Public Opinion Quarterly, Vol. 71, No. 3, Fall 2007, pp. 1–25

RESPONSE ORDER EFFECTS IN DICHOTOMOUS CATEGORICAL QUESTIONS PRESENTED ORALLY: THE IMPACT OF QUESTION AND RESPONDENT ATTRIBUTES

ALLYSON L. HOLBROOK
JON A. KROSNICK
DAVID MOORE
ROGER TOURANGEAU

Abstract Using data from 548 experiments in telephone surveys conducted by the Gallup Organization, we explored how attributes of questions and respondents moderate response order effects in dichotomous categorical questions. These effects were predominantly recency effects and occurred most in questions that were more difficult to comprehend (especially among respondents with the least education), with response choices that were more difficult to comprehend (because they were complete sentences instead of words or phrases and because they were not mutually exclusive), and that were asked after many prior questions. Recency effects were also more common in questions that explicitly or implicitly encouraged respondents to wait until they had heard all the answer choices before formulating a judgment than in questions that induced respondents to begin formulating a judgment before all the answer choices had been read (especially among the least educated respondents). A study of interviewer behavior revealed patterns of pausing between and within sentences that help to explain why some types of questions are especially prone to recency effects and others are not.

A sizeable body of evidence shows that small changes in the way a question is asked can substantially affect responses (see Krosnick 1999; Schuman and Presser 1981). In this article, we focus on one such question attribute: the order of presentation of categorical response options that are not inherently ordered

ALLYSON L. HOLBROOK is with University of Illinois, Chicago, USA. JON A. KROSNICK is with Stanford University, CA, USA. DAVID MOORE AND ROGER TOURANGEAU are with Institute for Social Research, University of Michigan, Ann Arbor, MI, USA and also with Joint Program in Survey Methodology, University of Maryland, College Park, MD, USA. Address correspondence to Allyson L. Holbrook; (e-mail: allyson@uic.edu) or Jon A. Krosnick (e-mail: krosnick@stanford.edu).

doi: 10.1093/opq/nfm024 Advanced Access publication xxx xx, 2007

(e.g., "Which is the more important problem facing the country: unemployment or crime?"). In many past studies of response order effects in such questions, some have observed primacy effects (wherein options are chosen more frequently when presented earlier in a list); others have observed recency effects (wherein options are chosen more frequently when presented later); and still others have found no order effects (for a review, see Krosnick 1999).

In this paper, we use a very large set of new experiments to explore the conditions under which these results occur in the course of telephone surveys tapping public opinion. We also explored the impact of the following factors that may govern order effects: the difficulty of comprehending the question and response options, whether the response options are mutually exclusive, whether the question induces respondents to begin generating an answer after hearing all the answer choices or at an earlier point, the cognitive skills of respondents, and the number of prior questions answered. Finally, we analyzed tape recordings of telephone interviews to explore whether interviewer pausing behavior might explain the patterns of response order effects we observed. Our findings provide justification for recommendations about best practices that many researchers will find unsurprising but nonetheless are rarely followed.

We begin below by outlining one theoretical account that has been offered to explain response order effects to date: satisficing theory (Krosnick 1991; Tourangeau 1984). We tested some predictions made by the theory, which we outline first. Then, we describe the potential impact of some variables not addressed specifically by satisficing theory, outlining reasons why these factors might have a moderating effect. Next, we test a variety of these predictions using data from hundreds of experiments conducted in recent telephone interviews. Finally, we describe our analysis of interviewer behavior and detail the implications of our findings.

Satisficing Theory

According to satisficing theory, response order effects are attributable to "weak satisficing," a process whereby respondents execute all four steps of optimal answering (interpreting the question, retrieving information from memory, integrating the information, and reporting an answer) but do so with a confirmatory bias (e.g., Koriat et al. 1980; Hoch 1984; Klayman and Ha 1987) and an inclination to stop considering alternatives as soon as an acceptable response has been identified (see Krosnick 1991). When response options are presented orally, respondents cannot think much about the first option they hear, because presentation of the second option interrupts this thinking. Similar interference occurs until after the last alternative is heard, at which point that option is the most salient and most likely to be the focus of respondents' thoughts. So confirmatory biased thinking and incomplete consideration of response options would yield recency effects. This tendency is reinforced by people's ability to

remember the last options read just after they have heard a list of response choices (Baddeley and Hitch 1977).

Satisficing theory posits that the likelihood and magnitude of response order effects depend on three classes of factors: respondents' ability to optimize, respondents' motivation to optimize, and the cognitive difficulty of optimizing inherent in the question. Satisficing is less likely among respondents with more cognitive skills and respondents who are more motivated to think carefully about questions. And satisficing is more likely when a question is especially difficult to comprehend, when a question demands an especially difficulty search of memory to retrieve information, when the integration of retrieved information into a summary judgment is especially difficult, or when translation of the summary judgment onto the response alternatives is especially difficult. Past research has yielded some evidence consistent with these claims about moderators (Payne 1949, 1950; McClendon 1986, 1991; Krosnick and Alwin 1987; Krosnick 1991; Krosnick, Narayan and Smith 1996; Narayan and Krosnick 1996; Bishop and Smith 2001).

In this paper, we used the largest set of response order experiments ever examined to test the moderating impact of several factors implicated by satisficing theory (comprehension difficulty of the question, the number of prior questions answered, and respondents' cognitive skills) separately and interactively (see Krosnick et al. 2002).

Linguistic Structure

We also explored whether an aspect of the linguistic structure of a question might govern response order effects. One linguistic structure common in surveys is what we call the *delayed processing question* (DPQ) structure, because the wording explicitly instructs respondents to wait until they have heard all response options before forming a judgment (e.g., "Which of the following . . ."). [1] Once respondents have heard all the answer choices, they can begin to retrieve information from memory and evaluate the response options to select one. If respondents begin thinking only after hearing all options and start by considering the last response option they heard, satisficing theory's account suggests that recency effects should predominate; that is, the last option read should be advantaged, because respondents begin by evaluating it with a confirmatory bias, and they give less attention to the other options due to fatigue.

Other linguistic structures do not include such explicit instructions to wait. In fact, some questions may encourage respondents to begin making a choice before all options have been read. One example of such a question is, "How do you feel about President Bill Clinton? Is he trustworthy or dangerous?" This question seems to ask for the requested judgment in an open-ended question. If interviewers pause just slightly after reading that first sentence, some

1. Graesser and colleagues (Graesser et al. 1996, 1992) call these "disjunctive questions."

respondents may, for just a moment, believe that the complete question has been read. We refer to such questions as *seemingly open-ended questions* (SOEQs), because they sound complete before any of the response options have been heard.[2]

If some respondents immediately and spontaneously generate thoughts in answer to the first sentence of a SOEQ, or even quickly retrieve a preconsolidated answer to the question from long-term memory (e.g., "I think Bill Clinton is a stupid man.") at that moment, respondents inclined to satisfice may listen to the response choices only until they hear one that matches the judgment they had generated. This would presumably undermine the usual tendency toward a recency effect, perhaps even inclining these respondents toward primacy effects. If most satisficing respondents form judgments early, we might even see primacy effects with such questions. But if only some satisficing respondents do this and others engage in more typical processing focused on the final alternatives read, these two processes could cancel each other out, yielding no overall response order effect.

