

3-1-2018

Evidence-Based Survey Design: The Use of Negatively Worded Items in Surveys

Seung Youn (Yonnie) Chyung
Boise State University

Julie R. Barkin

Jennifer A. Shamsy

This is the peer reviewed version of the following article:

Chyung, S.Y., Barkin J.R., & Shamsy, J.A. (2018). Evidence-Based Survey Design: The Use of Negatively Worded Items in Surveys. *Performance Improvement*, 57(3), 16-25.

which has been published in final form at doi: [10.1002/pfi.21749](https://doi.org/10.1002/pfi.21749) by Wiley on behalf of the International Society for Performance Improvement. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Evidence-Based Survey Design: The Use of Negatively Worded Items in Surveys

Seung Youn (Yonnie) Chyung, Ed.D.

Department of Organizational Performance and Workplace Learning
Boise State University

Julie R. Barkin, M.S.

Department of Organizational Performance and Workplace Learning
Boise State University

and

Jennifer A. Shamsy, M.S.

Department of Organizational Performance and Workplace Learning
Boise State University

Abstract

A close examination of the literature on including positively and negatively worded items in structured survey questionnaires revealed that contrary to the traditional wisdom, it is better not to use a mix of positively and negatively worded items as doing so can create threats to validity and reliability of the survey instrument. If mixing, it is recommended to use strategies derived from research to improve the quality of data and the instrument validity and reliability.

Two Pull Quotes

1. A majority of research studies we reviewed recommend against mixing positively and negatively worded items in a survey as it can create threats to validity and reliability of the survey instrument.
2. However, researchers also recommend that if mixing, negatively worded items be used sparingly and with caution. Furthermore, survey developers should consider using strategies derived from research to improve the quality of data and reporting.

Introduction

Performance improvement practitioners and researchers often develop survey questionnaires to collect data and make data-driven decisions. Survey questionnaires can be designed to be structured or unstructured. While *unstructured* survey questionnaires contain open-ended questions, survey items used in *structured* survey questionnaires are closed-ended, each consisting of a statement or a question to be answered with a response scale.

When you develop a battery of survey items to measure a specific performance improvement factor (or construct) with the intention of calculating an average score of the data, you generally use a *statement* format with the same response scale such as a Likert scale. Conversely, when using a *question* format, it is best to use different response scales that are tailored to individual survey questions. However, it can be difficult to calculate an average score of data obtained from multiple survey items if different questions employ different response scales that are not comparable. Examples can be seen in Table 1.

Table 1. Examples of Structured Survey Items Designed with Statement or Question Formats

Statement Format	Question Format
<p>S1. The workshop objectives were clearly stated.</p> <ul style="list-style-type: none"> ○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree 	<p>Q1. Were the workshop objectives clearly stated?</p> <ul style="list-style-type: none"> ○ None of them ○ Some of them ○ Half of them ○ Most of them ○ All of them
<p>S2. The quality of the workshop is satisfactory.</p> <ul style="list-style-type: none"> ○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree 	<p>Q2. How would you rate the quality of the workshop?</p> <ul style="list-style-type: none"> ○ Very low ○ Somewhat low ○ Average ○ Somewhat high ○ Very high

In addition to the selection of survey item format (statement or question), you as a survey developer will also need to consider several other issues regarding how to design the survey items with response scales, for example:

- Whether to use positively worded survey statements only or include negatively worded survey statements.
- Whether to include a midpoint on a Likert-type scale.
- Whether to use a discrete rating scale such as a Likert-type scale or a continuous rating scale such as a slider.
- Whether to use ascending or descending order when listing anchors in response scales.

These seemingly simple decisions that survey developers make, however, require a substantial amount of knowledge in measurements and research-based evidence, as the degrees of validity and reliability of structured survey instruments can be influenced by many factors. There are many studies conducted on these topics, and teams of researchers from the Organizational Performance and Workplace Learning department at Boise State University have been reviewing research articles and developing evidence-based recommendations for developing structured survey questionnaires. For example, see Chyung, Roberts, Swanson, and Hankinson (2017) on the topic of using a midpoint on the Likert scale. The authors' extensive literature review led them to develop a set of evidence-based recommendations and corresponding strategies on including or excluding a midpoint.

