Assessing Three Sources of Misresponse to Reversed Likert Items

Boston University
School of Management
Working Paper #2006-12

Scott Swain
Danny Weathers
Ronald W. Niedrich

Note: Forthcoming in Journal of Marketing Research. Scott D. Swain (swain@bu.edu) is an Assistant Professor of Marketing, School of Management, Boston University, Boston, MA 02215. Danny Weathers (weathers@lsu.edu) is an Assistant Professor of Marketing, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803. Ronald W. Niedrich (niedrich@lsu.edu) is an Associate Professor of Marketing, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803. The authors gratefully acknowledge the helpful comments and suggestions of the three anonymous JMR reviewers, Bill Bearden, David Hardesty, Elizabeth Miller, Richard Netemeyer, Jennifer Nevins, Subhash Sharma, and Russell Winer. They also thank the researchers listed in Table 2 for archival data and Suchi Chandran for assistance with Experiment 1. Please address all correspondence to Scott Swain.
ASSESSING THREE SOURCES OF MISRESPONSE TO REVERSED LIKERT ITEMS

ABSTRACT

Data collected using multi-item Likert scales that contain reversed items often exhibit problems, such as unexpected factor structures and diminished scale reliabilities. These problems arise when respondents select responses on the same side of the scale neutral point for both reversed and non-reversed items, a phenomenon we call misresponse. Across four experiments and an exploratory study using published data, we found that misresponse to reversed Likert items averaged nearly 20%, twice the level identified as problematic in previous simulation studies. Counter to prevailing thought, the patterns of misresponse and response latency across manipulated items could not be attributed to respondent inattention or acquiescence. Instead, the pattern supports an item verification difficulty explanation, which holds that task complexity, and thus misresponse and response latency, increases with the number of cognitive operations required for a respondent to compare a scale item with his or her belief. The observed results are well explained by the Constituent Comparison Model.
It is commonly recommended that multi-item Likert scales include both reversed and non-reversed items (e.g., Churchill 1979; Paulhus 1991). One aim of using such a mix of items is to alert inattentive respondents that item content varies (Drolet and Morrison 2001; Nunnally 1978; Schmitt and Stults 1985). Another is to reduce bias that may occur in scale scores due to acquiescent respondents (Ray 1983; Watson 1992). Further, when scales contain a mix of items, researchers can compute an indirect measure of acquiescence, which may then be used as a covariate in analyses involving the scales or as a tool for partialing acquiescence bias from scale scores (e.g., Baumgartner and Steenkamp 2001; Greenleaf 1992).

Despite the potential benefits of reversing scale items, some researchers have expressed concerns about the practice (e.g., Marsh 1996; Netemeyer, Bearden, and Sharma 2003). These concerns stem from evidence linking item reversal to problems such as unexpected factor structures (e.g., Babakus and Boller 1992), diminished scale reliabilities (e.g., Herche and Engelland 1996), and confounds in cross-cultural research (Steenkamp and Burgess 2002; Wong, Rindfleisch, and Burroughs 2003). Such problems may result from a phenomenon we call misresponse. Conceptually, misresponse occurs when a person selects a response option that is opposite to his or her beliefs. To illustrate, consider two items from the need for cognition scale (Cacioppo, Petty, and Kao 1984): “The notion of thinking abstractly is appealing to me” and “Thinking is not my idea of fun.” A respondent agreeing or disagreeing with both items would seem to have misresponded to one of them. Disturbingly, even low rates of misresponse can cause problems in basic analyses. For example, Schmitt and Stults (1985) demonstrate through simulation that when misresponse is as low as 10%, factor analysis produces a two-factor

---

1 Since acquiescent respondents tend to agree with all items, their responses inflate scale means when items are worded in one direction. Using a mix of items addresses this problem because responses to non-reversed items are biased in one direction while responses to reversed items are biased in the opposite direction. Although this method does not prevent acquiescent responding or remove bias from individual items, it can reduce bias in scale means.
solution even if a single factor is known to exist in the population.

While the potential problems associated with item reversal have often been noted, we are unaware of any prior attempts to quantify the frequency of misresponse or to test its underlying causes. Considering the importance of measurement and the pervasiveness of Likert scaling, a better understanding of misresponse is needed. In our investigation, we first discuss the connection between item reversal and linguistic negation. Second, we describe three potential sources of misresponse to reversed items: respondent inattention, respondent acquiescence, and item verification difficulty. Third, we illustrate the prevalence of misresponse by introducing and applying a measure of the phenomenon to a variety of previously published scale data. We find that the average rate of misresponse to reversed items is nearly twice the level identified as problematic in simulations by Schmitt and Stults (1985). Fourth, we conduct four experiments to assess the roles of inattention, acquiescence, and item verification difficulty in misresponse. In each experiment, we find that misresponse varies across manipulated items in accordance with the item verification difficulty explanation. We also find that response latencies, which reflect the underlying cognitive response process, support this explanation. Finally, we discuss the contributions and implications of this research.

**ITEM REVERSAL AND NEGATION**

In Likert scaling terminology, non-reversed items are those for which higher numbers indicate a higher level of the construct, whereas reversed items are those for which smaller numbers indicate a higher level (Baumgartner and Steenkamp 2001; Bearden and Netemeyer 1999). A commonly used method for reversing Likert items is to reverse their linguistic
Linguistic polarity refers to whether a statement is an affirmation (positive polarity) or a negation (negative polarity). Horn (1989) describes affirmations as assertions about something (e.g., “Tigers are ferocious”) and negations as denials of assertions (e.g., “Tigers are not ferocious”). In practice, negation generally manifests as particle words (e.g., “not,” “no”) or affixal morphemes (e.g., “un-,” “non-,” “-less”). For example, the affirmative statement “Orange is an attractive color” can be negated as “Orange is not an attractive color” (particle negation) or “Orange is an unattractive color” (affixal negation).

Because scales are generally designed to measure the existence, rather than the absence, of traits, beliefs, and attitudes, most non-reversed items will be stated as affirmations, and most reversed items will be stated as negations. Confirming this, our analysis of nearly 2,000 Likert items in Bearden and Netemeyer’s (1999) Handbook of Marketing Scales revealed that 89% of non-reversed items were affirmations while 81% of reversed items were negations. Accordingly, references to reversed items in this research pertain to items reversed using negation.

THREE POTENTIAL SOURCES OF MISRESPONSE

As a general framework for discussing misresponse, we draw on the four-step model of response selection developed in the survey research literature (e.g., Sudman, Bradbury, and Schwarz 1996; Tourangeau, Rips, and Rasinski 2000). This model describes the basic processes involved in responding to a scale item: comprehension, retrieval, judgment, and response. Comprehension pertains to perceptual analysis of an item to extract meaning from its text (Kintsch 1998), retrieval refers to the activation of a previously formed belief in long-term memory or the transfer to working memory of information used to construct a new belief,
judgment entails comparing the representation of an item to the representation of retrieved information, and response involves formatting a judgment to match the options available on a response scale. Misresponse can occur as a result of problems at one or more of these steps.

Because inattention, acquiescence, and item verification difficulty all predict that misresponse will be greater for reversed items than for non-reversed items, distinguishing among these sources of misresponse requires additional conceptualization. Thus, we note two important item properties: polarity and truth value. **Polarity**, as we discussed previously, refers to whether an item represents an affirmation or negation, whereas **truth value** indicates whether an item is true or false for the respondent. For Likert items, a person should select “agree” options for true items and “disagree” options for false items. As we show, the expected patterns of misresponse across polarity and truth value differ depending on the source of misresponse. In addition, because response latencies can provide insight into the underlying cognitive response process, we derive predictions about the patterns of response latencies across polarity and truth value.

**Respondent Inattention**

Prior research suggests that misresponse may reflect a response style that has been variously labeled as carelessness (Nunnally 1978; Schmitt and Stults 1985), mindless consistency (Drolet and Morrison 2001), or noncontingent responding (Baumgartner and Steenkamp 2001). Although the literature has focused more on the psychometric consequences of this response style than on formulating a behavioral account of it, researchers have suggested that attention may play an important role in determining misresponse.

