

Middle Response Scale Options are Inappropriate for Ideal Point Scales

Dev K. Dalal · Nathan T. Carter · Christopher J. Lake

Published online: 25 September 2013
© Springer Science+Business Media New York 2013

Abstract

Purpose The purpose of this study is to provide a theoretical rationale for the inappropriateness of middle response options on the response scales offered on ideal point scales, and to provide empirical support for this argument to assist ideal point scale development.

Design/Methodology/Approach The same ideal point scale was administered in three quasi-experimental groups varying only in the response scale offered: the three groups received either a four-, five-, or six-option response scale. An ideal point Item Response Theory model, the Generalized Graded Unfolding Model (GGUM), was fit to the response data, and model-data fit was compared across conditions.

Findings Responses from the four- and six-option conditions were fit well by the GGUM, but responses from the five-option condition were not fit well.

Implications Despite the scale being constructed to follow the tenets of ideal point responding, the GGUM was

unable to provide a reasonable probabilistic account of responding when the response scale contained a middle option. The authors find support for the argument that an odd-numbered response scale does not match the principles of ideal point responding, and can actually result in misspecifying the underlying response process.

Originality/Value Although a growing body of research has suggested that attitude and personality measurement is best conceptualized under the assumptions of ideal point responding, little practical advice has been given to researchers or practitioners regarding scale creation. This was the first study to theoretically and empirically assess the response scale on ideal point scales, and offer guidance for constructing ideal point scales.

Keywords Item Response Theory · Measurement · Ideal point measurement · Survey research

Parts of this manuscript were presented at the 2012 Annual Meeting for the Society of Industrial and Organizational Psychology, San Diego, California.

D. K. Dalal (✉)
Department of Psychology, University of Connecticut, U-1020,
Storrs, CT 06269, USA
e-mail: dev.dalal@uconn.edu

N. T. Carter
Department of Psychology, University of Georgia, 323
Psychology Building, Athens, GA 30602, USA
e-mail: ntcarter@uga.edu

C. J. Lake
Department of Psychology, University of Minnesota, Duluth,
1207 Ordean Ct., Duluth, MN 55812, USA
e-mail: lakec@d.umn.edu

The predominant means of measuring attitudes and personality in psychology has been the self-report survey, using scales of agreement to various statements (e.g., *agree*, *disagree*, *neutral*). Utilizing Likert's (1932) now-ubiquitous method of measurement, researchers knowingly or unknowingly assume that responses follow a *dominance* model, which assumes higher agreement to positively worded scale items indicates a higher standing on the measured attribute (see Fig. 1). Recently, researchers have begun to reconsider the use of Thurstone's (1925, 1928) idea that responses follow an *ideal point* model. For conventional, extreme-worded items, these models are nearly identical. For moderately worded items, however, the ideal point model would suggest a non-monotone function (see Fig. 2). Recent Item Response Theory (IRT) modeling research has suggested that the assumption of a Likert-

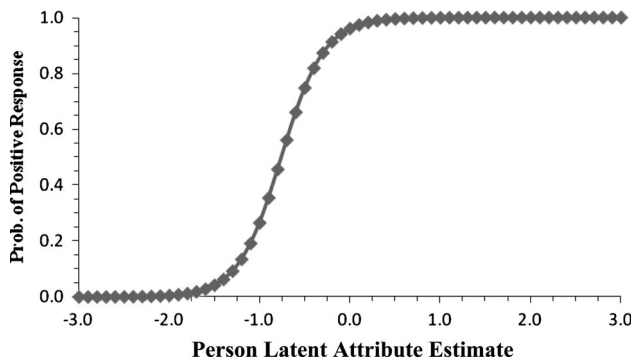


Fig. 1 Example of a dominance response process model. Probability of endorsing an item increases monotonically as the respondent's latent attribute standing increases