This article is one of a series of articles on evidence-based survey design, addressing the topic of whether survey developers should use all positively worded statements or a mixture of positively and negatively worded statements. The purpose of this article is twofold: 1) describe several issues to be aware of when developing positively and negatively worded survey statements with Likert-type response scales (e.g., response set bias, wording types, and assumptions behind reverse-coding) and 2) present research-based evidence and recommendations regarding the use of positively and negatively worded statements in structured survey instruments.

When Developing Positively and Negatively Worded Survey Items...

Be Aware of Response Set Bias

Any measurement tool including structured survey questionnaires must be valid and reliable. However, survey developers need to be aware of various types of *response set bias*, a tendency of survey respondents to respond to a given survey item untruthfully, threatening the validity and reliability of survey instruments. For example, a commonly observed response set bias is an *acquiescence bias*, which is also known as a *yea-saying bias*, referring to the tendency for respondents to agree with questionnaire statements regardless of the content (Cronbach, 1942). Such response set biases are a threat to the validity of the survey instrument and should be avoided (Cronbach, 1950). To help avoid them, Rensis Likert, an American social psychologist and the original developer of the Likert scale, recommended

designing one half of survey items to be associated with agreement and the other half to be associated with disagreement (Likert, 1932). This design would alert survey respondents to pay attention to the meaning of survey items while also helping researchers detect data with potential response set bias.

Although Likert (1932) suggested the use of “straight-forward statements” (p. 45) and did not specifically indicate the use of negatively worded statements, survey developers have widely adopted the strategy of mixing positively and negatively worded items in structured survey questionnaires to reduce response set bias. However, it is questionable whether inclusion of negatively worded items in an otherwise positively worded survey is an effective solution to the response set problem. It has been recognized that the characteristics of survey items themselves, including positively or negatively worded statements, can cause response set bias (Podsakoff, MacKenzie, Podsakoff, & Lee, 2003). All types of response set bias are potential threats to the validity and reliability of survey instruments by yielding inaccurate and inconsistent data. This is especially true if a set of multiple survey items is used to measure a single dimension (aka, a construct).

Thus, you as a survey developer have two dilemmas:

- Should I include a mix of positively and negatively worded statements to reduce potential acquiescence bias, or should I design all statements to be worded in the same direction (usually, all positively worded)?
- If I use a mixed format, which wording (e.g., *not clear* or *unclear*) should I use to design negatively worded statements?

Before we discuss the above dilemmas, it is important to first identify different ways of wording survey statements and potential problems when reverse-coding negatively worded items.

Be Aware of Different Ways of Wording Survey Statements

When looking more closely into the dichotomous categories of positively and negatively worded survey items, you find four ways of wording survey statements.

First, look at the main descriptor of each survey statement and group survey statements into two categories depending on whether the descriptor itself has a positive or negative meaning. For example, a descriptor such as *clear* has a positive meaning, while a counterpart descriptor *unclear* has a negative meaning. Then, each descriptor type is divided into two subgroups depending on whether a negated word such as ‘not’ is absent or present. For example, *clear* vs. *not clear* or *unclear* vs. *not unclear*. Thus, there are four ways of wording survey statements.

Among the four ways of wording survey statements, positively worded statements (also called direct positive mode or regular) are the ones written with a positive descriptor and without a negated word (‘not’). Two other ways of wording survey statements, negated positive mode (negated regular) and direct negative mode (polar opposite), are considered negatively worded items. The fourth way, the double-negative format (negative polar opposite), is a frequent source of confusion for respondents and should not be used in survey questionnaires. Examples of these statement wording are shown in Table 2.