Attention is a limited cognitive resource necessary for completing the mental processes involved in selecting focal stimuli for conscious thought while ignoring or filtering out other aspects of the environment (e.g., Hunt and Ellis 1999). We argue that misresponse can occur
when respondents allocate insufficient cognitive resources to item processing and instead rely on expectations regarding item content. The more general notion of inattentional blindness refers to the phenomenon where proximal stimuli go undetected in unattended regions of the perceptual field (Moore, Grosjean, and Lleras 2003). One explanation for inattentional blindness is the contingent-capture hypothesis (Folk, Remington, and Johnston 1992), which suggests that attention is contingent on observers’ expectations rather than the distinctiveness of stimuli. In the context of Likert scaling, we argue that respondents’ experiences with statements in everyday language lead to the expectation that items are stated affirmatively and contain similar content. Support for this argument comes from analyses of spoken and written language corpora, which find that affirmations comprise approximately 80–90% of speech clauses and 90–95% of written clauses (Givón 1993; Tottie 1991). We also analyzed the Likert items in Bearden and Netemeyer’s (1999) *Handbook of Marketing Scales* and found that 82% were affirmations.

Thus, while all respondents are likely to have the expectation that items are stated affirmatively and are similar in content, inattentive respondents rely on these expectations rather than engaging in additional item processing. Further, since inattentive respondents do not fully engage in comprehension, retrieval, or judgment processes, they are likely to resort to simple heuristics when selecting overt responses. Although a variety of response heuristics are possible, we focus on “straight-line” responding (i.e., selecting the same response for all items) because it produces systematic error when scales contain reversed items and is consistent with an expectation of affirmative items that contain similar content (Baumgartner and Steenkamp 2001; Drolet and Morrison 2001). Straight-line responding is also concordant with the status quo heuristic, a common, low elaboration decision rule whereby respondents complete sequential choice tasks by successively selecting the same option (cf. Bettman, Johnson, and Payne 1991).
In summary, based on the contingent-capture hypothesis, we predict that inattentive respondents overlook item negation and changes in item direction because they rely on expectations for affirmation. Thus, they will respond to affirmations and negations as if both are affirmations, resulting in a main effect of item polarity on misresponse; that is, inattentive respondents misrespond to negations but not to affirmations. Furthermore, because inattentive respondents rely on expectations that items possess similar affirmative content rather than determining the truth value of each item, truth value should have no effect on misresponse. Finally, since inattentive respondents invoke the same response heuristic for all items, response latencies will not vary by item polarity or truth value (Figure 1, Panels A and B).

---Insert Figure 1 about here---

**Respondent Acquiescence**

Respondent acquiescence has been characterized in survey research as a tendency toward uncritical agreement with items (Messick 1991; Paulhus 1991; Ray 1983). That is, though acquiescent respondents are attentive, they select agreement response options regardless of item content. Various psychological, social, and cultural factors may contribute to acquiescence (McClendon 1991; Rorer 1965), and it has been associated with greater extraversion and impulsiveness (Couch and Keniston 1960), lower education (Greenleaf 1992), lower income (Krosnick 1999), lower social status (e.g., Winkler, Kanouse, and Ware 1982), and greater collectivism (Javeline 1999; Wong, Rindfleisch, and Burroughs 2003).

Knowles and Condon (1999) suggest that the cognitive process underlying acquiescence is consistent with Gilbert’s (1991) dual-stage “Spinozan” model of belief. This model holds that initial comprehension of a statement involves the automatic acceptance of its content, followed by the retrieval of relevant information. From this perspective, acquiescence can be
conceptualized as a tendency to omit the second stage of the belief model, such that relevant and potentially contraindicating information is neither retrieved nor constructed (Knowles and Condon 1999; Krosnick 1999). Since the automatic acceptance associated with item comprehension does not depend on whether the item is affirmed or negated, acquiescence-based misresponse will not exhibit a main effect of item polarity. However, because automatic acceptance of item content leads to the selection of agreement responses for all items, acquiescence results in correct responses for true items but misresponses for false items; that is, acquiescence-based misresponse will exhibit a main effect of item truth value (Figure 1, Panel C).

For response latencies, we first note that prior research indicates that reading times for affirmative and negated versions of a given statement do not differ (Carpenter and Just 1975; MacDonald and Just 1989). Further, since the automatic acceptance process that characterizes acquiescence does not depend on item polarity, there should be no comprehension-based differences in response latencies across affirmations and negations. Thus, while acquiescent respondents are attentive to item content, response latencies should not differ across affirmative and negated items. Similarly, since acquiescent respondents do not attempt to determine item truth value by means of retrieving or considering other information, we do not expect differences in response latencies across true and false items (Figure 1, Panel D).

**Item Verification Difficulty**

It has been suggested that reversed items, relative to non-reversed items, may be more confusing or difficult to process and thereby result in greater misresponse (e.g., Netemeyer, Bearden, and Sharma 2003; Nunnally 1978). Although we are unaware of any research that directly examines this perspective, the sentence processing literature provides a basis for
developing a theoretical account of item verification difficulty.

Research on sentence processing finds that reading latencies and cued recall accuracies do not differ for negated and affirmed versions of sentences (Carpenter and Just 1975; Grant, Malaviya, and Sternthal 2004; Just and Clark 1973; MacDonald and Just 1989; Mayo, Schul, and Burnstein 2004). However, in studies where respondents are required to use comprehended information to perform subsequent tasks (e.g., congruency judgments, sentence completions, answering questions), negation is associated with greater response latencies and inaccuracy than affirmation (e.g., Gough 1966; Jacoby, Nelson, and Hoyer 1982; Just and Clark 1973; Wason 1961). Collectively, these findings suggest that errors associated with negation result from difficulty at judgment rather than comprehension or retrieval.

A Model of Item Verification. To explain such judgment errors, we introduce the Constituent Comparison Model (Carpenter and Just 1975). This model specifies the effects of item polarity and truth value on the number of cognitive operations that occur during the item verification process. Because the number of cognitive operations associated with an item type quantifies its verification difficulty, the Constituent Comparison Model can be used to predict both misresponse and response latency.

The inputs to the Constituent Comparison Model are the respondent’s mental representations of both the item and a relevant belief. These representations take the form of abstract propositions, comprised of two types of constituents: a predicate–argument construction and a polarity marker (Kintsch 1998). The argument is a concept that performs a specific semantic function, whereas the predicate describes a property or relation associated with the argument. Polarity markers indicate whether the predicate is affirmed or denied.\footnote{For simple declarative statements such as scale items, the argument is typically the subject of a clause or sentence, and the predicate is the main verb plus any words governed by or modifying it.} To illustrate, “I
do not like green eggs and ham” would be mentally represented as “Negated(like, green eggs and ham).” “Green eggs and ham” is the argument, “like” is the predicate, and “Negated” indicates that the predication is denied. Consistent with prior research, the Constituent Comparison Model asserts that external information (e.g., a scale item) can be either affirmed or negated, whereas retrieved information (i.e., a belief) is represented affirmatively (Clark and Chase 1972; Fiedler et al. 1996; MacDonald and Just 1989).

The item verification process described by the Constituent Comparison Model begins with a comparison of the two predicate-argument constituents and ends with a comparison of the two polarity marker constituents, though, as we discuss subsequently, recomparisons may be necessary. A true-false response index, initially set to “true,” tracks the constituent comparison process. When constituents do not match, a mental “tag” or note is generated, and the true-false response index toggles. Such mismatches also generate related memory details and operations (e.g., storing the tag, updating the response index), which join existing details in working memory (i.e., item and belief representations, a placeholder for each representation to track which constituents were last retrieved and compared, and the current value of the response index). Acknowledging the limitations of working memory, the model proposes that mismatches cause the comparison process to “forget” which constituents were last compared. As a result, the process reinitializes and returns to the first pair of constituents (i.e., the predicate-argument constituents). To prevent infinite looping, the mental tags applied to mismatched constituents allow the reinitialized process to treat them as matches during recomparisons. These recomparisons are necessary because subsequent constituents, such as negation markers, can reverse item truth value.5

5 Although several information verification models have been proposed and tested (Clark and Chase 1972; Shoben 1978; Singer 1977), the Constituent Comparison Model is arguably the best supported and most parsimonious. In
Illustration of the Constituent Comparison Model. In Table 1, we illustrate the item verification process. In the example, a respondent holds the belief that a brand is exciting, represented as “Affirmed(exciting, brand).” By crossing truth value and polarity, we can construct four types of Likert items: true affirmation (“The brand is exciting”), false affirmation (“The brand is boring”), false negation (“The brand isn’t exciting”), and true negation (“The brand isn’t boring”). When presented with one of these items, the respondent’s judgment task is to determine the correspondence between the item and his or her belief.

