

SUSCEPTIBILITY TO SURVEY ITEM ORDER EFFECTS AS A  
FUNCTION OF PERCEIVED CONTROL OVER CONTENT

By

JEFFREY MONROE MILLER

A THESIS PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF ARTS IN EDUCATION

UNIVERSITY OF FLORIDA

2004

Copyright 2004

by

Jeffrey Monroe Miller

## ACKNOWLEDGMENTS

I thank my committee members, Professors M. David Miller and Randall D. Penfield, for their input and feedback regarding this research. I thank Professors Anne Seraphine and Tracy Linderholm for their input regarding the psychological processes addressed in this research. I thank Professor James Algina for his input regarding statistical design and analysis. I thank the University of Florida and the Institutional Research Board for approving this research. I thank my parents, Gail and Daniel Jacobs and Jerry and Darnelle Miller, as well as the following individuals for their support and encouragement: Lee Anna Rasar, Patricia Quinn, Blaine Peden, Todd Liebert, Laura Batich, Terry Anger, Rachal Weidner, Rosaria Upchurch, and Elaine Green.

## TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGMENTS .....	iii
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
CHAPTER	
1 INTRODUCTION .....	1
2 REVIEW OF LITERATURE .....	3
Assimilation and Contrast Item Order Effects.....	3
Measurement of Item Order Effects .....	4
Explanations for Item Order Effects.....	6
Perceived Control as a Moderator of Item Order Effects .....	15
3 METHOD .....	18
Respondents.....	18
Materials .....	18
Procedures.....	22
Design .....	22
4 RESULTS .....	29
5 DISCUSSION .....	36
Interpretation of Results .....	36
Methodological Advantages and Limitations to Findings.....	39
Final Comments.....	45
APPENDIX	
A INVITATION TO PARTICIPATE.....	46

B	FORM 1 – GENERAL ITEM AFTER SET OF SPECIFIC ITEMS.....	47
C	FORM 2 – GENERAL ITEM BEFORE SET OF SPECIFIC ITEMS.....	48
	REFERENCE LIST .....	49
	BIOGRAPHICAL SKETCH .....	52

## LIST OF TABLES

<u>Table</u>	<u>page</u>
3-1 School Descriptives for 2001-2002 Academic Year .....	18
3-2 Reliability Descriptives—“General Item First” / High Control .....	20
3-3 Reliability Descriptives—“General Item First” / Low Control .....	20
3-4 Reliability Descriptives—“Specific Items First” / High Control .....	21
3-5 Reliability Descriptives—“Specific Items First” / Low Control .....	21
3-6 Responses Compared to Test for Interaction of Content, Position, and Type .....	25
3-7 Responses Compared to Test for Interaction of Content and Type .....	26
3-8 Responses Compared to Test for Interaction of Position and Type .....	28
4-1 Descriptive Statistics of Survey Results .....	29
4-2 Descriptive Statistics and Dependent Samples <i>t</i> -Test Statistic for Topic .....	29
4-3 ANOVA Summary Table .....	30
4-4 Descriptive Statistics by Group, Content, and Items .....	31

## LIST OF FIGURES

<u>Figure</u>	<u>page</u>
4-1 Line graph illustrating mean response to general item and mean of specific items as function of both content and group. ....	32
4-2 Line graph illustrating mean response to general item and mean of specific items as a function of content.....	33

Abstract of Thesis Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Master of Arts in Education

SUSCEPTIBILITY TO SURVEY ITEM ORDER EFFECTS AS A  
FUNCTION OF PERCEIVED CONTROL OVER CONTENT

By

Jeffrey Monroe Miller

May, 2004

Chair: M. David Miller  
Major Department: Educational Psychology

Survey item order effects occur whenever the response to one item influences the responses to prior or subsequent items. Such unintended artifacts affect the validity of the interpretation of the survey response scores. Mixed findings in previous research may be due to respondents' perceptions of control over the survey content.

Seventy-two teachers employed by three schools in a medium-sized school district in north-central Florida responded to a survey in which half of the items pertained to classroom climate and half of the items pertained to national educational accountability policy. For each content area, a general item either preceded or followed a set of specific items. Additional items measured perceived control over the two content areas.

The results suggest that surveys are relatively robust to item order effects when the respondents perceive high control over the content. However, when the respondents perceive low control over the content, the presence of item order effects depends on the placement of the general and specific items. Under the perception of low perceived

control, the difference between responses to the general item and the mean of the specific items was greater when the general item preceded the specific items than when the specific items preceded the general item. The results have implications for the development of content-specific surveys, item placement, and the valid interpretation of survey scores.

## CHAPTER 1 INTRODUCTION

An item order effect occurs whenever the response to one item influences the response or responses to a preceding or succeeding item or set of items (Schwarz, Strack, & Mai, 1991; Sigelman, 1981; Tourangeau & Rasinski, 1988; Willits & Ke, 1995). Item order effects have also been termed question order effects (e.g., Alspach & Bishop, 1991) and context effects (e.g., Tourangeau & Rasinski, 1988). Typically, they are explained as an artifact of experimental procedures (McFarland, 1981). Much research has been conducted to better understand these effects in surveys. Underlying most of these experiments is a paradigm in which the respondent does not participate actively in the response process. Rather, specific item features influence the respondent (Bradburn & Mason, 1964; Tourangeau & Rasinski, 1988).

This paradigm may be valid when the rationale for the survey is to gather information that is equally relevant to every member of the survey population. However, some surveys contain content that is personally relevant to the interests of a specific group of respondents (e.g., Schuman, Presser, & Ludwig, 1981). In these scenarios, it is difficult to imagine that item features alone are guiding the response process without the input of a conscious and aware respondent. The respondent is likely to transfer firm beliefs and attitudes into the personally relevant survey response experience (Tourangeau, Rasinski, Bradburn, & D'Andrade, 1989).

Hence, item order effects are influenced by the relevance that the subject matter holds for the respondent (Tourangeau & Rasinski, 1988). However, personal relevance may subsume other constructs. For example, the content of two surveys may be equally personally relevant but is differentiated by the degree of perceived control over the content area. Since much research of item order effects has ignored characteristics of the respondent, it is worthwhile to consider the relationship between respondents' perceived control over the survey content and survey item order effects.

In education, one subject matter that is personally relevant to teachers is their classroom climate (e.g., student conduct, organization). Another subject matter that is personally relevant to teachers is the educational accountability reform initiated in President George W. Bush's No Child Left Behind Act (Bush, 2001). Although teachers likely perceive both subject matters as personally relevant, they should perceive greater personal control over their classroom climate than over national education policy. Such differences in perceived control may then moderate survey item order effects.

The present research begins with a discussion of the different findings of and explanations for item order effects in the literature. This is followed by a description of the study, which considers respondents' perceived control over the survey content and resulting item order effects. The presentation of results is followed by the limitations and implications of this research.

## CHAPTER 2 REVIEW OF LITERATURE

### **Assimilation and Contrast Item Order Effects**

The most commonly studied item order effects involve the relationship between responses to an item of general content and responses to a set of items of specific content. This is known as a part-whole combination (Mason, Carlson, & Tourangeau, 1994; Willits & Ke, 1995). Items of general content tend to be broad and begin with phrases such as “In general” and “Overall,”. Items of specific content tend to be less vague. For example, a general item on a teacher evaluation may be worded, “In general, this was an excellent course”; a specific item may be worded, “The teacher incorporated multimedia on a weekly basis to enhance lectures.”

Respondents in McFarland’s (1981) study completed a survey containing one specific item and one general item on each of four topics—politics, religion, economics, and energy. The surveys were constructed such that the specific item preceded the general item for each topic on half of the surveys, and the general item preceded the specific item for each topic on half of the surveys. An item order effect in which the specific item prompted stronger responses on the following general item was found for the topics of religion and politics, but not for economics and energy. This was explained as a part-whole assimilation effect (Schwarz, Strack, & Mai, 1991) in which the respondent incorporated information from the prior specific item into the responses to the general items.

Item order effects become pronounced when more than one specific item precedes the general item. Tourangeau and Rasinski (1988) explained, “If the list of particulars is long enough, however, it may encourage respondents to interpret the general item as a summary of the particulars rather than as a residual category” (p. 303). The respondent is interpreting the general item as asking, “Taking these specific items together as a whole...”. Hence, they cognitively summarize the specific items when responding to the general item (Schwarz, Strack, & Mai, 1991). Willits and Ke (1995) confirmed this assimilation effect of multiple prior items using 19 specific items related to rural life and a general item related to overall quality of rural life.

Some researchers have found that responses to general items are weaker when following specific items. This effect is known as a part-whole contrast (Schwarz, Strack, & Mai, 1991). One explanation for the lower ratings of general items when following specific items is that the respondent is interpreting the general item as asking, “Aside from [the topic of the specific item]...” The respondent cognitively subtracts the response to the specific item when determining the response to the general item (Bradburn, 1964; Tourangeau & Rasinski, 1988; Tourangeau, Rasinski, & Bradburn, 1991).

One explanation for mixed findings (i.e., assimilation vs. contrast) is based on the quantity of specific items preceding a general item. Contrast is most likely to occur when only one prior item precedes the general item. Assimilation is most likely to occur when multiple specific items precede the general item (Willits & Ke, 1995).

