Correlates of Response Latency on a Web Survey

Benjamin Phillips[1], Stanislav Kolenikov[1],
Elaine Howard Ecklund[2], Allison Ackermann[1], and Ho Won Cheng[1]

[1]Abt SRBI, 55 Wheeler Street, Cambridge, MA 02138
[2] Department of Sociology, Sewall Hall 372, Rice University, Houston, TX 77251

Abstract
We examine factors associated with response latency in a web survey: the Rice University Religion and Science in International Context (RASIC) survey of members of biology and physics departments in Italian universities and research institutes. We found some evidence that respondents decrease their attention and start taking cognitive shortcuts with longer questions, as evidenced by a linear increase in latency with the question length, measured in the number words, giving way to sublinear increase beyond about 100 words in question stem. We also found evidence for decreased attention shown by lower latency beyond the first 15 or so minutes of the survey, followed by respondents getting tired beyond 60 minutes of the survey. Opinion items had greater latency than factual items. Items requiring averaging took longer than other items. Numeric and text entry items had greater latency than radio button items. Items in matrix (not measured separately) had higher latency than single items. Surveys taken in Italian (the native language of the survey population) had lower latency than those taken in English. These analyses provide important context for the perhaps simplistic interpretations of response latency: low latency being a desirable trait for items but undesirable for a respondent. Data collection utilized for this paper was funded by the Templeton World Charity Foundation, grant TWCF0033.AB14, Elaine Howard Ecklund, PI, Kirstin RW Matthews and Steven W. Lewis co-PIs.

Key Words: response latency, nonsampling error, total survey error, bilingual survey, mixed modeling, response time

1. Motivation

Among the paradata (Kreuter 2013) items that the recent developments in survey data collection technology allow harvesting is response latency. In web surveys, time stamps associated with the respondents' actions, such as progressing to the next page or selecting a response option, can be collected along with the primary survey data, and response latency on a given item can be computed as the time elapsed since the previous time stamp.

Response latency can be considered in terms of problematic respondents and problematic items. Respondents that are progressing through the survey too quickly may be taking cognitive shortcuts, e.g., not reading all of the question text, not reading all of the response categories, selecting uninformative categories such as "Don't know", and otherwise not putting much cognitive effort into the survey (Krosnick 1991). On the other hand, items with high latency may indicate problems, e.g., confusing language, incomplete response categories, etc. Also, controlling the survey burden is one of the commitments of research industry to the study participants, and analysis of high latency items may be one of the more effective strategies of reducing this burden.

The current research is aimed at gaining a broader understanding of latency as function of respondent and item characteristics using a rich dataset from a special population survey.

2. Data

We utilize the data from the Religion among Scientists in International Context (RASIC) survey of scientists in biology and physics departments in Italian universities and research institutes. The survey was fielded June 9-August 12, 2014, and administered in both Italian and English, with respondents selecting the language to complete the survey in. The final data had n=1,411 respondents (AAPOR response rate RR3=56.7%, AAPOR 2015). The survey instrument contained questions about educational background, work experiences as scientists, religion, and spirituality. Data were collected primarily in web mode; n=43 telephone interviews excluded from analysis. Web data collection allowed recording the timing of question start and end. RASIC survey was sponsored by Rice University and funded by Templeton World Charity Foundation Grant.

3. Methods

As opposed to the substantive analysis of survey data, analysis of survey paradata and other types of internal quality control analysis (Kolenikov and Pitblado 2014) typically call for a different representation of the data set. Rather than using items with their natural scales, the response latency analysis associates a unified variable of response time with each person-by-item survey turn, along with item characteristics. A natural representation of such data is a long data set where each line is a person-by-item combination. With the single dependent variable of response time, the data set at our disposal is a typical cross-classified (Goldstein 1994) data set, where one dimension of classification represents respondents (so that variation in response times can be explained by respondent characteristics such as age, race or gender) and another represents items (so that variation in response times can be explained by item characteristics, such as topic, format or length). After dropping missing items, the long cross-classified data set of respondent (n) × item (m) contained mn=89,096 observations. Outlier times were also dropped. These were identified as either the response time was greater than the item median + 7 × item IQR; or more than 180 seconds for standard items and 600 seconds for matrix, numeric entry, or text entry items. This procedure eliminated mn=402 long data set observations.