---Insert Table 1 about here---

Consider the comparison process for the true negation “The brand isn’t boring,” represented as “Negated(boring, brand).” As we show in Table 1, a mismatch occurs during the first constituent comparison because the predicate-argument constituent in the respondent’s belief (boring, brand) does not match that of the item (exciting, brand). Thus, the mismatch is tagged and the process reinitializes. During the next iteration, the predicate-argument constituents match because of the tag, but the polarity constituents produce another mismatch (Affirmed versus Negated). The mismatch is tagged and the process reinitializes for a second time. During the final iteration, all constituents match on the basis of their tags. Thus, the verification process requires five total comparisons, and the two mismatches toggle the true-false response index twice, resulting in a response of “true.”

Predictions. With Table 1, we show that the Constituent Comparison Model predicts that true affirmative items should cause the least verification difficulty, followed by false affirmative, false negated, and true negated items. Because increased difficulty is likely to translate into more errors and longer processing times, we posit an interaction between item polarity and truth value addition, studies employing brain imaging (Just et al. 1996; Reichle, Carpenter, and Just 2000) and eye tracking (e.g., Knoeferle et al. 2005) have provided process evidence supporting the Constituent Comparison Model.
for both misresponse (Figure 1, Panel E) and response latencies (Figure 1, Panel F). In addition, because the model assumes each constituent comparison employs the same mental operation, each comparison represents an equal increment in verification difficulty (Carpenter and Just 1975). Therefore, we further predict that misresponse and response latencies will be linear functions of the number of constituent comparisons required to verify an item.

**RESEARCH MEASURES**

In this section, we present the measures used in this research. Although misresponse and response latencies are the primary measures used to test our predictions, we also consider explicit measures of inattention and acquiescence for supplemental analyses.

**Truth Value, Misresponse, and Response Latencies**

Items for which we want to assess misresponse are target items. When we know respondents’ beliefs prior to measurement, as in Experiment 1, we can assess misresponse objectively. However, the purpose of measurement is usually to reveal respondents’ beliefs. We consider this case in the exploratory study and Experiments 2a, 2b, and 3 by using non-target, non-reversed (affirmative) items within a scale to assess the true direction of the respondents’ beliefs (i.e., truth value). Specifically, when a respondent answers all non-target items on one side of the scale neutral point, we treat this side as the respondent’s subjective truth value. Thus, if a target item is a reversed (non-reversed) item, misresponse occurs when a response falls on the same (opposite) side as the responses to the non-target items, excluding the neutral point.

To compute misresponse for a target item, we divide the number of people identified as misresponders by the sample size, less the number of neutral responders for that item (i.e., misresponders/[total responders – neutral responders]). For example, if there are 110 total respondents, 15 of whom are classified as misresponders and 10 of whom select the neutral
category, misresponse for the target item is 15% (i.e., 15/(110 − 10)).

In Experiments 1 and 2b, we collect response latencies via computer for individual target items as standard measures of elaboration (e.g., Fazio 1990). These latencies measure the time it takes for respondents to read and respond to the scale items.

**Inattention and Acquiescence**

*Inattention.* Consistent with prior research (Drolet and Morrison 2001; Schmitt and Stults 1985), the status quo decision heuristic (Bettman, Johnson, and Payne 1991), and our theoretical development of inattention, we use straight-line responding (i.e., selecting the same response for all items) as a measure of inattention. Unlike other response patterns which may indicate inattention, straight-line responding results in both misresponse and systematic effects on scale scores.

*Acquiescence.* We measure acquiescence by determining the degree to which a respondent agrees with both non-reversed and reversed items in the same scale (Baumgartner and Steenkamp 2001; Watson 1992). For five-point Likert scales (1 = strongly disagree; 5 = strongly agree), respondents earn one point for agreeing (i.e., selecting a response of “4”) with both a reversed and a non-reversed item in the same scale, two points for agreeing with a reversed item and strongly agreeing (i.e., selecting a response of “5”) with a non-reversed item (or vice versa), and three points for strongly agreeing with both a reversed item and a non-reversed item. To make the measure comparable for seven-point Likert scales, we code responses of “5” and “6” as agreeing with the item. We average all comparisons between non-reversed items and a given reversed item in a scale to create a measure of acquiescent response style (ARS1) based on that reversed item.

**EXPLORATORY STUDY OF SECONDARY DATA**
Purpose and Procedure

We first sought evidence to confirm that misresponse can be a substantive phenomenon in Likert scaling. Accordingly, we obtained archived data from authors of articles in marketing and psychology and computed misresponse. Although these data span a wide range of respondents, research settings, and construct domains, this study serves as an exploratory and motivational effort rather than a formal meta-analysis. We provide a brief description of each data set in Table 2, and additional details appear in the corresponding publications.

---Insert Table 2 about here---

Results and Discussion

Misresponse. In Table 2, misresponse to the target reversed items appears in the last column. If the study used multiple reversed items, the percentage presented represents the average misresponse across all reversed items in that data set. For the 48 reversed items, misresponse averaged 17.50%, which is significantly greater than the “spurious factor” level of 10% (Schmitt and Stults 1985; \( t_{47} = 5.81 \), one-tailed \( p < .001 \)).

Inattention and Acquiescence. Across the 31 scale administrations, an average of 3.18% of respondents displayed inattentive behavior, as indicated by straight-line responding. Nine of the data sets contained multiple reversed items, which enabled us to compute multiple measures of ARS1. Using these measures, an average of 1.73% of respondents were classified as acquiescent (i.e., scored above 0) for all reversed items in a given data set. Both of these averages were significantly lower than the average misresponse of 17.50% (inattention: \( t_{77} = 8.37 \), one-tailed \( p < .001 \); acquiescence: \( t_{55} = 5.23 \), one-tailed \( p < .001 \)). The mean correlation for pairs of ARS1 scores in studies containing multiple reversed items was .183, which indicates a lack of consistency in observed acquiescence.
Discussion. These data, which span a variety of research settings and respondent types, suggest that misresponse can be a substantial source of systematic measurement error, even when few respondents exhibit acquiescent or inattentive behavior. However, the data cannot eliminate the possibility that these examples of misresponse are due to item content unrelated to negation. Thus, to investigate the sources of misresponse more rigorously, we conducted four experiments.

**EXPERIMENT 1**

**Purpose, Design, and Procedure**

Although the exploratory study enabled us to assess acquiescence and inattention in a variety of data, its lack of affirmative control items restricted our ability to perform a complete test of item verification difficulty. Experiment 1 addresses this limitation. A 2 (item polarity) × 2 (truth value) × 2 (product feature) within-subjects design was conducted via computer so that we could collect response latencies. Polarity and truth value served as the primary manipulations, and product feature was a replicate. We manipulated item polarity by using either explicit negation (reversed items) or affirmation (non-reversed items). To manipulate truth value, we provided respondents with either a true or a false item pertaining to the product. Using a sports utility vehicle (SUV) as the study topic, we manipulated the product feature at two levels: automatic transmission and manual transmission. Subjects for Experiment 1 were 62 undergraduate students who participated in exchange for course credit and an opportunity to win a prize. To ensure subject involvement and reduce noise in the response latencies, we offered $50 prizes for the three best performances based on an equal weighting of accuracy and speed.