### **Measurement of Item Order Effects**

Item order effects are sometimes measured by comparing correlations. The correlation between general and specific items will be higher than normal under

conditions of assimilation and lower than normal under conditions of contrast<sup>1</sup> (Mason, Carlson, & Tourangeau, 1994). The research of Metzner and Mann (1953) and Schwarz, Strack, and Mai (1991) are examples of a statistical comparison of item correlations.<sup>2</sup>

A directional shift occurs to the extent that prior responses are similar between respondents resulting in group movement toward a similar response on the latter item (i.e., assimilation effect) or group movement toward a different response on the latter item (i.e., contrast effect) (Mason, Carlson, & Tourangeau, 1994; Schuman & Presser, 1981). Moore (2002) argues that past research using directional measurements have not explained the procedures for analyzing the response difference. His research compared the percentage difference in responses between adjacent items.<sup>3</sup>

Sometimes researchers measure item order effects by concentrating on the distribution of categorical responses. Bradburn and Mason (1964) compared response percentages without statistical analysis. Lorenz and Ryan (1996) concluded an item order effect based on a higher percentage of positive responses to a general item when it followed a set of specific items than when it preceded a set of specific items. Statistically, this was analyzed using a chi-square test of independence between item position conditions.<sup>4</sup> Tourangeau, Rasinski, Bradburn, and D'Andrade (1989) compared categorical responses using a logit model and a z-distribution.

---

<sup>1</sup> As explained further in this report, there should be no difference in correlations if item order systematically affects responses from all respondents.

<sup>2</sup> Mason, Carlson, and Tourangeau (1994) conducted an analysis of variance on correlation coefficients.

<sup>3</sup> Although not discussed, the statistical comparison most likely used a chi-square distribution.

<sup>4</sup> The following used similar statistical analysis: Alspach and Bishop (1991), Benton and Daly (1991), Colasanto, Singer, and Rogers (1992), McFarland (1981), Schuman, Presser, and Ludwig (1981), Sigelman (1981), Tourangeau, Rasinski, and Bradburn (1991).

On some occasions, researchers have measured item order effects by comparing means of responses. For example, Schwarz and Hippler (1995) found item order effects as a function of the mode of administration. Willits and Ke (1995) based their analysis on both the comparison of means and correlation coefficients.

More sophisticated analytic techniques have been applied to the study of item order effects. McClendon and O'Brien (1988) used regression analysis where a *t*-test determined the significance of the interaction term (viz., topic by item position). Shields (2003) recently applied structural equation modeling to analyze factor loadings and error variances from item order effects found in the 1986 and 1997 General Social Survey. Specifically, the author found that item order effects did not significantly affect the measurement of the constructs.

### **Explanations for Item Order Effects**

Cognitive theory drives most descriptions and explanations for item order effects. Different researchers have explained these effects using different terms and theories.

#### **Cognitive Accessibility**

Cognitive accessibility is predicated on the notion that memory and attention are limited resources. One explanation for item order effects is that respondents simply do not have access to all information and must rely on the information that is most accessible when a decision must be made (Schwarz, Strack, & Mai, 1991). It is more efficient to use information already present in memory than it is to create an original representation.

In the cognitive psychology literature, the context effect where something prior influences something present is known as priming (Leahey & Harris, 2001). Hence, survey responses may be primed directly by the preceding item. The effect of a string of

prior events influencing the present event can be explained in terms of a spreading activation of stored events in the associative memory network (Posner, 1978; Tourangeau & Rasinski, 1988).

Further, survey responses may be affected by events occurring before administration of the survey (Tourangeau, & Rasinski, 1988). McClendon and O'Brien (1988) refer to the "everyday" frame of reference as those cognitive activities occurring before the survey that may influence the response to general items that precede specific items as well as specific items that precede general items.<sup>5</sup>

### **Salience**

Salience serves to thrust the topic it into attention rendering it more cognitively accessible (Bradburn & Mason, 1964). The effect of a vehicle crashing into another vehicle illustrates priming; the impact of the effect depending on whether the first vehicle is a bicycle, car, or truck illustrates salience. Responses to items related to specific elements of political opinion may effect responses to a subsequent general item pertaining to political opinion because the general item has been primed by the prior specific items, but the extent of the effect may depend on the salience of the topic of political opinions to a particular respondent. The topic may be more salient three weeks before an election and less salient to a child.

McFarland (1981) hypothesized that salience would increase the relationship between general items (religion, politics, the economy, and the energy crisis) and specific items related to those topics when the specific items are posed before the general items. In other words, they proposed a part-whole assimilation effect based on correlations

---

<sup>5</sup> This may be especially problematic if prior events include prior familiarity with the survey (Metzner & Mann, 1953).

between item responses and due to topic salience. The results suggested item order effects for the topics of politics and religion but not for the economy and the energy crisis. The researchers explained this disparity in terms of differences in item characteristics. Some items were requests for levels of interest while others were requests for discriminating judgments.

Cognitive accessibility addresses limitations of cognition. Salience addresses limitations due to the interaction between respondent and topic. The next two explanations address limitations imposed through conscious awareness of the respondent.

### **Redundancy**

Tourangeau and Rasinski (1988) and Shields (2003) mention Grice's (1975) maxim that "one should be informative and avoid redundancy (p. 302)".<sup>6</sup> Redundancy of item content can create a low correlation between related items (Bradburn & Mason, 1964). If participants recognize that items are related they may choose to not answer in like manner in order to not be redundant<sup>7</sup>. Hence, an assimilation item order effect may be present but intentionally violated due to the participant's awareness and decision. A conscious decision to rebel against consistency interferes with the assimilative effect of limited-cognition priming.

Redundancy of item content also provides an explanation for contrast effects when multiple specific items precede a general item. The respondent is initially compelled to assimilate due to priming and spreading activation. The respondent then opts for an inconsistent response merely to avoid redundancy.

---

<sup>6</sup> This maxim is known in the psycholinguistics literature as the given-new contract (Schwarz, Strack, & Mai, 1991).

## **Consistency**

Participants' awareness and decisions can also create a consistency effect (Bradburn & Mason, 1964)<sup>8</sup>. The participant opts to respond in a manner that is uniform with prior responses regardless of the true response. Whether one opts to avoid redundancy or to retain consistency may be a function of the number of prior specific items, personality, mood, the survey topic, or other variables.

## **Social Desirability**

Social desirability is a hindrance to the interpretation of virtually any survey score. In item order studies, social desirability surfaces when the general item precedes the specific item. Since respondents have not had time to become acquainted with the topic, they may respond in a stereotypical manner so that the response does not deviate from the expected norm response (Sigelman, 1981)<sup>9</sup>. Interpreting responses affected by social desirability is especially difficult because the researcher must disentangle the desire to grant a favorable response from both the possibility of it being the respondent's true response as well as the possibility of it being a mechanism for dealing with uncertainty.

Moore (2002) found an effect of social desirability in survey responses and explained it as "the classic case of people trying to make their ratings of the two men [Clinton and Gore] more consistent" (p.83). However, Sigelman (1981) did not find an effect of social desirability when posing a general item regarding the current president's performance before or after 47 items specific to the current president's performance.

---

<sup>8</sup> Tourangeau & Rasinski (1988) describe consistency as a carryover effect whereas redundancy produces a backfire effect.

<sup>9</sup> Acquiescence is a related term where the tendency is to agree (Dillman, 2000).

**Noncommittal Response**

Sigelman (1981) also suggested the possibility of a noncommittal response as an alternative strategy for dealing with general items posed early in the survey. In this instance, the respondent is not yet acquainted with the topic, apparently has not formulated a judgment, and chooses to not respond to the item. Of course, there are other explanations (e.g., the participant accidentally skipped the item or the participant intended to return to the item). Regardless, this creates contradictory suggestions for item placement. One argument is that general items should precede specific items in the survey to avoid the aforementioned effects of salience, consistency, and redundancy. The counterargument is that general items posed first are more susceptible to the effects of social desirability and noncommittal responses.

**Personal Relevance**

One variable that may explain anomalous findings is the degree of personal relevance that the survey topic holds for the respondent. First, “how much effort is put into testing for order effects will depend on the uses for which the results of a particular survey are intended” (Bradburn & Mason, 1964, p.61).<sup>10</sup> Second, the principle of cognitive economy explains that a participant will only put as much effort into responding as is necessitated by the participant’s motives (Tourangeau & Rasinski, 1988).

---

<sup>10</sup> This is consistent with the National Council on Measurement in Education standards for testing which state, “It is the interpretations of test scores required by proposed uses that are evaluated, not the test itself” (AERA, APA, & NCMA, 1999).

Personal relevance includes such features as partisanship, issue familiarity, and belief entrenchment. Tourangeau and Rasinski (1988) provided the following explanation on pg.32:

Highly partisan respondents appear to answer a series of items . . . by using an absolute standard but respondents with mixed views appear to respond more flexibly, perhaps making comparisons among the individual items . . . when the issue is ambiguous or unfamiliar . . . respondents may have difficulty in identifying a relevant attitude structure; they must search for one, and context can bias this search . . . middle-of-the-road respondents (who are likely to have mixed views) are most vulnerable to context effects . . . expert and involved respondents [terms that characterize familiarity] respondents ought to be less prone to such effects.

If the survey topic is relevant to respondents, they may consistently place more time and thought into their responses yielding a more valid interpretation of responses as well as more identification of and interpretation of item order effects (Tourangeau & Rasinski, 1988). On the other hand, if the content is irrelevant then it is likely that the respondent has not formulated a solid belief regarding the topic. For example, survey results regarding automotive tools and machinery are much more valid when obtained from auto mechanics than when obtained from accountants (presumably with no auto mechanic interest, training, or experience.) Examples of survey topics considered personally relevant to particular respondents include abortion (Schuman, Presser, & Ludwig, 1981) and religion (McFarland, 1981).