We fit a multilevel cross-classified model that can generally be expressed as

$$\ln t_{ij} = \mu + \alpha_{0i} + \beta_{0j} + \alpha' z_i + \beta' x_j + u_{ij}$$

where i enumerates items, j enumerates persons, z_i are item characteristics such as number of words in the text of the question stem or location in the survey, x_j are the person characteristics such as gender or academic rank, α and β are regression coefficients, μ is the overall mean, and u_{ij} are observation level regression errors assumed uncorrelated. This model accounts for correlation of errors within items and respondents through the random effects of items α_{0i} and the random effects of persons β_{0j}. Response time in seconds was transformed into a natural logarithm to remove skewness. All models were estimated in Stata 14 software (Stata Corp 2015).

3. Results

Table 1 reports the results of the cross-classified model estimation. Only the final model is shown. We also tested number of response options, reading grade level of item (in English), item involved recall, item included don't know option, respondent gender, respondent in biology or physics department, respondent restarted survey. Neither of these variables were found significant at 5% level, and models involving them are not shown.

Figures 1 through 8 visualize the results as margin plots, i.e., predicted latency from fixed portion of model only, with pointwise 95% confidence intervals. To produce these plots, interval-level variables set to means, and nominal and ordinal variables set to modes for all variables other than the plotted one. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds).

Figure 1 shows that while latency increases approximately linearly for short and medium size items, respondents start speeding up in items with stems beyond 105 words (English) and 125 words (Italian). The effect of the number of words in the response categories (Figure 2) was linear; while the quadratic terms were included in regression and were significant, significant departures from linearity did not occur within the range of the data. Surveys taken in Italian language had lower latency; however English respondents may have been ESL speakers.

Figures 3 and 4 present the changes in latency throughout the instrument, with the question number in the sequence and time since start as the marginal explanatory variables. Figure 4 in particular shows that the respondents need about 10–15 minutes to get used to the instrument, as their response latency goes up during this time period, and then start speeding up. Interestingly, there is also a bump at the upper end of the plot, with respondents likely displaying fatigue after an hour into the survey.

As expected from the cognitive theories of psychology of survey response (Toureangeau et. al. 2000), items that require more cognitive processing had greater latency (slower responses). Along these lines, opinion items that require the respondents to formulate an opinion on the spot rather than just retrieve the information took longer to respond than factual items (Figure 5). Items requiring averaging took longer than other items. Items that required keyboard entry, like numeric and text entry items, had greater latency than radio button items. Items in matrix (not measured separately) had higher latency than single items. Figure 5 reports these results along with the mean marginal values within each group.

Demographic characteristics of respondents (who all were highly educated individuals due to the nature of the target population) had little explanatory power. Figure 6 shows that a somewhat greater latency was observed for the more senior scientists. Age was not specifically controlled for as it was not available on the frame. Respondents in institutions classified as elite responded slightly faster (12.5 sec) than those in nonelite institutions (13.1 sec; no figure shown).

Table 1: Cross-classified response latency model.