Table 2. Four Ways of Wording Survey Statements

Descriptor	Negated word, 'not'	Example	Respondents' perception	Colston (1999)	Schriesheim, Eisenbach, and Hill (1991)
Positive (e.g., clear)	Absent	The objectives were <i>clear</i> .	Positively worded	Direct positive mode	Regular
	Present	The objectives were <i>not clear</i> .	Negatively worded	Negated positive mode	Negated regular
Negative (e.g., unclear)	Absent	The objectives were <i>unclear</i> .	Negatively worded	Direct negative mode	Polar opposite
	Present	The objectives were <i>not unclear</i> .	Double-negative	Negated negative mode	Negated polar opposite

Be Aware of Assumptions behind Reverse-Coding

When including negatively worded items along with positively worded items in your survey instrument, you must reverse-code the data obtained from the negatively worded items to allow all data to be combined and statistically analyzed.

Reverse-coding includes the assumption that agreeing to a positively worded statement and disagreeing to its negatively worded counterpart are the same. However, there are problems associated with this assumption. To understand this, we need to put on a linguist's hat for a moment. Take a look at S2 and S2-1 in Table 3. To combine data obtained from S2-1 (negatively worded) with the data obtained from other positively worded items, you reverse-code the data obtained from S2-1. For example, a response of 'Somewhat disagree' to S2-1 is reverse-coded as 4, instead of 2, as if respondents would have selected 'Somewhat agree (4)' if its counterpart positively worded statement (S2) had been presented.

Table 3. An Assumption behind Reverse-Coding

Regular coding of a positively worded item	Reverse-coding of a negatively worded item
S2. The quality of the workshop is satisfactory. <ul style="list-style-type: none"> ○ Strongly disagree (coded as 1) ○ Somewhat disagree (2) ○ Neutral (3) ⊙ Somewhat agree (4) ○ Strongly agree (5) 	S2-1. The quality of the workshop is unsatisfactory. <ul style="list-style-type: none"> ○ Strongly disagree (reverse-coded as 5) ⊙ Somewhat disagree (4) ○ Neutral (3) ○ Somewhat agree (2) ○ Strongly agree (1)

However, "I somewhat disagree that the quality of the workshop is unsatisfactory" is not always the same as "I somewhat agree that the quality of the workshop is satisfactory." The respondents who somewhat disagreed that the quality was unsatisfactory (negatively worded S2-1) could have selected any option among *Neutral*, *Somewhat agree*, or *Strongly agree* if they had responded to the satisfactory (positively worded S2) statement.

Thus, development of negatively worded survey statements requires careful selection of an appropriate negative descriptor that can be correctly reversed to its intended counterpart during a reverse-coding process. In some cases, a negated positive mode (e.g., *not encourage*) and its direct negative mode (e.g., *discourage*) may not have the same meaning. Other examples are provided in Table 4.

Table 4. Examples of Four Types of Wording

Direct positive mode (regular)	Happy	Happy	Well done	Superior	Understand	Satisfied	Encourage
Negated positive mode (negated regular)	Not happy	Not happy	Not well done	Not superior	Do not understand	Not satisfied	Do not encourage
Direct negative mode (polar opposite)	Unhappy	Sad	Poorly done	Inferior	Misunderstand	Dissatisfied	Discourage
Negated negative mode (negated polar opposite)	Not unhappy	Not sad	Not poorly done	Not inferior	Do not misunderstand	Not dissatisfied	Do not discourage

Research Findings on the Use of Negatively Worded Items

What exactly has research shown when mixing positively worded items with negatively worded items that require reverse-coding? Evidence from the last several decades of research revealed the following findings.

Using a Mixed Format Can Create Threats to Validity and Reliability of the Survey Instrument

Research has indicated a concern for the *accuracy* of data obtained from survey instruments using a mix of positively and negatively worded items. In Schriesheim and Hill's (1981) research, 150 undergraduate students in the United States were asked to read a script describing a fictitious supervisor' behaviors and to indicate the behaviors by responding to a survey with three conditions: 1) 10 positively worded items, 2) a mix of five positively worded items and five negatively worded items, and 3) 10 negatively worded items. The researchers compared the scores between the three conditions to evaluate the effect of the wording conditions on accurate indication of the behaviors. They concluded that all positively worded survey items yielded significantly greater accuracy when compared with all negatively or mixed worded items. While the use of negatively worded items is sometimes employed to control acquiescence bias, the benefits may be outweighed by its effect on response accuracy and instrument validity.

