7

This chapter discusses an experimental study that shows that the order of items on a questionnaire and the response options for those items both affect the results of college student surveys.

Effects of Item Order and Response Options in College Student Surveys

Nicholas A. Bowman, Jonathon P. Schuldt

Given numerous pressures toward greater accountability and transparency in higher education, obtaining high-quality data about students' experiences and outcomes is more important than ever. A number of recent studies have examined issues related to college student surveys, including survey nonresponse (Adams & Umbach, 2013; Laguilles, Williams, & Saunders, 2011), survey completion (Barge & Gehlbach, 2012; Laguilles et al., 2011), socially desirable responding (Bowman & Hill, 2011; Gonyea & Miller, 2011), and satisficing (i.e., suboptimal cognitive processing that results in low-quality responses; Barge & Gehlbach, 2012; Chen, 2011). As a whole, these studies suggest that careful attention and inquiry is critical for drawing valid conclusions from college student surveys.

This study explores two issues that have received very limited attention in higher education. First, the order in which items are presented may play a role in shaping students' responses. For example, the Cooperative Institutional Research Program's College Senior Survey (CIRP CSS) asked students about their own learning and growth at the beginning of the questionnaire from 2010 to 2012, whereas these items were placed differently—toward the middle of the survey—in the preceding 17 years (see Higher Education Research Institute [HERI], 2014). The National Survey of Student Engagement (NSSE) asked about self-reported gains toward the end of the questionnaire from 2003 to 2013 (appearing before the demographics), whereas these were somewhat closer to the middle of the instrument in 2000–2002 (NSSE, 2014). To what extent might these item order differences within and across surveys affect student responses?

Second, the response options that are provided for a given item or set of items may also shape the results. The CIRP CSS and NSSE, along with many other surveys that are administered within and across institutions, ask about the number of hours per week in which students engage in various activities

New Directions for Institutional Research, no. 161 © 2014 Wiley Periodicals, Inc.
Published online in Wiley Online Library (wileyonlinelibrary.com) • DOI: 10.1002/ir.20070

(e.g., studying, working for pay, socializing). Both of these major national surveys provide eight response options; however, the maximum response on the CSS is "over 20" hours/week compared with "more than 30" for the NSSE. To what extent might these modest differences affect survey results?

Literature Review

A body of research in survey methodology and psychology demonstrates that seemingly trivial differences in questionnaire design can have a pronounced influence on the answers elicited (for reviews, see Schwarz, 1999; Tourangeau, Rips, & Rasinski, 2000). Survey respondents were once largely assumed to consider any given questionnaire item in near-perfect isolation from neighboring items and to be capable of providing fairly accurate reports concerning a range of personal behaviors. However, numerous research findings have challenged these assumptions. In a classic example, Schuman, Presser, and Ludwig (1981) hypothesized that self-reported opinions about a controversial political issue (i.e., abortion) may partly hinge on the nature of a preceding question. To test this hypothesis, the researchers varied the order of two questions that asked whether the respondents would support legal abortions in scenarios when a woman (1) "is married and does not want any more children" and (2) "there is a strong chance of serious defect in the baby." Results showed that a majority of participants (61%) responded "yes" to scenario 1 when that question came first; when scenario 2 came first, however, the proportion of "yes" responses to scenario 1 dropped to fewer than half (48%). In explaining this observation, Schuman and colleagues posited that the thoughts rendered cognitively accessible, or primed, by the more specific of the two scenarios (scenario 2) constrained opinions regarding the more general one (scenario 1) when that item was asked second. In other words, once thoughts of serious birth defects were brought to mind, respondents may have found it more difficult to endorse scenario 1.