A third linguistic structure we considered are *seemingly yes/no questions* (SYNQs). In such questions, respondents are first asked if they concur with one point of view, and a list of alternative points of view are offered following the word "or," as in, "Do you think that President Bill Clinton is trustworthy or dangerous?" If interviewers pause briefly after reading "trustworthy," respondents may think they have heard the whole question and may spontaneously generate a response to that question.[3] And if satisficing respondents manifest confirmatory bias in thinking about what sounds like a yes/no question, this would presumably incline them toward endorsing the first response option offered. If only some satisficing respondents are so inclined and others wait until the end of the question to begin evaluating, the two tendencies might cancel out one another (leaving no response order effect apparent); on the other hand, if most satisficing respondents generate judgments that are biased in the direction of confirmation only after hearing the whole question, this would presumably create recency effects.

Linguists believe that speakers are especially likely to pause for relatively long time periods after reading complete sentences, because they express complete thoughts, and their ends are natural places for speakers to take a breath (e.g., Clark and Clark 1977). Pausing at points of punctuation (e.g., commas) or co-ordinate transitions (e.g., "or") within sentences is less common and briefer (e.g., Boomer 1965; Goldman-Eisler 1972). If interviewers conform to these patterns when they read questions, then SOEQs are more likely to induce early cognitive processing (and thereby undermine typical recency effects) than SYNQs or DPQs.

2. These are what Graesser and colleagues called "judgmental questions."
3. Graesser and his colleagues call these "verification" questions.

We explored whether response order effects varied in magnitude or direction across these three question types in an exploratory investigation. We also assessed whether linguistic structure is especially consequential among people with the most limited cognitive skills by examining interactions between cognitive skills and linguistic structure. If linguistic structure influences confirmatory biased thinking (and therefore moderates response order effects), it seems most likely to do so among respondents for whom response order effects are most prevalent (i.e., those with the fewest cognitive skills).

Tests of These Hypotheses

DATA

We meta-analyzed 548 dichotomous response order experiments with categorical response options that were included in 149 RDD telephone surveys conducted by the Gallup Organization between 1995 and 2000. The topics addressed in these questions are listed in table 1. The surveys generally involved nationally representative samples, and the experiments involved samples ranging from 284 people to 1,727 people, with a mean of 853.[4]

All surveys involved list-assisted samples of telephone numbers obtained from Survey Sampling, Inc. (SSI; see Brick et al. 1995; Tucker, Lepkowski and Piekarski 2002 for discussions of list-assisted sampling).[5] Gallup surveys during this time had response rates between 20 percent and 50 percent, with an average response rate of 30.4 percent. About 16 percent of the surveys had response rates between 20 percent and 25 percent; about 40 percent had response rates between 25 percent and 30 percent; about one quarter had response rates between 30 and 35 percent; and about 20 percent had response rates of 35 percent or higher.[6]

4. In a number of the surveys where oversamples (e.g., of African-Americans) were used, these oversamples were not included in our analyzes. Of the 149 surveys, two targeted registered voters (9608PostDem and 9608wv1), and one (9608wv1) targeted people who had watched a particular presidential speech. All other surveys involved national, general population RDD samples. In most surveys, the youngest male/oldest female method was used to select an adult in the household who was at home at the time of the call. In the fall of 2000, Gallup changed its respondent selection procedure to the most recent birthday method (with gender quotas used to match the proportions of male and female respondents to the population).

5. Additional methodological detail can be found in Appendices A and B and by searching for Gallup Poll surveys in archives at the Roper Center for Public Opinion Research (http://www.ropercenter.uconn.edu).

6. These response rates most closely correspond to AAPOR's response rate 1 and were calculated as follows: total numbers dialed − businesses − other nonresidential − nonworking − busy − answering machine − no answer − break-off − refused)/(total numbers dialed − businesses − other nonresidential − nonworking).

Table 1. Summary of Question Topics

Topic	*N*	Response (%)
Presidential candidates; elections (elections or candidates)	152	27.74
President Bill Clinton scandals (questions about all Clinton scandals)	98	17.88
Racial or ethnic issues (affirmative action; race relations; immigration)	20	3.65
Justice Department investigations (excluding investigations connected to Clinton)	19	3.47
2000 Election (questions about the "special circumstances" of the 2000 election)	19	3.47
Iraq	17	3.10
Balance of power (who has responsibility or blame for outcomes – President or Congress)	16	2.92
IRS, taxes	16	2.92
Federal budget, balanced budget; budget surplus	15	2.74
Major political parties	13	2.37
Media, entertainment business	12	2.19
Economy, stock market, money	11	2.01
Relative importance of multiple political issues (e.g., environment versus economy)	10	1.82
Medical issues	9	1.64
Other questions about President Clinton	8	1.46
O. J. Simpson	7	1.28
Campaign finance reform	7	1.28
China	7	1.28
The death penalty	6	1.09
Former Yugoslavia, Serbia, Bosnia, Kosovo, Milosovich	6	1.09
Unions/Strikes	6	1.09
Sports	5	.91
The Middle East	2	.36
The environment	2	.36
Ireland	2	.36
Social Security	1	.18
Other political issues	34	6.20
Other nonpolitical topics	28	5.11
Total	548	100.00

From all Gallup Polls between 1995 and 2000, we selected all of the response order experiments in questions that were asked of the full sample of respondents or of a randomly selected subsample of respondents. None of the questions explicitly included "don't know" or "no opinion" responses, and respondents who volunteered a "don't know" response or refused to answer the question were excluded from our estimation of response order effects.

MEASURES OF RESPONSE ORDER EFFECTS

Approximately half of the respondents (selected randomly) were asked a question with the response options read in one order, and the remainder were asked the same question with the response options read in the reverse order. For each response order experiment, we calculated the effect size Somer's *d* to gauge the response order effect (Somers 1962). For dichotomous response order experiments, this is equal to the difference between the percent of respondents who chose a response option when it was presented last and the percent of respondents who chose that response option when it was presented first (Newson 2002). Thus, this statistics range from −100 to +100, with positive effect sizes indicating a recency effect, negative effect sizes indicating a primacy effect, and an effect size of zero representing no response order effect.

MEASURES OF COMPREHENSION DIFFICULTY

Question comprehension difficulty: The difficulty of comprehending each question was indexed by three indicators: (1) the number of sentences in the question, (2) the number of words per sentence, and (3) the number of letters per word. These three measures were each recoded to range from 0 to 1 (with 1 meaning the highest observed value for each variable and 0 meaning the lowest) and averaged to yield a composite measure of question comprehension difficulty.[7]

Response option comprehension difficulty: Some response options were single words; others were phrases; and still others were complete sentences. The latter may have been more difficult to comprehend than the former. Also, if the response options were mutually exclusive (e.g., "friendly" or "not friendly"), the respondent had only one concept to comprehend. But if the response options were not mutually exclusive (e.g., "Is Bill Clinton trustworthy or dangerous?"), respondents must have comprehended two different concepts, thereby perhaps increasing comprehension difficulty.

MEASURING NUMBER OF PRIOR QUESTIONS

We counted the number of questions asked before each target question. When a prior question was asked of only of a subset of the respondents, it was counted as one half of a question in this exercise.