McFarland's (1981) findings can be explained as artifacts of personal relevance. Initially, the premise was that item order effects would be differentiated by topic salience. Anomalous results were explained as due to differences in item verbiage (i.e., interest vs. discriminating judgments). Although all four topics may have been equally or differentially salient regardless of item characteristics, perhaps the topics of politics and religion were more personally relevant to the respondents than the topics regarding the

economy and the energy crisis. This leads to the notion that personal relevance of item content overrides both salience and item characteristics. Broadly speaking, respondent characteristics may be a larger determinant of susceptibility to item order effects than item characteristics and environmental conditions.

The respondents in McFarland's (1981) study may not have even had to possess a strong opinion regarding the topics for a relevance-based item order effect to occur. Tourangeau, Rasinski, Bradburn, and D'Andrade (1989) found that respondents were most susceptible to item order effects when their beliefs were conflicted or mixed and, simultaneously, were important. One interpretation is that when a topic is personally relevant, its presence as a cognitive entity is firm; the individual does not need excessive cognitive effort in order to generate a response, judgment, or evaluation in these domains. This process has also been referred to as attitude crystallization; based on the file drawer model, attitude and evaluative judgments are automatically retrieved (Shields, 2003; Tourangeau, Rips, & Rasinski, 2000). Hence, contrast item order effects are more prevalent when the respondent has to cogitate about the item, which minimizes cognitive accessibility and promotes responses based on other processes (Shields, 2003) (e.g., priming, salience, avoiding redundancy, and maintaining consistency).

### **Summary**

Item order effects tend to occur when a specific item or a set of specific items precedes a general item or a set of general items. In some situations, the result is assimilation in which the correlation between the general item and the specific items are higher than normal or follow the same trend. In other situations, the result is contrast in which the correlation between the general item and the specific items are lower than normal or follow a different trend.

From a psychometric perspective, there should be no difference between correlations regardless of assimilation or contrast. Suppose that one group completes a form containing only specific items and another group completes a form in which the first item is a general item. For the first group, the correlation is between the first item ( $X$ ) and the remaining items ( $Y$ ). For the second group, due to the contrast effect, responses to the remaining items are lower than the responses to these items made by the first group. However, all participants are assumed to respond lower in a systematic fashion. Hence, ( $Y$ ) is reduced by a constant resulting in ( $Y - C$ ). Under these circumstances, there should be no difference in correlations between the two groups (i.e.,  $\rho_{X,Y} = \rho_{X,Y-C}$ ). It is more appropriate to compare the mean responses between groups to determine the statistical significance of these differences.

When only one specific item precedes a general item, an assimilation effect may be due to (a) high topic salience, or (b) the desire to remain consistent in responding. When several specific items precede a general item, the assimilation effect may be due to (a) the perception that the general item is calling for a summative judgment, (b) priming prompted by diminished cognitive accessibility, (c) high topic salience, or (d) the desire to remain consistent in responding. However, regardless of the number of prior specific items, assimilation item order effects may be subsumed by a high degree of personal relevance of the content to the respondent.

When only one specific item precedes a general item, a contrast effect may be due to (a) the perception that the general item is calling for a subtractive response, or (b) the desire to avoid redundancy. When several specific items precede a general item, the contrast effect may be due to the desire to avoid redundancy. However, regardless of the

number of prior specific items, all of these possibilities may be subsumed by a low degree of personal relevance of the content to the respondent.

The supremacy of personal relevance suggests that conscious processing of survey content information occurs first. If the content is highly personally relevant then the response process becomes a conscious process with deliberate intent (Feldman & Lynch, 1988). If the content is not highly personally relevant, then it is speculated that content features such as salience prompt responses. In the event of low personal relevance and low content salience, cognitive accessibility is preserved by resorting to heuristics (e.g., maintaining consistency, avoiding redundancy).

It is important to make these distinctions because different surveys target both different populations and different breadths of topic. For example, many surveys are intended to reach a large population of people assumed to all be capable of answering the items but to not necessarily be highly personally invested (e.g., presidential approval ratings; Alspach & Bishop, 1991). In survey research, these often take the form of attitudes of presidential popularity, life satisfaction, and marital happiness. On the other hand, some surveys target a specific sample presumed to be highly personally invested in the survey topic (e.g., abortion; Schuman, Presser, & Ludwig, 1981).<sup>11</sup>

Varying degrees of content relevance may promote either an assimilation or contrast effect leading to varying degrees of the use of strategies and heuristics.

Furthermore, the respondent may utilize more than one

---

<sup>11</sup> The compromise is studies in which the respondent may not be highly personally invested in the topic, yet the scope is socially relevant such as people's misconceptions about AIDS (Colasanto, Singer, & Rogers, 1992).

strategy or heuristic. For example, when content is highly personally relevant, item order effects may be moderated to the extent that the topic is concurrently highly salient, the respondent tends to avoid redundancy, but the respondent tends to respond in a socially desirable manner.

### **Perceived Control as a Moderator of Item Order Effects**

Another factor that may subsume the personal relevance of survey content is the degree to which the respondent perceives control over the personally relevant subject matter. For example, respondents may find both the amount of violent programming at home and national policies on violent television programming to be highly personally relevant. However, most respondents have greater personal control over television viewing at home than over national policies. Hence, it is possible to theorize layers of relevance with each outer layer permitting less personal control.

Mason, Carlson, and Tourangeau's (1994) results may be interpreted in terms of personal relevance and perceived control. The specific item pertained to perceptions of the future economic situation in respondents' communities, and the general item pertained to perceptions of the future economic situation in the respondent's state (Oregon or Idaho). Both can be considered personally relevant to the respondents living in these communities and states. Hence, an assimilation effect should occur if the topic is highly salient (or the respondent resorts to promoting consistency of response); a contrast effect should occur if the respondent views the general items as calling for a subtractive response (or the respondent resorts to avoiding redundancy of response). The correlation between the specific (community) and general (state) item was consistently lower when the specific item preceded the general item suggesting a contrast item order effect.

The authors offered an interpretation consistent with resorting to avoid redundancy. This explanation is based on the least involvement of conscious decision-making. An additional interpretation regards personal relevance. However, this asserts that state economic concerns are not as personally relevant to the respondent as community economic concerns. Further, personal relevance should lead to assimilation effects not the obtained contrast effects.

An alternative explanation is based on the notion that citizens have more perceived control over community issues than over state issues. The presentation of the state item before the community item can be considered a general – specific and low control – high control presentation format. Lack of control over the initial item situated an attitude that then prompted a consistent response to the following item. The presentation of the community item can be considered a specific – general and high control – low control presentation format. The perception of control over the initial item inspired higher conscious processing for both items prompting the lower correlation between items. However, this explanation would be more valid if the general – specific format was disentangled from the high control – low control format and investigated systematically.

Lorenz and Ryan's (1996) study is an example of this manipulation. Two separate topics, respondents' communities and respondents' local governments, were included in both telephone and self-administered formats. Responses were more positive for the community general item when it preceded the specific items than when it followed the specific items; however, responses were more negative for the local government general item when it preceded the specific items than when it followed the specific items. Although not part of the original hypotheses, the authors explained the anomalous

findings. “A post-hoc interpretation of this unexpected result is that respondents may already have had well-defined and mostly positive images of their community, images that were undermined when they were asked to first examine the specific domains before making an overall evaluation (Lorenz & Ryan, 1996, p.612).” A similar interpretation is that the respondents perceived higher control over community than over local government. However, it is difficult to substantiate this interpretation without comparing the distribution of responses to the general item and the distribution of responses to the specific items.

Alspach and Bishop (1991) found an item order effect for the approval of former president, Ronald Reagan, but not for former Ohio governor, Richard Celeste. At first glance, this finding contradicts the notion that local topics are more resilient to item order effects than distant topics. However, the authors suggest that the findings are due to a wider awareness and firmer opinion of the president than the governor. Hence, it is imperative that awareness as well as personal relevance be considered when determining topics of differential perceived control. For the present study, it is hypothesized that item order effects are a function of both item position and perceived control over the survey topic.

## CHAPTER 3 METHOD

### Respondents

The sample was 250 teachers employed in one of three high schools (School A:  $n = 86$ , School B:  $n = 117$ , School C:  $n = 47$ ) in a medium-sized school district in north-central Florida during the 2002-2003 academic school year. Specifically, these three schools were selected due to (a) availability of teacher names on school websites<sup>1</sup>, and (b) desire for a representative sample of teachers in this school district. Table 3-1 displays several key demographic variables that highlight the schools' similarities and differences (School Advisory Council Report Results, 2001-2002). Seventy-two teachers (28.8%) returned the surveys (School A:  $n = 18$ , School B:  $n = 42$ , School C:  $n = 12$ ).

Table 3-1. School Descriptives for 2001-2002 Academic Year

Demographic	School A	School B	School C	State
Graduation Rate (%)	62.6	71.7	77.6	67.9
School Grade (Determined by State)	B	A	A	N/A
Students Absent 21+ days	20.2	12.0	16.5	15.1
White, Non-Hispanic (%)	38.8	69.6	77.8	54.6
Black, Non-Hispanic (%)	50.3	21.9	16.6	23.1
Hispanic	2.6	5.4	3.4	19.0
Other Race	8.3	3.0	2.2	3.3

### Materials

Appendix A contains the cover page mailed to all participants.

<sup>1</sup> The sample may not have included all teachers in the school due to the potential for teacher mobility between updates to the school website where names were obtained.

Two 14-item surveys, differing by position of the general and specific items, were developed to assess item order effects for two different content areas. Each survey contained six items assessing perceptions of classroom climate (items 1 – 6) and six items assessing perceptions of national accountability policy (items 8 – 13). An item assessing perceived control over the content area followed each set of six items (item 7 and item 14). Two additional open-ended items requested the course subjects taught by the teacher and the number of years of teaching experience.