This type of cognitive-accessibility explanation for item order effects is common in many studies since the 1980s that explore what are known as cognitive aspects of survey methods or "CASM" (see Schwarz, 2007; Strack, 1992). Over the past few decades, variable response patterns observed under different item orders have come to be seen less as mysterious, random, or haphazard, and more as systematic, predictable outcomes of the basic processes governing human cognition. An instructive example comes from the domain of surveys on subjective well-being, which sometimes ask respondents to provide an overall assessment of life satisfaction in addition to their satisfaction with specific domains of life (e.g., work, dating). An experiment by Strack, Martin, and Schwarz (1988) showed that the order of these questions mattered greatly for the results obtained: When respondents first reported on their overall life satisfaction (general) and then reported on their dating satisfaction (specific), the correlation between these

variables was small and nonsignificant ($r = .16$). In contrast, a strong and significant correlation emerged when their order was reversed ($r = .55$)—a finding that, considered in isolation, might lead an observer to conclude that one's dating life matters enormously for one's overall happiness in life. On the contrary, Strack and colleagues posit that when specific questions precede more general ones, response correlations are inflated because the thoughts rendered accessible by the first question can readily inform the more general judgment.

Similar insights have been gleaned about the impact of response option format on the distribution of survey responses. Although many behavioral frequency surveys tacitly assume that respondents' behavioral histories are well defined and readily accessible in memory, this is not always the case. Whereas respondents may have little trouble accurately recalling how often they have broken a leg in the past two years, more frequent and mundane behaviors are represented more abstractly, opening the door for a variety of context effects when respondents are asked about them. For instance, Smyth, Dillman, and Christian (2007) asked college students to report the number of hours per day that they studied on a six-point scale. They randomly assigned students to respond to either a low-range scale (i.e., the largest category was "more than 2½ hours") or a high-range scale (i.e., the smallest category was "2½ hours or less"). Because instances of such common behaviors likely blend together in memory, the researchers expected responses to be swayed by the social norm implied by the given scale (i.e., low vs. high studying). Indeed, 70% of students reported studying 2.5 hours or less in the low-response option condition, whereas only 29% reported studying 2.5 hours or less in the high-range condition. Presumably, such results reflect the assumption among respondents that researchers have provided a meaningful and accurate distribution of response options, thereby freeing respondents to rely on their subjective sense of studying relative to others as opposed to a "recall and count" strategy for generating a response. It is now widely accepted that respondents engage in these sorts of social and communicative processes in attempting to make sense of the questions asked of them, as they would in any other conversational contexts (e.g., Schaeffer & Presser, 2003; Schwarz, 1996).

Present Study

Given the potential importance of order effects and response options on survey responses, this study examined these two issues within questionnaire designs that are similar to those in the NSSE and CIRP CSS, which are the two most prominent U.S. college experience surveys. By doing so, we hope to inform the practices of institutional researchers, practitioners, and scholars who study college students. While it is clear that a major shift in response options affects the results, would a fairly modest shift also have statistically and practically significant effects on a variety of college

experiences? Similarly, although the order of specific versus general items seems to affect survey responses, would rearranging sets of commonly used questionnaire items alter the responses to these items?

Method

To examine possible effects of item order and response options in college student surveys, we conducted an experiment involving a diverse sample of undergraduates recruited from two U.S. universities. All participants were asked the same set of survey questions; only design features of the survey instrument (i.e., question order and response options) were varied depending on experimental condition.

Data Sources and Participants. Participants were recruited from the psychology subject pools at two public universities: a regional commuter institution in the West and a selective flagship institution in the Midwest. A total of 439 undergraduates completed the survey (61% female, 38% Latino/Hispanic, 32% White/Caucasian, 11% Asian American/Pacific Islander, 9% multiracial/multiethnic, 7% Black/African American, and 3% other race/ethnicity). Students received partial course credit by participating in this online survey.

Measures and Procedure. Across all experimental conditions, some indices were created from 2010 NSSE items (NSSE, 2009), which included academic engagement (20 items; 1 = never, to 4 = very often; $\alpha = .88$), diversity interactions (2 items; 1 = never, to 4 = very often; $\alpha = .87$), and self-reported gains (16 items; 1 = very little, to 4 = very much; $\alpha = .91$). Other indices were created using 2009–2010 CIRP CSS items (HERI, 2009), which included sense of belonging (7 items; 1 = strongly disagree, to 4 = strongly agree; $\alpha = .84$) and campus climate for diversity (3 items; reverse-coded so that 1 = strongly agree, to 4 = strongly disagree; $\alpha = .67$).

