# Scale format effects on response option interpretation and use

Elke Cabooter [a,*], Bert Weijters [b,1], Maggie Geuens [c,2], Iris Vermeir [c,2]

[a] IESEG School of Management (LEM-CNRS 9221), Rue de la digue 3, 59000 Lille, France
[b] Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium
[c] Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium

## ARTICLE INFO

## ABSTRACT

When designing questionnaires, researchers often use different scale formats, which vary on the dimensions polarity (unipolar versus bipolar endpoint labels) and the numbering of the response options (e.g., only positive numbers versus positive and negative numbers). This study uses survey experiments and cognitive interviews, to test specific hypotheses regarding how scale formats' polarity and their numbering might affect (1) observed response distributions and (2) the meaning of response options. In line with our theoretical predictions, the results unveil consistent differences in the interpretation and use of the scale formats as a function of their polarity and numbering. Therefore, alternative formats cannot be used interchangeably as the scale formats are differently interpreted and used. In addition, researchers need to choose scale formats and interpret scale responses in accordance with respondents' interpretations.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Questionnaires are a key source of data in business research. When designing questionnaires, researchers have to make many important decisions. Some of these decisions can be based on solid evidence from previous research. For instance, many scales have been extensively validated, such that complete scale inventories are at researchers' disposal (e.g., Bruner, James, & Hensel, 2001) and if no validated scales are at hand, well-established methodological guidelines for multi-item scale construction are available in the literature (Churchill, 1979; Peter, 1981; MacKenzie, Podsakoff, & Podsakoff, 2011; Rossiter, 2002). Once a researcher has formulated scale items or selected them from existing scales, an important question is how to format the response scale.

The most popular scale format used in questionnaires is the Likert-type scale in agree–disagree format (Liu, Lee, & Conrad, 2015). However, the data quality yielded by the Likert scale is controversial, and as measurement quality remains an ongoing concern (e.g., Rossiter, 2002; Sharma & Weathers, 2003; Strizhakova, Coulter, & Price, 2008), methodologists have suggested the need to consider the use of alternative scale formats (Rindfleisch, Malter, Ganesan, & Moorman, 2008; Zhang & Savalei, 2015). Clearly, more evidence is needed to help

researchers in deciding which scale format to employ. Therefore, in this study we compare the classic Likert-type scale format with three other scale formats (see Fig. 1) that differ from the Likert scale on one or two dimensions, that is polarity (i.e., the use of bipolar versus unipolar endpoint poles) and the numbering of the response options (i.e., the use of positive numbers only (1 to 7) or positive and negative numbers ($-3$ to $+3$)).[3]

Our aim is to compare response patterns for the four scale formats in Fig. 1. This question is important for two related reasons. First, the choice for a unipolar vs. bipolar format and the type of numbering used is typically made on an ad hoc basis. Second, there are good reasons to expect that this choice will affect research results and conclusions (Baumgartner & Steenkamp, 2001). As for the first point, researchers use the different scale formats, even for the same concepts, without knowing whether the choice for a specific scale format influences respondents' behavior. For instance, respondents' satisfaction has been measured via a unipolar scale format 'In general, I am satisfied with my life,' (Oliver, 1980; Chiou & Droge, 2006), but also via a bipolar scale format 'In general, I am dissatisfied with my life' to 'In general, I am satisfied with my life' (Prakash & Lounsbury, 1984; Marinova & Singh, 2014), and this is no exception. An analysis from the manuscripts published in the Journal of Business Research from January 2014 till January

---

* Corresponding author. Tel.: +33 320545892; fax: +33 320 574 855.
  E-mail addresses: e.cabooter@ieseg.fr (E. Cabooter), bert.weijters@ugent.be (B. Weijters), maggie.geuens@ugent.be (M. Geuens), iris.vermeir@ugent.be (I. Vermeir).
  [1] Tel.: +32 9 264 64 56, fax: +32 9 264 64 94.
  [2] Tel.: +32 9 264 35 21, fax: +32 9 264 42 79.

[3] Seven-point scale formats are very commonly used (Weijters, Cabooter & Schillewaert, 2010), which is in line with methodological recommendations, as seven-point scale formats allow for good information recovery without overburdening respondents (Green & Rao, 1970) and contain a midpoint which enhances measurement quality (Nowlis, Kahn, & Dhar, 2002). For these reasons, we will limit our attention to seven-point formats in the current study.

| Unipolar - positive | | Not at all<br>agree | | | | | | Strongly<br>Agree | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Unipolar - positive | I'm good at sports | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| Unipolar - positive and negative | | Not at all<br>agree | | | | | | Strongly<br>Agree | | |
| Unipolar - positive and negative | I'm good at sports | -3 | -2 | -1 | 0 | 1 | 2 | 3 | | |
| Bipolar - positive | I'm bad at sports | 1 | 2 | 3 | 4 | 5 | 6 | 7 | I'm good at sports | |
| Bipolar - positive and negative | I'm bad at sports | -3 | -2 | -1 | 0 | 1 | 2 | 3 | I'm good at sports | |

Fig. 1. Example of various scale formats.

