UNIVERSITEIT GENT

**FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE**

**TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92**

WORKING PAPER

The Proximity Effect:
The Role of Interitem Distance on Reverse-Item Bias[1]

Bert Weijters[2]

Maggie Geuens[3]

Niels Schillewaert[4]

December 2008

2008/546

[1] The authors would like to thank the editor, area editor and reviewers of IJRM for their constructive comments on an earlier version of this manuscript. Bert Weijters would like to thank the ICM (Belgium) for supporting his research, and Patrick Van Kenhove, Alain De Beuckelaer, Jaak Billiet and Hans Baumgartner for their feedback on his Phd version of this paper.
[2] Corresponding author. Vlerick Leuven Gent Management School, Ghent, Belgium, e-mail: bert.weijters@vlerick.be
[3] Ghent University and Vlerick Leuven Gent Management School, Ghent, Belgium, e-mail: maggie.geuens@ugent.be..
[4] Vlerick Leuven Gent Management School, Ghent, Belgium, e-mail: niels.schillewaert@vlerick.be

D/2008/7012/55

The Proximity Effect:

THE ROLE OF INTERITEM DISTANCE ON REVERSE-ITEM BIAS

ABSTRACT

On the basis of cognitive interviews and drawing on the belief-sampling model, we introduce the proximity effect model. The model explains the correlation between two items as a function of their conceptual relationship (nonreversed same-construct items, reversed same-construct items, unrelated items) and their proximity in the questionnaire. In a quantitative study using primary data (*N* = 3,114), the (positive) correlation between a nonreversed item pair decreases with increasing interitem distance. In contrast, the (negative) correlation between reversed item pairs decreases (i.e., become stronger) with increasing interitem distance. This is related to the way respondents tend to minimize retrieval of additional information when answering nearby nonreversed items and maximize retrieval of new and different information when answering nearby reversed items. Using two simulated correlation matrices that represent common measurement situations, we assess the impact of the proximity effect on factor structure and reliability. The resultant key recommendations pertain to the use of reversed items that are dispersed throughout the questionnaire and the use of a confirmatory factor analysis model specification including a response style factor.

The Proximity Effect: The Role of Interitem Distance on Reverse-Item Bias

1. INTRODUCTION

Questionnaires are an indispensable source of data in marketing research, and the Likert item format, in which participants rate their agreement with statements, is particularly popular. When Likert (1932, p. 46) introduced this measurement type, he recommended the use of reversed items. A reversed item i′ is intended to relate to the same construct as its nonreversed counterpart i, but in the opposite direction. To illustrate this with two items taken from an innovativeness scale (Steenkamp & Gielens, 2003), item i could be “I enjoy taking chances in buying new products,” and item i′ could be “I am very cautious in trying new and different products.”

The inclusion of reversed items in scales enhances scale validity and may be strategically used to make respondents attend more to the specific content of individual items (Barnette, 2000), to ensure more complete coverage of the underlying content domain (Bentler, 1969; Carroll, Yik, Russell, & Barrett, 1999), and/or to counter bias due to acquiescence response style (Paulhus, 1991). Unfortunately, the use of reversed items has its own methodological problems; in particular, reversed items tend to show lower factor loadings and lead to lower internal consistency because of their weaker correlations with other items that measure the same construct (Herche & Engelland, 1996; Motl & DiStefano, 2002; Quilty, Oakman, & Risko, 2006). We refer to this phenomenon as *reversed-item bias*. Whereas reversed-item bias has led some researchers to oppose the use of reversed items (Barnette, 2000; Marsh, 1996), others remain in favor of it (Baumgartner & Steenkamp, 2001; Paulhus, 1991). To sort out this debate, clearer insight is needed regarding when and how reversed-item bias arises.

Whereas reversed-item bias was initially believed to stem from respondent inattentiveness (Schmitt & Stults, 1985) or acquiescence response style defined as the tendency to agree to items regardless of their content (Mirowsky & Ross, 1991), more recent studies have focused on the way respondents cognitively process reversed items (e.g., Swain, Weathers, and Niedrich 2008). Wong, Rindfleisch, and Burroughs (2003) suggest that part of the problem may relate to the specific beliefs that respondents retrieve to respond to an item. In particular, Wong et al. suggest that reversed-item bias may be related to the way respondents think about what researchers intend to be opposite poles of a construct: respondents may think of two opposite statements as both containing some truth, resulting in responses to reversed items that are not necessarily opposite in valence.

We propose that respondents may indeed be able to think of apparently contradictory evidence, for example retrieving instances of how they behave in both innovative and noninnovative ways. Drawing on qualitative and quantitative research, we demonstrate that this is more likely to occur if nonreversed and reversed items related to the same construct are positioned close to each other. More generally, we propose and test a model of how correlations between items are in part determined by their proximity in the questionnaire. We call this *the proximity effect*[5] and specify this effect for nonreversed items,[6] reversed items, and unrelated items. Our contribution to the literature is threefold. First, we provide a conceptual framework for the proximity effect; this is based primarily on a combination of a review of the literature and qualitative research using cognitive interviews. Second, we provide quantified estimates of the expected correlation between two items as a function of their distance (number of intervening items) and their relationship (nonreversed, reversed, unrelated). Third, on the basis of an analysis of two simulated correlation matrices, we show the impact of the proximity effect on factor structure and reliability.

The proximity effect model is an important contribution to the literature because it shows how reversed-item bias and, more generally, interitem correlations can be managed by manipulating questionnaire design, specifically either by grouping items that measure the same construct or by dispersing items that measure the same construct throughout the questionnaire.

2. Conceptual Background

In the literature review, we first focus on how item positioning within the questionnaire guides respondents' processing of items. Next, we introduce the belief-sampling model (Tourangeau, Rips, & Rasinski, 2000) and discuss how it is related to item positioning. To complement our literature review, we then present a qualitative study. Building on both the literature review and the qualitative research, we then formulate hypotheses to be tested in a quantitative study.

2.1. Literature review

2.1.1. Item spacing as information

Researchers usually apply one of two methods for positioning conceptually related items within a questionnaire (Ostrom, Betz, & Skowronski, 1992). In the first, the researcher positions items that measure the same construct together in blocks. In the second, the researcher disperses same-construct items over the questionnaire, mixing them with other-construct items. In marketing research, both approaches are commonly used, though the debate is still open as to which approach is preferable (Bradlow & Fitzsimons, 2001). Grouping items that measure the same construct leads to higher interitem consistency (Budd, 1987; McFarland, Ryan, & Ellis, 2002). To the respondent, positional organization

[5] We thank the area editor for suggesting this term and the title for the current paper.

of the items is often a clear indication of conceptual organization (Ostrom, Betz, & Skowronski, 1992). Responses to an item are not merely a function of the item itself but are also affected by the presence and proximity of other items that measure the same construct (Budd, 1987; Knowles, 1988; Knowles et al., 1992; Ostrom et al., 1992). Such effects have been shown to dissipate over increasing interitem distances (Feldman & Lynch, 1988; Tourangeau, Rasinski, Bradburn, & D'Andrade, 1989; Tourangeau, Singer, & Presser, 2003). Whereas the literature provides evidence for a positive proximity effect between nonreversed items, the underlying cognitive process has not been specified in detail. This is remarkable because understanding this process is necessary information for assessing whether the higher consistency is desirable. Furthermore, it is not clear whether and how the proximity effect generalizes to reversed item pairs. Finally, the exact form and strength of the effect has not been quantified. We aim to address both these issues using qualitative and quantitative research, respectively. Before specifying the proximity effect model, we discuss the theoretical framework on which it draws.