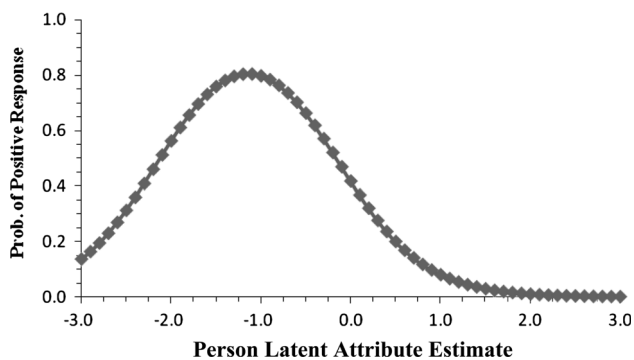


Fig. 2 Example of an ideal point response process model. Probability of endorsing an item increases nonmonotonically as the respondent's latent attribute standing increases

based dominance model may not be the most appropriate approach to scale construction as it might impede high reliability for attribute estimates across the full range of respondents due to the elimination of moderately worded items (e.g., Chernyshenko et al. 2007; Drasgow et al. 2010; Stark et al. 2006).

Despite this call to reconsider Thurstone scaling, two categories of barriers to the widespread adoption of ideal point measurement in applied measurement are (1) the scoring and (2) the creation of ideal point scales (e.g., Carter et al. 2010; Dalal et al. 2010). Issues that have been raised within this second category include writing ideal point items and the number of items needed to achieve adequate measurement precision. In this article, we raise an additional concern that has received virtually no attention with respect to ideal point measurement; namely, the type of response scale to offer on ideal point scales.

Indeed, although the role item content plays in the response process has been considered, little is known about how response scales fit into the discussion. This is problematic insofar as inappropriate response scales can limit the reliability and validity of the psychological measures (e.g., Guion 2011; Hinkin 1998). Researchers generally

construct measures which follow Likert's (1932) recommendation of using a middle anchor such as *Neutral* or ? options. In this research note, we offer a rationale that supports using even-numbered response scales, excluding a middle response option, based on the tenets of ideal point responding as originally presented by Thurstone (1928), and refined by Coombs (1964), Andrich (1996), and Roberts et al. (2000). In short, we argue that because the dominance and ideal point response processes make fundamentally different assumptions regarding how attribute standing is indexed, different response scales are needed.

In what follows, we first briefly review the fundamentals of ideal point and dominance models of psychological scaling. We then discuss some of the issues raised with creating ideal point measures noting that the response scale on ideal point measures has received no attention. Continuing, we highlight some of the past research on middle response options, which have primarily focused on Likert scales. These discussions will serve as the backdrop for our arguments regarding how the different response processes implied by dominance and ideal point measurement inform the type of response scale to offer. We then demonstrate that an ideal point IRT model fits responses to the same scale obtained using an even-numbered response scale better than responses obtained from an odd-numbered version. Implications for scale construction in research and practice are discussed.

Ideal Point Responding

Historically, two basic approaches to constructing self-report scales have been considered by psychologists. One method, presented by Thurstone (1925, 1928), represents an ideal point response process (Coombs 1964). According to the Thurstonian approach to measurement, items are written to cover the entire range (i.e., positive, negative, and moderate regions) of an attitude or personality continuum. The principles of ideal point responding suggest that respondents will endorse items that are close to their position on the latent attribute distribution, and reject items that are further away from their position (Andrich and Styles 1998). That is, an assumption of the ideal point response process is that respondents either endorse or reject an item (Roberts et al. 1999; Thurstone and Chave 1929). This has been termed ideal point responding because items are endorsed when the difference between the respondent's location and the item's location is ideal (i.e., minimized). This process holds for all items across the latent attribute distribution. According to the ideal point response process, an item that is on the negative end of the continuum, for example, is more likely to be endorsed by individuals at that end of the continuum (i.e., difference close to 0).

Importantly, this process of responding implies that individuals with opposite standing on the latent attribute continuum can disagree to the same item for different reasons. That is, there are different *subjective* responses elicited for the same *objective* responses—this point is discussed in greater detail below.