	Coefficient	S.E.
Fixed Portion		
Item length		
Words in item stem	.019***	.001
Normed square of words in item stem	-.054***	.007
Words in response options	.013***	.002
Normed square of words in options	-.024***	.007
Language × Item length		
Italian	-.064**	.021
Italian × Words in item stem	-.002***	.000
Italian × Normed sq. of words in item stem	.014***	.002
Italian × Words in response options	-.002***	.001
Italian × Normed sq. of words in response options	.012***	.002
Type of item		
Factual (vs. opinion)	-.422***	.064
Averaging	.178*	.084
Numeric entry	.375***	.095
Text entry	.674***	.139
Matrix	.340**	.112
Respondent Characteristics		
At elite institution	-.041*	.017
Respondent rank from frame (vs. junior)	χ^2=42.964***	
Middle (postdocs, assistant professors, etc.)	.048*	.021
Senior (associate and full professors, etc.)	.129***	.021
Unknown: junior to middle	.010	.089
Unknown: middle to senior	-.065	.096
Unknown: nothing known about rank	.090***	.028
Time		
Time of day in Italy in hours (0 to 23)	χ^2=25.364***	
First cubic spline	-.004	.003
Second cubic spline	.201*	.055
Third cubic spline	-.848***	.227
Fourth cubic spline	1.653***	.510
Fifth cubic spline	-1.849*	.916
Day of week	χ^2=22.797***	
Monday	.002	.047
Tuesday	.047	.047
Wednesday	.070	.049
Thursday	-.000	.050
Friday	.075	.052
Saturday	.126*	.060
Relative time		
Time from start in minutes	.008***	.001
Normed square of time from start	-.070***	.008
Normed cube of time from start	.007***	.001
Item sequence		
Sequence in survey for respondent	-.010**	.004
Normed square of sequence in survey	.170*	.078
Constant	2.533***	.108

Table 1, continued: Cross-classified response latency model.

Random Portion	Final model		Intercept-only model	
Variance				
Item	.066***	.011	.282***	.044
Respondent	.081***	.003	.079***	.003
Residual	.230***	.001	.232***	.001
Intraclass correlation				
Item	.175	.023	.477	.039
Respondent	.215	.009	.133	.011

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$ for the null hypothesis of zero coefficient vs. two-sided (regression slopes) or one-sided (variance components) alternatives.

Notes: Wald tests shown for sets of dummy variables. Normed values for squares and cubes are used to ensure coefficients ≥ .001 for display and are calculated as $(x/\bar{x})^a$, where $a = 2,3$ is the exponent. Time from start was reset for breaks ≥ 15 minutes between items.

Time of survey completion had a small effect on latency. Time of day had a bimodal distribution of latency with peaks mid-afternoon and around midnight (Figure 7). Day of week has bimodal distribution latency with the slowest response on Saturday and, to a lesser extent, on Wednesday (Figure 8).

Even after controlling for the available item and person characteristics, unexplained variance in response latencies remains. The intraclass correlations, showing the remaining variance explained due to item and person characteristics, were 17.5% and 21.5%, respectively, indicating that adding further information and item and/or respondent level could have provided improved fit.

Compared to the intercept-only cross-classified model, the item random effect variance is reduced by 77% from 0.282 to 0.066. Thus the model was relatively successful in explaining the item-related variability. Contrasted to that, the respondent level variance and the residual variance did not change.

4. Conclusions

This paper provides at attempt to characterize response latency in a web survey of a specialized, highly educated population. It represents a complementary view to the existing literature on speeding in web surveys. It also helps quantifying the thresholds of respondent burden at which respondents start taking cognitive shortcuts and speeding through the survey, represented by decreases in/leveling off of latency for items with longer stem and/or for later items in survey. At the same time, we can argue that linearity of increase in latency with length of response option text is not consistent with satisficing.

Other findings of the paper are consistent with psychology of survey response theories. For instance, cognitively difficult items such as averaging were found to have longer latency.

References

The American Association For Public Opinion Research (2015). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 8th edition.

Goldstein, H. (1994). Multilevel Cross-Classified models. *Sociological Methods & Research* **22** (3), 364-375.

Kolenikov, S., and J. Pitblado (2014). Analysis of Complex Health Survey Data. Ch. 29 in Johnson, T. P. (ed), *Handbook of Health Survey Methods*. Wiley, Hoboken, NJ.

Kreuter, F., editor (2013). *Improving Surveys with Paradata: Analytic Uses of Process Information.* Wiley Series in Survey Methodology, Wiley, Hoboken, NJ.