Subjects first read a description of the task and completed a sample trial. In each trial, the initial computer screen presented information about the SUV’s transmission (e.g., “The SUV has an automatic transmission”), thus creating the belief. Subjects then clicked a “Continue” button,
and a new screen presented one of four items pertaining to the SUV’s transmission, along with seven buttons arranged to form a seven-point Likert response scale. The items contained the polarity and truth value manipulations. For example, in the automatic transmission condition, an item was either a true affirmation (“The SUV does have an automatic transmission”), false affirmation (“The SUV does have a manual transmission”), false negation (“The SUV doesn’t have an automatic transmission”), or true negation (“The SUV doesn’t have a manual transmission”); parallel items constituted the manual transmission condition. The presentation of the initial statement and the Likert item, along with the subject’s selection of a response, constituted a trial. Subsequent trials began after subjects had selected one of the response buttons. We recorded the time between the presentation of the response screen and the selection of a response option to represent response latency. We also randomized the order of the eight trials. As respondents were provided with factual information prior to responding, we were able to objectively code responses as correct or incorrect. No subjects selected the middle response category (i.e., 4).

Results and Discussion

Misresponse. A repeated measures loglinear model with misresponse as the dependent variable indicated a significant polarity main effect ($\chi^2 = 4.20, 1 \text{ df, } p = .041$), an insignificant truth value main effect ($\chi^2 = 0.72, 1 \text{ df, } p = .395$), and a significant interaction between polarity and truth value ($\chi^2 = 14.09, 1 \text{ df, } p < .001$). Because we found no significant effects involving product feature, we pooled the results across this factor to compute misresponse. Respondents misresponded to true affirmative items at the lowest rate (.81%), followed by false affirmations (5.65%), false negations (8.40%), and true negations (19.83%). We found significant differences

---

6 We provide the target items for all experiments in the Web Appendix.
at the .05 level between each pair of proportions except for false affirmations and false negations
\( z = 0.84, \) one-tailed \( p = .201 \). This pattern (Figure 2, Panel A) is consistent with the Constituent Comparison Model’s predictions (Figure 1, Panel E). Furthermore, a regression of overall misresponse on the predicted number of constituent comparisons for each item type yielded an \( R^2 \) value of .92, in support of the model’s prediction of a linear increase.

**Response Latencies.** A repeated-measures ANOVA with response latency as the dependent variable indicated significant polarity (Wilks’ lambda = .423, \( F_{1, 54} = 73.63, p < .001 \)) and truth value (Wilks’ lambda = .920, \( F_{1, 54} = 4.72, p = .034 \)) main effects and a significant interaction between these factors (Wilks’ lambda = .643, \( F_{1, 54} = 29.94, p < .001 \)). Because we found no significant effects involving product feature, we pooled the results over the two levels of this factor. Average latencies, in seconds, were 1.81 for true affirmations, 2.41 for false affirmations, 2.82 for false negations, and 3.05 for true negations. Paired comparison t-tests indicated significant differences between each pair of mean response latencies (largest one-tailed \( p = .078 \)). Again, this pattern of results (Figure 2, Panel B) is consistent with the Constituent Comparison Model’s predictions (Figure 1, Panel F), and a regression of the mean response latencies on the predicted number of constituent comparisons for each item type yielded an \( R^2 \) value of .96.

---Insert Figure 2 about here---

**Discussion.** A significant interaction occurred between item polarity and truth value for both misresponse and response latencies, and the pattern of results suggests that true affirmations were the least difficult to verify, followed in increasing order by false affirmations, false negations, and true negations. Furthermore, both misresponse and response latencies increased linearly with the number of constituent comparisons predicted for each item type. Thus,
Experiment 1 provides support for the Constituent Comparison Model and the item verification difficulty explanation. Although Experiment 1 confirms that misresponse is well predicted by the Constituent Comparison Model when it is due to item verification difficulty, acquiescence or inattention may contribute more to misresponse in non-incentive, non-objective contexts. Therefore, we conducted Experiments 2a and 2b to address this possibility. Because the procedures and results of these experiments were very similar, we present them together.

**EXPERIMENTS 2A AND 2B**

**Design and Procedure**

In Experiments 2a and 2b, we randomly assigned subjects to one of two versions of a survey that contained previously validated, five-item scales that measured a variety of constructs. In one version, each scale contained one negated item, whereas in the other, the scales contained no negated items. The experiments differed only in the scales used (i.e., five in Experiment 2a, three in Experiment 2b) and the mode of data collection (i.e., Experiment 2a was paper based, Experiment 2b was computer based). Two hundred sixty-one undergraduate student subjects participated in Experiments 2a (n = 143) and 2b (n = 118) in exchange for course credit.

**Results and Discussion**

**Misresponse.** The results for the two experiments were very similar, so we pooled the misresponse data over the five target items from Experiment 2a and the three target items from Experiment 2b.\(^7\) We first compared the between-subjects misresponse for the negated and affirmative target items and found misresponse to be significantly greater for negated items

---

\(^7\) To determine whether it was appropriate to aggregate the data, we conducted a pooling test by including a “study” factor in the misresponse analysis. This factor did not have a significant main effect ($\chi^2 = 0.79$, 1 df, $p = .376$), nor did it significantly interact with the polarity or truth value factors (study $\times$ polarity: $\chi^2 = 2.63$, 1 df, $p = .105$; study $\times$ truth value: $\chi^2 = 0.00$, 1 df, $p = .949$; study $\times$ polarity $\times$ truth value: $\chi^2 = 0.87$, 1 df, $p = .351$).
(19.37%) than for the same items stated affirmatively (4.34%; \( z = 7.03 \), one-tailed \( p < .001 \)). Thus, it appears that misresponse in these data is due primarily to negation, not to item content unrelated to negation.

As described previously, we determined the target item’s truth value by examining responses to the scale’s non-target items. Thus, we further classified misresponses as occurring for true affirmations, false affirmations, false negations, or true negations. A loglinear model with misresponse as the dependent variable indicated a significant main effect of polarity (\( \chi^2 = 39.29, 1 \text{ df, } p < .001 \)), a significant main effect of truth value (\( \chi^2 = 4.33, 1 \text{ df, } p = .037 \)), and a significant interaction between these factors (\( \chi^2 = 8.14, 1 \text{ df, } p = .004 \)). Misresponse was 3.74\% for true affirmations, 9.84\% for false affirmations, 26.47\% for false negations, and 39.02\% for true negations, and there were significant differences between each pair of proportions (all \( p \)-values \(< .05 \)). A regression of overall misresponse on the predicted number of constituent comparisons for each item type yielded an \( R^2 \) value of .97. This pattern of misresponse (Figure 2, Panel C) is consistent with the Constituent Comparison Model’s predictions (Figure 1, Panel E).

**Response Latencies.** In Experiment 2b, we used three scales and a computer-based approach to collect response latencies. The latencies were pooled across the three target items, and an ANOVA indicated a significant main effect of polarity (\( F_{1,321} = 35.07, p < .001 \)), an insignificant main effect of truth value (\( F_{1,321} = 0.06, p = .802 \)), and a significant interaction between the factors (\( F_{1,321} = 5.10, p = .025 \)). Mean latencies, in seconds, were 5.12 for true affirmations, 5.94 for false affirmations, 7.21 for false negations, and 7.97 for true negations (Figure 2, Panel D). The four mean latencies were significantly different from each other (largest one-tailed \( p = .079 \)). A regression of mean response latencies on the predicted number of constituent comparisons for each type of item yielded an \( R^2 \) value of .99, consistent with the
Constituent Comparison Model’s predictions (Figure 1, Panel F).

*Inattention and Acquiescence.* Considering only the surveys that contained negated items, we determined that the overall percentage of straight-line respondents was 2.98%, significantly lower than the 19.37% misresponse rate for negated items ($z = 8.37$, one-tailed $p < .001$). In addition, the ARS1 measure indicated that no subjects were consistently acquiescent (i.e., none scored greater than 0) across the five negated items in Experiment 2a, and 3.82% of the subjects were consistently acquiescent across the three negated items in Experiment 2b. For the two studies, the average correlation for all within-study pairs of ARS1 measures was .067. These findings together indicate a lack of consistency in observed acquiescence.