Appendix B contains the survey form in which five specific items (items 1 –5) precede one general item (item 6) for the high control content area and five specific items (items 8 – 12) precede one general item (item 13) for the low control content area.

Appendix C contains the survey in which one general item (item 1) preceded five specific items (items 2 – 6) for the high control content area and one general item (item 8) preceded five specific items (items 9 – 13) for the low control content area. The same items were included on both surveys; the only difference between the surveys was the position of general and specific items. Responses were made using a 10-point (0-9) Likert scale. All forms were coded to identify the teacher's school.

Scale reliability was addressed separately for the “general item first” form and the “specific items first” form.<sup>2</sup> Analyses of reliability distinguished between the sections of the form pertaining to high control items (items 1 – 7) and low control items (items 8 – 14). For the “general item first” group, Cronbach's alpha based on high control items was

---

<sup>2</sup> Calculation of reliability across forms is contrary to the purposes of the study. If item order effects are different depending on whether the general item is presented before or after the specific items, the results should be manifested in varying consistency of responses. Hence, disregarding item placement by collapsing across forms would unjustifiably compromise the reported reliability.

.82. Table 3-2 summarizes the results. The first item did not discriminate adequately between respondents most likely to respond in agreement or disagreement to a particular item. Although deletion of this item would not appreciably increase reliability, rewording may improve discrimination.

Table 3-2. Reliability Descriptives—"General Item First" / High Control

Item	n	M	SD	Corrected Item Total Correlation	Reliability if Item Deleted
1	31	7.45	1.65	0.38	0.84
2	31	8.23	1.02	0.51	0.80
3	31	7.97	1.38	0.56	0.79
4	31	7.42	1.63	0.70	0.77
5	31	7.90	1.04	0.51	0.80
6	31	8.13	0.92	0.57	0.80
7	31	7.81	1.11	0.85	0.75

For the "general item first" group, Cronbach's alpha based on low control items was .92. Table 3-3 summarizes these results. Inspection of the item-total correlations revealed that all items discriminated adequately between respondents most likely to respond in agreement or disagreement to a particular item.

Table 3-3. Reliability Descriptives—"General Item First" / Low Control

Item	n	M	SD	Corrected Item Total Correlation	Reliability if Item Deleted
8	33	4.73	2.50	0.68	0.92
9	33	3.67	2.35	0.78	0.91
10	33	2.97	2.46	0.87	0.90
11	33	3.52	2.53	0.76	0.91
12	33	3.06	2.77	0.85	0.90
13	33	3.61	2.86	0.78	0.91
14	33	1.42	2.08	0.57	0.93

For the “specific items first” group, Cronbach’s alpha based on the high control items was .51. Almost 50% of the variance in these responses was due to unsystematic error. Although the mean response to all items was similar, responses to item 2 and item 7 did not discriminate adequately between respondents most likely to respond in agreement or disagreement to a particular item. Table 3-4 summarizes these results.

Table 3-4. Reliability Descriptives—“Specific Items First” / High Control

Item	n	M	SD	Corrected Item Total Correlation	Reliability if Item Deleted
1	36	8.58	0.81	0.36	0.45
2	36	8.00	1.67	0.04	0.61
3	36	8.36	1.10	0.36	0.43
4	36	7.83	1.38	0.22	0.49
5	36	8.08	0.97	0.04	0.54
6	36	8.17	0.88	0.58	0.37
7	36	8.11	1.30	0.40	0.40

For the “specific items first” group, Cronbach’s alpha based on the low control items (items 8 – 14) was .94. Table 3-5 summarizes these results. This section displayed the strongest levels of discrimination between those respondents most likely to respond in agreement or disagreement to a particular item.

Table 3-5. Reliability Descriptives—“Specific Items First” / Low Control

Item	n	M	SD	Corrected Item Total Correlation	Reliability if Item Deleted
8	34	2.82	2.49	0.90	0.92
9	34	2.76	2.51	0.89	0.92
10	34	3.21	2.90	0.70	0.94
11	34	2.29	2.49	0.85	0.92
12	34	3.38	2.89	0.83	0.93
13	34	3.41	2.57	0.81	0.93
14	34	1.32	2.08	0.62	0.94

## Procedures

During September of 2003, all surveys were hand-delivered to the three schools. For each school, half of the teachers received a form in which the general item preceded the specific items; half of the participants received a form in which the specific items preceded the general item. Random assignment was used to determine which teachers would receive which forms. Self-addressed envelopes were included to increase survey response.<sup>3</sup> The scores and demographic information from returned surveys were entered into SPSS 11.0 for data management and statistical analysis.

## Design

The design was based on establishing three independent variables. Item content is a within-subjects variable with two levels (high control and low control). This variable represents the two sections of the survey presented to all respondents. Item type is also a within-subjects variable with two levels (response to general item and mean response to specific items). This variable represents the two types of items presented to all respondents. Finally, item position is a between-subject variable with two levels (general item first or specific items first). This variable represents the two groups of participants who completed different forms.

The analysis of survey data was dependent on the following specific hypotheses:

**Hypothesis 1:** Teachers will indicate significantly higher personal control over their classroom climate than over national accountability policy. This hypothesis served as a manipulation check for the assumption that the topics differ in perceived control from the onset of the study. Intuitively, teachers should perceive more control over their classes

---

<sup>3</sup> Return of surveys within the school district was free to teachers through use of the “truck-mail” delivery system.

than over national education policy; the analysis of responses to these items provides a more rigorous test of this assumption. Specifically, there should be a significant difference in the mean response to item 7 (“I have personal control over the environment in my classes.”) and item 14 (“I have personal control over national education accountability issues.”) indicating that teachers perceive more control over the topic in item seven than over the topic in item fourteen.

Since, this hypothesis was not concerned with item position, the responses to item 7 as well as the responses to item 14 were aggregated across the two forms. This hypothesis was tested using a nondirectional dependent samples *t*-test with the Type I error rate set at  $\alpha = .05$ . Rejection of the hypothesis of no difference between means would suggest that the difference in perception of control between the two topics is statistically different from zero.

**Hypothesis 2:** There will be an interaction of content, type, and position. Teachers perceiving low control over the content area will have a greater difference in responses between the general and specific items than when they perceive high control over the content area especially when the general item precedes the specific items. Based on the test of item 7 and item 14, the high control content area was classroom climate (items 1 – 6), and the low control content area was national accountability policy (items 8 – 13). The rationale for this hypothesis was to add evidence for or against the premise that item order effects are a function of perceived control over the content.

The specific item responses compared for this analysis are summarized in Table 3-6. For the group in which the general item preceded the specific items pertaining to classroom climate, the analysis compared the responses to item 1 and the mean of

responses to items 2 - 6. For national accountability policy, the analysis compared the responses to item 8 and mean of responses to items 9 – 13. For both topics, the preceding items (item 1 and item 8) began with the phrase, “In general...” and were followed by items with content specific to classroom climate.

For the group in which the specific items precede the general item pertaining to classroom climate, the analysis compared the responses to item 6 and the mean of responses to items 1 – 5. For national accountability policy, the analysis compared the responses to item 13 and the mean of responses to items 8 – 12. For both topics, the preceding items (items 1-5 and items 8-12) were specific to that content area and were followed by an item that began with the phrase, “In general...” (item 6 and item 13).

To illustrate the hypothesized interaction, suppose that the group responding to the “general item first” survey had a mean difference of 2 for the high control content area (i.e., the group’s difference in responses between item 1 and the mean of items 2 - 6) and a mean difference of 5 for the low control content area (i.e., the group’s mean difference between item 8 and the mean of items 9 - 13). Further suppose that the group responding to the “specific items first” survey had a mean difference of 3 for the high control content area (i.e., the group’s mean difference between item 6 and the mean of items 1 - 5) and a mean difference of 4 for the low control content area (i.e., the group’s mean difference between item 13 and the mean of items 8 - 12). Statistical significance of these fictitious results would be explained as the hypothesized interaction. For both groups, the mean difference in responses was greater for the low control content area than for the high control content area (i.e., 5 vs. 2 for the “general item first” group; 4 vs. 3 for

the “specific items first” group). However, the difference was greater for the “general item first” (i.e.,  $5 - 2 = 3$ ) group than for the “specific items first group” ( $4 - 3 = 1$ ).

Table 3-6. Responses Compared to Test for Interaction of Content, Position, and Type

Topic	Group	
	“General Item First”	“Specific Items First”
High Control (Classroom Climate)	Item 1 vs Mean of Items 2 – 6	Mean of Items 1 – 5 vs. Item 6
Low Control (Accountability)	Item 8 vs Mean of Items 9 – 13	Mean of Items 8 – 12 vs. Item 13

This hypothesis was tested with 2 X 2 X 2 mixed analysis of variance using the independent variables of item content, item type, and item position. The Type I error rate was set at  $\alpha = .05$ .

**Hypothesis 3:** There will be a significant interaction of content and type. The mean difference in responses between the general item and the specific items is greater when teachers perceive less control over the content area than when they perceive high control over the content area regardless of item placement. This hypothesis ignored the distinction between the “general item first” group and “specific items first” group. The rationale for this hypothesis was to examine differences between ratings of items when they differ in perceived control over the item content. This served to add more evidence to the claims of the first hypothesis while exploring whether or not differences due to content vary as a function of whether the item is general or a mean of specifics.

The items compared in this analysis are summarized in Table 3-7. For the high control content area, the responses to the general item for the “general item first” group (item 1) and the responses to the general item for the “specific items first” group (item 6) were combined. The responses to the specific items for the “general item first” group

(items 2 - 6) and the responses to the specific items for the “specific items first” group (items 1 - 5) were combined. This resulted in a total “general item” score and a total mean “specific items” score that disregarded item placement in the high control content portion of the survey.