Weem, Onwuegbuzie, and Lusting (2003) conducted a study with 185 undergraduate students in the United States who completed three anxiety scales with a five-point Likert response scale. For each anxiety scale, three items were positively worded and three items were negatively worded; negatively worded items were reverse-coded. Researchers found that scores on the positively and negatively worded items were not consistent. That is, strongly disagreeing to a positively worded statement is different from strongly agreeing to a negatively worded statement. The inconsistent scores suggest that survey respondents may not read negatively worded items carefully (carelessness), or they may process them differently than they process positively worded items.

This carelessness or difference in cognitive processing was also found in research with graduate-level students. Weem, Onwuegbuzie, and Collins (2006) conducted a study with 153 graduate students in the United States to examine the role of reading ability in responding to negatively worded items. They found that positively worded items produced higher means than negatively worded items. Just as with the undergraduate students, graduate students may not read

negatively worded statements as carefully or process them the same as positively worded statements. These studies provide evidence against using a mix of positively worded and negatively worded items in the same survey questionnaire.

Another concern with using a mixed format is that negatively worded items may cause a *method factor* (method effect) that is irrelevant to the characteristics or traits (constructs) being measured. Ibrahim (2001) analyzed the data obtained from a 21-item course evaluation questionnaire (with only one negatively worded item) submitted by 20,164 college students in Oman. Two factors emerged from the data: the first factor with 19 positively worded items and the second factor with one positively worded item and one negatively worded item. Upon analyzing the wording of the two items that loaded onto the second factor, the researcher's interpretation was that the two items likely loaded onto the same factor due to ambiguity. Thus, the fact that all the positively worded items loaded onto one factor (except one) while the negatively worded item loaded onto another separate factor implies a method effect.

Similar results were found in Greenberger, Chen, Dmitrieva, and Farruggia's (2003) research with 741 undergraduate students in the United States using three versions of the 10-item Rosenberg Self-Esteem Scale measured with a six-point Likert scale. The original version with five positively worded and five negatively worded items resulted in a two-factor model, measuring positive and negative self-image. A revised version with 10 positively worded items resulted in a one-factor model, measuring only positive self-image. Likewise, a revised version with 10 negatively worded items resulted in a one-factor model, measuring only negative self-image. The researchers concluded that the two-factor structure of the instrument was created by item wording difference (mixing), which proves a threat to construct validity.

Likewise, in Salazar's (2015) study, 699 Spanish people were surveyed over the telephone using one of three versions of the 15-item Keyes Social Well-being Scale with a five-point Likert scale. The first version contained all positive items (e.g., *honest*), the second version contained a combination of eight positive and seven negated positive items (e.g., *not honest*), and the third version contained a combination of seven positive, three negated positive, and five polar opposite items (e.g., *dishonest*). The research revealed positively worded items had higher scores than those of their negatively worded counterparts. Furthermore, positively worded items' scores (*honest*) were more like the negated positives' reversed scores (*not honest*) than the polar opposites' reversed scores (*dishonest*). In addition, this research found that positively worded items promoted acquiescence bias and that mixing positively worded with negatively worded items helped reduce acquiescence bias. However, this research also showed mixing could cause a method effect, impair factorial validity, and hurt internal consistency. The results of the study indicate that it is better to use all positively worded items in a questionnaire since its major weakness of potential acquiescence bias can be offset by forewarning respondents of the importance of providing valid responses.

Survey respondents' carelessness in responding to negatively worded items is also a cause for creating a separate factor and a threat to construct validity. The results of a study by Schmitt and Stults (1985) indicate that with only 10% of the respondents ignoring the wording of negatively worded items, a negative factor will appear regardless of the substantive meaning of the items. Similarly, Woods (2006) conducted a simulation study with 1,000 replications for each of 15 conditions using a 23-item survey with 13 positively worded and 10 negatively worded items. The study found:

- With 5% of respondents responding carelessly, the one-factor model still fits fairly well.
- With 10% of respondents responding carelessly, there is a noticeable decline in fit of the one-factor model.
- With 20% of respondents responding carelessly, fit is poor for the one-factor model, but excellent for the two-factor model.
- With 30% of respondents responding carelessly, fit is abysmal for the one-factor model but excellent for the two-factor model.