*Discussion.* Experiments 2a and 2b replicated the results of Experiment 1 in a more typical Likert scaling context. First, we observed greater misresponse and response latencies for negation than for affirmation. Second, misresponse and response latencies increased linearly across types of items as follows: true affirmations, false affirmations, false negations, and true negations. As in Experiment 1, this pattern matches the predictions of item verification difficulty but not those of inattention or acquiescence. Third, measures of inattention and acquiescence revealed low levels of each, further indicating that these response styles are not necessary for misresponse to occur.

**EXPERIMENT 3**

**Purpose and Predictions**

In Experiment 3, we advance our investigation in three ways. First, we assess whether the type of negation used to reverse an item affects misresponse. Second, to examine further the potential for acquiescence and inattention to produce misresponse, we consider additional measures of each. Third, we explore the effect of need for cognition on misresponse.
Explicit Versus Implicit Negation. In the previous experiments, we considered only explicit negation, which occurs in two main forms: particle (e.g., “not”) and affixal (e.g., “un-”). However, the psycholinguistics literature suggests that, during comprehension, readers convert some explicit affirmations into implicit negations (Givón 1993). For example, after reading the affirmative item “I believe the advertisement is vague,” a respondent may form the initial representation “Affirmed(vague, advertisement)”); however, if the respondent’s activated knowledge structures support a negated form of the predicate “vague” (e.g., “not clear”), the affirmative representation will yield to the representation “Negated(clear, advertisement)”.” In this example, the implicit negation of “vague” indicates that the respondent best understands vagueness as the absence of clarity rather than as a meaningful property in itself.

We theorize that implicit negation requires an additional mental operation during comprehension to convert the affirmative item into an implicitly negated representation. Thus, implicit negations should be more difficult to process than explicit negations and thereby result in greater misresponse. Because the conversion from explicit affirmation to implicit negation occurs prior to item verification, it should increase misresponse by the same amount for both true and false negations. Consequently, the Constituent Comparison Model predicts the same form of interaction between polarity and truth value for implicit negation as it does for explicit negation (Figure 1, Panel E). Additionally, there should be no differences in misresponse between affixal and particle negation because these types of explicit negation yield the same propositional form.

Accounts of both inattention and acquiescence predict no differences in misresponse across implicit and explicit negation. Specifically, inattentive respondents expect items to be stated affirmatively, and thus, negation type does not affect their responses. Similarly, acquiescent respondents automatically accept items at comprehension, independent of item
content or polarity; thus, their responses should not vary with negation type (e.g., Gilbert 1991; Grant, Malaviya, and Sternthal 2004; Mayo, Schul, and Burnstein 2004).

Additional Measures of Inattention and Acquiescence. Because ARS1 uses reversed items to identify acquiescence, it can overstate acquiescence if people misrespond to reversed items for other reasons. Therefore, we consider a second measure of acquiescence (ARS2) that does not use reversed items but instead is based on the extent to which respondents agree with items that are heterogeneous in content (Baumgartner and Steenkamp 2001; Greenleaf 1992). In calculating ARS2, strong agreements yield a score of 2, agreements result in a score of 1, and all other responses receive a score of 0. The final ARS2 measure averages these scores across heterogeneous items.

We also consider a measure of inattention that reflects random, rather than straight-line, responses. Although random responses generally do not contribute to systematic error, we can quantify this behavior in an effort to identify inattentive respondents. We measure random responding by computing the sum of the absolute difference between pairs of items with correlations greater than .5 and means that are not significantly different (Baumgartner and Steenkamp 2001). The larger this sum is, the more a respondent exhibits random responding. Because items must be stated in the same direction to compute this measure, we examine only non-target, non-reversed items.

Need for Cognition. The item verification difficulty explanation of misresponse implicates both item content and individual cognitive response processes. Thus, a reasonable question is whether the observed misresponse patterns hold across traits that may affect the response processes. One such trait, which has received significant attention from researchers in marketing and other social sciences, is the need for cognition, defined as “a stable individual
difference in people’s tendency to engage in and enjoy effortful cognitive activity” (Cacioppo et al. 1996, p. 198). On the basis of this definition, we expect misresponse to be inversely related to the need for cognition. Less clear is whether the predictions of the Constituent Comparison Model, and thus the item verification difficulty explanation, hold across groups that differ in need for cognition. Consequently, we investigate need for cognition as both an exploratory variable and an additional test of the Constituent Comparison Model.8

Design and Procedure

We manipulated target item wording within subjects using four treatments: three types of negation (particle, affixal, implicit) plus affirmation (the control). In addition, we assigned subjects randomly to one of four survey versions, each of which employed a different Williams Latin square design (Williams 1949). The four designs controlled for potential first-order carryover effects and resulted in the presentation of each target item to a portion of the subjects at each item wording level. Thus, subjects received 16 multi-item scales, divided into four groups of four scales, with the order of the scales constant for all subjects. Within each group of four scales, we provided four target items, or one per scale. Three of the target items were reversed by affixal, particle, or implicit negation, whereas the fourth was stated affirmatively.

We measured the truth value of each target item using the same approach as in Experiments 2a and 2b. The scales used in Experiment 3 have been published in peer-reviewed journals. Six scales contained four items, six contained five, three contained six, and one contained eight items. Subjects were 483 undergraduate students who participated in exchange for course credit.

Results and Discussion

Misresponse. We first analyzed only the negated items using a loglinear model that

8 We thank two anonymous reviewers whose suggestions led to this analysis.
included misresponse as the dependent variable and type of negation, truth value, and their interaction as the independent variables. This analysis revealed significant main effects of the type of negation ($\chi^2 = 16.06, 2 \text{ df}, p < .001$) and truth value ($\chi^2 = 73.15, 1 \text{ df}, p < .001$) but a nonsignificant interaction ($\chi^2 = 0.52, 2 \text{ df}, p = .771$). Because the interaction was not significant, we combined all types of negation and included the affirmative items in the subsequent analysis.

We ran a second loglinear model with misresponse as the dependent variable and item polarity (i.e., negation or affirmation), truth value, and their interaction as the independent variables. We found significant main effects of item polarity ($\chi^2 = 139.65, 1 \text{ df}, p < .001$) and truth value ($\chi^2 = 77.79, 1 \text{ df}, p < .001$), as well as a significant interaction ($\chi^2 = 32.68, 1 \text{ df}, p < .001$).

Misresponse was 5.99% for true affirmations, 10.61% for false affirmations, 18.11% for false negations, and 32.23% for true negations, and each pair of proportions was significantly different at the .05 level. This pattern of results (Figure 2, Panel E) is consistent with the predictions derived from item verification difficulty (Figure 1, Panel E).

We computed misresponse for each of the four treatments to examine the main effect of item wording. Misresponse was 8.12% for affirmative items, 20.70% for affixal negations, 20.90% for particle negations, and 27.71% for implicit negations. Tests of proportions indicated that misresponse for affirmative items was significantly lower than that for all types of negation (all p-values < .05). As we predicted, misresponse for implicit negation was significantly higher than misresponse for the other types of items (all p-values < .05). Also as predicted, misresponse did not differ for affixal and particle negation (two-tailed p = .910). Additionally, the pattern of misresponse was consistent across each type of negation, with greater misresponse for true negations than for false negations (particle negation: false = 16.88%, true = 30.03%; affixal negation: false = 16.42%, true = 29.34%; implicit negation: false = 22.44%, true = 38.71%).
Consistent with the prediction that implicit negation introduces difficulty prior to the constituent comparison process, and thereby increases item verification difficulty equivalently for both true and false items, we found a nonsignificant interaction between truth value and a factor that indicated whether the negation was implicit or explicit ($\chi^2 = .58$, 1 df, $p = .445$).

*Inattention and Acquiescence.* Recall that subjects responded to 16 scales. To the extent misresponse is due to acquiescence and/or inattention, measures of these response styles from one group of scales should predict misresponse in the other scales. To test this, we computed ARS1 and ARS2 for the first eight scales and then for the last eight scales. For inattention, we created a dummy variable for the first group of eight scales, coded as “1” if a subject exhibited a straight-line response pattern for any of the scales and coded as “0” otherwise. We applied the same measure to the last eight scales. We also measured random responding using the continuous measure (described earlier) for the first and last groups of eight scales and found 27 (10) pairs of items from the first (last) eight scales that met this measure’s criteria.