2.1.2. The belief-sampling model

The belief-sampling model (Tourangeau et al., 2000) provides a valuable theoretical framework for better conceptualizing the influence of an item on responses to subsequent items. This model begins with the assumption that respondents perform a set of cognitive processes when answering questionnaire items: (1) comprehension (they attend to the question and interpret it), (2) retrieval (they generate a retrieval strategy and then retrieve relevant beliefs from memory), (3) judgment (they integrate the information into a conclusive judgment), and (4) response (they map the judgment onto the response categories and report it). The belief-sampling model posits that in response to a

[6]We use the term *nonreversed* to refer to items that assess the same construct and in the same direction.

questionnaire item, respondents retrieve beliefs that are related to the item under consideration and that are accessible at that moment. In this context, a belief sample refers to the assortment of considerations that a respondent retrieves in memory in order to respond to a questionnaire item (Tourangeau, Rips and Rasinski, 2000). Accessibility of different beliefs in memory may vary across time and across different contexts. For example, if respondents encounter two items on a related topic, the beliefs they retrieved in answering the first item are likely to be accessible when answering the second item. This might suggest that proximity automatically leads to high overlap in belief samples and, thus, to greater consistency in responses to nearby items. Yet accessible beliefs may be disregarded if they are considered irrelevant or redundant (Schwarz, 1997). In particular, respondents may disregard accessible beliefs because conversational norms and expectations lead the respondents to assume that they are supposed to provide new or different information (Schwarz, Strack, & Mai, 1991). We argue that the polarity of items combined with the distance between them determine the extent to which respondents base their response on deliberately maximally or minimally overlapping belief samples. On the aggregate level, this directly affects interitem correlations for nonreversed and reversed item pairs. Before detailing our hypotheses, we present qualitative research that fine-tunes our conceptual development of these phenomena.

2.2. Qualitative research using cognitive interviews

Because studies on the effect of interitem proximity on nonreversed versus reversed item pairs are rather scarce, we complement our literature review with qualitative exploratory research. In particular, we conducted cognitive interviews (DeMaio & Rothgeb, 1996; Jobe & Mingay, 1989) to study respondents' processing of unrelated items, nonreversed items, and reversed items in a controlled setting. Ericsson and Simon (1980) show that verbalizing cognitive processes provides valid data if respondents report their thoughts as they occur (i.e., concurrently rather than retrospectively). Using the same questionnaire as in the quantitative study, we asked respondents to think aloud as they processed items and responded to them. In particular, we focused on the extent to which respondents referred to previous items and/or previously retrieved beliefs and how respondents used the polarity of the items (nonreversed or reversed) and interitem distance as cues for steering the retrieval process.

A trained interviewer who was not informed about the hypotheses of the study (to avoid demand effects) conducted the interviews. Initially, we told the interviewer that the interviews were part of a prestudy with the aim of optimizing a questionnaire. After the interviews, we debriefed the interviewer on the actual purpose of the study.

Audio recordings were made of all interviews, and the interviewer was instructed to make notes for each questionnaire item, indicating whether any problems were apparent in terms of comprehension and interpretation of the item and, most importantly, whether the respondent related the current item to any of the previous items in the questionnaire (and, if so, which item, if specified). None of the items led to interpretation problems for more than two respondents, so notwithstanding random error, we can safely assume that there were no biases in the comprehension process.

2.2.1. Measurement instrument

To maximize compatibility, we used the same questionnaire in the qualitative and quantitative studies, though we used a paper-and-pencil version in the qualitative study and an online version in the quantitative study. The questionnaire contains a wide variety of items—76 in total—consisting of the following item sets: We randomly selected 10 pairs of reversed items (totaling 20 items) from the scales that Bruner, James, and Hensel (2001) compiled. We positioned each of the items randomly throughout the questionnaire, resulting in different distances between the respective pairs of reversed items. Furthermore, we dispersed the items of two balanced multi-item scales throughout the questionnaire: *dispositional innovativeness*, consisting of 3 nonreversed and 5 reversed items, and *market mavenism*, consisting of 2 nonreversed and 2 reversed items (Steenkamp & Gielens, 2003).[7] We also dispersed the items of one unbalanced multi-item scale throughout the questionnaire: *susceptibility to normative influence*, consisting of 8 positively scored items (Steenkamp & Gielens, 2003). We also included one unbalanced scale, the items of which were placed together as a block of items: *trust* and *loyalty* in a clothing retail context, consisting of 4 positive trust and 4 positive loyalty items (Sirdeshmukh, Singh, & Sabol, 2002); these scales were treated as one construct because they were closely related. Finally, we randomly selected 28 filler items from the same set of scales (see Bruner et al., 2001). More specifically, we used a two-step sampling procedure. First, we drew a random sample of scales, after which we randomly sampled one item from each scale. If two scales related to the same content domain (e.g., price sensitivity), we excluded one from the sample. Consequently, the selected items were not conceptually related. In addition, they were randomly dispersed throughout the questionnaire.

All items are shown in Appendix 1 and were presented in a seven-point Likert format, with response categories ranging from *strongly disagree* to *strongly agree*. The pages of the questionnaire contained 8 items, except for the last page, which contained 4 items. (Page length was the same in the qualitative paper-and-pencil version and the quantitative online version.)

---

[7] The item sample contained only two particle negations (i.e., Items in which the main proposition was negated by including the word *not*), one in the mavenism scale and one in the innovativeness scale. A sensitivity analysis revealed that omitting the two negated items from the analysis yielded nearly identical results.

2.2.2. Respondents

In total, 36 cognitive interviews were conducted in respondents' homes. The respondents were recruited by going door-to-door (in the regions around Ghent and Antwerp, Belgium; response rate: 51.4%) and asking for people's participation in an academic study that aimed to optimize a questionnaire.

The resultant sample consisted of 21 male respondents and 15 female respondents. The average age was 35.7 years (SD = 17.1), with a minimum age of 19 years and a maximum age of 80 years. The respondents were also highly heterogeneous in terms of professional background (including students, retirees, housekeepers, blue-collar workers, employees, and artists). All respondents stated they had participated in questionnaire research before.

2.2.3. Analysis

We listened to the audio recordings of all interviews, verified the interviewer's notes, and made additional notes when relevant. We checked whether a respondent referred to a previous item when he or she responded to an item and, if so, which item. When the coders differed in opinion, we (easily) reached a consensus by discussing the audio fragment.

2.2.4. Discussion of the results of the qualitative study

Overall, respondents were accurate in identifying items as belonging to the same scale. Across all respondents, not a single false alert occurred; that is, none of the respondents mentioned a content relationship between items in the questionnaire that was not intended. However, if we were to label as an omission every case in which respondents did not mention an intended content relationship, many such omissions would have occurred. Of the 36 respondents, 27 referred to a previous item at least once. Among those 27 respondents, the average number of times a reference was made was 4 (SD = 3). However, comments from the respondents made it clear that all 36 respondents believed that it was obvious that some of the items in the questionnaire were related. Some respondents referred to this with terms such as “control questions”; “variations to a general theme, but with different nuances”; and “repeating the same question in different words.” This practice seems to be experienced as common in questionnaire research, but it can also be irritating because the items are viewed as essentially redundant, making the questionnaire administration less efficient (e.g., “It’s really a waste of time to ask me the same question four times”).

Respondents were remarkably accurate in detecting the relationships between items, even for items positioned far away from each other (e.g., when responding to an item that came 30 items after its nonreversed counterpart, several respondents said something along the lines of “we had this question in the beginning of the questionnaire already”).

Importantly, the nature of recall seems to be different depending on the interitem distance. Specifically, if the items are close to each other, respondents tended to remember the previously encountered item (almost) verbally and thus were aware of its polarity (i.e., whether the item is nonreversed or reversed). For example, a respondent would point out, “Well, obviously, I disagree, as I just said when you asked me this other question just then.” In other cases, the verbal memory was apparent from remarks that respondents made on some specific words that were similar or dissimilar (e.g., “Well, that’s the same question basically, but now it’s specifically about products”).