Table 3-7. Responses Compared to Test for Interaction of Content and Type

<p style="text-align: center;">High Control (Classroom Climate)</p>	<p style="text-align: center;">Item 1 vs. Mean of Items 2 – 6 if “General Item First” and Mean of Items 1 – 5 vs. Item 6 if “Specific Items First”</p>
<p style="text-align: center;">Low Control (Accountability)</p>	<p style="text-align: center;">Item 8 vs. Mean of Items 9 – 13 if “General Item First” and Mean of Items 8 – 12 vs. Item 13 if “Specific Item First”</p>

For the low control content area, the responses to the general item for the “general item first” group (item 8) and the responses to the general item for the “specific items first” group (item 13) were combined. The responses to the specific items for the “general item first” group (items 9 - 13) and the responses to the specific items for the “specific items first” group (item 8 - 12) were combined. This resulted in a total “general item” score and a total “specific items” score that disregarded item placement in the low control content portion of the survey. According to this hypothesis, the mean difference between these scores is greater than the mean difference between the aforementioned scores for the high control content portion of the survey. This hypothesis was tested by examining the two-way interaction of item content and item type in the aforementioned 2 X 2 X 2 mixed analysis of variance. The Type I error rate was set at  $\alpha = .05$ .

**Hypothesis 4:** There will not be a significant interaction of type and position. The mean difference in responses between the general item and the mean of the specific items is equal when the general item precedes the specific items and when the specific items precede the general item regardless of perceived control over the content area. In contrast to the previous hypothesis, this hypothesis ignored the distinction between content. The rationale for this hypothesis was to add evidence to the premise that interactions are primarily due to differences in perceived control over content. A nonsignificant interaction of item type and item position while controlling for item content suggests that ratings for the general item and the mean of specific items do not vary as a function of item order after disregarding item content.

The items compared in this analysis are summarized in Table 3-8. For the “specific items first” group, the responses to the general item for the high control content area (item 6) and the responses to the general item for the low control content area (item 13) were combined. The responses to the specific items for the high control content area (items 1 - 5) and the responses to the specific items for the low control content area (item 8 - 12) were combined. This resulted in a total “general item” score and a total mean “specific items” score that disregarded perceived control over content area for the group in which the specific items precede the general item.

For the “general item first” group, the responses to the general item for the high control content area (item 1) and the responses to the general item for the low control content area (item 8) were combined. The responses to the specific items for the high control content area (items 2 - 6) and the responses to the specific items for the low control content area (items 9 - 13) were combined. This resulted in a total “general item”

score and a total mean “specific items” score that disregarded perceived control over content area for the group in which the general item precedes the specific items.

This hypothesis was tested by examining the interaction of item position and item type in the aforementioned 2 X 2 X 2 mixed analysis of variance. The Type I error rate was set at  $\alpha = .05$ .

Table 3-8. Responses Compared to Test for Interaction of Position and Type

“General Item First”	Item 1 – Item 2 if “High Control Content Area” and Item 8 – Item 9” if “Low Control Content Area”
“Specific Item First”	Item 5– Item 6 if “High Control Content Area” and Item 12 – Item 13 if “Low Control Content Area”

## CHAPTER 4 RESULTS

Table 4-1 displays the descriptive statistics resulting from this study. The first hypothesis stated that teachers would indicate significantly higher personal control over their classroom climate than over national accountability policy. The nondirectional dependent samples *t*-test rejected the null hypothesis of no mean difference,  $t(67) = 22.99$ ,  $p < .001$ . Teachers perceived significantly greater control over classroom content than over national educational accountability policy. Table 4-2 displays these results.

Table 4-1. Descriptive Statistics of Survey Results

Topic	Condition					
	General First			Specifics First		
	n	Mean	SD	n	Mean	SD
Classroom Climate						
Personal Control	33	7.64	1.62	36	8.11	1.30
General Item	31	7.45	1.65	37	8.16	0.87
Specific Items	33	7.89	0.97	37	8.17	0.61
National Policy						
Personal Control	33	1.42	2.08	38	1.34	2.00
General Item	33	4.73	2.50	37	3.49	2.50
Specific Items	34	5.27	1.79	38	3.21	2.47

Table 4-2. Descriptive Statistics and Dependent Samples *t*-Test Statistic for Topic

Topic	N	Mean	SD	<i>t</i>	df	<i>p</i>
Classroom Climate	67	7.90	1.48	22.99	67	.000
Accountability	67	1.43	2.05			

The second hypothesis stated that there would be a significant interaction of item content (high control or low control), item position (general item first or specific items first), and item type (general item and mean of specific items). The three-way mixed ANOVA rejected the null hypothesis of no interaction,  $F(1, 64) = 8.56, p = .005$ . Table 4-3 presents the summary of the analysis of variance.

Table 4-3. ANOVA Summary Table

Source	df	SS	MS	<i>F</i>	<i>p</i>
Between Subjects					
Item Position (P)	1	2.609	2.609	.396	.530
Subjects (S) / P	64	421.61	6.588		
Within Subjects					
Item Content (C)	1	1114.463	1114.463	223.368	.000
CP	1	31.132	31.132	6.240	.015
SC/P	64	319.318	4.989		
Item Type (T)	1	6.321	6.321	5.472	.022
TP	1	1.949	1.949	1.687	.199
ST/P	64	73.935	1.155		
CT	1	20.501	20.501	15.931	.000
PCT	1	11.022	11.022	8.564	.005
SCT/P	64	82.362	1.287		

The interaction of item content and item position as a function of item type was of key interest in this study. Table 4-4 displays the means, standard errors, and 95% confidence intervals. Figure 4-1 is a line graph illustrating that teachers perceiving low control over the content area have lower responses for both the general item and the mean of specific items. The figure further illustrates that teacher perceiving low control over the content rated the general item higher than the mean of the specific items when the general item preceded the specific items. Although they also rated the general item higher

than the mean of the specific items when the specific items preceded the general item, the difference was not as large.

Table 4-4. Descriptive Statistics by Group, Content, and Items

Group / Content	Item	M	SE	95% Confidence Interval	
				Lower Bound	Upper Bound
Specifics First / High Control	General	8.19	0.22	7.76	8.63
	Mean of Specifics	8.21	0.12	7.96	8.45
General First / High Control	General	7.47	0.24	7.00	7.94
	Mean of Specifics	7.95	0.14	7.68	8.22
Specifics First / Low Control	General	3.53	0.41	2.71	4.34
	Mean of Specifics	3.24	0.40	2.44	4.04
General First / Low Control	General	5.00	0.45	4.11	5.90
	Mean of Specifics	3.55	0.44	2.67	4.24

The third hypothesis stated that there would be a significant interaction of item type and item content. Items were combined across forms; hence, this hypothesis ignored the factor of item position. In reference to Figure 4-1, this hypothesis tested the difference between the mean of the top two lines and the mean of the bottom two lines. Based on the original analysis displayed in Table 4-3 as Source CI, the two-way interaction was significant,  $F(1, 64) = 15.931, p < .001$ . Specifically, teachers rated the general item higher when they perceived high control over the topic ( $M = 7.83, SE = 0.16$ ) than when they perceived low control over the topic ( $M = 4.26, SE = 0.30$ ). These results were