7. Number of words per sentence is perhaps the most widely used indicator of text difficulty (see, e.g., Flesch 1948; Bormuth 1968; Gunning 1968; Smith and Kincaid 1970; Coleman and Liau 1975; Kincaid et al. 1975; Greenfield 2003). Number of letters per word is also used in many readability indices (Bormuth 1968; Smith and Kincaid 1970; Coleman and Liau 1975). Number of sentences is an indicator of the number of ideas or thoughts that respondents had to remember when considering their response to the question, an aspect of difficulty not typically considered in readability indices.

LINGUISTIC STRUCTURE CLASSIFICATION

Each question was categorized as a DPQ, a SOEQ, or a SYNQ independently by two coders who were not informed about the hypotheses being tested in this study. The coders agreed on 89.9 percent of their decisions. When the two coders disagreed, we used the judgment of the coder with whom the researchers agreed most (these judgments of agreement were made without consulting the results of the particular experiments in question).

MEASURES OF COGNITIVE SKILLS

Education: In keeping with prior investigations of response order effects, cognitive skills were measured by years of formal education among adults (e.g., Krosnick and Alwin 1987; Narayan and Krosnick 1996), which is an indicator very strongly correlated with scores on direct tests of cognitive skills (Ceci 1991). People who had less than a high school education were categorized into a "low" education group; people with only a high school degree and those with a high school degree and some technical, trade, or business school after high school were categorized into a "medium" education group; and people with at least some college education were categorized into a "high" education group.[8]

Age: We also explored whether age could be used to measure the cognitive skills required by optimizing, because working memory capacity (an element of cognitive skills relevant to satisficing) generally declines sharply at the end of the life-cycle (see, e.g., Craik and Jennings 1992). To reflect the expected sharp decline in cognitive skills among the elderly, we categorized respondents under age 65 into the "young" group and respondents age 65 and older into the "old" group.[9]

RESULTS

Patterns of effects: Across the 548 experiments, recency effects predominated: 19.2 percent showed significant recency effects, 42.7 percent showed nonsignificant recency effects, 1.1 percent showed no difference at all due to

8. Education was categorized in this fashion to permit detection of an expected nonlinear effect. Narayan and Krosnick (1996) found that low education respondents showed stronger response order effects than did medium or high education respondents, and that response order effects among medium and high education respondent did not differ significantly. This pattern has emerged as typical of weak satisficing response effects (see Krosnick 1999).

9. The age 65 cut point was chosen to make our results compatible with those of Knäuper (1999), who chose this cutoff because working memory capacity declines especially rapidly just after age 65 (cf. Schaie 1996). Other researchers have also used approximately 65 as a threshold to categorize respondents as elderly or older (e.g., Herzog, Rodgers and Kulka 1983; Herzog and Rodgers 1988; O'Rourke et al. 1999).

response choice order, 35.2 percent showed nonsignificant primacy effects, and only 1.8 percent showed significant primacy effects. Thus, there were almost no significant primacy effects, more than ten times as many significant recency effects, and more nonsignificant recency effects than nonsignificant primacy effects. The mean response order effect across all experiments was a significant recency effect (average percent shift = 2.19, $F(1,148) = 51.40$, $p < .001$; see the bottom row of column 1 of table 2).[10]

Question attributes: We conducted OLS regressions predicting response order effects with question comprehension difficulty, response option type (dummy variables represented questions with sentence response options and phrase response options; single word response options were the comparison group), mutual exclusivity of the response options, number of prior questions, and linguistic structure (dummy variables represented SOEQs and SYNQs; DPQs were the comparison group).

As expected, greater question comprehension difficulty was associated with larger response order effects ($b = 18.47$, $p < .01$; see column 1 of table 2). Consistent with the hypothesis that greater difficulty would lead to stronger response order effects, response order effects were significantly stronger when each of the response options was a complete sentence than when they were each a word ($b = 1.92, p < .01$), but not when the response options were phrases ($b =$ .69, ns; see rows 2 and 3 of column 1 of table 2, respectively). Questions with mutually exclusive response options also had marginally significantly weaker recency effects ($b = -.87$, $p < .10$; see row 4 of column 1 of table 2). Also as expected, later question placement was significantly associated with increased response order effects ($b = .03, p < .01$; see row 5 of column 1 of table 2). And SOEQs manifested significantly less recency than did DPQs ($b = -1.54$, $p <$ .01), but response order effects among SYNQs were not significantly different from those among DPQs ($b = -.59$, ns; see rows 6 and 7 of column 1 of table 2).

Education: As expected, the low education group's average recency effect (average $d = 2.89$, $F(1,148) = 15.14$, $p < .001$) was slightly larger than that of the medium education group (average $d = 2.62$, $F(1,148) = 35.48$, $p < .001$.), which was quite a bit larger than that of the high education group (average

10. Except where otherwise specified, all statistical tests were conducted controlling for nonindependence between multiple response order experiments from a single survey by setting the survey as the clustering variable, or PSU, in STATA and using "svy" commands which control for the psu clustering, which are specifically designed for analyzing survey data with complex sampling designs (see Stata Press Publishing Staff 2003). Specifically, we used the "svyreg" command, which enables the user to estimate the parameters of OLS regression equations with survey data with complex sampling designs, and the "svytest" command, which uses the adjusted Wald test to control for clusters in tests of means. Because our hypotheses were directional, one-tailed significance tests are reported unless otherwise specified.

Table 2. Unstandardized Regression Coefficients Predicting Response Order Effects in Various Respondent Groups

Predictor	All respondents	Low education	Medium education	High education	Younger	Older
Question comprehension	18.47**	28.05**	21.34**	.95	13.19**	30.04**
Difficulty	(3.54)	(11.60)	(5.84)	(4.96)	(3.88)	(8.26)
Sentence response options[a]	1.92**	3.42*	1.64†	1.34†	2.21**	1.23
	(.69)	(1.86)	(1.08)	(.91)	(.72)	(1.48)
Phrase response options[a]	.69	.20	.78	1.00†	.96**	−.18
	(.56)	(1.84)	(.86)	(.68)	(.56)	(1.49)
Mutually exclusive	−.87†	−2.01†	−.32	−.21	−.78†	−1.32
Response options	(.58)	(1.56)	(.92)	(.78)	(.59)	(1.34)
Number of prior questions	.03**	.04	−.001	.03†	.03**	.02
	(.01)	(.04)	(.02)	(.02)	(.01)	(.03)
SOEQ[b]	−1.54**	−4.45**	−2.73**	.20	−1.55**	−1.42
	(.58)	(1.92)	(1.03)	(.68)	(.62)	(1.34)
SYNQ[b]	−.50	.02	−.57	−.13	−1.28**	1.39
	(.62)	(1.64)	(.83)	(.78)	(.62)	(1.19)
Constant	−1.44†	−1.25	−.84	−.90	−.59	−3.04†
	(.97)	(3.29)	(1.51)	(1.33)	(.98)	(2.21)
R^2	.17	.09	.09	.02	.14	.06
N	548	548	548	548	548	548
Mean response order effect[c]	2.19**	2.89**	2.62**	.57*	2.21**	2.07**

NOTE.—The dependent variable is d, capturing the effect size of the response order effect, for which larger numbers mean a stronger recency effect.
[a]One word response options are the comparison group.
[b]DPQs are the comparison group.
[c]Significance tests compare each mean to 0 using an adjusted Wald test to correct for nonindependence of questions within surveys.
†$p < .10$, *$p < .05$, **$p < .01$.