More importantly, this study's results are not unique to negatively worded items. If enough respondents (10% or more) carelessly respond to a survey regardless of the item's wording (positive or negative), the same results would be observed.

Better Not to Mix, but If You Do, Use These Strategies

The evidence presented above challenges the widely adopted practice of using negatively worded statements in structured survey questionnaires to reduce respondent's carelessness and resulting response set bias. Some researchers still advocate the use of negatively worded items in surveys while others caution that they should be used with care (e.g., Weijters & Baumgartner, 2012). One of the most important areas that you as a survey developer should address is the survey respondents' comprehension: Do the respondents understand the statement enough to appropriately respond to it? This becomes especially important when the statement is negatively worded. To ensure respondents' comprehension, survey developers need to use effective strategies in their survey design to elicit accurate responses.

Cognitive information processing theory tells us that people generally store information regarding the presence or absence of positive attributes (e.g., *clear* or *not clear*) as opposed to negative attributes (e.g., *unclear*). Thus, survey respondents may find it difficult to retrieve information based on negative attributes. This is exactly what Schriesheim et al. (1991) found in their study with 280 college students in the United States who completed a survey including four different forms: regular, negated regular, polar opposite, and negated polar opposite items. Both the regular and negated regular formats produced higher levels of reliability when compared with the polar opposite and negated polar opposite formats (Cronbach alpha of 0.90 and 0.83 versus 0.57 and 0.45, respectively). This study, as well as earlier work by Schriesheim and Hill (1981), suggests using regular items (e.g., *clear*) or negated regular items (e.g., *not clear*) in a survey while avoiding polar opposites (e.g., *unclear*) or negated polar opposites (e.g., *not unclear*).

Negatively worded statements, especially, double negatives, also require additional cognitive resources to process and often cause confusion. In Johnson, Bristow, and Schneider's (2004) study, 253 college students in the United States completed a seven-item survey using a six-point Likert scale. Although a unidimensional factor structure emerged regardless of positive or negative wording, internal consistency did decrease with negatively worded items. When double-negative items were presented, not only did internal consistency further decrease (overall Cronbach's alpha decreasing from 0.84 when positively worded, to 0.66 when negatively worded, to 0.26 with double negatives), but there was also an adverse impact to the factor structure. This indicates that the survey respondents became confused by the presence of double negatives. Therefore, it is recommended that negatively worded statements be converted to positively worded statements, or if negatively worded statements are used, survey developers should avoid the use of any double negatives.

Though the use of negated regular statements (e.g., *not clear*) is supported in some research (e.g., Schriesheim & Hill, 1981), survey respondents' cognitive load may be affected by how you word the negative statements. Weijters and Baumgartner (2012) caution against using the word 'not' to negate regular statements. They suggest that this may cause the survey respondents to retrieve information that is not needed in processing the statements and can make the judgment process more difficult. In addition, other complex forms of negation can be confusing to respondents and should be avoided to reduce cognitive load and errors during judgment. Instead, the researchers suggest the use of carefully selected polar opposites (true opposites of the construct being measured) to support a more robust information retrieval process.

As pointed out earlier, the purpose of using a mix of positively and negatively worded items in surveys is to help decrease potential response set bias such as acquiescence bias. However, research suggests that there are other strategies you can use to reduce respondents' careless responses. A forewarning method is one such strategy. In Matthews and Shepherd's (2002) study, the researchers found that forewarning participants about the presence of negatively worded items did reduce the number of careless responses and the amount of negative factor loading. However, it did not eliminate carelessness or the method effect. Another strategy derived from Roszkowski and Soven's (2010) study of 3,605 undergraduate students' course evaluations is that if you do use a mixed format, you should group the same statement types (positively or negatively worded) together so that the respondents' attention is drawn to the different nature of each group. By focusing attention, you are increasing the probability that the respondents will avoid mental shortcuts and more deeply process the statements. However, the researchers did indicate that forewarning or grouping does not eliminate undesirable outcomes such as response set bias.

