Additionally, individual characteristics of survey respondents can also affect the time it takes them to answer questions in general. In particular, people who read at different rates will take more or less time to answer questions in general. These baseline differences must also be accounted for when analyzing response timings.

Finally, an additional issue in self-administered online questionnaires is the fact that participants can take breaks during the completion of the survey and have no limits to the length of time they

can take to answer a question. In the CCES 2010-2014 Panel Study respondents were asked if they engaged in other activities during the time they answered the survey. Forty-five percent of the total 9,500 respondents in that study said they engaged in activities like doing chores, taking a break, dealing with children, and talking on the phone, among others. Ansolabehere and Schaffner (2015) find that these frequent distractions and interruptions during the completion of the survey do not affect the quality of the data collected by the CCES, but they do affect the length of page timings. If a respondent walks away from a survey while taking a break, the page timing for that question no longer represents a good measure of cognitive effort. Thus, one must be careful to address extreme outliers in the page timing data.

4 Analyzing Page Timing Data

The creation of a baseline timing measure is recommended in the response latency literature in order to control for individual differences in response speed. Mayerl (2005) defines baseline timing "as the general mental speed that a person needs to answer questions independent of the content of the question" (p.4). Baseline timing can be computed by the mean response times of filler questions. Filler questions are any questions that are "not connected theoretically or thematically to the interesting target questions" (Mayerl, 2005: p. 4). Those using the CCES may choose questions from both the Pre Election and Post Election waves of the study to calculate a baseline page timing that reflects an individual's general response speed for the survey. Using response latencies from questions that have a similar construction to that of the questionnaire item of interest will yield a baseline page timing that is indicative of the average time it takes a respondent to answer a certain type of question. For example, if the response latency of interest corresponds to a single choice question with its own individual questionnaire page, then the items in the baseline timing should correspond to other single choice questions with their own individual questionnaire page.

Because respondents sometimes take a break from the survey and come back to complete it later, latencies of minutes, hours or even days are sometimes recorded for the CCES. Data cleaning and/or transformation is not uncommon with regard to response latencies, because this data is often skewed by extremely long outliers that "can increase the mean, inflate the standard deviation, and change the measures that shape" this data (Ratcliff, 1993). The goal of data cleaning and transformation is to "reduce the effects of potential outliers while eliminating as little as possible of the data of real interest" (Ratcliff, 1993). If researchers choose not to transform or clean the data, then the best central tendency measure for response latencies is the median and quantile regression is most appropriate for correlational analysis. Still, if researchers want to use the mean as the central tendency measure and Ordinary Least Squares regression they must transform the response latency data.

Data transformation can be achieved by trimming outliers from the data to minimize their effect. Trimming involves eliminating all observations above a determined cutoff point from the response latencies of interest. The cutoff point employed can be values in the top percentile of response latencies, values a certain number of standard deviations above the mean, or values above a specific time with meaning to the questionnaire item of interest (Fazio, 1990; Ratcliff,1993). Another method for transforming response latency data is windsorizing. Windsorizing involves replacing the values above the cutoff point with the values at the cutoff point. This method allows the researcher to avoid losing the observations that would be excluded from analysis with the implementation of a trimming technique (Fazio, 1990).

For more information on best practices for dealing with response timings, see Ratcliff (1993) and Mayerl (2005).

5 Connecting Page Timing and Common Content Datasets

The two available CCES page timing datasets include page timing data corresponding to the responses of every individual respondent in the CCES, be it the 2012 or the 2014 CCES. Page timings in the CCES are measured in seconds (s), with precision to the third decimal point or millisecond

(0.000). These data sets can be easily merged with their corresponding CCES Common Content datasets with a "one-to-one" merge on the "Case ID" variable named V101 in both the 2012 and 2014 CCES. Once the response latency data is merged with its corresponding CCES Common Content researchers should be able to identify whether their questionnaire item of interest has a corresponding page timing item fairly easily. Page timing items commonly have the variable name of their corresponding question included in their own variable name. For example, the question labeled "National Economy" has the variable name "CC302" and its corresponding page timing has the variable name "page_CC302_page_timing ".

Please note that page timing data is not available for every item in the CCES, but it is available for any questionnaire item that was answered while the CCES was in the field. Some questionnaire items that do not have page timings may be demographic questions that are part of the YouGov profile battery.

Missing values or values of zero (0) within page timing items indicate that the corresponding questionnaire item was skipped by the respondent. Those response latencies with a value of zero (0) should be set to missing and excluded from analysis.

6 Reference

Schaffner, Brian F., and Cameron Roche. "Misinformation and Motivated Reasoning Responses to Economic News in a Politicized Environment." Public Opinion Quarterly (2016): nfw043.

Fazio, Russell H. "A practical guide to the use of response latency in social psychological research." Research methods in personality and social psychology 11 (1990): 74-97.

Mayerl, Jochen. "Controlling the baseline speed of respondents: an empirical evaluation of data treatment methods of response latencies." In Proceedings of the Sixth international Conference on Logic and Methodology, pp. 1-20. 2005.

Mulligan, Kenneth, J. Tobin Grant, Stephen T. Mockabee, and Joseph Quin Monson. "Response latency methodology for survey research: Measurement and modeling strategies." Political Analysis (2003): 289-301.

Ratcliff, Roger. "Methods for dealing with reaction time outliers." Psychological bulletin 114, no. 3 (1993): 510.