Two randomized experiments were conducted simultaneously within the same sample, and participants could have been in any combination of experimental conditions. First, about half of participants were asked to provide self-reported gains before their college experiences and perceptions, and the other half were asked about self-reported gains after their college experiences and perceptions (demographics were assessed at the end of the questionnaire for all students). Second, participants were also randomly assigned to one of two response option conditions when reporting the hours per week spent engaging in 16 college behaviors, which were taken from the 2009–2010 CIRP CSS. One condition used the NSSE response options, which are about evenly distributed (0, 1–5, 6–10, 11–15, 16–20, 21–25, 26–30, more than 30); the other condition used the CSS response options, which are skewed toward smaller frequencies and with a lower maximum value (0, less than 1, 1–2, 3–5, 6–10, 11–15, 16–20, over 20). These responses were then recoded so that they could be compared across conditions. Since the CSS responses had a single category for "over 20," the NSSE

responses of 21–25, 26–30, and more than 30 were combined into a corresponding single category. Similarly, the CSS categories of less than 1, 1–2, and 3–5 were combined for comparison with the NSSE category of 1–5. Although "less than 1" technically does not fit with any of the NSSE choices, we assumed that respondents to the NSSE items who engaged for less than an hour per week would not self-report that they never participated (i.e., by responding 0 hours/week). Thus, the six corresponding categories were 0, 1–5, 6–10, 11–15, 16–20, and over 20. Preliminary factor analyses showed that these 16 experience variables generally could not be combined into any coherent indices with one exception: the socializing, partying, and social network items could be combined into an overall socializing index ($\alpha = .64$).

Analyses. To examine the effect of the item order manipulation, *t*-tests were conducted for each of the indices (self-reported gains, sense of belonging, overall socializing, academic engagement, diversity interactions, and campus climate for diversity). Moreover, 16 chi-square analyses were conducted for the response option condition and each of the college experiences. Preliminary analyses showed that the assumption of an expected count of at least five for each cell was often violated. Therefore, the college experience variables were recoded by combining the three adjacent response options that generally had low frequencies so that there were four categories (0, 1–5, 6–20, over 20). The substantive empirical results for the four- and six-category variables were quite similar. Preliminary analyses also explored whether any experimental effects might be moderated either by the university that students attended or by each other (i.e., whether a certain combination of item order and response options might interact to yield unique results). No significant interactions were identified, so only the analyses of main effects are provided here.

Limitations. Some limitations should be noted. First, the sample only included two public universities. Although these institutions differ notably in terms of selectivity, region, and residential status, additional studies are needed to explore the generalizability of these results. Second, all participants completed the study online. While this survey mode has become quite popular for college student surveys, the results may differ for paper-and-pencil questionnaires. Third, only one order manipulation was conducted (i.e., moving self-reported gains to a different part of the questionnaire), so it is unclear whether and how other sequences would affect students' responses.

Results and Discussion

Participants who completed the self-reported gain items at the beginning of the questionnaire reported significantly higher academic engagement ($M = 2.68$ vs. 2.48, $t = 3.98$, $p < .001$, Cohen's $d = 0.38$), sense of belonging ($M = 2.99$ vs. 2.88, $t = 2.05$, $p = .04$, $d = 0.20$), and self-reported