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, **5**, 213–236.

StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Tourangeau, R., Rips, L. J., and K. Rasinski (2000). *The Psychology of Survey Response*. Cambridge University Press.

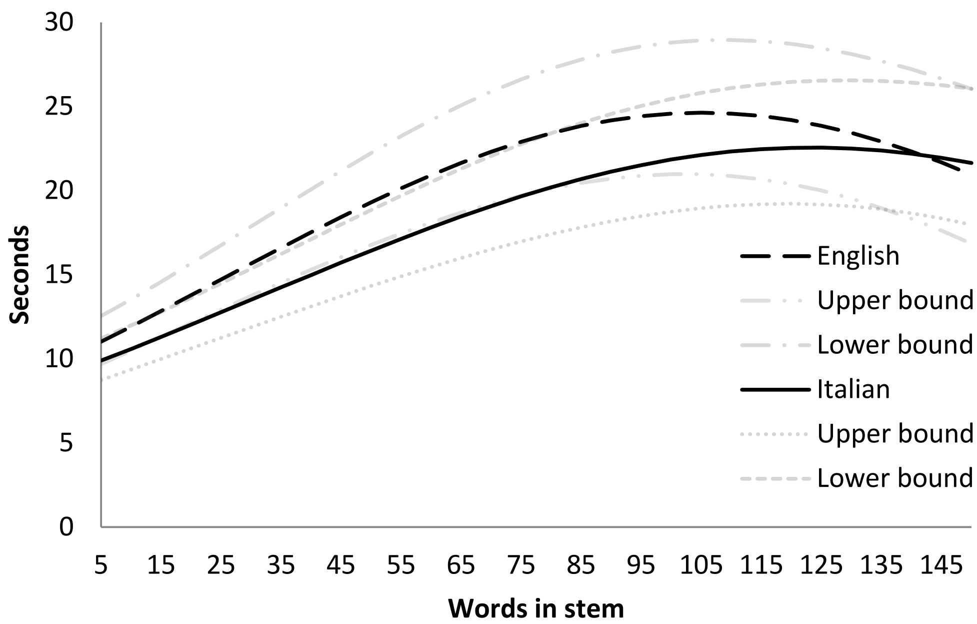

Figure 1. Predicted latency by words in question stem. Greater numbers represent slower response. Predicted latency from fixed portion of model only. Interval-level variables set to means, nominal and ordinal variables set to modes. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds). 95% confidence intervals shown.