Dominance Responding

An alternative to Thurstonian measurement, presented by Likert (e.g., 1932; Likert et al. 1934) utilizes what has come to be termed the dominance response process (Roberts et al. 1999; Stark et al. 2006). This is an easier method of indexing personality and attitudes because researchers need only write items that tap the high and low end of the attribute continuum, with neutral items being purposely avoided. That is, Likert's and Thurstone's approaches to scale development differed with respect to the inclusion of moderately worded or neutral items. More specifically, with Likert's method, "...certain statement could not be used in this scoring method because it was found impossible to determine whether to assign a value of 1 or 5 to the 'strongly agree' alternative" (p. 230, Likert et al. 1934). Stated differently, in Likert's method of scaling, the items needed to be written such that responses could clearly be coded as either pro- or anti-attitude sentiment. For example, three items from the Droba Attitudes Toward War Scale (1930), a scale developed in the Thurstone tradition, are: "The benefits of war outweigh its attendant evils;" "It is difficult to imagine any situation in which we should be justified in sanctioning or participating in another war;" and "Compulsory military training in all countries should be reduced but not eliminated." The first two items correspond to the extreme pro-war and anti-war attitudes, respectively. The third item, however, would represent a more moderate standing, and was given by Likert et al. (1934, p. 230) as a specific example item that Likert's method could not accommodate. Likert and colleague stated that "In general, those statements whose scale values in the Thurstone method of scoring fell in the middle of the scale...were the statements that were found to be unsatisfactory for the simpler method of scoring" (p. 231). Clearly, Likert and Thurstone scales differ in that items implying moderate standing on the attribute were excluded from Likert scales.

In the typical application of Likert's method, respondents are asked to respond to the extreme items using a numeric 1–5 scale; variations include using a 1–3, 1–7, or *yes, no, ?* response scales. Regardless of the response scale (and anchors) used, a numeric representation of the person's standing on the attribute is calculated by simply adding the selected response option values together (items

at the low end are reverse-scored). The process assumed in this type of scale is referred to as the dominance response process—borrowed from the ability testing terminology, the probability of getting an item correct increases as the individual dominates (i.e., has a higher ability level than) the item. A similar process holds for personality and attitude measurement: For any particular item, if the individual is higher on the latent attribute than the item implies he/she will endorse the item. Again, this process holds for all items on a measure; for individuals following the dominance response process, the probability of endorsing an item increases monotonically as the difference between the person's location and the item's location is more positive.

Likert's method of scale development has been the more popular approach because the scoring method is simple and does not require item location calibration like Thurstonian scales do (Roberts et al. 1999). In addition, Likert's method of measurement facilitates the use of item-total correlations, factor analyses, and internal consistency reliability estimates (Stark et al. 2006), as indicators of a scale's reliability and internal validity; these are tools with which scale developers have become quite comfortable (see also Zickar and Broadfoot 2009). Although more difficult than the Likert approach to scaling, however, one of the greatest strengths of ideal point measures may be their ability to accurately assess the attribute standing for individuals across the whole of the attribute continuum. Indeed, some have argued that when researchers are interested in assessing the full spectrum of a particular attitude or personality trait, they should consider using an ideal point (Thurstonian-type) scale; this argument has since sparked renewed interest in ideal point measurement (see Drasgow et al. 2010 for a discussion).

Creating Ideal Point Scales

Two potential limitations to the widespread use of ideal point scales that have been noted are (1) the difficulty with creating ideal point scales, and (2) the difficulty associated with scoring ideal point measures. Regarding scoring of ideal point scales, traditional methods of scoring were cumbersome and required that all items be administered to a calibration group before being administered to respondents of interest (e.g., Roberts et al. 1999). Scoring ideal point measures has been simplified by Roberts and colleagues (Roberts et al. 2000; Roberts and Laughlin 1996) who have developed a series of IRT models that allow for non-monotonic item response functions, a seemingly accurate method of scoring ideal point scales.