$d = .57$, $F(1,148) = 2.79$, $p < .05$; see the last row of table 2). The difference between the low and medium education groups was not significant ($F(1,148) = .15$, ns), but the difference between the low and high education groups was highly significant ($F(1,148) = 7.52$, $p < .001$), as was the difference between the medium and high education groups ($F(1,148) = 12.42$, $p < .001$).

Age: The average response order effect was a significant recency effect among younger respondents (average d = 2.21, $F(1,148) = 50.35$, $p < .001$) and, surprisingly, was slightly weaker among older respondents (average d = 2.07, $F(1,148) = 17.52$, $p < .001$; see the last row of table 2). These two effects were not significantly different from one another, indicating that age did not moderate the size of response order effects ($t(547) = .10$, ns).

Interactions between question attributes and education: To explore whether the effects of question characteristics varied according to respondents' cognitive skills, we first estimated the regression separately for low, medium, and high education respondents (see the second, third, and fourth columns, respectively, of table 2). Then, for each predictor listed in table 2, we tested whether the coefficients differed significantly across the education groups (see Cohen and Cohen 1983, p. 111).

Question comprehension difficulty had more impact among less educated respondents (see row 1 of columns 2–4 in table 2). The effect of question comprehension difficulty was significantly stronger in the medium education group ($b = 21.34$, $p < .01$) than in the high education group ($b = .95$, ns; difference between medium and high education groups: $z = 2.66$, $p < .01$) and significantly stronger in the low education group ($b = 28.05$, $p < .01$) than in the high education group (difference between low and high education groups: $z = 2.15$, $p < .01$). The effect of question comprehension difficulty was not significantly different in the low and medium education groups (difference: $z = .52$, ns).

The effect of the dummy variable representing sentence response options manifested a similar pattern (strongest in the low education group and weakest in the high education group), but these differences were not statistically significant (difference between low and medium education groups: $z = .83$, ns; difference between low and high education groups: $z = .21$, ns; difference between medium and high education groups: $z = 1.00$).

Similarly, the effect of mutual exclusivity was also strongest among the least educated respondents. This effect was marginally significantly stronger among low education respondents ($b = -2.01$, $p < .01$) than among medium education respondents ($b = -.32$, ns; difference between low and medium education groups: $z = 1.37$, $p < .10$), and marginally significantly stronger among low education respondents than among high education respondents ($b = -.21$, ns; difference between low and medium education groups: $z = 1.53$, $p < .10$). The effect of mutual exclusivity did not differ between medium and high education

485 respondents (difference between medium and high education groups: $z = .20$, ns).

The effect of the number of prior questions was also strongest among the least educated respondents, though its effects did not vary significantly between the education groups (difference between low and medium education groups:
490 $z = .20$, ns; difference between low and high education groups: $z = .14$, ns; difference between medium and high education groups: $z = .09$).

The difference between SOEQs and the other question types was largest among low education respondents ($b = -4.45$, $p < .01$), somewhat smaller among medium education respondents ($b = -2.73$, $p < .01$), and smaller
495 and nonsignificant among high education respondents ($b = .02$, ns; see row 6 of columns 2–4 of table 2). The effect of the SOEQ dummy variables was significantly stronger in the low education group than in the high education group ($z = 2.28$, $p < .05$) and significantly stronger in the medium education group than in the high education group ($z = 2.37$, $p < .01$) but was
500 not significantly different in the low and medium education groups ($z = .79$, ns).

Interactions between question attributes and age: The effect of question comprehension difficulty was significantly stronger among older respondents
505 ($b = 30.04$, $p < .01$) than among younger respondents ($b = 13.19$, $p < .01$; test of difference: $z = 1.85$, $p < .05$; see row 1 of columns 5 and 6 in table 2). The effect of sentence response options, phrase response options, mutual exclusivity, and number of previous questions did not differ across age groups (tests of differences: sentence response options: $z = .60$, ns; phrase response options:
510 $z = .72$; mutual exclusivity: $z = .37$; and number of previous questions: $z = .32$, ns). Similarly, the effect the SOEQ dummy variable did not differ across age groups ($z = .09$, ns, see row 6 of columns 5 and 6 of table 2). The SYNQ effect was significant and negative among younger respondents ($b = -1.28$, $p < .01$) and nonsignificant and positive among older respondents ($b = 1.39$,
515 ns), a significant difference ($z = 1.99$, $p < .05$).

Study of Interviewer Behavior

520 The finding that SOEQs manifested weaker recency effect than SYNQs is consistent with the notion that interviewers pause longer after reading the first sentence of an SOEQ than they do after reading the first response option in an SYNQ. To test this possibility directly, we analyzed recordings of 71 telephone interviews conducted by the Gallup Organization as part of a questionnaire
525 pretest. The questionnaire for this pretest included one SOEQ (asked of 65 respondents) and one SYNQ (asked of 14 respondents).

The SOEQ was asked with one of two different response choice orders, and each respondent was randomly assigned to hear one of the two orders.

One version was: "Just your best guess, how do you think the war against terrorism will be fought over the next several years? Will it rely mainly on economic, diplomatic, and intelligence efforts with little or no military force, or will it require an extensive use of military force in addition to economic, diplomatic, and intelligence efforts?" The other version was: "Just your best guess, how do you think the war against terrorism will be fought over the next several years? Will it require an extensive use of military force in addition to economic, diplomatic, and intelligence efforts, or will it rely mainly on economic, diplomatic, and intelligence efforts with little or no military force?" We expected that interviewers would pause between the two sentences of the question, which would encourage respondents to begin cognitive processing to generate answers early on.

The SYNQ was also asked with one of two different orders of response choices, with respondents being randomly assigned to each order. One version was: "Do you think the Bush administration was generally accurate in describing the threat Iraq posed to the U.S. but exaggerated some of the specific details or do you think the Bush administration greatly overstated the threat Iraq posed to the U.S. in order to justify a war in Iraq?" The other version was: "Do you think the Bush administration greatly overstated the threat Iraq posed to the U.S. in order to justify a war in Iraq or do you think the Bush administration was generally accurate in describing the threat Iraq posed to the U.S. but exaggerated some of the specific details?" We measured the length of the pause interviewers made before saying the word "or" to assess whether they gave respondents less time at that moment to begin generating answers than they gave after the first sentence of SOEQs.

To measure the lengths of pauses, we used the software program called *Audacity*, which visually displayed the volume of the sounds recorded at each moment of each interview and allowed us to mark the time of each event during each interview. The duration of time between the beginning and ending of each pause in the questions of interest were recorded in this fashion, and the beginning time was subtracted from the ending time to yield the length of each pause.

Interviewers paused for .38 seconds on average before reading the response options in the SOEQs, whereas they paused for less time (.32 seconds on average) between the response options in the SYNQs, a marginally significant difference, $t(77) = 1.56$, $p = .07$.[11] This finding is consistent with the hypothesis that pauses after question stems in SOEQs are typically longer than pauses between response options in SYNQs. This could explain why the SOEQ question format reduced recency effects more than the SYNQ format did.