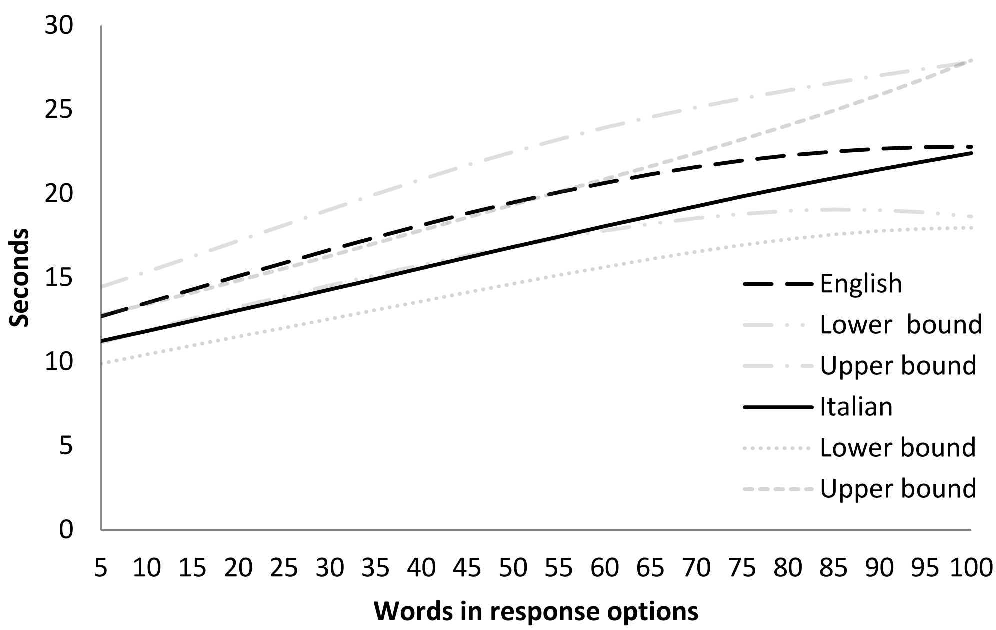

Figure 2. Predicted latency by words in response options. Greater numbers represent slower response. Predicted latency from fixed portion of model only. Interval-level variables set to means, nominal and ordinal variables set to modes. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds). 95% confidence intervals shown.

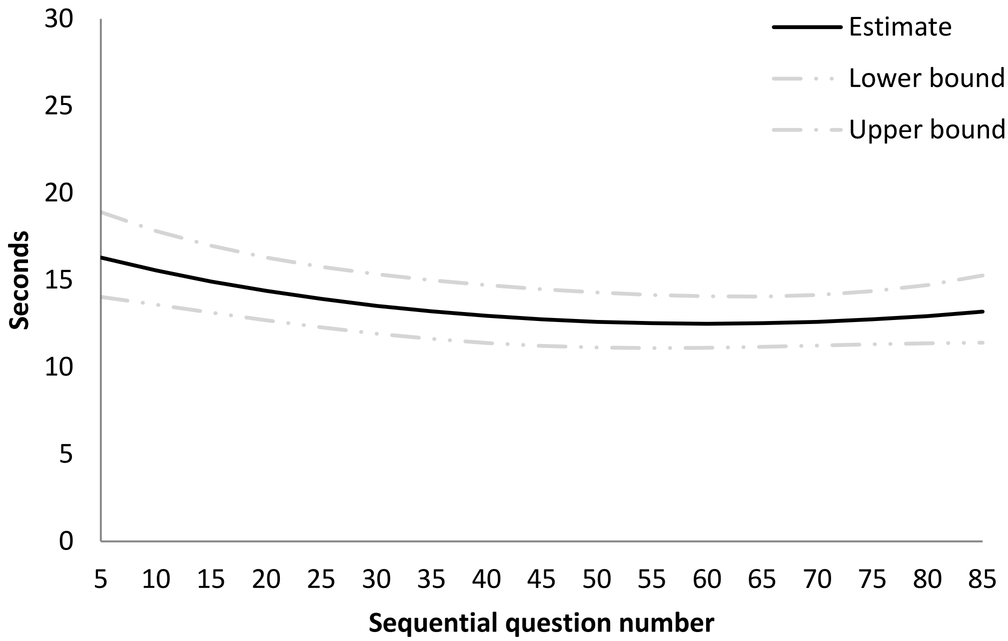

Figure 3. Predicted latency by sequence in instrument. Greater numbers represent slower response. Predicted latency from fixed portion of model only. Interval-level variables set to means, nominal and ordinal variables set to modes. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds). 95% confidence intervals shown.