Regarding scale creation, some researchers have found it difficult to create items that follow the ideal point

response process—indeed, Carter et al. (2010) and Dalal et al. (2010) comment that further research is needed to understand how to create items that reliably induce ideal point responding (see also Chernyshenko et al. 2007). Moreover, the number of items needed for accurate indexing of attribute standing with ideal point scales is also unknown (Dalal et al. 2010).

In addition to these known issues, one important but often overlooked component of ideal point scale development is the selection of a response option scale. Although this consideration may seem trivial, it is, in fact, quite important to the quality of a psychological instrument. Indeed, as Guion (2011) noted, “Responses define the measurement purpose of the scale” (p. 43). Moreover, ambiguity in the response scale can negatively impact scale reliability and validity conclusions (Hinkin 1998). Clearly, insufficient attention to the response scale will impede quality measurement. Therefore, if researchers and practitioners are to effectively utilize ideal point scales, research on the type of response scale to offer respondents is necessary.

Unfortunately, no research has examined the role the response option scale can play in the ideal point response process—though the question has been raised (see Oswald and Schell 2010). Response scales used in research on ideal point models are seldom discussed by the authors, and often little or no justification is given for preferring an even versus odd response option scale. We believe that giving consideration to this matter is not only important but necessary for the development of high-quality ideal point scales because, as is discussed next, the ideal point response process suggests that one response scale is more appropriate than the other.

A Response Scale Based on the Item Response Process

Although the effects of a middle response option have been extensively studied, these studies investigated scales that followed the dominance response process. For example, Kalton et al. (1980) and Bishop (1987) demonstrated that simply offering a middle option on the scale resulted in individuals shifting their responses toward the neutral point and away from neighboring options on the scale (i.e., saying *neutral* rather than *slightly agree*).

Hernández et al. (2004) used a mixed-measurement approach to investigate respondents’ use of the middle response option and determined that two latent classes fit the data better than a single latent class; they showed that, whereas one class was unlikely to use the middle option, the other class had a high probability of using the middle option. Based on IRT analyses, Kulas et al. (2008) observed that the middle option was used not to indicate

neutrality, but instead as a “default” option when respondents were unwilling to select other options. Hanisch (1992) showed that the middle response option used on the Job Descriptive Index (JDI; a ? option) was actually indicative of negative sentiment as opposed to a neutral or not applicable response. Finally, similar to Hernández et al., Carter et al. (2011) applied a mixed model-IRT methodology to further understand how the middle option is used on the JDI. They found that three classes emerged with one class more likely to use the middle option than the other classes—that is, these individuals were more likely to respond with the ? than either *Yes* or *No*. The authors found that individuals were more likely to use the middle option when they did not want to divulge their true opinions. Although it is thought that the middle response option is useful for conveying neutral standing on the attribute being assessed, the results of these studies suggest that individuals are using the middle option for reasons other than indicating neutral standing on the attribute, thus raising questions about the appropriateness of the middle response option.

Moreover, in a series of experiments, Nowlis et al. (2002) showed that the distribution of responses is systematically altered when individuals are ambivalent (i.e., simultaneously holding positive and negative sentiment) rather than univalent. They also demonstrated that respondents’ tolerance for ambiguity and need for cognition moderated the impact ambivalence had on response scale use. Presser and Schuman (1980) showed that the middle option was used by individuals that lack strong, crystallized opinions. Finally, Carter et al. (2012) showed that people in cultures characterized by high uncertainty avoidance used the middle option substantially *less* frequently than people in cultures characterized by low uncertainty avoidance. Again, these results show that the use of the middle option is related to factors other than conveying neutral standing on the attribute being measured.