11. No respondents ever interrupted an interviewer to answer before all response choices had been read.

Discussion

Significant response order effects occurred in a minority of orally-presented dichotomous categorical questions, and when such effects appeared, they were predominantly recency effects. Such recency effects were more common when questions were more difficult to comprehend, when answer choices were complete sentences (versus words or phrases), when question stems encouraged respondents to begin processing the response options only after hearing all of them, when more prior questions had been answered, and among respondents with more limited cognitive skills. We discuss each of these findings and their implications.

QUESTION AND RESPONSE OPTION COMPREHENSION DIFFICULTY

Response order effects were more common among questions that were more difficult to comprehend, when the response options were complete sentences instead of single words or phrases, and among questions involving response options that were not mutually exclusive. This is consistent with previous research showing that response order effects are stronger for longer questions, questions with longer response options, and those with more difficult language (Bishop and Smith 2001; Payne 1949, 1950; although McClendon 1986 found no such relation).

The effects of question comprehension difficulty were especially strong among the least educated respondents. This is the first research to provide support for the hypothesis derived from satisficing theory (Krosnick 1991) that respondent ability and task difficulty would interact to influence response order effects.

QUESTION PLACEMENT

Later question placement was associated with stronger recency effects. This finding is consistent with evidence that other forms of satisficing (e.g., no opinion effects; Krosnick et al. 2002) are more likely when questions are asked after more prior questions. This finding might at first seem to conflict with a study by Bishop and Smith (2001) of response order experiments in orally-presented, dichotomous questions from older Gallup experiments, which found no such relation. However, these experiments involved no more than 20 prior questions (Bishop and Smith 2001, Appendix B). In contrast, the number of prior questions in the surveys we examined was as large as 78.5 questions. In the data we examined, when the model shown in column 1 of table 2 was run only with the 285 experiments having 20 or fewer prior questions, number of prior questions was not significantly related to the response order effect ($b = -.08$, SE $= .06$, ns). This suggests that the difference in the range of number of prior questions between our study and Bishop and Smith's (2001) accounts for the difference between results.

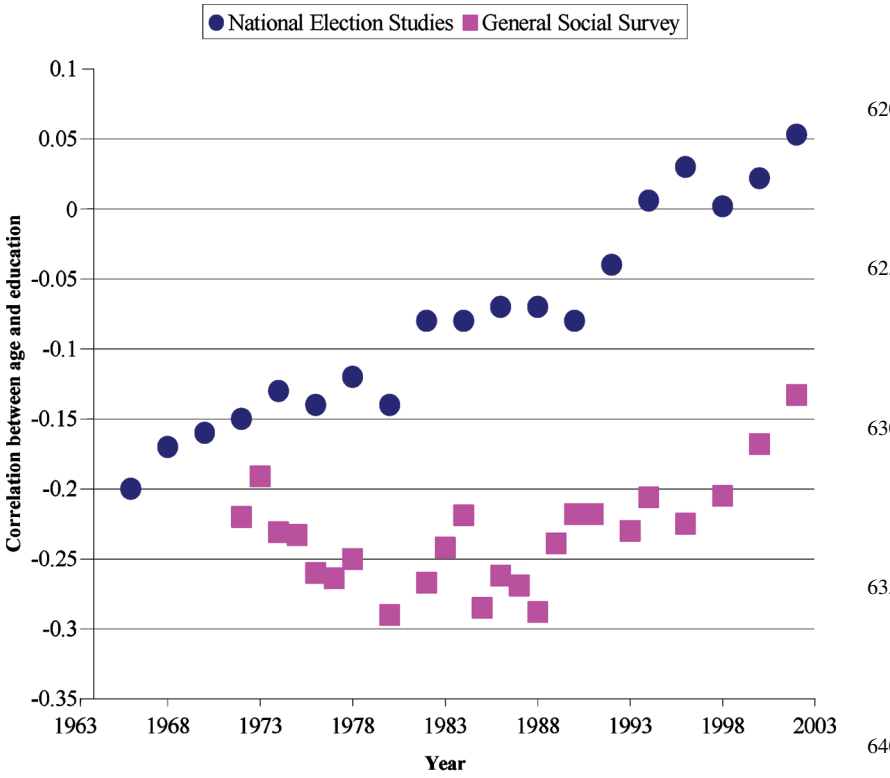

Figure 1. Association between age and education.

AGE

Our finding that age did not moderate response order effects is inconsistent with Knäuper's (1999) evidence that age did so in Schuman and Presser's (1981) experiments. The discrepancy between her results and ours may be due to changes over time in the correlation between age and education. Schuman and Presser's (1981) studies were conducted between 1971 and 1980, whereas the data we analyzed were collected between 1995 and 2000. Age and education were reliably correlated at the time of Schuman and Presser's studies in the American Public, but these variables are now much less correlated with one another. As shown in figure 1, the correlation between education (coded into three groups as in the current investigation) and age (coded into two groups as in the current investigation and in Knäuper's 1999 study) in presidential election year American National Election Studies changed from −.20 in 1968 to .05 in 2002 and in the General Social Survey from −.19 in 1972 to −.13 in 2002. In both sets of studies, the association between the correlation coefficients shown in figure 1 and year was statistically significant (NES: $r = .97$, $p < .01$,

$N = 19$; GSS: $r = .47$, $p < .05$, $N = 24$). This raises the possibility that Knäuper (1999) may have found evidence that age moderated response order effects because in the older data, age and education were confounded, and education was responsible for the apparent moderation of age.[12]

The differences we observed in the association of education and age with response order effects raises the possibility that these purported moderators may be associated with different *aspects* of cognitive skills. For example, old age is associated with declines in processing speed, working memory capacity, and cued and free recall, but not in vocabulary knowledge (Park 1999). Education is also associated with a wide range of cognitive skills (see Ceci 1991 for a review), though not necessarily with the cognitive skills that are associated with age. For example, education (coded into low, medium, and high as it was in our research) is strongly correlated with scores on the General Social Survey's vocabulary test (called "WORDSUM"): $r = .47, p < .001, N = 20{,}195$, whereas age (coded young and old as in our research to reflect the expected nonlinear effect of age) correlated with vocabulary test scores much more weakly: $r = -.05$, $p < .001$, $N = 20{,}191$.[13] Because vocabulary ability is one component of the cognitive skills needed to optimize when answering survey questions (see Krosnick and Alwin 1987 for evidence that vocabulary scores moderate response order effects), this suggests that education is a better proxy for at least some cognitive skills than age. Our evidence suggesting that age and education are becoming less strongly correlated suggests that it is important for researchers to understand the differences in ability reflected in age and educational attainment.

LINGUISTIC STRUCTURE

This paper provides the first evidence that warnings to wait and illusory endings of questions affect the likelihood and size of response order effects. Recency effects were most pronounced among DPQs, a bit less apparent in the SYNQs, and almost invisible in the SOEQs. This suggests that linguistic structure can play a role in determining the direction of response order effects by affecting whether respondents begin by considering the first or the last response option. The effects of linguistic structure were most pronounced among the least educated respondents, consistent with the notion that these individuals are most susceptible to the impact of question attributes that determine processing focus.

12. Although Knäuper (1999) examined the effect of age within groups of low education and high education respondents, there was still variation in education within each group that was not controlled for and may have been correlated with age and responsible for the apparent effects of age. In logisitic regressions, Knäuper (1999) represented the impact of education on response order effect magnitude as linear, whereas Narayan and Krosnick (1996) found that this impact was nonlinear in those data, so this analysis may also not have fully controlled for education.