Another issue relating to respondents' careless responses is fatigue. Survey developers should be aware that respondents may overlook the presence of negatively worded statements when they are fatigued. They may already be fatigued when they start the survey or become fatigued by the number or type of survey items within the survey questionnaire. Decremental performance has been noted to occur only 12 minutes after respondents start a survey

questionnaire. Merritt's (2012) collection of five studies revealed a consistent pattern that two factors emerged when the survey respondents were fatigued and negatively worded items were present. Furthermore, when participants were fatigued, efforts to draw their attention to negatively worded statements by bolding, underlining, or capitalizing the negated element ('not') were insufficient. Thus, surveys should be administered when respondents have adequate cognitive resources to effectively process negatively worded statements. For example, you may present negatively worded items early in the survey, and/or provide respondents with a mental rest period/break during the survey.

Survey developers should also pay attention to the development and presentation of a response scale. As noted above, negatively worded items are reverse-coded, so it is important to use a symmetrical response scale to maintain accuracy (Locker, Jokovic, & Allison, 2007). For example, when using a symmetrical Likert scale consisting of Strongly disagree (SD), Disagree (D), Neutral, Agree (A), and Strongly agree (SA), reverse-coding of SD is SA, and vice versa. However, with a non-symmetrical scale such as *Never* on one extreme side and *Almost all the time* on the other extreme side, *Almost all the time* is reverse-coded as *Never*, when it should be *Almost never*, and *Never* is reverse-coded as *Almost all the time*, when it should be *Always*.

Finally, instead of focusing on the design of survey statements or response scales, survey developers might also look at how results are reported. As an alternative to reverse-coding negatively worded items, Hartley (2013) suggests that researchers present survey results obtained from negatively worded items separately from positively worded items, in lieu of reverse-coding the data obtained from negatively worded items and combining them with the data obtained from positively worded items.

Summary

Survey developers may use a mix of positively and negatively worded items in structured survey questionnaires as a means of safeguarding against acquiescence bias. However, due to expectations, biases, statement wording, reading levels/intellectual capacity, carelessness, and/or fatigue, survey respondents may not appropriately comprehend negatively worded statements. When mixing positively and negatively worded items, negatively worded items often emerge as a separate factor (construct) regardless of the content of the items, creating a threat to construct validity and reliability. However, simply excluding negatively worded items from a survey instrument does not make the instrument problem-free. Researchers and practitioners should still be concerned with potential response set bias when using all positively worded items. Also, in some surveys, the use of negatively worded items is inevitable as the attributes to be measured are negative in nature: for example, depression.

The key is to make design choices that result in the most valid responses whether that includes using all positively worded items or a mix of positively and negatively worded items. A majority of research studies we reviewed recommend against mixing positively and negatively worded items in a survey as it can create threats to validity and reliability of the survey instrument. However, researchers also recommend that if mixing, negatively worded items be used sparingly and with caution. Furthermore, survey developers should consider using strategies derived from research to improve the quality of data and reporting.

Table 5 presents a summary of the evidence-based recommendations based on the literature we reviewed. Table 6 is a summary of the research evidence used in generating the recommendations.

Table 5. Evidence-Based Recommendations on the Use of Negatively Worded Statements in Surveys