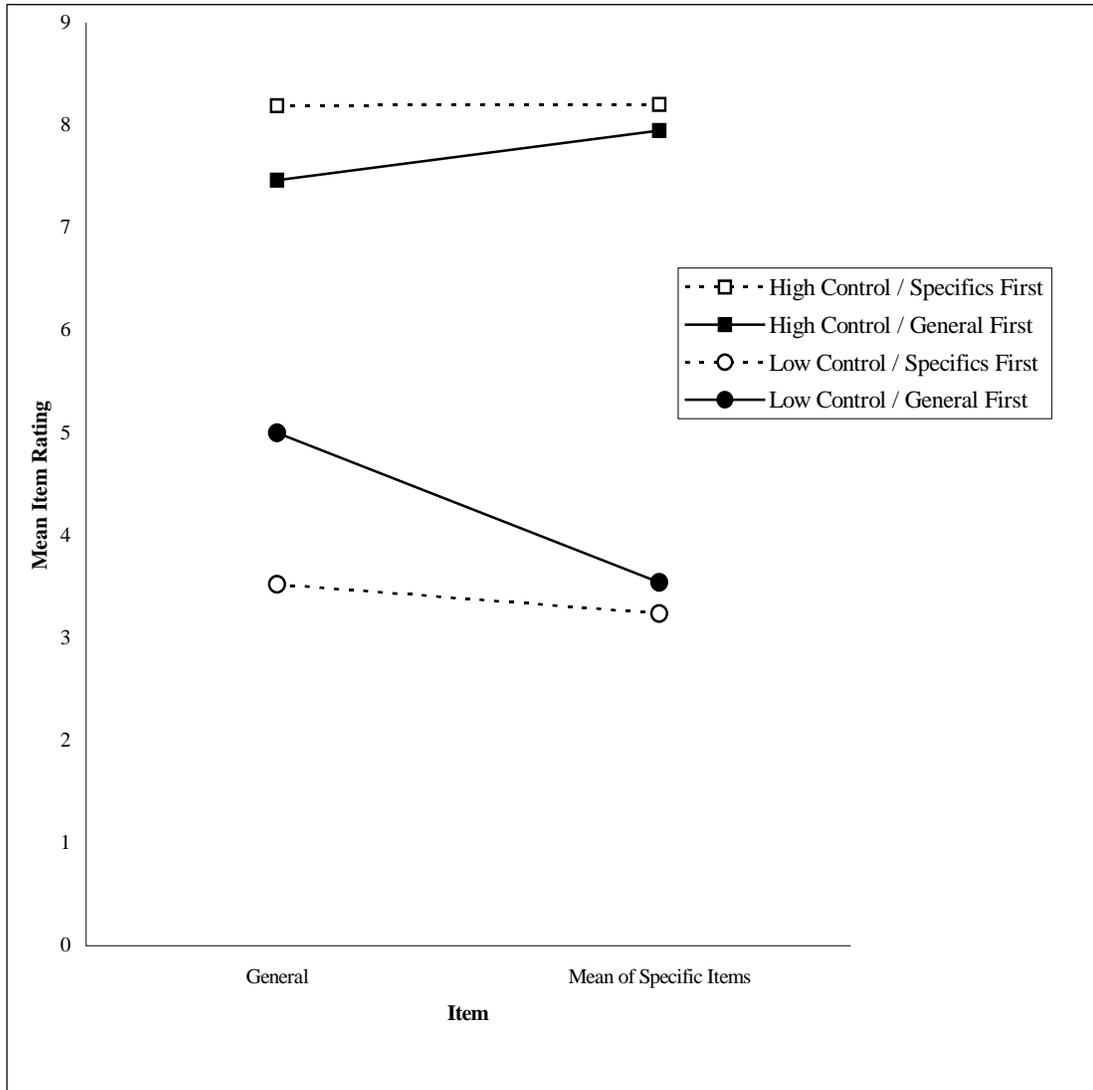


Figure 4-1. Line graph illustrating mean response to general item and mean of specific items as a function of both content and group.

consistent for the mean of the specific items. However, when they perceived high control over the topic, the mean rating was even higher than before ( $M = 8.08$ ,  $SE = 0.09$ ); when they perceived low control over the topic, the mean rating was even lower than before ( $M = 3.40$ ,  $SE = 0.30$ ). These results are illustrated in Figure 4-2.

The fourth hypothesis stated that there would not be a significant interaction of

item type and item position. Items were combined across content areas; hence, this hypothesis ignored the factor of item content. In reference to Figure 4-1, this hypothesis

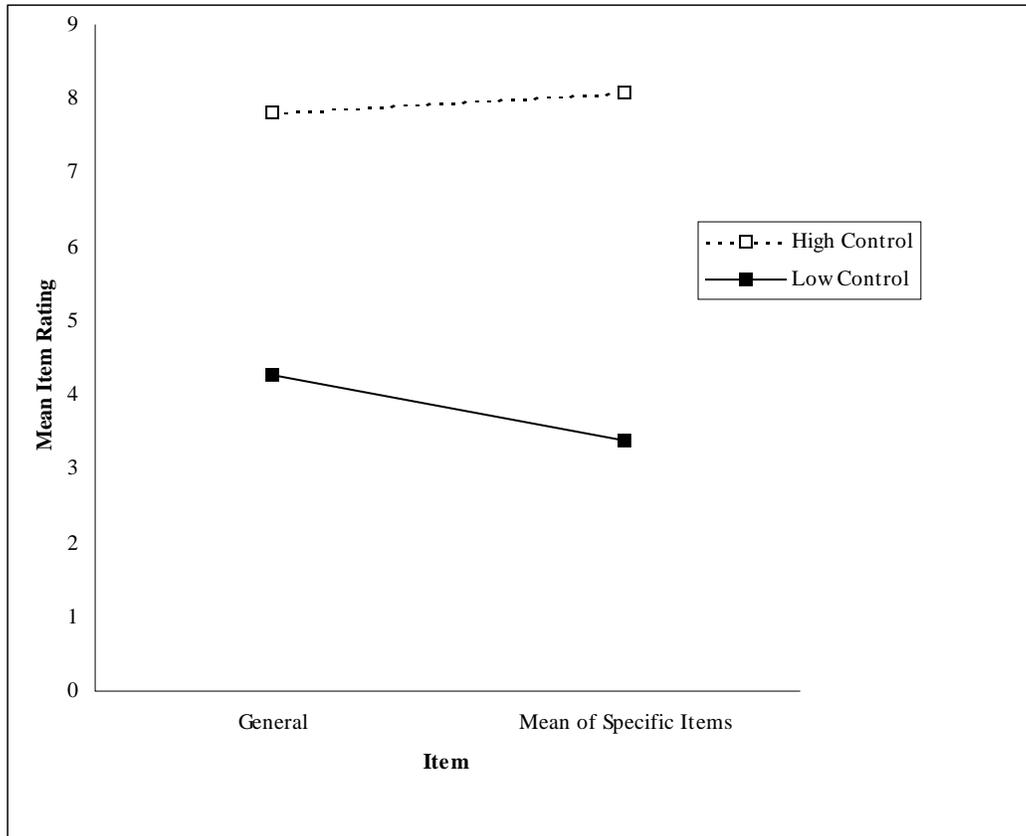


Figure 4-2. Line graph illustrating mean response to general item and mean of specific items as a function of content.

tested the difference between the mean of the two dashed lines (specific items first) and the mean of the two solid lines (general item first). Based on the original analysis displayed in Table 4-3 as Source IG, the two-way interaction was not significant,  $F(1, 64) = 1.687, p = .20$ . Specifically, teachers rated the general item and the mean of the specific items similarly. For the group completing a survey in which the specific items preceded the general item, the rating for the general item ( $M = 5.86, SE = 0.25$ ) was similar to the mean rating for the specific items ( $M = 5.72, SE = 0.38$ ). For the group completing a survey in which the general item preceded the specific items, the rating for the general

item ( $M = 6.23$ ,  $SE = 0.27$ ) was similar to the mean rating for the specific items ( $M = 5.75$ ,  $SE = 0.24$ ). In fact, all four mean ratings were similar, which supports the lack of significant results for this hypothesis.

In order to further explain the results of this study, several additional tests were conducted. Since the three-way interaction was significant, and there was a significant difference between the mean ratings of the content areas, it was of interest to examine the two-way interactions for the two content areas separately. In other words, there was interest in examining the interaction involving only the upper two lines in Figure 4-1, and there was interest in examining the interaction involving only the lower two lines in Figure 4-1. A 2 X 2 mixed analysis of variance using item position and item type did not result in a significant interaction of items as a function of item content for high control items,  $F(1, 66) = 2.381$ ,  $p = .128$ . However, there was a significant interaction of items as a function of item content for low control items,  $F(1, 68) = 6.02$ ,  $p = .017$ . Hence, much of the three-way interaction can be explained by (a) lower responses to items when low control is perceived over the content area than when high control is perceived over the content area, and (b) higher responses to the general item when it precedes the set of specific items than when it follows the set of specific items (when low control is perceived over the content area). In fact, an independent-samples  $t$ -test confirmed the latter explanation,  $t(68) = 2.071$ ,  $p = .042$ , indicating that the general item was rated higher when the general item preceded the specific items ( $M = 4.73$ ,  $SE = 0.44$ ) than when the specific items preceded the general item ( $M = 3.49$ ,  $SE = 0.41$ ).<sup>1</sup>

---

<sup>1</sup> Discrepancies between these means and those produced by the three-way ANOVA are attributed to differences in sample size due to missing data. However, the relative difference between means is nearly identical, and the interpretation is the same.

It was important to consider if any differences were due to a teacher's particular school. A one-way between subjects ANOVA conducted to determine differences in the perception of control over the content area between schools revealed no significant main effects for both perceived high control,  $F(2, 67) = 1.186, p = .182$  and perceived low control,  $F(2, 68) = .029, p = .866$ .

## CHAPTER 5 DISCUSSION

### **Interpretation of Results**

In this research, item order effects were shown to depend on the respondents' perception of control over the content of the survey. Perceived control was based on the locality of the issue to the respondent. Since classroom climate is more local to teacher than national accountability policy, teachers should perceive more control over classroom climate than national accountability policy. Possibly, this is because actions to alter classroom climate are directly verifiable by the teacher while actions to alter federal policy require the transfer of opinions and beliefs through the channel from teacher to state legislature to federal legislature.<sup>1</sup>

Past research has shown assimilative or contrast item order effects depending on the manipulation of item placement (i.e., general first or specific first). Ignoring perceived control, the placement of a general item after a set of specific items may result in an assimilative item order effect. Alternatively, the placement of a general item before a set of specific items may result in a contrast effect.

It appears that the judgment of an assimilative or contrast effect depends on the interpretation of results based on a particular methodology. It is difficult to validate either effect since, in practice, survey respondents complete only one form. As discussed previously, this is especially problematic when choosing to base interpretations on

---

<sup>1</sup> The results of the first hypothesis disconfirmed the perception of equal control over both content areas.

comparisons of correlations. The present study avoided this problem by using an analysis of variance to compare the response to the general item and the mean of the responses to the specific items. Although one can argue for the presence of both assimilative and contrast item order effects in this study, a contrast effect is most evident where the mean of responses to specific items are markedly lower than the response to the general item when the general item appears before the specific items and when the respondent perceives low control over the content area. These results support Tourangeau and Rasinski's (1988) explanation in terms of difficulty in attitude retrieval; however, it sheds light on the importance of disentangling perceived personal control over the content from personal relevance of the content.

One could argue that the present study did not truly measure item order effects since any effects should appear as a trend between two adjacent items with an effect that dissipates as the respondent proceeds through the survey. However, if this is true, then a comparison of the response to the mean of the specific items provides a more robust finding. Furthermore, it suggests implications for the design of the entire survey, not just several specific items.

The primary contribution of this study to the body of survey methodology literature was the examination of perceived control over the survey content. In this study, no item order effect was observed when collapsing items from both content areas (i.e., ignoring the perceived control factor). Apparently, perceptions of control over the content were a major contributor to the item order effects.