If the items were located further apart, respondents tended to remember (1) the general topic the item referred to (e.g., “Ha, that’s one more on how I’m influenced by others’ opinions”) and (2) beliefs (e.g., memories, feelings, events) they retrieved in responding to the item (e.g., “Well, as I told you before, I had this discussion with my partner the other day”) but not (3) the wording of the item or (4) the polarity of the item. Overall, the shift from detailed (surface) recall to more generic (deep structure) recall is gradual, but in general, respondents did not seem to be able to repeat an item verbally after responding to approximately 5 to 10 intervening items.

To conclude, respondents are very aware of the fact that questionnaires usually contain several items that relate to the same content domain. The belief sample retrieved by respondents in response to a specific item is influenced by the proximity of related items. Also, if related items are positioned nearby, respondents are able to recall the previous item in detail. This allows them to distinguish between nonreversed and reversed items if these items are close to one another. We will discuss how this influences the response in the next section, where we focus on this distinction between nonreversed and reversed items.

2.3. Hypotheses development

We now integrate the findings from our qualitative study and the literature review to formulate hypotheses. To guide the subsequent discussion, we consider any two items i and i′ and their correlation in different situations. Item i is always positioned before item i′ in the questionnaire. Subsequently, we discuss the proximity effect for nonreversed and reversed item pairs.

2.3.1. Nonreversed items

If two nonreversed items are positioned close to each other, respondents are likely to perceive them as being related (Schwarz et al., 1991). Moreover, respondents usually still have access to item i and their response to it in their short-term memory when responding to i′. The chance that this will be the case decreases as a function of the distance between i and i′ (Feldman & Lynch, 1988).

Overall, we expect that respondents will be able to retrieve a previously encountered item from short-term memory only for a limited time. How many items it takes for the item (and the related response) to become unavailable in short-term memory is an empirical question that we will address based on the data from our quantitative study, although the exact distance may vary across situations and questionnaires[8].

When respondents identify items as highly similar, they are unlikely to retrieve any additional beliefs for two reasons: (1) Respondents believe that the item is redundant (given its high similarity) and do not invest the effort of retrieving new information (as a way of coping with cognitive demands; Krosnick, 1991), or (2) if retrieval is reinitiated, it tends to activate the same beliefs because of the beliefs' high accessibility (Tourangeau et al., 2000). Respondents who answered an item they perceived as repetitive indicated this as follows: "Same line of reasoning"; "As I just said, I agree"; "Ha, that's actually the same question, so—well—same response too"; and "Well, let me think.… Yes, this refers to the same example."

Once the surface memory (for the item and the related beliefs) fades, only deep structure memory remains. This means that respondents still hold a generic memory of the general topic of the item and some of the related beliefs they retrieved. Consequently, when a subsequent item is encountered on a related topic, it is likely that the belief sample generated in response will overlap without being identical. Respondents referred to this as follows: "Similar question as somewhere in the beginning, but I don't remember what I said exactly back then. So, er, let's think again."

[8] Generally speaking, it is unlikely that respondents still hold a verbal memory of the item and their response to it in their short-term memory if 5 to 9 items or more have been positioned in between, because in general, the capacity of short-term memory is estimated to be 7 ± 2 chunks of information (Bavelier et al., 2006; Miller, 1956). In our questionnaire, the maximum page length was 8 items, so the page length almost coincides with the maximal memory span and with the suggested range in which the breakpoint is expected to lie. We thank an anonymous reviewer for drawing our attention to this fact.

Therefore, consistency in responses to nonreversed items will decrease with increasing interitem distance, and the decrease will be strongest in the short-distance range. Because consistent responses to nonreversed items result in higher correlations, we posit the following:

H1: The positive correlation between (nonreversed) items measuring the same construct (a) decreases as a function of the items' distance in the questionnaire in the short-distance range and (b) remains stable afterward (i.e., in the long-distance range).

As indicated, the breakpoint between short-distance and long-distance range will be determined based on the data from the quantitative study and need not be generalizable to other settings (depending on page length and item difficulty, among other things).

2.3.2. Reversed items

Reversed items have been argued to not necessarily tap exactly the same content as their nonreversed counterparts (Barnette, 2000; Marsh, 1996; Schriesheim & Eisenbach, 1995; Watson, 1988). We argue that the extent to which reversed and nonreversed items are seen as overlapping depends on the distance between the two items. The reason is that respondents look at the way items are positioned in the questionnaire to interpret how the items relate conceptually (Ostrom, Betz, & Skowronski 1992).

Imagine a situation in which i and i′ are reversed items[9] that are positioned close to each other. Whereas respondents tend to view items that are related and worded in the same direction as highly redundant, this appears not to be the case for reverse-worded items. In

[9]Because the scaling of a latent construct is essentially arbitrary, we consider the attribute of being reversed as a characteristic of an item–item pair rather than an item–construct pair. As McPherson and Mohr (2005, p. 129) note, " the keying direction of an item is entirely relative to the definition of the construct of interest:

particular, respondents seem to perceive i′ in such a situation as contrasting with i and, thus, as inviting them to provide additional information (Schwarz et al., 1991). So, whereas the beliefs related to i are still highly accessible when responding to i′, respondents tend to decide against voicing the same beliefs when answering i′. In our qualitative research, we often observed this phenomenon, as illustrated by the following respondent, who gave a positive answer to both items in a reversed item pair measuring innovativeness (respectively, ”In general, I am among the first to buy new products when they appear on the market,” and “If I like a brand, I rarely switch from it just to try something new”): “Well, yes, actually, [hesitates] on the other hand, I’m also someone who values certainty, in a way.” When responding to reversed items right after encountering a nonreversed counterpart, respondents commonly retrieve beliefs that might be considered counterevidence to a previous response.

Again, this effect depends on (1) the verbal accessibility of item i (either in memory or visually) to notice the deep structure similarity but surface dissimilarity (the latter cueing the respondent to come up with new information) and (2) the perception that the two items are part of the same set of items and thus are intended to be related (Schwarz et al., 1991). Therefore, this effect will strongly decrease in the short-distance interitem range and remain absent afterward.

For reversed item pairs that were positioned far apart in the questionnaire, many respondents commented on the similarity of the second item with the previously encountered question. Remarkably, however, if interitem distance was high, there were no indications that respondents were aware of the reversed nature of the second item. In other words, respondents noticed the similarity in content but did not notice the dissimilarity in

---

For example, positively keyed items from a depression scale may resemble negatively keyed items from a happiness scale.”

form (i.e., with higher interitem distance, the reactions would be identical to reactions to nonreversed items; e.g., “Yes, same question as before”). Thus, when responding to reversed items at greater interitem distances, the contrast effect disappears, and respondents tend to retrieve belief samples that overlap with the previously encountered nonreversed item.

In summary, when responding to reversed items that are close to each other, belief samples will overlap, but additional counterevidence will be retrieved as a result of the perceived contrasting effect of the reversed item. Thus, reversed item pair correlations will be weak (i.e., close to zero) in such cases. This effect fades over short interitem distances, after which it remains absent.

H2: The correlation between a pair of reversed items (a) becomes more strongly negative with increasing interitem distance in the short-distance range, and (b) remains stable with increasing interitem distance in the long-distance range.

3. QUANTITATIVE STUDY

We now present quantitative results that (1) test our hypotheses and (2) clarify the implications of the model estimates for factor analysis. We realize the former objective by conducting an empirical study and the latter by conducting a simulation study.

In this section, we subject the hypotheses to empirical testing by means of regression analysis. We specify a regression model that explains interitem correlations as a function of distance and relationship. We discuss the model and the variables, the respondent sample, and the results of the quantitative study, as well as the main implications. Afterward, we assess how our findings are related to confirmatory factor analysis (CFA) based on two simulated correlation matrices.