The two acquiescence and two inattention measures from the last (first) eight scales served as independent variables in separate logistic regression analyses for each of the first (last) eight scales, and misresponse to the target item served as the dependent variable. As we analyzed each independent variable separately, we estimated four models for each of the 16 target items. At the .05 level, ARS1 predicted misresponse for two of the 16 target items, ARS2 predicted misresponse for three target items, the straight-line response dummy variable predicted misresponse for one target item, and the random response measure predicted misresponse for two

---

9 To assess the appropriateness of the items for the ARS2 measure, we computed the average absolute correlation and the range of correlations across all non-target affirmative items for both the first and the last eight scales. The average absolute correlation was .262 for the first eight scales and .202 for the last eight scales, and they ranged from -.538 to .857 and -.538 to .874, respectively. Therefore, the items were heterogeneous and appropriate for ARS2. Principle components analysis of the 64 non-target items also yielded 16 eigenvalues greater than 1, indicating that the items probe many different constructs (cf. Greenleaf 1992).
target items. Thus, these measures were poor predictors of misresponse, suggesting that inattention and acquiescence were not prevalent in these data.¹⁰

*Need for Cognition.* One of the 16 scales measured need for cognition with five items (Cacioppo, Petty, and Kao 1984). We computed the average response to the four non-target items to obtain a measure of each subject’s need for cognition (coefficient $\alpha = .86$) and then performed a tertiary split to create categories of subjects who ranked low (less than 4.3 on a seven-point scale), medium (4.3–5.4), and high (greater than 5.4) in their need for cognition (cf. Inman, McAlister, and Hoyer 1990). For each subject, we also created a count of the total number of misresponses across the 16 scales. As conjectured, the high need for cognition group exhibited significantly fewer misresponses than the low need for cognition group ($M_{\text{High}} = 1.86$, $M_{\text{Low}} = 2.22$; $z = 2.36$, one-tailed $p = .010$). Further, we observed the misresponse pattern predicted by the Constituent Comparison Model for both groups of subjects. For the low need for cognition group, misresponse was 8.42%, 11.23%, 22.93%, and 32.32% for true affirmatives, false affirmatives, false negations, and true negations, respectively, and the corresponding misresponse for the high group was 4.07%, 8.55%, 15.39%, and 33.33%.

*Discussion.* As in Experiments 1, 2a, and 2b, misresponse in Experiment 3 exhibited a pattern consistent with the Constituent Comparison Model and the item verification difficulty explanation. We hypothesized additional difficulty associated with implicit negation as a result of an extra cognitive operation that converts explicit affirmations into implicit negations during the comprehension step and found support for this claim; as we predicted, greater overall misresponse occurred for implicit negation compared with affixal and particle negations, but

---

¹⁰ We also compared misresponders (i.e., subjects who misresponded to any of the 16 target items) to non-misresponders (i.e., subjects who did not misrespond to any of the 16 target items) for four measures: acquiescent responding (ARS1 and ARS2), random responding, and the number of scales for which they exhibited straight-line responding. We found no significant differences across these groups for any of the measures (all $p$-values $>.05$).
misresponse did not differ between affixal and particle negations. We also considered additional measures of acquiescence (ARS2) and inattention (random responses), which proved to be poor predictors of misresponse, as did the prior measures (ARS1 and straight-line responding).

Finally, our exploratory analysis of need for cognition revealed that people with a low need for cognition were more likely to misrespond to all types of items except true negations, the most difficult items. It is possible that true negations may be sufficiently difficult so as to obscure the effects of individual differences in need for cognition (Cacioppo et al. 1996). Nevertheless, respondents both low and high in need for cognition exhibited misresponse patterns that were consistent with the Constituent Comparison Model.

**GENERAL DISCUSSION**

We examined responses to reversed Likert items and found that the average rate of misresponse approached 20%, lending credence to recent concerns about item reversal (e.g., Marsh 1996; Netemeyer, Bearden, and Sharma 2003). Further, we consistently found substantial misresponse even when levels of inattention and acquiescence were minimal. In contrast, patterns of misresponse and response latency across item type were well predicted by the Constituent Comparison Model of item verification difficulty.

**Summary of Theoretical and Empirical Contributions**

We began by providing theory-based accounts for inattention, acquiescence, and a third explanation – item verification difficulty. Recognizing that misresponse may have multiple sources, we linked the mechanisms of each explanation to the broader conceptual framework described by the four-step model of item response (Sudman, Bradburn, and Schwarz 1996; Tourangeau, Rips, and Rasinski 2000). Furthermore, we identified item polarity and item truth value as variables that enable researchers to empirically disentangle the effects of inattention,
acquiescence, and item verification difficulty on misresponse.

For the item verification difficulty explanation, we introduced the Constituent Comparison Model (Carpenter and Just 1975). We used this model to describe how respondents judge the correspondence between Likert items and their beliefs. Whereas prior speculation regarding the difficulty of processing reversed Likert items has pointed towards item comprehension (e.g., vague or complex semantic content), we theorized and found that substantial difficulty arises during item judgment. Also as predicted, the difficulty across item type was a linear function of the number of cognitive operations required to verify an item.

Our empirical investigation began with an exploratory study involving previously published data that span a diverse set of respondents, research settings, and construct domains. Misresponse to reversed items was substantial, even though acquiescence and inattention were not prevalent in these data. We then conducted Experiment 1 to investigate the possibility of item verification difficulty as a source of misresponse. Both misresponse and response latencies exhibited an interaction between item polarity and truth value, and the pattern of results suggested a linear increase from true affirmative to false affirmative to false negated to true negated items. These results support the item verification difficulty account, as described by the Constituent Comparison Model. Experiments 2a and 2b addressed the concern that inattention and acquiescence may have been suppressed by the monetary incentives and simple, objective nature of the items in Experiment 1. However, inattention and acquiescence were again non-prevalent, and the patterns of misresponse and response latency conformed to the predictions of the Constituent Comparison Model. In Experiment 3, we extended the investigation of misresponse in three ways. First, we predicted and found that implicit negation led to greater misresponse than did affixal or particle negations (i.e., explicit negations) and that there was no
difference in misresponse between affixal and particle negation. Second, we introduced additional measures of inattention and acquiescence and found them to be poor predictors of misresponse. Third, we determined that higher (versus lower) need for cognition was associated with lower overall misresponse, but that misresponse patterns across item type did not differ. This suggests that the judgment process underlying item verification is robust with respect to an important cognition-based trait.

**Limitations and Further Research**

We note several potential limitations of this research. First, though the exploratory study included diverse respondents, the experiments used only student subjects. Accordingly, our results may understate the problem of misresponse for non-student samples that consist of less educated or lower social class respondents because of the elevated risk of acquiescence. Second, the levels of observed inattention and acquiescence were low. Accordingly, future research could manipulate inattention and acquiescence to more closely examine the patterns of misresponse and response latencies associated with these response styles. Such research also could explore the possibility of relationships between sources of misresponse. For example, acquiescent respondents may be more likely than non-acquiescent respondents to become inattentive when processing large sets of items since their response (agreement) is the same for all items (Krosnick 1999). Third, we only considered scales with fewer reversed items than non-reversed items. Although this choice is consistent with the norm for multi-item Likert scales, Baumgartner and Steenkamp (2001) demonstrate that the extent to which a scale is unbalanced (i.e., an unequal proportion of reversed and non-reversed items) influences the extent to which acquiescence bias contaminates scale scores. Similarly, the proportion of reversed items may affect misresponse. To predict the nature of such an effect, it would be necessary to specify how
the proportion of reversed items relates to response selection (Sudman, Bradbury, and Schwarz 1996).