When ignoring item placement, responses for the high control items increased between the general item and the mean of the specific items; however, responses for the

low control items decreased. The reason for this, in the absence of varying item placement, is unclear. Perhaps, this merely indicates that teachers displayed high overall responses for classroom climate items and low overall responses for national accountability items. Yet, they rated their general impression of their classroom climate slightly lower, and they rated their general impression of national accountability slightly higher. Future research could investigate a potential tendency to differ in opinions of high control and low control topics depending on whether the item is general or specific and regardless of item order.

The most interesting finding was that teachers provided lower ratings for the low control items than for the high control items. For the low control items, teachers rated the general item highest when it preceded the set of specific items. The interpretation is consistent with the previous discussions of personal relevance and perceived control over survey content. Teachers perceived high control over their classroom climate. Hence, they held firmly entrenched beliefs and attitudes about this topic. Retrieving attitudes and opinions about this topic from memory did not require excessive cognitive accessibility. Therefore, there was no need to resort to less conscious or unconscious response heuristics leading to a susceptibility to any item order effects.

On the other hand, the teachers did not perceive high control over the topic regarding national education accountability policy. Lacking control over the topic, the teachers were less able to retrieve firm beliefs and attitudes from memory. This became exacerbated when the initial item was asking, "In general...". Lacking a clear cognitive representation for a general response to this topic, the teachers responded in a more haphazard way. The following items, being more specific, yielded responses more likely

to vary from the response to the general item. This interpretation is further substantiated by the fact that item responses were similar when the specific items were presented before the general item for this topic. A contrast effect was evident between the general item and the specific items when the general item was presented first as the teacher attempted to situate and retrieve attitudes about a low control topic.

There has been much debate as to whether or not humans respond as fully conscious decision-makers or as automatons guided by less conscious heuristics (Leahey and Harris, 2001). It is highly possible that humans select both conscious strategies and unconscious heuristics depending on a particular task. In this case, the task was affected by the placement of a general, and possibly vague, survey item. The conclusion is that the teachers in this research behaved in a fully conscious manner when they felt they had control over the content. On the other hand, the teachers in this research behaved less consciously when they felt they had low control over the content. In this latter situation, the teachers became susceptible to an item order effect when a general item was presented before a set of specific items.

### **Methodological Advantages and Limitations to Findings**

#### **Item Grouping**

In item order studies and experiments where there are multiple groupings of items, a dependency between the preestablished groups of items may surface. For example, Metzner and Mann (1953) delimited four groups of items (i.e., job, supervision, wages and promotions, and work group.) Administering a work-related survey with four work-related groups should blur the distinction between the four groups of items. If this is the

case, then the results are explained not necessarily as a lack of item order effects, but as a lack of effects due to overlapping constructs.<sup>2</sup>

The present research clearly delimited two topics. The distinction may be justified in terms of the differences in perceived control over the two topics as well as vast differences in responses to items included in the two topics. However, the construction of the two forms precipitated a clear limitation in the analysis of results. Item-order effects should be most prevalent in the comparison of adjacent general and specific items. In this study, although the general item was content equivalent between forms, the adjacent specific item differed between forms. Hence, there was no valid method for comparing item order effects based on adjacent general and specific items. This limitation necessitated a comparison of the general item and the mean of the specific items. Future research should ensure that the general item can be compared to the same individual specific items across forms.

The Gestalt Law of Proximity states that elements placed close together tend to be grouped together (Koffka, 1935); hence, Dillman (2000) recommends spacing. The survey in this study used one thin and one thick line along with an extra single space above and below the lines to demarcate the two subject matters. Although, this intuitively supports proximity between two separate subject matters, no research was found to suggest that this use of lines diminishes item order effects between two groupings. Future research could counterbalance the surveys for both forms differing by item placement such that half of the forms introduce the high perceived control topic before

---

<sup>2</sup> In this case, a Type II error (failing to find results that truly exist) occurs not because of sample size or other statistical considerations but because of an inadequate operational definition of the construct and its facets.

the low perceived control topic, and half of the forms introduce the low perceived control topic before the high perceived control topic. However, the differences in perceived control and item responses to the two topics suggest that this improvement is not necessary especially when considering the increase in sample size that would be required to detect item order effects.

It is incumbent on the researcher to clearly delimit groupings of items that are defined as discrete, provide a theoretical basis for their discreteness, and grant the reader with evidence of their lack of dependency or interdependency<sup>3</sup>. Although the comparison of means in this study is an improvement to past research comparing correlations, any comparison of one item to other items in a repeated measures design assumes that the two items measure the same construct and do so using the same metric. With a large enough sample size, item-specific methods of analysis such as item response theory (IRT) and differential item functioning (DIF) would present potential improvements to the research of item order effects.

### **Mode of Presentation**

Many early studies of item order effects used an oral administration (i.e., interview) mode (e.g., Bradburn & Mason, 1964). This design shares the features of both a qualitative and a quantitative design; however, the studies report only quantitative results. The influence of the interviewer may introduce subjective bias that affects the validity of the quantitative results. Furthermore, some of these oral surveys were face-to-face interviews while others were telephone interviews (e.g., McFarland, 1981).

---

<sup>3</sup> Further, Dillman (2000) recommends item grouping to ease the respondents' cognitive burden.

Indeed, Tourangeau, Rasinski, and Bradburn (1991) argued that unsuccessful attempts at replicating item order effects may have been due to differences when using personal interviews and when using self-administered questionnaires. Respondents may answer items in a non-sequential order when the survey is self-administered. Schwarz and Hippler (1995) found order effects in a telephone survey condition but not in a self-administered survey condition.

A second argument is that an orally administered survey is more constrained by conversational norms than a self-administered survey. Dillman (2000) argued that interviews might lead to more culturally acceptable responses, different responses due to the aural modality, and different responses due to interviewer control of administration. However, he explained that the evidence of differential effects between telephone and self-administered surveys is not yet conclusive. Lorenz and Ryan (1996) found that responses were more positive when the survey was conducted as a telephone interview than as a self-administered mail questionnaire. However, Willits and Ke's (1995) results suggest "although subjects can review previous and subsequent questions and take whatever time they wish in a self-administered survey, they may not do so" (p.400). Regardless, the present research avoided the potential for bias by relying solely on a computer-produced mail survey.

### **Sampling Error**

Sampling error affects inferences made to a population due to sample characteristics (Dillman, 2000). The survey in this study targeted teachers and was limited to teachers in one school district in Florida<sup>4</sup>. It is possible that teachers in other

---

<sup>4</sup> No significant effects for schools in this district were found for the topic of perceived personal control.

districts and states have different perception of control and are susceptible to different item order effects than those found here.

### **Coverage Error**

Coverage error occurs when the sampling frame does not include all possible members (Dillman, 2000). A sampling frame of teachers whose names were publicly listed on school websites limited the coverage for this study. If teacher names are posted on the Internet as requests are received from the website administrator, then it may be the case that teachers who respond to such requests concurrently perceive control over their classroom climate.

### **Measurement Error**

Dillman (2000) discussed the potential for measurement error in survey analysis. Item order effects can be considered a systematic source of measurement error. Respondents perceiving low control over survey content are assumed to provide a non-true response that is systematic across respondents when the general item precedes the specific items. Hence, the investigation of measurement error in survey responses was central this study. However, future developers of surveys should consider the potential for measurement error and seek ways for correction (i.e., disattenuation of correlation coefficients.) especially since, in the case of item ordering, such error is considered systematic (Shields, 2003).

### **Reliability**

An important component of measurement error is the reliability of the survey. To this end, reliability indices were reported; however, these results can be generalized only to this sample. When disaggregating by topics, reliability was not acceptable for the high control portion of the survey in which the general item followed the specific items.

Although, this may be due to the small number of items per topic, future replications should seek to increase reliability, especially for this section of this form.

### **Nonresponse Error**

Nonresponse error affected the results of this study to the extent that these nonrespondents potentially would have had different response characteristics than the respondents (Fowler, 1993). Teachers who did not respond may not have perceived high control over either area; frustration may have been the cause of their nonresponse. However, the usual difficulty in nonresponse affecting the interpretation of results is not as much of a threat to the present research as the topic is primarily targeted at the response patterns of surveys respondents. Hence, nonresponse is considered representative of nonresponse in all survey administrations.

### **Context**

Mason, Carlson, and Tourangeau (1994) argue, “In earlier work, the part-whole questions seem to have come later in the interview” (p.577). If the response to one item affects the response to subsequent items, then prior experiences before responding to the item (e.g., other prior items, the survey introduction, experiences before beginning the survey) may affect the response to this first item and, possibly, future items (Tourangeau & Rasinski, 1988).

An example of this effect is evident in the Bradburn and Mason (1964) study. They state, “Because we wished to establish the same orientation toward the interview for all respondents and give the interviewer time to establish rapport with the respondent, we began each interview with a series of questions on social participation” (p.59). If the items intended to measure order effects indeed confirm order effects on subsequent items, it is incumbent on the researcher to validate such claims by providing a description of

how the original item relates to previous items or how the researcher attempted to create a blank cognitive slate (e.g., direct introduction of a new topic, counterbalancing of item groups).

### **Final Comments**

The hypothesis that the susceptibility to item order effects is moderated by perceived personal control over survey content was not disconfirmed. However, future research should attempt to improve upon the aforementioned limitations. Further, there is some difficulty in how these interpretations are grounded in theory. For example, although control is both promoted (e.g., sales motivation books) and discouraged (e.g., twelve-step recovery groups), no research could be found to specifically address this construct.<sup>5</sup> Furthermore, the distinction of the constructs used to explain item order effects is not clear. To what extent are salience, personal relevance, and perceived control related? To what extent are they distinct from the desire to remain consistent, the desire to avoid redundancy, and social desirability? If the distinction is based on conscious and unconscious processes, at what point does a respondent shift from one strategy to another? Answering these types of questions requires continued advancement of psychological model development and survey analysis methods.