3.1. Empirical study: regression analysis

3.1.1. Regression model specification

We translated the hypotheses into the following regression equation:

$$r_{ij} = \beta_0 + \beta_1 \times DIST^{short}_{ij} + \beta_2 \times DIST^{long}_{ij} + \beta_3 \times NONREVERSED_{ij} + \beta_4 \times REVERSED_{ij} + \beta_5 \times DIST^{short}_{ij} \times NONREVERSED_{ij} + \beta_6 \times DIST^{long}_{ij} \times NONREVERSED_{ij} + \beta_7 \times DIST^{short}_{ij} \times REVERSED_{ij} + \beta_8 \times DIST^{long}_{ij} \times REVERSED_{ij} + \varepsilon_{ij}, \quad (1)$$

where r_{ij} is the correlation between item i and j.

For the independent variables, we created a distance variable by coding the number of intervening items between the two focal items in the correlations. To differentiate the effect of distance in the short-distance range from the effect in the higher distance range, we used a spline regression specification. The breakpoint will be determined based on the data, as we discuss in the results section.

First, the initial distance measure is centered around the breakpoint (i.e., the breakpoint is subtracted from the initial distance variable). Second, we created a separate distance variable for all values below the breakpoint and one for all values above the breakpoint. These variables are $DIST^{short}_{ij}$ and $DIST^{long}_{ij}$, respectively. For $DIST^{short}_{ij}$, values equal to or above the breakpoint are set to zero, for $DIST^{long}_{ij}$ values equal to or below the breakpoint are set to zero. Thus, the intercept β_0 corresponds to the expected interitem correlation for two items at a distance equal to the breakpoint (we explain why this intercept term is expected to be positive when discussing the main effects).

Furthermore, we created two dummy variables. The first dummy variable marks item pairs assumed to tap a same latent construct in the same direction (NONREVERSED), and the second dummy variable flags item pairs that tap a same latent construct in the reverse direction (REVERSED). These dummy variables serve as an additional intercept term for nonreversed and reversed item correlations respectively (i.e., this term has to be added to the overall intercept to obtain the intercept for nonreversed or reversed item correlations). Because we hypothesize a differential effect of distance for reversed and nonreversed items (H1 and H2), we include the interactions of the NONREVERSED and REVERSED dummies with the distance measures $DIST^{short}_{ij}$ and $DIST^{long}_{ij}$.

Whereas our conceptualization of the proximity effect for nonreversed and reversed item pairs posits a content-driven response process, directional response bias (due to individuals response styles) has been shown to bias correlations (Paulhus, 1991; Fischer, 2004; Greenleaf, 1992; Billiet & McClendon, 2000).

Because directional bias is a source of common variance in measures, it leads to a general spurious increase in observed correlations (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Thus, the net result is that the correlation between any two random items is expected to take on a positive value rather than zero (Green, Goldman, & Salovey, 1993; Paulhus, 1991).

As Hui and Triandis (1985) show, nearby items share more directional response style bias than items that are further apart. Hui and Triandis's further suggest that correlations due to directional response style bias tend to decrease over interitem distance at a decreasing rate. The main effect of $DIST^{short}_{ij}$ and $DIST^{long}_{ij}$ on r corresponds to the notion that nearby items may share more common directional response style bias than items that are further apart and that the rate of the decline is different in the short-distance and the long-distance range. Finally, the disturbance term ($\square_{ij}$) captures the variance in interitem correlations that has not been accounted for by the preceding variables, including specifics in content and/or form shared by an item pair. Some main descriptive statistics for the independent and dependent variables appear in Table 1.

As for the dependent variable, we used observed interitem Pearson correlations as the data points in the analysis. Thus, we computed Pearson correlations between all 76 items. To account for missing data (all item pairs had at least 3,000 valid observations), we estimated the correlation matrix by means of the E-M (expectation–maximization) algorithm in NORM (Schafer, 1999), using age, sex, and education level as covariates for estimating the correlation matrix (Schafer & Graham, 2002).

Of 2,850 correlations, 29 were based on reversed item pairs, and 71 were based on nonreversed item pairs.[10] The other correlations were based on items that had no conceptual relationship to each other. The absence of a conceptual relation between the latter items was confirmed by the low average inter-item correlation (see Table 1). Furthermore, an investigation of the scree plot indicated the presence of one high eigenvalue (supporting the expected common variance due to directional bias) followed by a gradually declining series of eigenvalues (the first ten eigenvalues were 3.02, 1.78, 1.69, 1.51, 1.28, 1.21, 1.14, 1.07, 1.00, 0.99).

We regressed the observed correlations on the independent variables that reflect questionnaire design and content factors that varied across the item pairs under study. Studying correlations as the dependent variable was relevant because studies in the domain of reversed items have focused on interitem correlations (e.g., Wong et al., 2003), or methods based on correlations (e.g., Billiet & McClendon, 2000), because interitem correlations capture the variance shared by the items and indicate both the strength and the

[10] The apparent imbalance of conceptually unrelated items to conceptually related items is not problematic. As we clarify subsequently, we estimated separate effects for the different categories of items. Consequently, all estimates had their own appropriate standard errors. At the same time, we controlled for directional response style bias in a highly reliable way, based on the many unrelated item pair correlations, such that the main effect of directional bias and its effect moderated by distance could be assessed independently of the reversal main and interaction effects. Furthermore, the correlations were based on a large number of respondents (>3,000), which enhanced their stability and reliability (Zimmerman, Zumbo, & Williams 2003), and the items were randomly assigned to positions in the questionnaire. These factors made it possible not to include extraordinarily large numbers of reversed items in the questionnaire, which might have led to demand effects (respondents perceiving the task as a reversed-item examination).

direction of their association. For these reasons, the unit of analysis was not the respondent but rather the interitem correlation (computed across respondents).

3.1.2. Data

We took a sample from the general online population by recruiting respondents on multiple major Belgian portal Web sites in the Dutch language[11]. We collected data by means of an online questionnaire that did not allow respondents to scroll back to previous pages. The Web pages of the questionnaire each contained 8 items (with the exception of the last page, which had only 4 items). Respondents were told that the online survey was part of an academic study that measured the opinions of the population with regard to a variety of issues. We obtained 3,114 valid responses. In this sample, 51.6% of the respondents were male, and 48.4% were female; 37.9% had a higher education (i.e., formal education after secondary school), and 62.1% did not have a higher education; and 17.4% were between the ages of 16 and 25, 21.1% between the ages of 25 and 34, 25.9% were between the ages of 35 and 44, 20.4% between the ages of 45 and 55, 11.2% were between the ages 55 and 64, and 4.0% were age 65 or older.

3.1.3. Discussion of the results of the regression analysis

We now discuss the main results of the regression analysis. In the "General Discussion" section, we provide a more extended discussion of theoretical and practical implications. As a first step, we ran the regression model using different break points for the spline regression. Specifically, we varied the breakpoint between 3 and 11 (a range that is two units below and above the theoretically expected range of 7 +/- 2; see footnote 4). We

[11] Due to the confidentiality of some Web site traffic information, we unfortunately were unable to establish the response rate.

compared the resulting R^2 values to determine the optimum point where the model captures the maximum amount of variance in the dependent variable. R^2 reached its optimum when the breakpoint equaled 6.

With an R^2 of 0.47, the regression model explained a sizeable proportion of the variance in the observed correlations. The key assumptions of the multiple linear regression model were met: (1) All condition indexes were below 7, indicating that there was no problem of collinearity, and (2) the standardized residuals showed approximately normal distributions (as revealed on plots of the regression standardized residuals).

Table 2 lists the results of the regression analysis. Figure 1 is based on the estimates in Table 2 and depicts the regression-implied predicted values of interitem correlations over interitem distance.