13. These correlations were generated combining GSS data collected in 1974, 1976, 1978, 1982, 1984, 1987, 1988, 1989, 1990, 1991, 1993, 1994, 1996, 1998, and 2000.

Our findings have a number of implications for survey research practice, though they are hardly surprising. First and foremost, our findings suggest that survey organizations should routinely rotate the order of response choices to guard against creating bias in results. Although the Gallup Organization does so in its Gallup Poll, the vast majority of survey organizations do not do this. It is advisable.

Second, because response order effects are stronger in questions that are more difficult to comprehend, researchers should strive to keep their language as simple as possible. Advice of this sort is offered in numerous research methods textbooks, but there is still room for researchers to heed this advice more faithfully: questions are often laden with social science jargon and long sentences involving many multisyllabic words. Because survey researchers are usually well-educated, they may find such questions to be easily understandable. It is rare indeed to hear of a questionnaire designer calculating the reading difficulty level of a question or using a dictionary or thesaurus to find synonyms that are simpler and easier to understand. Our findings suggest that this sort of effort may be worthwhile to minimize response order effects.

Our evidence that response order effects were more common after many prior questions had been asked has an obvious implication as well: keep questionnaires short. This advice is heeded by some survey organizations, but many survey organizations routinely administer very long questionnaires, presumably because they require large amounts of data. When long questionnaires are used, researchers need to be especially aware of the likelihood of response order effects. Interestingly, though, we found a good number of response order effects even in the short interviews conducted by Gallup (averaging 18 minutes or less). So minimal length is not sufficient to assure these effects will not occur.

The practical implications of our findings regarding linguistic structure are less clear. We found the weakest response order effects among SOEQs. This might be viewed as evidence that response quality was the highest for these questions, because answers were minimally contaminated by a source of systematic measurement error. Therefore, perhaps researchers should strive to use this format as often as possible, stating the judgment to be made generally in an open-ended format before offering specific answer choices. This advice seems in keeping with the general finding that open-ended questions yield more reliable and valid data than close-ended questions (for a review, see Krosnick 1999).

However, it is also possible that the weak response order effects typical of SOEQs were not the result of less satisficing. Instead, the amount of satisficing might have been the same when answering SOEQs as when answering DPQs and SYNQs, but the focus of confirmatory thinking may have differed. Whereas DPQs and SYNQs may have focused most satisficing respondents' attention

on the final response alternatives and thereby induced recency effects, SOEQs may have caused some satisficing respondents to focus on the first response alternative while others focused on the second alternative. Because response order effects can only be observed in the aggregate, such effects could have cancelled each other out, leading to the appearance of no effect of response option order. The data we have do not permit us to assess whether the SOEQ format did indeed yield more valid responses, so we must wait for future research to explore this issue.

Conclusion

The findings reported here suggest that response order effects in dichotomous categorical questions administered orally are affected by a variety of question and respondent attributes, both individually and interactively. Such evidence helps to move us closer to understanding when and why response order effects occur and how best to prevent them. We look forward to future research testing additional hypotheses about the roles that linguistic structure and other factors play in regulating response order effects not only for the sorts of questions we examined here but also in questions with more than two response options.

Appendix A: Survey Methodology

Note: In this appendix, the names of all the surveys we used are shown (these correspond to the survey names used in Appendix A). In the first section, surveys archived at the Roper Center (http://www.ropercenter.uconn.edu/) are listed with the study number under which they are archived. The archive contains the questionnaire for these surveys, the dates during which the survey was conducted, and information about the sample size and composition (most surveys were of the general adult population, but some of those conducted about elections used only registered voters and this information is also specified in the archive). In the second section, methodological information is listed for the studies not archived at the Roper Center.

METHODOLOGICAL INFORMATION ABOUT STUDIES ARCHIVED AT ROPER CENTER

Survey	Study number
9501wv1	USAIPOCNUS1995–5001007
9503wv1	USAIPOCNUS1995–5001012
9503wv2	USAIPOCNUS1995–5001014
9504wv2	USAIPOCNUS1995–5001017

Survey	Study number
9504wv3	USAIPOCNUS1995–5001018
9505wv1	USAIPOCNUS1995–5001020
9506wv1	USAIPOCNUS1995–9506004
9507wv2	USAIPOCNUS1995–9507024
9508M&Gr	USAIPOCNUS1995–9508009
9508wv2	USAIPOCNUS1995–9508010
9509wv1	USAIPOCNUS1995–9509012
9510O.J.	USAIPOCNUS1995–9509015
9511Flsh	USAIPOCNUS1995–9511021
9511SHTD	USAIPOCNUS1995–9511020
9512wv1	USAIPOCNUS1995–9512023
9602wv1	USAIPOCNUS1996–9602006
9603wv2	USAIPOCNUS1996–9603008
9604wv1	USAIPOCNUS1996–9604009
9605wv1	USAIPOCNUS1996–9605011
9605wv2	USAIPOCNUS1996–9605012
9606wv1	USAIPOCNUS1996–9606013
9606wv2	USAIPOGNS1996–9606015
9607GPNS	USAIPOGNS1996–9607017
9607wv1	USAIPOCNUS1996–9607018
9608wv1	USAIPOCNUS1996–9608020
9701wv1	USAIPOCNUS1997–9701001
9702BDRM	USAIPOCNUS1997–9702006
9703wv1	USAIPOCNUS1997–9703007
9704wv1	USAIPOCNUS1997–9704008
9705wv1	USAIPOCNUS1997–9705009
9705wv2	USAIPOCNUS1997–9705012
9706wv2	USAIPOCNUS1997–9706016
9707wv1	USAIPOCNUS1997–9707017
9708wv1	USAIPOCNUS1997–9708018
9708wv2	USAIPOCNUS1997–9708019
9709DIANA	USAIPOCNUS1997–9709020
9709wv1	USAIPOCNUS1997–9709021
9710CHIN	USAIPOCNUS1997–9710025
9710wv1	USAIPOCNUS1997–9710023
9710wv2	USAIPOCNUS1997–9710026
9711wv2	USAIPOCNUS1997–9711029
9712RENO	USAIPOCNUS1997–9711031
9801INTN	USAIPOCNUS1998–9801004
9801PSTU	USAIPOCNUS1998–9801006
9801wv1	USAIPOCNUS1998–9801001
9801wv2	USAIPOCNUS1998–9801002
9801wv4	USAIPOCNUS1998–9801007
9802IRAQ	USAIPOCNUS1998–9802010
9802wv1	USAIPOCNUS1998–9802008
9802wv2	USAIPOCNUS1998–9802009