	Use Positively Worded Statements Only	Mix Positively and Negatively Worded Statements
Benefit	<ul style="list-style-type: none"> There is no need to reverse-code some data. It helps improve construct validity and reliability of the survey instrument. 	<ul style="list-style-type: none"> It may help reduce acquiescence bias. It can be used to detect data with acquiescence bias.
Problem	<ul style="list-style-type: none"> It may increase acquiescence bias. 	<ul style="list-style-type: none"> Negatively worded items may emerge as a separate factor (aka, a method effect). Careless respondents may misunderstand negatively worded statements and provide erroneous data.
When	<ul style="list-style-type: none"> There is minimal concern for the presence of acquiescence bias. There is high confidence in preventing careless responses. 	<ul style="list-style-type: none"> The attributes to be measured are negative in nature (e.g., depression). There is a need to safeguard against acquiescence bias.
Strategies to Use	<ul style="list-style-type: none"> Use straight-forward statements to prevent respondents from making careless responses. Administer the survey when respondents are not fatigued and minimize fatigue during survey completion. 	<ul style="list-style-type: none"> Use negated regular items or carefully selected polar opposites. Never use negated polar opposites (double negatives). Alert respondents of negatively worded items by using a forewarning method and/or grouping negatively worded items together. Administer the survey when respondents are not fatigued and minimize fatigue during survey completion. Use symmetrical response scales to allow appropriate reverse-coding. Report the results of negatively worded items separately, instead of combining them with positively worded items.

Table 6. Research Evidence for the Use of Negatively Worded Items

Focus	Authors (Year)	Recommendations Based on Research Findings
Do not mix as mixing can create a threat to construct validity and reliability	Greenberger et al. (2003)	<ul style="list-style-type: none"> Do not mix as it creates a two-factor structure of the instrument based on the item wording difference (positively and negatively worded items), which is a threat to construct validity.
	Ibrahim (2001)	<ul style="list-style-type: none"> Do not mix because it may cause a method effect, which is when positively worded items and negatively worded items are loaded onto separate factors.
	Salazar (2015)	<ul style="list-style-type: none"> Do not mix because although mixing can reduce the acquiescence bias, it causes a method effect, impairs factorial validity, and hurts internal consistency.
	Schmitt and Stults (1985) Woods (2006)	<ul style="list-style-type: none"> Be aware that even a small number (10%) of survey respondents carelessly responding to negatively worded items can create a separate factor and a threat to construct validity.

	Schriesherim and Hill (1981)	<ul style="list-style-type: none"> Do not mix as it decreases response accuracy; all positively worded items yield significantly greater accuracy than all negatively worded or mixed items.
	Weem et al. (2006)	<ul style="list-style-type: none"> Do not mix even for highly educated samples such as graduate-level students; positively worded items produce higher means than negatively worded items.
	Weem et al. (2003)	<ul style="list-style-type: none"> Do not mix because survey respondents may not read negatively worded items carefully and may process them differently than they process positively worded items, which can create a threat to validity and reliability of the instrument.
Better not to mix, but if mixing, use strategies	Hartely (2013)	<ul style="list-style-type: none"> Do not mix. However, if mixing, present results obtained from negatively worded items separately, instead of reverse-coding the data and combining them with the data obtained from positively worded items.
	Johnson et al. (2004) Schriesheim et al. (1991)	<ul style="list-style-type: none"> Do not mix. However, if mixing, do not use polar opposite (e.g., <i>unclear</i>) and negated polar opposite (e.g., <i>not unclear</i>). Use negated regular items (e.g., <i>not clear</i>).
	Locker et al. (2007)	<ul style="list-style-type: none"> When using a mixed format with the intent to reverse-code negatively worded items, make sure to use a symmetrical response scale with an equal number of anchors on the positive and negative sides of the scale.
	Mathews and Shepherd (2002)	<ul style="list-style-type: none"> When mixing, the potential acquiescence bias can be reduced by using a forewarning method (warning respondents to look out for negative wording), although the forewarning method does not always work perfectly.
	Merritt (2012)	<ul style="list-style-type: none"> Do not mix. However, if mixing, administer the survey when respondents are not fatigued; simply warning respondents about negatively worded items by bolding, underlining, or capitalizing the word 'not' is insufficient.
	Roszkowski and Soven (2010)	<ul style="list-style-type: none"> Do not mix. However, if mixing, group the negatively worded items together and alert respondents to the nature of the statements being changed from positive to negative.