Insert Table 2 about here

Insert Figure 1 about here

As Table 2 and Figure 1 show, two random, unrelated items are expected to have a correlation of 0.04 when they are positioned 6 items apart. The correlation is even higher if the items are positioned right next to each other ($r = 0.09$). In other words, the expected correlation reduces to half merely by positioning two items only a few items further apart. Whereas the interitem distance also has a significant impact in the range above 6, its effect becomes small; for example, adding another 6 intervening items would only reduce the correlation from 0.038 to 0.036. So, whereas the linear effect is still significant in the inter-item distance range above 6, its impact is no longer meaningful for inter-item distances commonly encountered in practice.

For nonreversed item pairs, with increasing interitem distances, the expected correlation significantly and substantially decreases. When positioned next to each other, the expected correlation between a pair of nonreversed items is 0.62 (see Figure 1). When 6 items or more are positioned in between, the correlation drops to 0.35. After the sharp initial decline, the distance effect becomes nonsignificantly negative.

When positioned next to each other, a reversed item pair has an expected correlation of –0.06. This drops to –0.26 if 6 or more items are positioned in between (see Figure 1). Thus, for nonreversed items, the correlation becomes weaker, whereas for reversed items, it becomes stronger the further the items are positioned from each other (in the low distance range). Remarkably, in the long-distance range, a nonreversed correlation is typically very close in size (absolute value) to a reversed correlation (0.32 and –0.28 on average respectively).

To conclude, the results of the regression analysis support the hypotheses. Positioning nonreversed items directly next to each other in a questionnaire leads to a strong increase in the expected correlation. In other words, there is a reinforcing proximity effect for nonreversed items (H1). The effect of distance becomes nonsignificant in the distance range beyond 6.

On the contrary, for reversed items, placing the items next to each other leads them to be weakly correlated (i.e., the correlation approaches zero). In other words, there is a weakening proximity effect for reversed items (H2). Here, too, the effect of distance becomes nonsignificant in the distance range beyond 6.

Finally, the correlation between any two items in a questionnaire is expected to be significantly positive, especially if the items are positioned close to each other. This indicates that items share common variance as a result of directional bias and that this directional bias is locally stable (and evolves over the course of the questionnaire administration).

3.2. Simulation study: the impact of the proximity effect on factor structure

To assess the impact of the proximity effect on factor structure and composite reliability, we simulate correlation matrices for two scenarios that represent common measurement situations in marketing research. In both scenarios, imagine that a researcher measures two uncorrelated constructs with two four-item scales. One scale (Construct A) has three nonreversed items and one reversed item, and the other scale (Construct B) has two nonreversed and two reversed items. Construct A is measured by items A1, A2, A3, and A4r (r indicates that the item is reversed), and Construct B is measured by Items B1, B2r, B3, and B4r. The scales are included in a longer questionnaire, consisting of 32 items. In Scenario 1, the researcher positions the items at the beginning of the questionnaire, grouping them by construct; we refer to this as the traditional approach. In Scenario 2, the researcher positions the items dispersed throughout the questionnaire in such a way that the distance between items measuring the same construct is maximized. Thus, in Scenario 1 (the traditional approach), items A1, A2, A3, A4r, B1, B2r, B3, and B4r are placed in positions 1, 2, 3, 4, 5, 6, 7, and 8, respectively. In Scenario 2 (the dispersed approach), the items are placed in positions 1, 11, 21, 31, 2, 12, 22, and 32, respectively.

3.2.1. Method for the simulation study

We construct two correlation matrices by plugging in the correlation values that our regression model would predict on the basis of the items content relationship (unrelated/reversed/nonreversed) and their distance (ranging from 0 through 30). The resultant correlation matrices appear in Table 3.

Insert Table 3 about here.

Assuming a sample size of $N = 800$, we fit three alternative models to these two correlation matrices to assess the factor structure that results from the two different measurement situations. A first model, the base model, is a basic CFA in which Constructs A and B are freely correlating factors with four items each. In a second mdel, the response style factor model, we add a response style factor, as Billiet and McClendon (2000) propose (i.e., a factor on which all items have a positive unit loading). In a third model, the error covariance model, the reversed items' residuals are allowed to correlate freely with one another, as Marsh (1996) proposes.

For each model, we evaluate (1) model fit, (2) the estimated value and significance of the correlation between Construct A and Construct B (the true value of which is zero), (3) the average factor loadings for nonreversed and reversed items in the model, and (4) the composite reliabilities (C.R.) of Factor A and Factor B. We ran the CFAs using the maximum likelihood (ML) estimator in MPlus 5 (Muthén & Muthén, 2004), and the results appear in Table 4.

3.2.2. Discussion of the results of the simulation study

In the traditional approach, reversed items clearly lead to problems in the CFA. The base model shows unacceptable fit, low loadings for the reversed items, and spurious significance for the correlation between A and B ($r_{A,B}$). It does not seem possible to solve the issues caused by the presence of reversed items post hoc by means of modeling solutions. The modeling solution that Billiet and McClendon (2000) propose (i.e., including a response style factor on which both reversed and nonreversed items load) leads to some improvement, in that $r_{A,B}$ is no longer significant in this model and the reversed items' loadings are somewhat higher. However, the model does not provide optimal fit to the current data. The approach that Marsh (1996) advocates, in which residual terms of reversed items correlate freely, provides a good fit to the data but leads to biased estimates; that is, $r_{A,B}$ becomes significant again (whereas the true value is zero). The loadings of the reversed items are low, resulting also in low C.R. for Factor B.

Overall, the dispersed data lead to more valid inferences than the traditionally structured data. First, both the response style factor model and the correlated residuals model fit the data rather well. Second, regardless of the model specification, we observe no spurious correlation between Factors A and B; in none of the models does $r_{A,B}$ reach significance. Furthermore, regardless of the model specification, the factor loadings of the reversed and nonreversed items are all above 0.40, and the C.R. values for both factors are above 0.60. Although both the response style factor model (Billiet & McClendon, 2000) and the correlated residuals model (Marsh 1996) perform well in the dispersed condition, the response style factor model is preferable because (1) it fits the data slightly better, (2) it is particularly efficient in that it uses only one degree of freedom to capture the unwanted common variance in the items, and (3) the response style factor has a conceptual meaning that is more clearly defined than the residual correlations (for details, see Billiet & McClendon, 2000).

4. General discussion

In this final section, we summarize our main findings and relate them to the literature. On the basis of this discussion, we provide recommendations for questionnaire design. We conclude by pointing out some limitations of the current study that may offer routes for further research.

4.1. Summary of main findings

The current study explains correlations between items as a function of their content relationship and their proximity in the questionnaire. We provide evidence for a proximity effect that is different for items that are nonreversed, reversed, or unrelated.

4.1.1. The proximity effect for nonreversed item pairs

For nonreversed items, we find that proximity leads to higher positive correlations. This is in line with previous findings (Budd, 1987; McFarland et al., 2002), but our combined qualitative and quantitative approach enables us to understand more clearly the process behind this phenomenon and quantify the impact more precisely. In particular, based on the thought processes that respondents verbalized in cognitive interviews, these high correlations are not desirable, because they indicate redundancy in the information obtained. Specifically, when nonreversed items are positioned within close range of each other (in the current data this means fewer than 6 items in between them), respondents base their replies to the items on the same belief samples, and the resultant responses are rather thoughtless reiterations of information that was already provided (for a related critique, see Rossiter, 2002). This results in correlations of approximately 0.60 for adjacent nonreversed items (see Figure 1; Table 2). When nonreversed items are positioned at greater distances from each other (i.e., 6 items or more in between them), the beliefs typically retrieved by respondents are more diverse. This phenomenon results in slightly lower correlations (typically in the range 0.28 to 0.35; see Figure 1) but it also results in better coverage of the content domain of interest.