Survey	Study number
9803WILL	USAIPOCNUS1998–9803012
9803wv1	USAIPOGNS1998–9803011
9803wv2	USAIPOCNUS1998–9803013
9804JONE	USAIPOCNUS1998–9804014
9804wv1	USAIPOCNUS1998–9804015
9805wv1	USAIPOCNUS1998–9805016
9806wv1	USAIPOCNUS1998–9806017
9806wv2	USAIPOCNUS1998–9806018
9807wv1	USAIPOCNUS1998–9807019
9807wv2	USAIPOCNUS1998–9807020
9808wv1	USAIPOCNUS1998–9808023
9808wv2	USAIPOCNUS1998–9808027
9809wv2	USAIPOCNUS1998–9809033
9809wv3	USAIPOCNUS1998–9809036
9810wv1	USAIPOCNUS1998–9810039
9810wv3	USAIPOCNUS1998–9810041
9811_02	USAIPOGNS1998–9811042
9811_13	USAIPOCNUS1998–9811043
9811_19	USAIPOCNUS1998–9811044
9811_20	USAIPOGNS1998–9811045
9812_04(s	USAIPOGNS1998–9812046
9812_12	USAIPOCNUS1998–9812053
9812_15	USAIPOCNUS1998–9812054
9812_28	USAIPOGNS1998–9812057
9901_15	USAIPOCNUS1999–9901003
9901_22	USAIPOCNUS1999–9901005
9902_12L	USAIPOCNUS1999–9902011
9902_19x	USAIPOGNS1999–9902012
9902_26x	USAIPOGNS1999–9902013
9903_04x	USAIPOCNUS1999–9903014
9903_05	USAIPOGNS1999–9903015
9903_12x	USAIPOCNUS1999–9903016
9903_30x	USAIPOCNUS1999–9903019
9904_06	USAIPOCNUS1999–9904020
9904_13	USAIPOGNS1999–9904021
9904_30	USAIPOCNUS1999–9904025
9905_07	USAIPOGNS1999–9905026
9906_04x	USAIPOCNUS1999–9906028
9906_11x	USAIPOCNUS1999–9906031
9906_25	USAIPOCNUS1999–9906032
9907_13x	USAIPOGNS1999–9907033
9907_16x	USAIPOCNUS1999–9907034
9907_23x	USAIPOGNS1999–9907035
9908_16x	USAIPOCNUS1999–9908037
9909_10x	USAIPOCNUS1999–9909039
9909_23x	USAIPOGNS1999–9909040

Survey	Study number
9910_08x	USAIPOCNUS1999–9910041
9910_21x	USAIPOCNUS1999–9910042
9912_9x	USAIPOCNUS1999–9912046
2000_01_07	USAIPOCNUS2000–01
2000_01_13	USAIPOCNUS2000–02
2000_01_17	USAIPOCNUS2000–03
2000_01_25	USAIPOGNS2000–05
2000_02_04	USAIPOCNUS2000–06
2000_02_14	USAIPOGNS2000–08
2000_02_20	USAIPOCNUS2000–10
2000_02_25	USAIPOCNUS2000–11
2000_03_10	USAIPOCNUS2000–12
2000_03_30	USAIPOGNS2000–14
2000_04_07	USAIPOCNUS2000–16
2000_04_28	USAIPOCNUS2000–19
2000_05_23	USAIPOCNUS2000–23
2000_06_06	USAIPOCNUS2000–24
2000_06_22	USAIPOGNS2000–25
2000_06_23	USAIPOCNUS2000–26
2000_07_06	USAIPOGNS2000–27
2000_07_14	USAIPOCNUS2000–28
2000_07_25	USAIPOCNUS2000–30
2000_08_04	USAIPOCNUS2000–32
2000_08_11	USAIPOCNUS2000–34
2000_08_18	USAIPOCNUS2000–35
2000_08_24	USAIPOGNS2000–36
2000_09_11	USAIPOGNS2000–39
2000_10_06	USAIPOGNS2000–42
2000_10_25	USAIPOGNS2000–45
2000_11_11	USAIPOCNUS2000–48
2000_11_19	USAIPOCNUS2000–51
2000_11_26	USAIPOCNUS2000–52

METHODOLOGICAL INFORMATION ABOUT STUDIES NOT YET ARCHIVED AT ROPER CENTER

Survey	Dates of data collection	Sample
9608PostDem	8/30/1996–9/1/1996	1210 Registered Voters
9701wv3	1/31/1997–2/2/1997	1056 National Adults
9702OJ CIV SU	2/5/1997	438 National Adults
9705FLINN	5/22/1997	643 National Adults

Survey	Dates of data collection	Sample
9706ADLT	6/10/1997	6 National Adults
9706MCVE	6/2/1997	600 National Adults
9709wv2	9/25/1997–9/28/1997	1000 National Adults
9801SOTU	1/27/1998	400 National Adults who watched President Clinton's 1/27/1998 speech
9801wv3	1/23/1998–1/26/1998	1767 National Adults
9812_19	12/19/1998–12/20/1998	852 National Adults
9901_08	1/8/1999–1/10/1999	1014 National Adults
9903_19x	3/19/1999–3/21/1999	1018 National Adults
9904_15	4/15/1999–4/18/1999	1011 National Adults
9911_18x	11/18/1999–11/21/1999	1010 National Adults
2000_05_18	5/18/2000–5/21/2000	1011 National Adults
2000_08_29	8/29/2000–9/2/2000	1000 National Adults
2000_12_02	12/2/2000–12/4/2000	1026 National Adults
2000_12_10	12/10/2000	735 National Adults
2000_12_13	12/13/2000	633 National Adults
2000_12_15	12/15/2000–12/17/2000	1011 National Adults (Black oversample of 300 not included in our analyzes)

Acknowledgement

The research reported here was conducted partly while the first author was a university fellow at the Ohio State University and partly while she was a predoctoral trainee of the National Institute for Mental Health (Grant #T32-MH19728). Jon Krosnick is university fellow with resources for the future. The authors thank the Gallup Organization for providing the data analyzed in this paper and Joanne Miller, George Bizer, Young Cho, and members of the Political Psychology Research Interest Group at the Ohio State University for their comments and suggestions.

References

Baddeley, Alan D., and Graham J. Hitch. 1977. "Recency Re-examined." In *Attention and Performance IV*, ed. Stanislav Dornie. Hillsdale, NJ: Erlbaum.

Bishop, George, and Andrew Smith. 2001. "Response-Order Effects and the Early Gallup Split-Ballots." *Public Opinion Quarterly* 65:479–505.

Bormuth, John R. 1968. "Close Readability Criterion Reference Scores." *Journal of Educational Measurement* 5:189–196.