**Recommendations for Reducing the Likelihood of Misresponse**

Knowledge of the underlying processes that generate misresponse provides a basis for taking actions to reduce it. Because extant literature on questionnaire design contains guidance for reducing inattention and acquiescence, we focus our recommendations on item verification difficulty.\(^{11}\)

*Consider Antonymic Likert Items.* Negation consistently led to greater misresponse than affirmation, which suggests that researchers should consider avoiding negation. This may appear to be at odds with the measurement literature, where it is recommended that scales contain both reversed and non-reversed items to help control acquiescence bias (Baumgartner and Steenkamp 2001; Greenleaf 1992). However, items can also be reversed by using affirmations that carry opposite (i.e., antonymic) meaning to that of the focal construct. This approach retains the benefits of item reversal and can avoid the hazards of explicit negation. With this advice, however, comes a word of caution: As we note in Experiment 3, respondents may represent some antonymic affirmations as implicit negations, which can result in greater misresponse than if an explicit negation had been used. To mitigate this concern, we recommend that researchers consider pretesting items for implicit negation. This can be accomplished by, first, instructing respondents to read all non-target items in the focal scale, thereby establishing a semantic context in which to comprehend the target item(s). Next, respondents should be presented with each target antonymic affirmative item with its predicate overtly highlighted (e.g., underlined).

\(^{11}\) Although inattention and acquiescence were not significant sources of misresponse in our data, researchers should remain vigilant. Such response styles are known to vary across content domains (Baumgartner and Steenkamp 2001; Ray 1983; Rorer 1965; Winkler, Kanouse, and Ware 1982), research designs (Feldman and Lynch 1988), time (Hui and Triandis 1985; Piedmont et al. 2000), and cultures (Johnson et al. 2005; Steenkamp and Baumgartner 1998).
Respondents should then be asked to generate a synonym for the highlighted predicate. If a large proportion of the generated synonyms are explicit negations, the antonym is an implicit negation and thus should be avoided (cf. Klima’s [1964] methodology for identifying negation).

Consider Alternative Item Formats. The observed pattern of misresponse across all Likert items in this research—approximately 5% for true affirmations, 10% for false affirmations, 22% for false negations, and 35% for true negations—raises two concerns. First, respondents who are low (high) in a construct should disagree (agree) with affirmative items and agree (disagree) with negated items measuring that construct. Because false affirmative items lead to greater response difficulty than true affirmative items, whereas true negated items lead to greater response difficulty than false negated items, respondents low in a construct will experience greater difficulty responding to all scale items. Consequently, data quality may differ across groups based simply on the extent to which they possess a measured construct. This holds important implications for analyses involving the estimation or comparison of means and multi-group correlational analyses. Second, though we expected that negation would prove problematic, misresponse to false affirmative Likert items also reached the threshold of 10% identified by Schmitt and Stults (1985). Collectively, our empirical findings suggest that researchers should consider alternative scaling formats that may be less susceptible to misresponse. We believe interrogative scaling (Javeline 1999) holds promise.

Consider the Likert item, “The notion of thinking abstractly is not appealing to me” (strongly disagree/strongly agree), which can be reformatted as the interrogative item, “How do you feel about the notion of thinking abstractly?” (not very appealing/very appealing). The stem of the Likert item yields a complete propositional representation, “Negated(appealing, thinking abstractly),” whereas the stem of the interrogative item yields an incomplete representation,
“__(__, thinking abstractly).” This incompleteness requires respondents to pair the attitude object (“thinking abstractly”) with a negated predicate (“not very appealing”) and an affirmed predicate (“very appealing”), yielding two distinct item representations to be compared with their belief. This results in two parallel chances to respond correctly, which should lower the likelihood of misresponse.

An added benefit of interrogative scaling is that it may reduce acquiescence (Javeline 1999; Wong, Rindfliesch, and Burroughs 2003), because the item stem lacks a predicate and thereby deters respondents from forming a judgment until they have considered both response options. In addition, the two anchors represent predicates with opposite meanings and thus should disrupt tendencies toward automatic acceptance because the response task requires a choice between assertions rather than an endorsement of a single assertion. Furthermore, whereas Likert items confound the nature of the acquiescence response style (agreeableness) with the nature and intensity of the measured response (relative agreement), interrogative items do not suffer from this limitation.

We also speculate that interrogative scaling, relative to Likert scaling, can reduce inattention. Support for this comes from research showing that interrogation (i.e., asking questions) leads to enhanced cognitive elaboration relative to assertion (Petty, Cacioppo, and Heesacker 1981; Petty, Rennier, and Cacioppo 1987). Accordingly, interrogative items may be more naturally engaging than Likert items, thereby encouraging inattentive respondents to allocate greater resources to processing item content. Additionally, interrogative items employ item-specific predicates, which discourage expectations for similar item content. The combined effects of increased elaboration and predicate uniqueness may help reduce respondents’ reliance on expectation-driven response heuristics.
Recommendations for Diagnosing and Treating Misresponse

Although carefully designed research may minimize the likelihood of misresponse, no set of techniques can fully prevent it. Accordingly, we present recommendations for identifying, quantifying, classifying, and analytically adjusting for misresponse.

Identifying and Quantifying Misresponse. Assessing misresponse for objective items is straightforward because the true response is known. For subjective items, we advocate the binary measure of misresponse used in this research. An alternative approach involves examining interitem correlations. Such correlations are typically large for reflective, multi-item scales. However, as misresponse increases, correlations involving the target item rapidly diminish in magnitude and ultimately reverse signs. Factor analysis is sensitive to the relationship between misresponse and interitem correlations. Specifically, given typical item loadings and sample sizes, factor analysis consistently exhibits a misresponse-driven factor when misresponse is as low as 10% (Schmitt and Stults 1985). Although this approach does not identify individual misresponders, it represents an efficient means of assessing the practical impact of misresponse (Marsh 1996).

Classifying Misresponse by Source. If researchers wish to make analytical adjustments to data that have been contaminated by misresponse, it is important to identify the source(s) of misresponse. Failure to do so may lead to adjustments to the data that introduce, rather than remove, bias. The source(s) of misresponse in a data set can be identified using the measures described herein. That is, researchers may identify misresponse due to inattention through straight-line responding12 and misresponse due to acquiescence through the ARS1 and ARS2 measures (e.g., Baumgartner and Steenkamp 2001). Item verification difficulty is most clearly

12 Straight-line responding may also indicate problems with item redundancy if a scale contains no reversed items (Drolet and Morrison 2001). The random response measure of inattention may identify additional inattentive respondents who contribute to random error (though not misresponse).
identified by an interaction between truth value and item polarity on misresponse, though verification difficulty is also suggested when misresponse is substantial and inattention and acquiescence are absent. Unfortunately, it is typically infeasible to include multiple versions of an item in a questionnaire. Thus, a more pragmatic approach is to determine a target negated item’s truth value using responses to the non-target items and then compare misresponse for respondents who should have disagreed with the target negated item (i.e., false negation) to that of respondents who should have agreed with the same item (i.e., true negation). Item verification difficulty is implicated when misresponse is significantly greater for the true negation. A similar approach applies for affirmative items, such that item verification difficulty is implicated if misresponse is significantly greater for the false affirmation than for the true affirmation. When response latency data are available, researchers may apply the same procedure, such that item verification difficulty is implicated when latencies are greater for true versus false negations and false versus true affirmations.

Analytical Adjustments. Having quantified and identified the source(s) of misresponse, researchers may then consider analytical adjustments. In the case of inattention, it may be advisable to delete respondents from the sample because it is unclear whether their responses contain any information about their true scores. However, because deletion causes non-response, researchers should consider whether deletion might bias the research results (Little and Rubin 2002).

In the case of acquiescence, it is not appropriate to delete respondents because their responses contain information about their true scores (Greenleaf 1992). Instead, we recommend using the acquiescence measures (ARS1 or ARS2) and implementing the regression techniques developed by Baumgartner and Steenkamp (2001). That is, researchers can regress responses on
either acquiescence measure (or an index of the two) such that the residuals represent purified construct scores. The acquiescence measures can also serve as covariates in analyses that involve the constructs from which the measures are derived. An alternative approach is to employ structural equation modeling with ARS1 or ARS2 as indicators that reflect an acquiescent response style factor (see Podsakoff et al. 2003, p. 896).