gains ($M = 2.93$ vs. 2.81, $t = 2.16$, $p = .03$, $d = 0.21$), as well as a marginally better campus climate for diversity ($M = 3.11$ vs. 3.01, $t = 1.80$, $p = .07$, $d = 0.17$) and more diversity interactions ($M = 2.87$ vs. 2.71, $t = 1.73$, $p = .08$, $d = 0.17$) than those who completed the self-reported gains toward the end of the questionnaire. No significant difference for overall socializing was observed ($M = 2.69$ vs. 2.61, $t = 0.83$, $p = .41$). These patterns suggest that students used their responses to previous items to inform their subsequent responses. Specifically, students tend to report that they make fairly substantial gains during college (e.g., NSSE, 2007), so those who completed self-reported gains first may be influenced to report levels of engagement and belonging that are consistent with those initial responses. On the other hand, many general college student surveys ask about a variety of experiences in which most students have not participated (e.g., student government, varsity athletics) or done so rarely (e.g., talking with faculty during office hours), which was also true within this study. Therefore, when students initially respond to college experience items, they may infer that they have not gained as much, since they previously reported having not engaged (or engaging very little) in many possible experiences. On the NSSE, self-reported gains are intentionally presented toward the end so that the earlier college experience and perception questions may inform students' responses to self-reported gains (Gonyea & Miller, 2011). This order likely improves responses for the college experience items that appear toward the beginning of the instrument (and therefore are not biased by any preceding items).

Table 7.1 contains the results for the chi-square analyses examining the relationship between the response option condition and the reported number of hours per week. Of the 16 analyses, 12 effects were significant (ps $< .05$) and two were marginally significant (student clubs/groups and partying, ps $< .10$); the effects were only nonsignificant for working on campus and working off campus (ps $> .66$). Because the hours that students work are systematically tracked—and are identical from week to week for many students—it is not surprising that changing the survey response options has no effect on these self-reports. The college experiences for which significant effects were observed ranged from those that might vary substantially across weeks and are difficult to keep track of (e.g., socializing, watching TV, and social networking) to those that would seem to be fairly consistent and predictable across weeks (e.g., attending class, praying, and commuting). The consistency of these effects across almost all experiences suggests the pervasive influence of response options for shaping self-reports.

To illustrate these effects in more detail, a typical example of the cross-tabulations (for time spent social networking) is shown in Table 7.2. As expected, a notably larger number of participants report spending 1–5 hours/week when this timeframe is represented by three response categories rather than just one. Similarly, a greater number of participants

Table 7.1. Results for Chi-Square Analyses of Survey Response Options and Self-Reported Time Spent Engaging in College Experiences

College Experience	*Chi-Square*	*Significance*
Attending classes/labs	7.96	*
Career planning (job searches, internships, etc.)	36.66	***
Commuting to class (driving, walking, etc.)	46.00	***
Exercising/sports	22.07	***
Online social networks (Facebook, MySpace, etc.)	23.46	***
Partying	7.48	+
Prayer/meditation	18.72	***
Socializing with friends	10.43	*
Student clubs/groups	6.78	+
Studying/homework	14.95	**
Talking with faculty during office hours	16.06	**
Talking with faculty outside of office hours	16.93	***
Volunteer work	13.23	**
Watching TV	11.66	**
Working (for pay) off campus	1.58	
Working (for pay) on campus	1.20	

Notes: Analyses each had 3 degrees of freedom. For all 16 experiences, more participants reported 1–5 hours/week in the low-maximum than the high-maximum condition, whereas more participants reported 6–20 hours/week in the high-maximum than the low-maximum condition. In addition, more participants reported over 20 hours/week in the high-maximum than the low-maximum condition for 14 of the experiences (both conditions had the same number of participants in this highest category for faculty interactions in office hours and out of class).
$^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

report spending more than 20 hours/week when this timeframe is represented by three categories rather than one. The fact that there are a reasonably small number of students in this highest group is common across most experiences; with only a few exceptions (attending class, studying, socializing, and working off campus), the largest response option had the fewest students of any category. Interestingly, far more students reported spending 6–20 hours/week in the high-maximum condition (i.e., more than

Table 7.2. Cross-Tabulations for Survey Response Options and Self-Reported Time Spent Social Networking