4.1.2. The proximity effect for reversed item pairs

Most important, for reversed items, we find that high proximity is problematic in that it results in weak negative correlations, evolving from -0.06 to -0.23 over the distance range of 0 through 6 (see Figure 1). As became apparent in our cognitive interviews, respondents tended to perceive reversed items as cues to provide information that had not yet been provided. Consequently, as a reaction, respondents tended to retrieve new beliefs and discount the previously reported beliefs when they came up with a response. This resulted in upwardly biased correlations (i.e., biased toward zero and, thus, weaker). In the close-distance range (0 through 6 intervening items), this effect rapidly diminished because (1) the items came to be perceived as belonging less to the same context/item group and (2) the first item became less accessible in short-term memory (and the visual field if the second item was on the next page of the questionnaire). Typically, respondents did not recall the polarity of previously encountered items when answering items on the same topic that were positioned farther away (i.e., > 6 items in between). Accordingly, respondents no longer perceived reversed items at these distances as providing a cue to retrieve new information and discount previous information. This results in overlapping belief samples and stronger (negative) correlations. Specifically, in the distance range of 6 through 76, correlation remained largely stable at approximately -0.26 through -0.31. This implies that at greater interitem distances, correlation strengths of reversed item pairs are comparable to nonreversed item pairs.

A major contribution to the literature is our conceptualization and empirical validation of reversed-item bias as a proximity effect. Recent theorizing about reversed items has contributed greatly to the understanding of reversed-item responding (Swain et al., 2008; Wong et al., 2003) but has largely ignored this aspect. Both Wong et al. (2003) and Swain et al. (2008) provide clear evidence that the way items are formulated and formatted is of crucial importance in optimizing item validity. In particular, Swain et al. (2008) demonstrate that negated items lead to item verification difficulty. We further broaden the scope from item formulation to questionnaire design. In a conceptualization that is closer to ours, Wong et al. (2003) also suggest that reversed items are not necessarily perceived as opposite poles of a one-dimensional continuum. Our results clarify that item positioning within a questionnaire can directly affect the extent to which respondents retrieve beliefs that are purposefully retrieved and filtered to be nonredundant. Consequently they may be more likely to violate the conceptualization of constructs as single bipolar dimensions. Thus, we extend the scope from item to item positioning within the questionnaire. This contribution is particularly useful given the scarcity of advice on how to design questionnaires, especially with regard to whether to group or disperse items that measure the same construct. Such conditions are both easy to manipulate and rather impactful, making them ideal tools for measurement optimization.

4.1.3. The proximity effect for unrelated item pairs

A final finding was that, in line with previous research, unrelated items were positively correlated (Mirowsky & Ross, 1991; Paulhus, 1991). However, our study adds an essential qualifying condition to this generalization: The positive, spurious correlation drops sharply as a function of interitem distance in the range of 0 through 6 intervening items and keeps on decreasing more gradually (i.e., significantly but nonsubstantially) over greater interitem distances. Specifically, for adjacent items, a baseline correlation of approximately 0.09 is expected, dropping to 0.04 at an interitem distance of 6 and further to 0.02 at a distance of 75. Whereas the size of the baseline correlation is not that large, a correlation of 0.09 is worrisome in light of the range of effect sizes of correlations and regressions commonly reported in social sciences (Green, 1991). Directional response style bias can lead to the overestimation of internal consistency of scales (Green & Hershberger, 2000) and relationships between scales (Podsakoff et al., 2003). The current results highlight that this problem should not be neglected and that researchers should account for this bias in their analyses (as detailed below). However, we provide clear evidence that directional response style bias significantly contributes to reversed-item bias but that it is not the main contributor. As for the former point, this is indicated by the significant, positive bias in correlations between items that do not share a common construct. In contrast to this observation, Swain et al. (2008) downplay the relevance of directional response style bias. However, note that their data are collected from student samples, who may be less prone to stylistic responding (Greenleaf 1992; Mirowsky & Ross 1991). Note also that we find clear evidence that directional bias is local in its effect and dissipates with interitem distance. Although Hui and Triandis (1985) pointed this out more than two decades ago, their observation has not been translated in the way directional response style bias is usually measured. Response style research might benefit from measurement and correction procedures that account for the instability of directional response styles.

4.1.4. Domain-sampling versus belief-sampling models

Item construction and questionnaire design decisions in marketing are often guided by the domain-sampling model (Churchill, 1979), which essentially views items as mutually interchangeable elements sampled from a broader universe of items that measure the same content domain. As our results show, the domain-sampling model is limited in scope, in that it does not account for the influence of items on each other. As an undesirable side effect of its dominance in the marketing research field, the focus on internal consistency may motivate researchers to use items that repeatedly tap the same belief sample by using similar items that are positioned close to each other. We favor the view that the marketing research field must move away from the uncritical maximization of the alpha coefficient, often at the expense of content validity (for a broader perspective on this issue, see Rossiter, 2002). This brings us to some recommendations on item construction, questionnaire design, and data analysis.

4.1.5. Recommendations

On the basis of the current qualitative and quantitative research, we advise researchers (1) to use reversed items and (2) to disperse items that measure the same construct (reversed or nonreversed) over the questionnaire. We recommend the use of reversed items because it leads respondents to consider a wider variety of relevant beliefs in responding to a questionnaire. This results in information that is less specific to the associations of a respondent at one point in time in response to one particular item (after which the information would be largely repeated in response to adjacent related items).

Moreover, researchers should try to position same-construct items apart from each other, preferably by positioning 6 or more other-construct items in between them. Otherwise, item response to subsequent items will be either artificially consistent (for nonreversed items) or artificially inconsistent (for reversed items). Different concrete scenarios are conceivable. For example, when measuring two constructs with four items each, researchers could use the approach we described in Scenario 2 in the simulation study. More generally, items can be separated by means of other-construct items in the same or other formats (including, for example, demographic information) or by means of dedicated buffer items (Wänke & Schwarz, 1997). If resource or other constraints make it impossible to use interitem distances of 6 items or more, it should be kept in mind that even inserting only a few items (e.g., 4) has a relevant effect (see Figure 1).

In addition, special measures must be taken in the analysis of the data. Researchers should be aware that the approach we advocate may result in lower factor loadings and composite reliabilities. Nonetheless, we believe that this is a price worth paying for improved content validity. The CFA model should account for sources of covariance other than the intended factors. A model that seems to perform particularly well in the dispersed condition with reversed items is Billiet and McClendon's (2000) response style factor model. By estimating an additional factor variance and fixing loadings for all items to this factor at one, we invest only one degree of freedom, and the results are optimized in several ways: (1) Model fit improves substantially, (2) factor loadings of reversed items become more comparable in strength to those of nonreversed items, and (3) factor correlation bias is reduced.

4.2 Limitations and directions for further research

Whereas the current study contributes to knowledge on how people respond to questionnaires and how that knowledge can be used, it also has limitations that may open up some routes for further research. The quantitative study used data that were collected online. This raises the question whether results can be generalized to other methods of data collection. Given their similarity in encoding and recall, it seems safe to assume that the current findings apply to questionnaire data collection via the visual mode (paper-and-pencil and online). Note that we deliberately limited page length to a maximum of 8 items per pages in both data collection methods. In other settings, differences in page length might lead to differences in the proximity effect because page breaks provide cues to respondents on whether items belong together and because items on a previous page cannot be as easily looked up (none of the respondents in our cognitive interviews looked up items on a previous page).

Related to the data collection method, the current study used seven-point rating scales. The specific rating scale format used in a questionnaire has been found to lead to interitem relationships that are not identical but are highly similar (Ferrando, 2000). Therefore, it is likely that our results will generalize to other scale formats.

Another question of generalizability pertains to cultural differences. In particular, it would be worthwhile investigating how the current results, which were obtained from a European sample (more specifically, a Belgian sample), would generalize to other cultures because cultural differences have been found in the way respondents process reversed items (Wong et al., 2003).

Finally, our conceptualization is complementary to the theories that Wong et al. (2003) and Swain et al. (2008) put forward. An important challenge for further research lies in formulating an integrated framework on reversed-item bias that simultaneously accounts for respondent characteristics, item characteristics, questionnaire design factors, and their interaction effects.