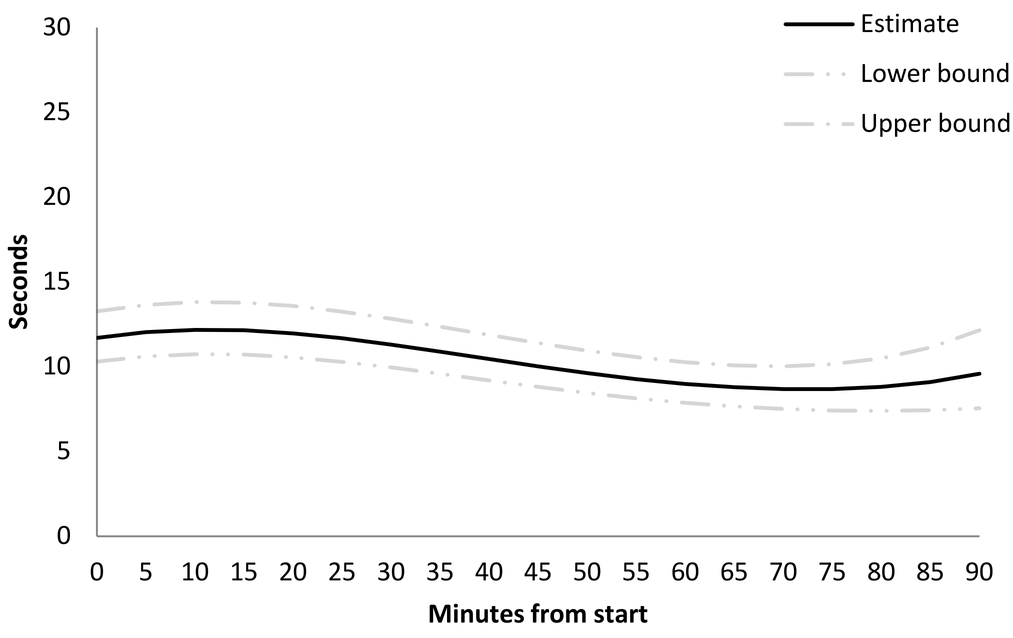

Figure 4. Predicted latency by time from start. Greater numbers represent slower response. Predicted latency from fixed portion of model only. Interval-level variables set to means, nominal and ordinal variables set to modes. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds). 95% confidence intervals shown.

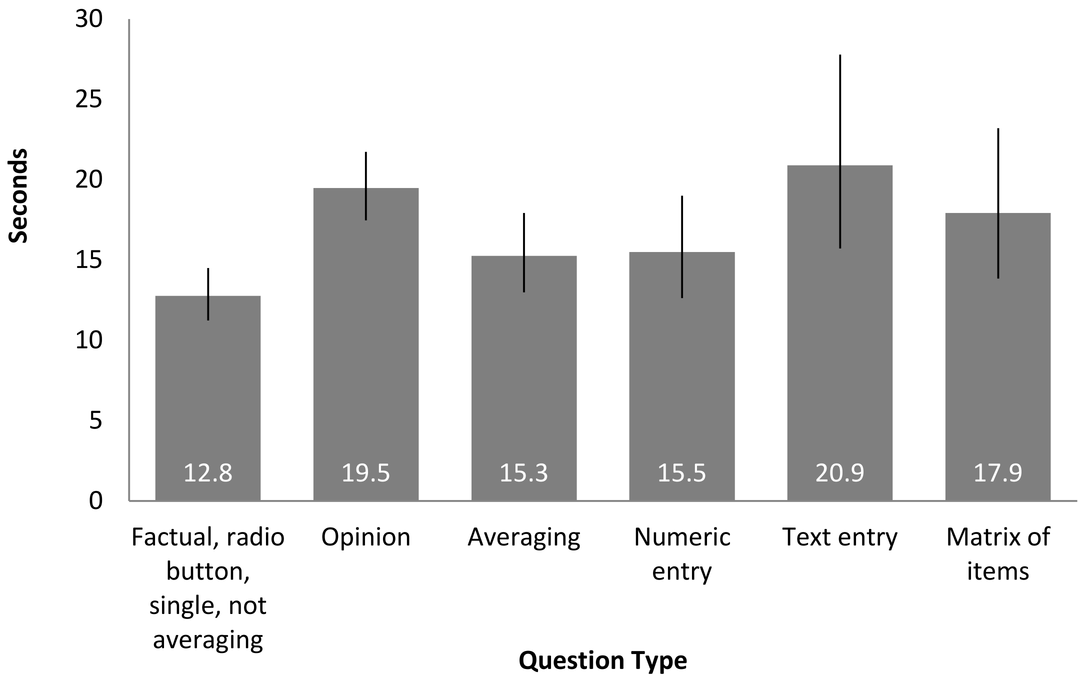

Figure 5. Predicted latency by question type. Greater numbers represent slower response. Predicted latency from fixed portion of model only. Interval-level variables set to means, nominal and ordinal variables set to modes. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds). 95% confidence intervals shown.