		Number of Hours per Week				
Experimental Condition	*Value*	*0*	*1–5*	*6–20*	*Over 20*	*Total*
Low-maximum condition (over 20 hours/week)	Count	16	139	51	11	217
	Column %	39.0	60.4	37.0	36.7	49.4
High-maximum condition (more than 30 hours/week)	Count	25	91	87	19	222
	Column %	61.0	39.6	63.0	63.3	50.6

30) even though the categories were the same for this timeframe across the two conditions; this general pattern occurred for all 16 experiences. In the high-maximum condition, the categories for 6–10, 11–15, and 16–20 hours/week were among the middle out of the eight possible options (3rd, 4th, and 5th from the left, respectively), so these responses likely seemed more normative in this condition than in the low-maximum condition (in which they were 5th, 6th, and 7th, respectively).

Conclusion and Implications

This study showed that both item order and response options can affect college students' survey responses, which has important implications for institutional research and higher education scholarship. Many colleges and universities are interested in determining the extent of students' overall engagement, satisfaction, and intentions. Because the means and distributions for numerous constructs may vary depending upon the order in which items are administered and the response options that are provided, institutional decision makers might draw divergent conclusions depending upon seemingly trivial aspects of the questionnaire design. Moreover, in the controversial book *Academically Adrift* (Arum & Roksa, 2011), the authors argued that students generally spend too little time studying to achieve substantial learning gains. However, the conclusions about the average amount of time studying—and therefore whether that amount is "sufficient"—may depend considerably on the scale used to measure that learning. With the modest response option differences in this study, the proportion of students who reported spending 11 or more hours per week studying was 43% in the high-maximum condition, but only 31% in the low-maximum condition. Thus, these statistically significant effects are also practically meaningful.

The impact of response options broaches an important question: How should researchers and practitioners design questionnaires that yield the most accurate data? For most college experiences, the vast majority of students reported a fairly low number of hours per week (10 or fewer), which suggests that providing smaller categories would be more appropriate to differentiate among student responses and not to push students to report (overly) high values. However, many students do engage for extended periods of time in some experiences, such as attending class, studying/homework, socializing, and working for pay. Therefore, providing higher possible values for these experiences might be helpful and would be less likely to lead to "engagement inflation." For longer surveys, it would probably be best to include two sections with different response options for hours per week; these should appear far enough apart in the questionnaire so that participants are not confused by the shift in available choices.

Another approach would be to ask for open-ended responses so that students can enter the number of hours per week that they spend, which

could subsequently be grouped into meaningful and appropriate categories. The benefit would be that students would not be affected by existing response categories, since the students would create their own value. However, there are at least two potential problems. First, for some experiences (e.g., social networking and watching TV), students may have so little idea of how long they spend that they would skip over this question entirely. These same students might be more willing to provide an educated guess when broad options are provided, which would lead to less missing data. Second, the potential for typographical errors could lead to substantially flawed results. In this situation, the researcher may have no idea whether the student intended to give this response or not; thus, s/he is faced with the unappealing possibilities of (a) removing a potentially correct and informative response, or (b) including a substantial, incorrect outlier that might bias the overall results.

As another practical question, in what order should survey items be presented to yield the most accurate responses? Dillman, Smyth, and Christian (2009) provide an excellent discussion of order effects; they group potential biases into cognitive-based effects (i.e., when early questions affect the processing of later questions) and normative-based effects (i.e., when early questions elicit a social norm that influences responses to later questions). The patterns in this study may have been influenced by both types of effects. That is, answering college experience questions first may have brought these experiences and related considerations in mind when responding to self-reported gains, which would be a cognitive-based priming effect. In addition, these participants may have also been seeking to align their level of engagement with their self-reported gains, which would be a normative-based consistency effect. Several of Dillman et al.'s specific guidelines for question order are also relevant for college student surveys: group related questions that cover similar topics, begin with questions that are likely to be salient to all participants, make the first question(s) interesting (which will also lead to higher completion rates), and ask questions about events in the order in which they occurred (when applicable). Satisfaction questions are particularly susceptible to order effects, so these should generally be asked early in the questionnaire. Moreover, overall satisfaction should be assessed first in a series of satisfaction questions, since the responses for overall satisfaction can be notably affected by reporting satisfaction in specific domains (e.g., Strack et al., 1988). It is also common practice to ask about demographics at the end of the survey; demographic responses are unlikely to be influenced by previous items, but they have the potential to shape participants' thinking regarding other types of items.