References

Barnette, J. Jackson (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*, 361–370.

Baumgartner, Hans, & Steenkamp, Jan-Benedict E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(May), 143–156.

Bavelier, Daphne, Newport, Elissa L., Hall, Matthew L., Supalla, Ted, & Boutla, Mrim (2006). Persistent difference in short-term memory span between sign and speech: Implications for cross-linguistic comparisons. *Psychological Science*, *17*(12), 1090–1092.

Bentler, Peter M. (1969). Semantic space is (approximately) bipolar. *Journal of Psychology*, *71*, 33–40.

Billiet, Jaak B., & McClendon, McKee J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, *7*(4), 608–628.

Bradlow, Eric T., & Fitzsimons, Gavan J. (2001). Subscale distance and item clustering effects in self-administered surveys: A new metric. *Journal of Marketing Research*, *38*(May), 254–261.

Bruner, Gordon C., James, Karen E., & Hensel, Paul J. (2001). *Marketing scales handbook: A compilation of multi-item measures* (Vol. 3).Chicago: American Marketing Association.

Budd, Richard J. (1987). Response bias and the theory of reasoned action. *Social Cognition*, *5*(1), 95–107.

Carroll, James M., Yik, Michelle S. M., Russell, James A., & Barrett, Lisa Feldmann (1999). On the psychometric principles of affect. *Review of General Psychology*, *3*(1), 14–22.

Churchill, Gilbert A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64–73.

DeMaio, Theresa J., & Rothgeb, Jennifer M. (1996). Cognitive interviewing techniques: In the lab and in the field. In Norbert Schwarz & Seymour Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 177–195). San Francisco: Jossey-Bass.

Drolet, Aimee, & Morrison, Donald G. (2001). Do we really need multiple-item measures in service research? *Journal of Service Research*, *3*(3), 196–204.

Ericsson, K. Anders, & Simon, Herbert A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251.

Feldman, Jack M., & Lynch, John G., Jr. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, *73*(3), 421–435.

Ferrando, Pere J. (2000). Testing the equivalence among different item response formats in personality measurement: A structural equation modeling approach. *Structural Equation Modeling*, *7*(2), 271–286.

Fischer, Ronald (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in *JCCP*. *Journal of Cross-cultural Psychology*, *35*(3), 263–282.

Green, Samuel B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, *26*(3), 499–510.

Green, Donald Philip, Goldman, Susan Lee, & Salovey, Peter (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, *64*(6), 1029–1041.

Green, Samuel B., & Hershberger, Scott L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, *7*(2), 251–270.

Greenleaf, Eric A. (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(2), 176–188.

Herche, Joel, & Engelland, Brian (1996). Reversed-polarity items and scale unidimensionality. *Journal of the Academy of Marketing Science*, *24*(Fall), 366–374.

Hui, C. Harry, & Triandis, Harry C. (1985). The instability of response sets. *Public Opinion Quarterly*, *49*, 253–260.

Jobe, Jared B., & Mingay, David J. (1989). Cognitive research improves questionnaires. *American Journal of Public Health*, *79*, 1053–1055.

Knowles, Eric S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, *55*(2), 312–320.

Knowles, Eric S., Coker, Michelle C., Cook, Deborah A., Diercks, Steven R., Irwin, Mary E., Lundeen, Edward J., Neville, John W., & Sibicky, Mark E. (1992). Order effects within personality measures. In Norbert Schwarz & Seymour Sudman (Eds.), *Context effects in social and psychological research* (pp. 221-236). New York: Springer-Verlag.

Krosnick, Jon A. (1991). Response strategies for coping with the cognitive demands of measures in surveys. *Applied Cognitive Psychology, 5*, 213-236.

Likert, Rensis (1932). A technique for the measurement of attitudes. *Archives of*

Psychology, *140*, 44–53.

Marsh, Herbert W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*(4), 810–819.

McFarland, Lynn A., Ryan, Ann Marie, & Ellis, Aleksander (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment*, *78*(2), 348–369.

McPherson, Jason, & Mohr, Philip (2005). The role of item extremity in the emergence of keying-related factors: An exploration with the life orientation test. *Psychological Methods*, *10*(1), 120–131.

Miller, George A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Mirowsky, John, & Ross, Catherine E. (1991). Eliminating defense and agreement bias from measures of the sense of control: A 2x2 index. *Social Psychology Quarterly*, *54*(2), 127–145.

Motl, Robert W., & DiStefano, Christine (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling*, *9*(4), 562–578.

Muthén, Linda K., & Muthén, Bengt O. (2004). *Mplus: Statistical analysis with latent variables: User guide* (3rd ed). Los Angeles: Muthén & Muthén.

Ostrom, Thomas M., Betz, Andrew L., & Skowronski, John J. (1992). Cognitive representation of bipolar survey items. In Norbert Schwarz & Seymour Sudman (Eds.), *Context effects in social and psychological research* (pp. 297-311). New York: Springer-Verlag.

Paulhus, Delroy L. (1991). Measurement and control of response bias. In John P. Robinson, Phillip R. Shaver, & Lawrence S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego: Academic Press.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Quilty, Lena C., Oakman, Jonathan M., & Risko, Evan (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling*, *13*(1), 99–117.

Rossiter, John R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, *19*(4), 305–335.

Schafer, Joseph L. (1999). NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2 (software for Windows 95/98/NT). Retrieved from http://www.stat.psu.edu/~jls/misoftwa.html

Schafer, Joseph L., & Graham, John W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

Schmitt, Neal, & Stults, Daniel M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, *9*, 367–373.

Schwarz, Norbert (1997). Questionnaire design: The rocky road from concepts to answers. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 29–45). New York: Wiley.

Schwarz, Norbert, Strack, Fritz, & Mai, Hans-Peter (1991). Assimilation and contrast effects in part–whole question sequences: A conversational logic analysis. *The*

Public Opinion Quarterly, *55*(1), 3–23.

Sirdeshmukh, Deepak, Singh, Jagdip, & Sabol, Barry (2002). Consumer trust, value, and loyalty in relational exchanges. *Journal of Marketing*, *66*(1), 15–37.

Steenkamp, Jan-Benedict E. M., & Gielens, Katrijn (2003). Consumer and market drivers of the trial probability of new consumer packaged goods. *Journal of Consumer Research*, *30*(4), 368–384.

Swain, Scott D., Weathers, Danny, & Niedrich, Ronald W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*(February), 116–131.

Tourangeau, Roger, Rasinski, Kenneth A., Bradburn, Norman, & D'Andrade, Roy (1989). Carryover effects in attitude surveys. *Public Opinion Quarterly*, *53*, 495–524.

Tourangeau, Roger, Rips, Lance J., & Rasinski, Kenneth (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

Tourangeau, Roger, Singer, Eleanor, & Presser, Stanley (2003). Context effects in attitude surveys: Effects on remote items and impact on predictive validity. *Sociological Methods and Research*, *31*(4), 486–513.

Wänke, Michaela, & Schwarz, Norbert (1997). Reducing question order effects: The operation of buffer items, In L. Lyberg, P. Biemer, M.Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds), *Survey measurement and process quality* (pp. 115–139). New York: Wiley.

Wong, Nancy, Rindfleisch, Aric, & Burroughs, James E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of Consumer Research*, 3(1), 72–91.

Zimmerman, Donald W., Zumbo, Bruno D., & Williams, Richard H. (2003). Bias in estimation and hypothesis testing of correlation. *Psicologica*, 24, 133–158.