Finally, in the case of item verification difficulty, it is not appropriate to delete respondents or adjust all items or scale scores because only a subset of respondents have provided inaccurate responses, and these respondents have only misresponded to a subset of items. Accordingly, we recommend deleting individual misresponses and leveraging missing data techniques (see Little and Rubin 2002). For example, it may be possible to use the values of non-target items in a scale as a basis for imputing values to target items.

In summary, this research illustrates the importance of item wording effects on the quality of data obtained from multi-item scales. Although marketing researchers have made great strides in developing techniques to produce valid and reliable measures, our findings indicate that these techniques can be further refined. Such refinements may require researchers to explore new approaches and integrate insights from cognitive science, psycholinguistics, and survey methodology. And though the present research provides a useful starting point, much work remains.
Table 1  
Illustration of Likert Item Verification Difficulty as a Function of Item Truth Value and Item Polarity

A. Explanation of Item Verification Process
- Respondents represent Likert items and beliefs as abstract propositions in the form $\text{polarity}(\text{predicate, argument})$; the $(\text{predicate, argument})$ constituent conveys an assertion, and the $\text{polarity}$ constituent indicates whether the assertion is affirmed or negated.
- Respondents judge or verify items by comparing the constituents that comprise their belief and the item.
- In the example below, **bold font** indicates the focus of the comparison process at each stage. “Matches” allow the process to continue. “Mismatches” are mentally tagged and cause the process to restart. A true-false response index (Index) is initialized to “true” and toggles for each mismatch. Tagged mismatches, indicated by “*,” are treated as matches in subsequent comparisons.

B. Respondent Belief: “The brand is exciting”

C. Four Types of Likert Items, Given the Respondent’s Belief

<table>
<thead>
<tr>
<th>True Affirmation</th>
<th>False Affirmation</th>
<th>False Negation</th>
<th>True Negation</th>
</tr>
</thead>
<tbody>
<tr>
<td>“The brand is exciting”</td>
<td>“The brand is boring”</td>
<td>“The brand isn’t exciting”</td>
<td>“The brand isn’t boring”</td>
</tr>
</tbody>
</table>

D. Item-Belief Comparison Process, by Item Type

<table>
<thead>
<tr>
<th>Item</th>
<th>Belief</th>
<th>Result</th>
<th>Item</th>
<th>Belief</th>
<th>Result</th>
<th>Item</th>
<th>Belief</th>
<th>Result</th>
<th>Item</th>
<th>Belief</th>
<th>Result</th>
<th>Item</th>
<th>Belief</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td>Item</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Affirmed(\text{boring, brand})</td>
<td>Negated(\text{exciting, brand})</td>
<td>Negated(\text{boring, brand})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Belief</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Affirmed(\text{boring, brand})</td>
<td>Match, Index: false, Restart</td>
<td>Match, Index: true, Continue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Result</td>
<td>Match, Index: true, Stop</td>
<td>Mismatch, Index: true, Stop</td>
<td>Mismatch, Index: false, Restart</td>
<td>Mismatch, Index: false, Restart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>Item</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Affirmed(\text{boring, brand})</td>
<td>Negated(\text{exciting, brand})</td>
<td>Negated(\text{boring, brand})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Belief</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Affirmed(\text{boring, brand})</td>
<td>Match*, Index: false, Continue</td>
<td>Match*, Index: true, Continue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Result</td>
<td>Match, Index: true, Stop</td>
<td>Mismatch, Index: true, Stop</td>
<td>Mismatch, Index: false, Restart</td>
<td>Mismatch, Index: false, Restart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td>Item</td>
<td>Affirmed(\text{boring, brand})</td>
<td>Negated(\text{exciting, brand})</td>
<td>Negated(\text{boring, brand})</td>
<td>Negated(\text{exciting, brand})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Belief</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Affirmed(\text{boring, brand})</td>
<td>Match, Index: true, Stop</td>
<td>Match, Index: true, Stop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Result</td>
<td>Match, Index: false, Stop</td>
<td>Mismatch, Index: false, Stop</td>
<td>Mismatch, Index: false, Stop</td>
<td>Mismatch, Index: false, Stop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 4</td>
<td>Item</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Negated(\text{exciting, brand})</td>
<td>Negated(\text{boring, brand})</td>
<td>Negated(\text{boring, brand})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Belief</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Affirmed(\text{boring, brand})</td>
<td>Match*, Index: false, Stop</td>
<td>Match*, Index: true, Stop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Result</td>
<td>Match, Index: true, Stop</td>
<td>Mismatch, Index: true, Stop</td>
<td>Mismatch, Index: false, Restart</td>
<td>Mismatch, Index: false, Restart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Step 5</td>
<td>Item</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Negated(\text{boring, brand})</td>
<td>Negated(\text{boring, brand})</td>
<td>Negated(\text{boring, brand})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Belief</td>
<td>Affirmed(\text{exciting, brand})</td>
<td>Affirmed(\text{boring, brand})</td>
<td>Match*, Index: true, Stop</td>
<td>Match*, Index: true, Stop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Result</td>
<td>Match*, Index: true, Stop</td>
<td>Mismatch, Index: true, Stop</td>
<td>Mismatch, Index: false, Restart</td>
<td>Mismatch, Index: false, Restart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E. Verification Difficulty

<table>
<thead>
<tr>
<th>True Affirmation</th>
<th>False Affirmation</th>
<th>False Negation</th>
<th>True Negation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 comparisons</td>
<td>3 comparisons</td>
<td>4 comparisons</td>
<td>5 comparisons</td>
</tr>
</tbody>
</table>


### Table 2
**Exploratory Study of Secondary Data: Data Source, Characteristics, and Misresponse to Reversed Likert Items**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Respondents</th>
<th>Sample Size</th>
<th>Number of Scales&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Items per Scale&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Misresponse (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnold and Reynolds (2003)</td>
<td>Non-student adults</td>
<td>520</td>
<td>4</td>
<td>9(3), 10, 10, 6</td>
<td>20.31</td>
</tr>
<tr>
<td>Bearden et al. (2000)</td>
<td>College students</td>
<td>184</td>
<td>2</td>
<td>3, 5(3)</td>
<td>25.09</td>
</tr>
<tr>
<td>Bhattacharya, Rao, and Glynn (1995)</td>
<td>Non-student adults</td>
<td>299</td>
<td>1</td>
<td>3</td>
<td>40.10</td>
</tr>
<tr>
<td>Hardesty, Carlson, and Bearden (2002)</td>
<td>College students</td>
<td>244</td>
<td>2</td>
<td>4, 5</td>
<td>18.01</td>
</tr>
<tr>
<td>McKee, Simmers, and Licata (2006)</td>
<td>Non-student adults</td>
<td>444</td>
<td>2</td>
<td>10(2), 4</td>
<td>17.15</td>
</tr>
<tr>
<td>Niedrich and Swain (2003)</td>
<td>College students</td>
<td>280 / 280&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1</td>
<td>5</td>
<td>29.93</td>
</tr>
<tr>
<td>Weathers, Sharma, and Niedrich (2005)</td>
<td>College students</td>
<td>595</td>
<td>1</td>
<td>4</td>
<td>41.00</td>
</tr>
</tbody>
</table>

<sup>a</sup> Scales and target items can be found in the respective publications.

<sup>b</sup> One negated item per scale, except as noted in parentheses.

<sup>c</sup> Sample sizes for each of the four years that the survey was administered.

<sup>d</sup> Sample sizes for each of two administrations of the survey.
Figure 1
Predicted Misresponse and Response Latencies as Functions of Item Polarity and Truth Value

Predicted Misresponse

Panel A

Inattention

Panel C

Acquiescence

Panel E

Item Verification Difficulty

Predicted Response Latencies

Panel B

Panel D

Panel F
Figure 2
Actual Misresponse and Response Latencies as Functions of Item Polarity and Truth Value

Actual Misresponse

Panel A

Experiment 1

Experiments 2a and 2b

Experiment 3

Actual Response Latencies

Panel B

Panel C

Panel D

Panel E
REFERENCES


Herche, Joel and Brian Engelland (1996), “Reversed-Polarity Items and Scale


——— and Herbert H. Clark (1973), “Drawing Inferences from the Presuppositions and


(June), 78-90.