Future research should examine these issues in more detail. For instance, how might rearranging items about college satisfaction, student intentions (e.g., regarding retention/persistence), and campus climate affect student responses to these questions and others? Moreover, for both

experimental conditions, this study examined changes in participants' responses to individual items and indices of those items. However, order manipulations can also lead to participants' providing answers that are either more similar or more different from one another across questions (Dillman et al., 2009); these dynamics among commonly used college student survey items should be explored. A strong understanding of the impact of questionnaire design is essential for supporting evidence-based decision making in higher education.

References

Adams, M. J. D., & Umbach, P. D. (2013). Nonresponse and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education*, *53*, 576–591.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*, 182–200.

Bowman, N. A., & Hill, P. L. (2011). Measuring how college affects students: Social desirability and other potential biases in self-reported gains. In S. Herzog & N. A. Bowman (Eds.), *New Directions for Institutional Research: No. 150. Validity and limitations of college student self-report data* (pp. 73–85). San Francisco, CA: Jossey-Bass.

Chen, P.-S. D. (2011). Finding quality responses: The problem of low-quality survey responses and its impact on accountability measures. *Research in Higher Education*, *52*, 659–674.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: Wiley.

Gonyea, R. M., & Miller, A. (2011). Clearing the AIR about the use of self-reported gains in institutional research. In S. Herzog & N. A. Bowman (Eds.), *New Directions for Institutional Research: No. 150. Validity and limitations of college student self-report data* (pp. 99–111). San Francisco, CA: Jossey-Bass.

Higher Education Research Institute (HERI). (2009). *2009–2010 College Senior Survey*. Retrieved from http://www.heri.ucla.edu/researchers/instruments/FUS_CSS/2010CSS.PDF

Higher Education Research Institute (HERI). (2014). *Survey instruments, codebooks, and participation history*. Retrieved from http://www.heri.ucla.edu/researchersToolsCodebooks.php

Laguilles, J. S., Williams, E. A., & Saunders, D. B. (2011). Can lottery incentives boost web survey response rates? Findings from four experiments. *Research in Higher Education*, *52*, 537–553.

National Survey of Student Engagement (NSSE). (2007). *Experiences that matter: Enhancing student learning and success—Annual Report 2007*. Retrieved from http://nsse.iub.edu/NSSE_2007_Annual_Report/index.cfm

National Survey of Student Engagement (NSSE). (2009). *National Survey of Student Engagement 2010*. Retrieved from http://nsse.iub.edu/pdf/US_paper_10.pdf

National Survey of Student Engagement (NSSE). (2014). *Survey instrument*. Retrieved from http://nsse.iub.edu/html/survey_instruments.cfm

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, *29*, 65–88.

Schuman, H., Presser, S., & Ludwig, J. (1981). Context effects on survey responses to questions about abortion. *Public Opinion Quarterly*, *45*, 216–223.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and the logic of conversation*. Hillsdale, NJ: Erlbaum.
Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93–105.
Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, *21*, 277–287.
Smyth, J. D., Dillman, D. A., & Christian, L. M. (2007). Context effects in web surveys: New issues and evidence. In A. Joinson, K. McKenna, T. Postmes, & U. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 427–443). New York, NY: Oxford University Press.
Strack, F. (1992). "Order effects" in survey research: Activation and information functions of preceding questions. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 23–34). New York, NY: Springer.
Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology*, *18*, 429–442.
Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

NICHOLAS A. BOWMAN is an assistant professor of higher education and student affairs at Bowling Green State University.

JONATHON P. SCHULDT is an assistant professor of communication at Cornell University.