2015[4] shows that about 16% of the scale formats used deviate from the originally validated scale format and about 20% of the articles used different scale formats (i.e., unipolar and bipolar scale format) within the same study. Even though most publications do report *which* scale format is used (82.4% of the papers analyzed), they do not explain why a specific format is selected. A likely reason for this lack of consistency and rigor is that data quality of the alternative scale formats has not been investigated so far and researchers lack evidence-based guidelines to guide their choice for a specific scale format. This is unfortunate, because choosing an optimal scale format does not take valuable questionnaire space (as opposed to, say, adding more items) and can thus be an efficient way of improving data quality.

As for the second point, because scale formats can influence responses (Tourangeau, Rips, & Rasinski, 2000), they can change research results and conclusions in several ways. First, scale formats can affect means, so the type of scale format that is used can change the extent to which respondents seem to like a brand or ad, seem to be satisfied with services, intend to buy new products, etc. (Weijters, Cabooter, & Schillewaert, 2010). Second, scale formats can affect item variances, which in turn affect the extent to which samples of respondents are seen as relatively homogeneous/heterogeneous, and the way in which they are clustered and segmented (Greenleaf, 1992a). Consider, for example, a study that identifies respondents in the highest and lowest deciles on a measure of attitude towards a product. Many of the respondents in the highest or lowest deciles may be classified as extreme because of their response behavior induced by the scale format, but may actually belong in more moderate segments and vice-versa. Third, scale formats can affect the correlation between items, and thus the factor structure and internal consistency of scales (Baumgartner & Steenkamp, 2001). For instance, responses to reversed items are usually recoded by reflecting the initial response around the midpoint (e.g., a six becomes a two on a seven-point scale), but this recoding is based on the assumption that the midpoint is also interpreted as such by respondents. Finally, differences in the interpretation and usage of scale formats will alter relationships with non-scaled variables, like demographics. For instance, in the data reported in the current paper, the

variable age shows a stronger relationship[5] with the item 'I am confident that I can learn technology-related skills' when the scale format is a bipolar positive scale (i.e. scale format 3 in Fig. 1) compared to a unipolar positive scale (i.e. scale format 1 in Fig. 1).

To the best of our knowledge, only one study has investigated the combination of polarity and numbering before (Tourangeau, Couper, & Conrad, 2007). However, this study does not allow to draw conclusions because the effect of questionnaire content and scale format cannot be disentangled. Therefore, with this study, we seek to contribute to a better understanding of how respondents' interpretation and use of response options depend on rating scale polarity and numbering. We begin with a quantitative survey experiment to test whether rating scale polarity and numbering affect response patterns (n = 337). Next, we further clarify our findings by means of a qualitative (n = 16) and a quantitative study (n = 393) in which we investigate how respondents interpret the response options in rating scale formats that vary in polarity and numbering.

The paper demonstrates that the choice for a specific scale format should be given serious consideration as changes in polarity and their numbering influence response behavior. We explain the differential response behavior in terms of two mechanisms: the symmetry effect and the intensity effect. The findings have implications for survey research and for some types of scales, which will be explained in the discussion section of the paper.

## 2. Conceptual development

### 2.1. Scale formats

Marketing research uses scale formats that differ in polarity and in the way that the response options are numbered. A typical unipolar scale format contains positive numbers; a typical bipolar scale format features both negative and positive numbers (e.g., Menezes & Elbert, 1979; Osgood, Suci, & Tannenbaum, 1957). The unipolar scale format (see the first and second format in Fig. 1) is easy to construct, because each item consists of a single statement that respondents rate in terms of their agreement with it. Each item also taps into one pole of the underlying construct (e.g., This brand is good), not both poles (e.g., This brand is bad–This brand is good). In an alternative measurement approach, respondents consider both poles of the construct underlying

---

[4] The analysis in JBR was based on 675 constructs used in 205 papers. All constructs or single items rated on a scale format where taken into account. There were 36 papers where the scale format was unknown. 109 (16%) of the 675 scale formats used a scale format that was different from the validated scale (that is the scale format originally used and tested when the scale was created). 40 (20%) out of the 205 papers used bipolar and unipolar constructs/items within the same paper.

[5] A regression with the unipolar positive scale as reference scale was executed. Age has a stronger relationship with the DV when the scale format is bipolar positive compared to unipolar positive, which is indicated by the β of the interaction term age ∗ bipolar positive: $\beta = .196$, $t(332) = 2.76$, $p < .01$. Similar results were found for 36 of the 52 items.