Table 1

Main descriptive statistics and correlation matrix of the regression variables (*N* = 2,850)

						Correlations			
	Scale/Coding	Minimum	Maximum	Mean or proportion	s	r_{ij}	Distance	NR	R
r_{ij}	Pearson correlation	–0.56	0.78	0.04	0.11	1.00	–0.16	0.59	–0.27
Interitem distance	Number of intervening items	0.00	074	24.67	17.79	–0.16	1.00	–0.08	0.00
NONREVERSED (NR)	1 = Nonreversed item pair 0 = other			0.03	0.16	0.59	–0.08	1.00	–0.02
REVERSED (R)	1 = Reversed item pair 0 = Other			0.01	0.10	–0.28	0.00	–0.02	1.00

Table 2

Results of regression analysis

$R^2 = .47$	B [a]	SE	t	p
Intercept	0.038	0.003	14.29	0.000
$DIST^{<6}_{ij}$	-0.008	0.001	-7.04	0.000
$DIST^{>6}_{ij}$	0.000	0.000	-2.82	0.005
NONREVERSED	0.312	0.022	14.39	0.000
REVERSED	-0.294	0.027	-10.77	0.000
$DIST^{<6}_{ij} \times NONREVERSED_{ij}$	-0.036	0.005	-6.60	0.000
$DIST^{>6}_{ij} \times NONREVERSED_{ij}$	-0.001	0.001	-0.86	0.388
$DIST^{<6}_{ij} \times REVERSED_{ij}$	-0.024	0.010	-2.30	0.022
$DIST^{>6}_{ij} \times REVERSED_{ij}$	-0.001	0.001	-0.53	0.595

Dependent variable: Pearson correlation

[a] Unstandardized regression weights

Table 3

Simulated correlation matrices for traditional and dispersed item positioning

	Scenario 1: Traditional								Scenario 2: Dispersed							
	A1	A2	A3	A4r	B1	B2r	B3	B4r	A1	A2	A3	A4r	B1	B2r	B3	B4r
A1	1.00								1.00							
A2	0.62	1.00							0.35	1.00						
A3	0.57	0.62	1.00						0.34	0.35	1.00					
A4r	-0.13	-0.10	-0.06	1.00					-0.27	-0.27	-0.26	1.00				
B1	0.06	0.07	0.08	0.09	1.00				0.09	0.04	0.03	0.03	1.00			
B2r	0.05	0.06	0.07	0.08	-0.06	1.00			0.04	0.09	0.04	0.03	-0.26	1.00		
B3	0.05	0.05	0.06	0.07	0.57	-0.06	1.00		0.03	0.04	0.09	0.04	0.34	-0.27	1.00	
B4r	0.04	0.05	0.05	0.06	-0.13	0.57	-0.06	1.00	0.03	0.03	0.04	0.09	-0.27	0.34	-0.26	1.00

Table 4

Model fit for the simulated correlation matrices

	χ^2	d.f.	p	CFI	TLI	RMSEA	$r_{A,B}$	$p(r_{A,B} = 0)$	λ_r	λ_{nr}	C.R.$_A$	C.R.$_B$
Traditional item positioning												
Base model	340.66	19	< 0.001	0.79	0.69	0.15	0.10	0.03	0.77	-0.12	0.74	0.53
Response style factor model	152.87	18	< 0.001	0.91	0.86	0.10	0.01	0.87	0.66	-0.43	0.69	0.65
Correlated reversed item residuals	21.86	16	0.148	1.00	0.99	0.02	0.08	0.04	0.79	-0.11	0.74	0.56
Dispersed item positioning												
Base model	64.06	19	< 0.001	0.94	0.91	0.05	0.02	0.74	0.57	-0.51	0.64	0.62
Response style factor model	12.80	18	0.803	1.00	1.01	0.00	0.02	0.69	0.55	-0.56	0.64	0.64
Correlated reversed item residuals	32.43	16	0.009	0.98	0.96	0.04	0.02	0.68	0.59	-0.47	0.65	0.60

λ_{nr} refers to the average loading of the nonreversed items in the model, and λ_r refers to the average loading of the reversed items in the model.

Appendix 1: Items

Scale	Position	Item
Buffer items	1	On a free evening I like to see a nice movie.
	5	I am a sensitive person.
	15	School children should have plenty of discipline.
	16	Communication is very important in a relationship.
	17	I prefer to avoid taking extreme opinions.
	18	I like to collect things.
	19	I am very curious about how things work.
	20	The clothing I buy gives a glimpse of the type of person I am.
	22	A mother working out of home is still a good mother.
	24	I like to speed in my car.
	27	I would forgive my family for practically anything.
	28	I enjoy clipping coupons out of the newspaper.
	30	I am confident that I can learn technology related skills.
	35	Feelings are more important than facts.
	36	I like to be in charge.
	48	I consider myself a brand loyal consumer.
	49	I find it very important to organize my grocery shopping well.
	51	Air pollution is an important global problem.
	52	I don't buy products that have excessive packaging.
	53	We experience a drop in living quality.
	54	Television is my primary form of entertainment.
	55	I am an animal lover.
	56	To me, it's very important what my friends think of me.
	69	I shop at more than one supermarket.
	70	I am very concerned about my health.
	72	There should be a gun in every house.
	73	Human contact in service encounters makes the process enjoyable for the consumer.
	74	Before buying a product, I check the price.
Pair 1	2	Most product advertising is believable.
	21	I often feel misled by advertising.
Pair 2	7	The work I do is useless.
	75	The work I do is valuable.
Pair 3	11	I find it difficult to compliment or praise others.
	46	I often give compliments to people.
Pair 4	12	My family is selfish.
	29	My family is very social.
Pair 5	26	Most products I buy are overpriced.
	65	In general, I am satisfied with the prices I pay.
Pair 6	31	I am satisfied with my income.
	68	I think that I deserve a higher income.
Pair 7	33	I am good at negotiating.
	63	I do not posses the skills that are required to be a good negotiator.
Pair 8	34	Poor people deserve our sympathy and support.
	66	It is a waste of time feeling sorry for the poor.
Pair 9	37	I often am the center of attention.
	61	In a group of people I am rarely the center of attention.
Pair 10	47	I seldom am under time pressure.
	58	I have the impression of continuously being under time pressure.
Scale 1	38	I feel that the employees of this store are very dependable.
	39	I feel that the employees of this store are very competent.

	40	I feel that the employees of this store are of very high integrity.
	41	I feel that the employees of this store are very responsive to customers.
	42	It is very likely that I will frequent this store in the future.
	43	It is very likely that I will recommend this store to friends, neighbors and family.
	44	It is very likely that I will go to this store the next time I need clothes.
	45	It is very likely that I will spend more than 50% of my clothing budget in this store.
Scale 2	62	I like introducing new brands and products to my friends.
	67	I don’t talk to friends about the products that I buy.
	71	My friends and neighbors often come to me for advice.
	76	People seldom ask me for my opinion about new products.
Scale 3	3	In general, I am among the first to buy new products when they appear on the market.
	4	If I like a brand, I rarely switch from it just to try something new.
	6	I am very cautious in trying new and different products.
	23	I am usually among the first to try new brands.
	25	I rarely buy brands about which I am uncertain how they will perform.
	50	I do not like to buy a new product before other people do.
	57	I enjoy taking chances in buying new products.
	64	When I see a new product on the shelf, I’m reluctant to give it a try.
Scale 4	8	If I want to be like someone, I often try to buy the same brands that they buy.
	9	If other people can see me using a product, I often purchase the brand they expect me to buy.
	10	When buying products, I generally purchase those brands that I think others will approve of.
	13	I achieve a sense of belonging by purchasing the same products and brands that others purchase.
	14	I like to know what brands and products make good impressions on others.
	32	It is important that others like the products and brands I buy.
	59	I rarely purchase the latest fashion styles until I am sure my friends approve of them.
	60	I often identify with other people by purchasing the same products and brands they purchase.

Scale 1 was adapted from Sirdeshmukh, Singh, and Sabol (2002). Scales 2, 3, and 4 were taken from Steenkamp and Gielens (2003). Other items were adapted from Bruner, James and Hensel (2001).