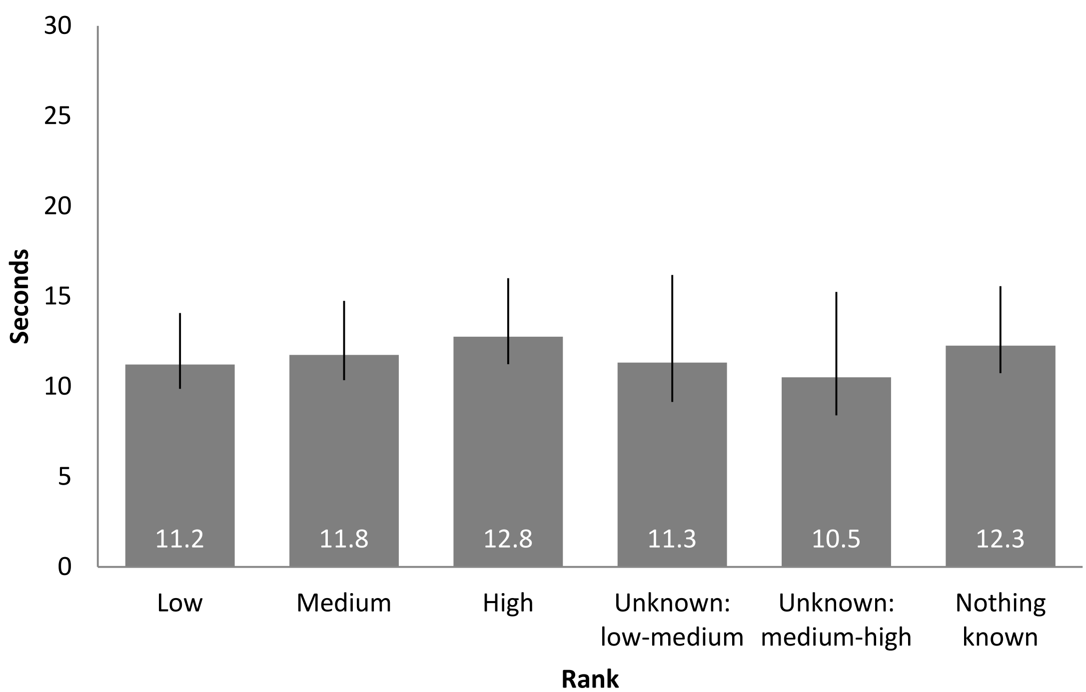

Figure 6. Predicted latency by respondent rank. Greater numbers represent slower response. Predicted latency from fixed portion of model only. Interval-level variables set to means, nominal and ordinal variables set to modes. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds). 95% confidence intervals shown.