References

- Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, 56(10), p-p. (page numbers to be inserted in October)
- Colston, H. L. (1999). "Not good" is "bad," but "not bad" is not "good": An analysis of three accounts of negation asymmetry. *Discourse Processes*, 28(3), 237-256. doi:10.1080/01638539909545083
- Cronbach, L. J. (1942). Studies of acquiescence as a factor in the true-false test. *Journal of Educational Psychology*, 33(6), 401-415. doi:10.1037/h0054677
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10(1), 3-31.
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, 35(2003), 1241-1254. doi:10.1016/S0191-8869(02)00331-8
- Hartley, J. (2013). Some thoughts on Likert-type scales. *International Journal of Clinical and Health Psychology*, 13, 83-86. doi:10.1080/13645570802648077
- Ibrahim, A. M. (2001). Differential responding to positive and negative items: The case of a negative item in a questionnaire for course and faculty evaluation. *Psychological Reports*, 88, 497-500. doi:10.2466/pr0.2001.88.2.497
- Johnson, J. M., Bristow, D. N., & Schneider, K. C. (2004). Did you not understand the question or not? An investigation of negatively worded questions in survey research. *Journal of Applied Business Research*, 20(1), 75-86.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 (1932-33), 5-55.
- Locker, D., Jokovic, A., & Allison, P. (2007). Direction of wording and responses to items in oral health-related quality of life questionnaires for children and their parents. *Community Dentistry and Oral Epidemiology*, 35(4), 255-262. doi:10.1111/j.1600-0528.2007.00320.x
- Mathews, B. P., Shepherd, J. L. (2002). Dimensionality of Cook and Wall's (1980) British Organizational Commitment Scale revisited. *Journal of Occupational and Organizational Psychology*, 75, 369-375. doi:10.1348/096317902320369767
- Merritt, S. M. (2012). The two-factor solution to Allen and Meyer's (1990) Affective Commitment Scale: Effects of negatively worded items. *Journal of Business Psychology*, 27, 421-436. doi:10.1007/s10869-011-9252-3
- Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Lee, J-Y. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. doi:10.1037/0021-9010.88.5.879
- Roszkowski, M., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 117-134. doi:10.1080/02602930802618344
- Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema*, 27(2), 192-199. doi:10.7334/psicothema2014.266
- Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9, 367-373. doi:10.1177/014662168500900405.
- Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, 41, 1101-1114. doi:10.1177/001316448104100420
- Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, 51(1), 67-78. doi:10.1177/0013164491511005
- Weem, G. H., Onwuegbuzie, A. J., & Collins, K. M.T. (2006). The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Evaluation & Research in Education*, 19(1), 3-20. doi:10.1080/09500790608668322
- Weem, G. H., Onwuegbuzie, A. J., Lustig, D. (2003). Profiles of respondents who respond inconsistently to positively- and negatively-worded items on rating scales. *Evaluation & Research in Education*, 17(1), 45-60. doi:10.1080/14664200308668290
- Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, 49(5), 737-747. doi:10.1509/jmr.11.0368

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factory analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 189-194. doi:10.1007/s10862-005-9004-7

Biography

Seung Youn (Yonnie) Chyung, Ed.D., is a professor of the Department of Organizational Performance and Workplace Learning in the College of Engineering at Boise State University (<http://opwl.boisestate.edu/faculty-staff/faculty/yonnie-chyung/>). She teaches graduate courses on Program Evaluation and Quantitative Research in Organizations. She runs a Workplace-Oriented Research Central (WORC) lab where teams of practitioners and researchers conduct research on various topics in the HPI context.

Julie R. Barkin is the assistant manager of Administrative Services at a not-for-profit organization in the domestic nuclear power industry. She has developed e-learning programs and instructor-led training for improving workplace performance. She completed her Master of Science degree in Organizational Performance and Workplace Learning from Boise State University in December 2017. Julie may be reached at [juliebarkin@gmail.com](mailto:julierbarkin@gmail.com).

Jennifer A. Shamsy, M.S., is a commercial airline pilot and adjunct online instructor. She graduated in Spring 2017 with a Master of Science degree in Organizational Performance and Workplace Learning and a Workplace Instructional Design Graduate Certificate from Boise State University. Jennifer may be reached at jshamsy@gmail.com.