---

<sup>5</sup> The most similar term, “spheres of influence”, is particular to political science. The term “locus of control” in psychology regards internal and external control, not disparate degrees of the localization of external control.

APPENDIX A  
INVITATION TO PARTICIPATE

UNIVERSITY OF FLORIDA  
College of Education  
Department of Educational Psychology

Norman Hall  
Gainesville, FL 32601  
1-352-392-0723

April 2, 2003

Dear Teacher:

Enclosed is a short one-page survey that is part of research being conducted by Jeffrey Miller, a graduate student in the University of Florida's Department of Educational Psychology. It concerns topics that are relevant to you as a teacher. Your responses will increase scientific understanding of how you respond to these topics.

All answers will be kept confidential. Results will be released only as summaries. Your identity will be kept confidential to the extent provided by law.

Participation is completely voluntary. There are no anticipated risks, discomforts, direct benefits, or compensation for participation. If you choose not participate, please either discard the survey or return it blank using the included addressed and stamped envelope. You may also submit survey responses by email ([millerjm@ufl.edu](mailto:millerjm@ufl.edu)) if so desired. You have the right to withdraw consent at any time without consequence. You are not required to answer any question that you do not wish to answer.

The group results may be published. You are entitled to a copy of the report if so desired.

It is estimated that the completing this survey will take 1-2 minutes. After completing the survey, please return it in the enclosed addressed by truck mail as soon as possible.

The faculty supervisor for this research is Dr. David Miller. If you have any questions or comments, we would be happy to talk to you. The number is 352-392-0723.

Alternatively, you can write to the address on this letterhead. You may also contact the University of Florida Institutional Review Board with any questions or concerns at: UFIRB Office, Box 112250, University of Florida, Gainesville, FL 32611-2250; phone 375-392-0433.

Thank you so much for helping with this important study!

Sincerely,

**APPENDIX B**  
**FORM 1 – GENERAL ITEM AFTER SET OF SPECIFIC ITEMS**

Please circle your response (from 0 to 9) with 0 meaning that you completely disagree with the statement and 9 meaning that you completely agree with the statement. Please circle responses using a black or blue ink pen. Please answer the items in the order that they are presented.

I have received the cover sheet, read the cover sheet, and voluntarily agree to participate.

Participant: \_\_\_\_\_ Date: \_\_\_\_\_

Principal Investigator: (B) \_\_\_\_\_ Date: \_\_\_\_\_

	Disagree									Agree										
1. I encourage my students to learn and seek new ideas.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
2. Cooperation among students is promoted in my classroom.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
3. I implement effective discipline procedures in my classes.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
4. When I identify a problem, I immediately try to identify several solutions.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
5. My students feel comfortable talking to me when they have an academic problem.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
6. In general, the environment in my classes is excellent.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
7. I have personal control over the environment in my classes.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
8. The Department of Education is succeeding in implementing new reforms.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
9. The No Child Left Behind Act is succeeding.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
10. Legislators are working hard to ensure accountability for all involved in education.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
11. National standardized test mandates are making sure that educators and administrators do a good job at work.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
12. The government should be applauded for its efforts to increase standards for school accountability.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
13. In general, national standards for school accountability are excellent.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
14. I have personal control over national education accountability issues.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9

Subject(s) I teach: \_\_\_\_\_. This is my \_\_\_\_\_ year teaching.

APPENDIX C  
FORM 2 – GENERAL ITEM BEFORE SET OF SPECIFIC ITEMS

Please circle your response (from 0 to 9) with 0 meaning that you completely disagree with the statement and 9 meaning that you completely agree with the statement. Please circle responses using a black or blue ink pen. Please answer the items in the order that they are presented.

I have received the cover sheet, read the cover sheet, and voluntarily agree to participate.

Participant: \_\_\_\_\_ Date: \_\_\_\_\_

Principal Investigator: (B) \_\_\_\_\_ Date: \_\_\_\_\_

	Disagree									Agree										
1. In general, the environment in my classes is excellent.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
2. Cooperation among students is promoted in my classroom.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
3. I implement effective discipline procedures in my classes.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
4. When I identify a problem, I immediately try to identify several solutions.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
5. My students feel comfortable talking to me when they have an academic problem.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
6. I encourage my students to learn and seek new ideas.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
7. I have personal control over the environment in my classes.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
8. In general, national standards for school accountability are excellent.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
9. The No Child Left Behind Act is succeeding.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
10. Legislators are working hard to ensure accountability for all involved in education.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
11. National standardized test mandates are making sure that educators and administrators do a good job at work.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
12. The government should be applauded for its efforts to increase standards for school accountability.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
13. The Department of Education is succeeding in implementing new reforms	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
14. I have personal control over national education accountability issues.	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9

Subject(s) I teach: \_\_\_\_\_ . This is my \_\_\_\_\_ teaching.

## REFERENCE LIST

- Alspach, S. E. & Bishop, G. F. (1991). Question-order effects of presidential approval ratings on gubernatorial approval ratings: a research note. *Social Forces*, 69, 1241-1248.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Benton, J. E. & Daly, J. L. (1991). A question order effect in a local government survey. *Public Opinion Quarterly*, 55, 640-642.
- Bradburn, N. M. & Mason, W. M. (1964). The effect of question order on responses. *Journal of Marketing Research*, 1, 57-61.
- Bush, G. W. (2001). *No child left behind: A blueprint for education reform*. Washington, DC: U. S. Department of Education. Available December 28, 2003: <http://www.whitehouse.gov/news/reports/no-child-left-behind.html>
- Colasanto, D., Singer, E., & Rogers, T. F. (1992). Context effects on responses to questions about AIDS. *Public Opinion Quarterly*, 56, 515-518.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method*. (2nd ed.). New York: John Wiley & Sons.
- Feldman, J. M. & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, 73, 421-435.
- Fowler, F. J. Jr. (1993). *Survey research methods*. (2nd ed.). Newbury Park, CA: Sage.
- Grice, H. (1975). Logic and conversation. In P. Cole, J. Morgan (Eds.), *Speech acts*. New York: Academic Press.
- Koffka, K. (1935). *The principles of Gestalt psychology*. New York: Harcourt, Brace.
- Leahey, T. H., & Harris, R. J. (2001). *Learning and cognition* (5<sup>th</sup> ed.). Upper Saddle River, NJ: Prentice Hall.
- Lorenz, F. O. & Ryan, V. D. (1996). Experiments in general/specific questions: Comparing results of mail and telephone surveys. *Proceedings of the Survey Methodology Section of the American Statistical Association, USA*, 2, 611-613.

- Mason, R., Carlson, J. E., & Tourangeau, R. (1994). Contrast effects and subtraction in part-whole questions. *Public Opinion Quarterly*, *58*, 569-578.
- McClendon, M. J. & O'Brien, D. J. (1988). Question-order effects on the determinants of subjective well-being. *Public Opinion Quarterly*, *52*, 351-364.
- McFarland, S. G. (1981). Effects of question order on survey responses. *Public Opinion Quarterly*, *45*, 208-215.
- Metzner, H. & Mann, F. (1953). Effects of grouping related questions in questionnaires. *Public Opinion Quarterly*, *17*, 136-141.
- Moore, D. (2002) Measuring new types of question-order effects. *Public Opinion Quarterly*, *66*, 80-91.
- Posner, M. (1978). *Chronometric explorations of the human mind*. Hillsdale, NJ: Erlbaum.
- School Advisory Council Report Results (2001-2002). Retrieved January 30, 2004 from <http://www.fldoe.org>
- Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and content*. New York: Academic Press.
- Schuman, H., Presser, S., & Ludwig, J. (1981). Context effects on survey responses to questions about abortion. *Public Opinion Quarterly*, *45*, 216-223.
- Schwarz, N. & Hippler, H. (1995). Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opinion Quarterly*, *59*, 93-97.
- Schwarz, N., Strack, F., & Mai, H. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, *55*, 3-23.
- Shields, J. A. (2003). Examining context effects using structural equation modeling. *Dissertation Abstracts International*, *64*(03), 1546B. (UMI No. AAT 3083123)
- Sigelman, L. (1981). Question-order effects on presidential popularity. *Public Opinion Quarterly*, *45*, 199-207.
- Tourangeau, R. & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, *103*, 299-314.
- Tourangeau, R., Rasinski, K. A., & Bradburn, N. (1991). Measuring happiness in surveys: A test of the subtraction hypothesis. *Public Opinion Quarterly*, *55*, 255-266.

- Tourangeau, R., Rasinski, K. A., Bradburn, N., & D'Andrade, R. (1989). Carryover effects in attitude surveys. *Public Opinion Quarterly*, 53, 495-524.
- Tourangeau, R., Rips, J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge, MA: Cambridge University Press.
- Willits, F. K. & Ke, B. (1995). Part-whole question order effects: Views of rurality. *Public Opinion Quarterly*, 59, 392-403.

## BIOGRAPHICAL SKETCH

Jeffrey Monroe Miller began his postsecondary education at the University of Wisconsin – Eau Claire where he studied psychology, music, and music therapy. After receiving a Bachelor of Arts in psychology in 2001, he briefly pursued graduate work at the University of North Florida before beginning the study of research and evaluation methodology at the University of Florida in 2002. His current research interests include construct validation and the applications of measurement theory to large-scale high-stakes assessment.