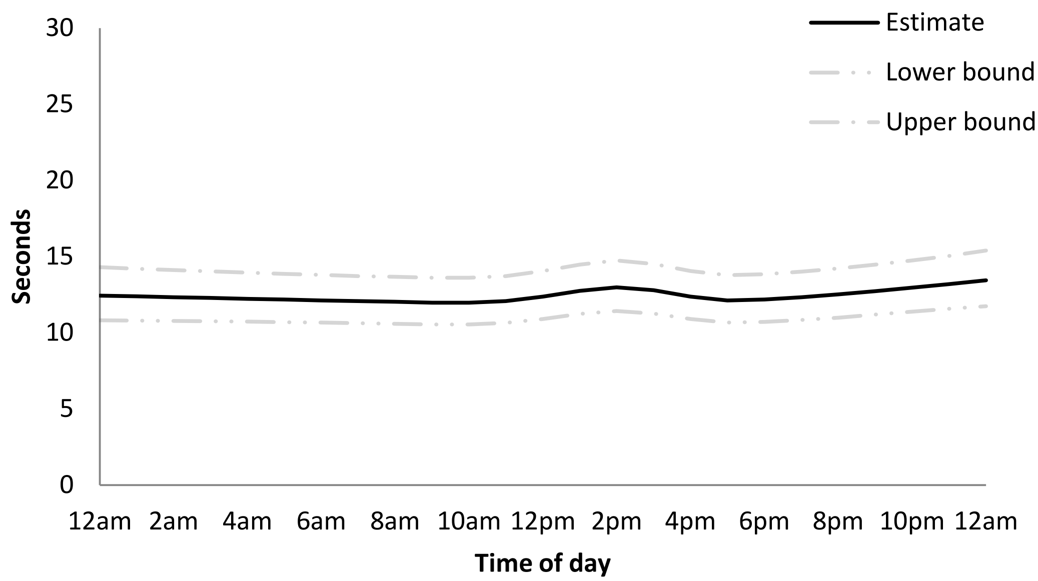

Figure 7. Predicted latency by time of day. Greater numbers represent slower response. Predicted latency from fixed portion of model only. Interval-level variables set to means, nominal and ordinal variables set to modes. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds). 95% confidence intervals shown.

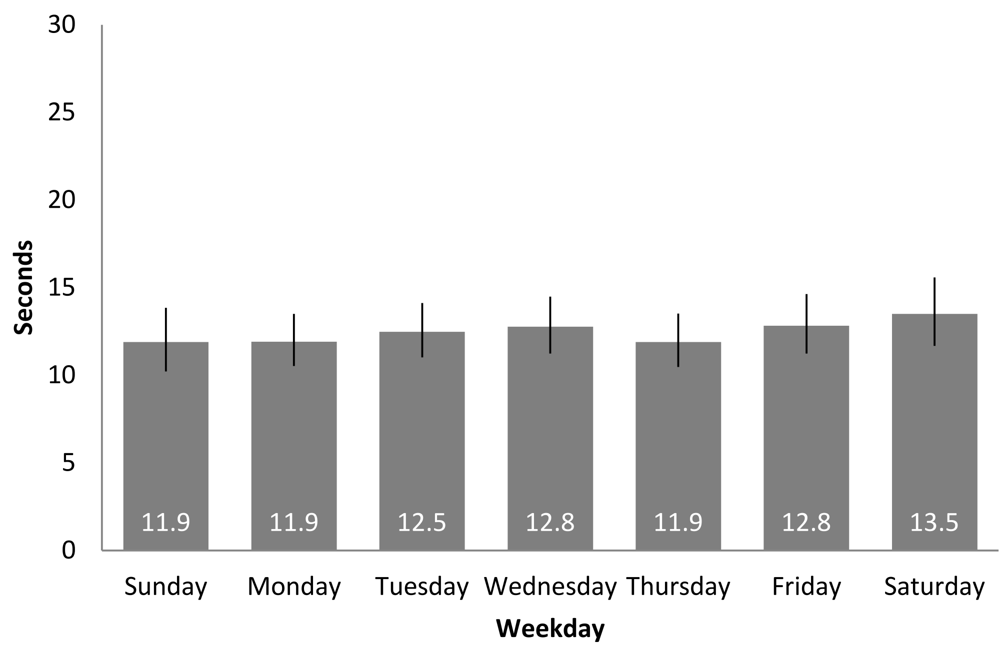

Figure 8. Predicted latency by day of the week. Greater numbers represent slower response. Predicted latency from fixed portion of model only. Interval-level variables set to means, nominal and ordinal variables set to modes. Estimates have been exponentiated to display as seconds (vs. the natural log of seconds). 95% confidence intervals shown.