Brick, J. Michael, Joseph Waksberg, Dale Kulp, and Amy Starer. 1995. "Bias in List-Assisted Telephone Samples." *Public Opinion Quarterly* 59:219–235.
Casady, Robert J., and James M. Lepkowski. 1993. "Stratified Telephone Survey Designs." *Survey Methodology* 19:103–113. **A8**
Ceci, Stephen J. 1991. "How Much Does Schooling Influence General Intelligence and Its Cognitive Components?: A Reassessment of the Evidence." *Developmental Psychology* 27:703–722.
Cohen, Jacob, and Patricia Cohen. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.
Coleman, Meri, and T. L. Liau. 1975. "A Computer Readability Formula Designed for Machine Scoring." *Journal of Applied Psychology* 60:283–284.
Craik, Fergus I. M., and J. M. Jennings. 1992. "Human Memory." In *The Handbook of Aging and Cognition*, ed. Fergus Craik and Timothy A. Salthouse, pp. 51–110. Hillsdale, NJ: Erlbaum.
Flesch, Rudolph. 1948. "A New Readability Yardstick." *Journal of Applied Psychology* 32:221–233.
Giesbrecht, Lee H. 1997. "Coverage Bias in Various List-Assisted RDD Sample Designs." Paper presented at the 52nd Annual Conference of the American Association for Public Opinion Research, Norfolk, Virginia, May 16, 1997. **A9**
Graesser, Arthur C., Sailaja Bommareddy, Shane Swamer, and Jonathan M. Golding. 1996. "Integrating Questionnaire Design with a Cognitive Computational Model of Human Question Answering. In *Answering Questions*, ed. Norbert Schwarz and Seymour Sudman, pp. 143–174. San Francisco: Jossey-Bass.
Graesser, Arthur C., Natalie K. Person, and John Huber. 1992. "Mechanisms that Generate Questions." In *Questions and Information Systems*, ed. Thomas W. Lauer, Eileen Peacock, and Arthur C. Graesser. Hillsdale, NJ: Erlbaum.
Greenfield, Gerald. 2003. "The Mihazaki EFL Readability Index." *Comparative Culture* 9:41–49.
Gunning, Robert. 1968. *The Technique of Clear Writing*. New York: McGraw-Hill.
Herzog, A. Regula, and Willard L. Rodgers. 1988. "Interviewing Older Adults: Mode Comparison Using Data from a Face-to-Face Survey and a Telephone Resurvey." *Public Opinion Quarterly* 52:84–99.
Herzog, A. Regula, Willard L. Rodgers, and Richard A. Kulka. 1983. "Interviewing Older Adults: A Comparison of Telephone and Face-to-Face Modalities." *Public Opinion Quarterly* 47:405–418.
Hoch, Stephen J. 1984. "Availability and Interference in Predictive Judgment." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:649–662.
Holbrook, Allyson L., Jon A. Krosnick, Richard T. Carson, and Robert Cameron Mitchell. 2000. "Violating Conversational Conventions Disrupts Cognitive Processing of Attitude Questions." *Journal of Experimental Social Psychology* 36:465–494. **A10**
Jacoby, L. L. 1991. "A Process Dissociation Framework: Separating Automatic From Intentional Uses of Memory." *Journal of Memory and Language* 30:513–541. **A11**
Just, Marcel A., and Patricia A. Carpenter. 1987. *The Psychology of Reading and Language Comprehension*. Newton, MA: Allyn & Bacon. **A12**
Just, Marcel A., and Patricia A. Carpenter. 1992. "A Capacity Theory of Comprehension: Individual Differences in Working Memory." *Psychological Review* 99:122–149. **A13**
Kincaid, J. Peter, R. P. Fishburne, R. L. Rogers, and B. S. Chissom. 1975. "Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel." *Research Branch Report* 8–75. Millington, TN.
Klayman, Joshua, and Young-Won Ha. 1987. "Confirmation, Disconfirmation, and Information in Hypothesis Testing." *Psychological Review* 94:211–228.
Knäuper, Barbel. 1999. "Age and Response Order Effects." *Public Opinion Quarterly* 63:347–370.
Koriat, Asher, Sarah Lichtenstein, and Baruch Fischhoff. 1980. "Reasons for Confidence." *Journal of Experimental Psychology: Human Learning and Memory* 6:107–118.
Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213–236.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50:537–67.

Krosnick, Jon A., and Duane F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement." *Public Opinion Quarterly* 51:201–219.

Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, V. Kerry Smith, Wendy R. Moody, Melanie C. Green, and Michael Conaway. 2002. "The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly* 66:371–403.

Krosnick, Jon A., Sowmya Narayan, and Wendy R. Smith. 1996. "Satisficing in Surveys: Initial Evidence." *New Directions for Evaluation* 70:29–44.

Krosnick, Jon A., and Howard Schuman. 1988. "Attitude Intensity, Importance, and Certainty and Susceptibility to Response Effects." *Journal of Personality and Social Psychology* 54:940–952.

McClendon, McKee J. 1986. "Response-Order Effects for Dichotomous Questions." *Social Science Quarterly* 67:205–211.

McClendon, McKee J. 1991. "Acquiescence and Recency Response-Order Effects in Interview Surveys." *Sociological Methods and Research* 20:60–103.

Moore, David W. 2002. "Measuring New Types of Question-Order Effects." *Public Opinion Quarterly* 66:80–91.

Narayan, Sowmya, and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60:58–88.

Newson, R. 2002. "Parameters Behind "Nonparametric" Statistics: Kendall's tau, Somers' *D* and Median Differences." *The Stata Journal* 2:45–64.

Oliver, Laurel W. 1987. "Research Integration for Psychologists: An Overview of Approaches." *Journal of Applied Social Psychology* 17:860–74.

O'Rourke, Diane, Seymour Sudman, Timothy Johnson, and Jane Burris. 1999. "Cognitive Testing of Cognitive Functioning Questions." In *Cognition, Aging, and Self-Reports*, ed. Norbert Schwarz, Denise Park, Barbel Knauper, and Seymour Sudman, pp. 285–302. Philadelphia, PA: Psychology Press.

Park, Denise. C. 1999. "Cognitive Aging, Processing Resources, and Self-Report." In *Cogjition, Aging, and Self-Reports*, ed. Norbert Schwarz, Denise Park, Barbel Knauper, and Seymour Sudman, pp. 285–302. Philadelphia, PA: Psychology Press.

Payne, Stanley L. 1949. "Case Study in Question Complexity." *Public Opinion Quarterly* 13:653–658.

Schaie, K. Warner. 1996. "Intellectual Development in Adulthood." In *Handbook of the Psychology of Aging*, ed. James E. Birren, and K. Warner Schaie, pp. 266–80. Sand Diego, CA: Academic Press.

Schuman, Howard, and Graham Kalton. 1985. "Survey Methods." In *Handbook of Social Psychology*, ed. Gardner Lindzey and Elliott Aronson, pp. 635–698. New York: Random House.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys*. New York: Academic Press.

Smith, E. A. and Peter Kincaid. 1970. "Derivation and Validation of the Automated Readability Index for Use with Technical Materials." *Human Factors* 12:457–464.

Somers, Robert H. 1962. "A New Asymmetric Measure of Association for Ordinal Variables." *American Sociological Review* 27:799–811.

Stata Press Publishing Staff. 2003. *Stata Base Reference Manual 8.0*. College Station, TX: Stata Press.

Tourangeau, Roger, and Kenneth A. Rasinski. 1988. "Cognitive Processes Underlying Context Effects in Measurement." *Psychological Bulletin* 103:299–314.

Tourangeau, Roger, Lance J. Rips, and Kenneth A. Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Tschirgi, Judith E. 1980. "Sensible Reasoning: A Hypothesis About Children." *Child Development* 51:1–10.

Tucker, Clyde, James M. Lepkowski, and Linda Piekarski. 2002. "The Current Efficiency of List-Assisted Telephone Sampling Designs." *Public Opinion Quarterly* 66:321–338.

Wason, Peter C., and Philip N. Johnson-Laird. 1972. *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press. **A20**

Waters, Gloria S., and David Caplan. 1996. "Processing Resource Capacity and the Comprehension of Garden Path Sentences." *Memory and Cognition* 24:342–355. **A21**

Yzerbyt, Vincent Y., and Jacques-Phillipe Leyens. 1991. "Requesting Information to Form an Impression: The Influence of Valence and Confirmatory Status." *Journal of Experimental Social Psychology* 27:337–356. **A22**