Each of the aforementioned studies was conducted using scales that followed a dominance style response process rather than an ideal point process. Dominance style responding avoids the use of neutral items; therefore, having a middle response option on the response scale (e.g., a *neither agree nor disagree* response option) is preferable in that it allows people with neutral attributes to convey their neutrality (e.g., Kalton et al. 1980; Likert 1932). Stated differently, because moderately worded items are excluded on Likert scales, neutrality on the attribute continuum is inferred from repeatedly selecting the middle option in response to a series of extreme items. When a middle response option is not provided on Likert scales, researchers have argued, respondents may be forced to falsely represent their attribute level (e.g., Kalton et al.

1980). Based on the characteristics of the dominance response process, therefore, an odd-numbered response option scale—allowing for a middle option—is most appropriate for Likert-type scales.

In contrast to the dominance response process, ideal point responding infers neutrality from the items, not the response scale. Items that are located near the neutral point of the latent attribute continuum should be endorsed by people who are neutral on the attribute. As Thurstone (1925, 1928) noted, and Roberts et al. (1999) echoed, a person's attitude or personality is assessed by *which* items they endorse. Stated differently, when responding to ideal point measures, a person's standing on the attribute continuum is inferred when that person agrees to items that are close to their standing on the latent attribute and disagrees with items that are further away from their standing.

Continuing, Roberts and colleagues (see also Andrich 1996) noted that responding to ideal point measures follows a particular pattern. As stated above, the probability of endorsing an item increases as the difference between the item's location and the respondent's location on the attribute continuum is minimized. Therefore, Roberts and Laughlin (1996) observed that items are rejected from above the item's location (i.e., the item is too low on the continuum for a given respondent to endorse) or from below the item's location (i.e., the item is too high on the continuum for a given respondent to endorse). This led them to posit that every observed response option on a scale has two subjective meanings that are differentially inferred by the relation between a respondent's location and an item's location on the latent attribute continuum. For example, suppose a respondent disagrees with the item “Although I like going to parties, I also enjoy time alone” by selecting the response option *disagree*. In traditional Likert scoring, this response suggests that the person is low on the trait. In ideal point scales, however, the respondent might have selected this response option because he/she is higher on the trait (i.e., he/she only likes to go to parties and does not like spending time alone) or is lower on the trait (i.e., he/she never likes to go to parties and only likes spending time alone) than the item's location implies. Hence, there is a sort of ambiguity present when attempting to interpret the meaning of an observed response given that the same *disagree* response can be used to indicate different levels of the trait.

In some instances the item's location and the person's location are nearly identical. In these cases, the respondent should—to some extent—agree with the statement. Therefore, individuals with extreme positive and extreme negative standing on the latent attribute will agree to items that are close to the high and low end of the latent attribute continuum, respectively. These individuals would disagree with items that are on the opposite end of the latent

attribute continuum because, as noted above, the item and person locations are too different. Individuals with attribute levels near neutral will agree to items that are neutral, and disagree with items that have locations further away (in both directions) from neutrality. Like the situation for disagreement, an individual can agree to an item from above and below the item.

To clarify the agreement from above and below concept, consider a scenario in which two respondents select the same *agree* response option to the item noted above (i.e., “Although I like going to parties, I also enjoy time alone”). A respondent who indicates *agree* from above may like to go to parties slightly more often than spending time alone; conversely, a respondent who indicates *agree* from below may like to spend time alone slightly more often than going to parties. Here, these two individuals indicated the same observed response but from slightly different points on the latent attribute continuum. In the dominance style of responding, agreeing to this item would imply higher standing on the trait; however, the ideal point style of responding indicates a more moderate position on the trait. Implicit in the ideal point response process is the idea that the person is either above or below the item's location when agreeing or disagreeing with the item (Roberts and Laughlin 1996; Roberts et al. 2000). Importantly, this process holds for all items on an ideal point scale. Subjective response options exist for agreement and disagreement to extreme-worded items as well; that is, according to the tenets of ideal point responding, respondents can “agree/disagree from above/below” extreme-worded items. For these extreme items middle response options are also inappropriate.

The process of ideal point responding suggests that offering a middle response option is inappropriate for two reasons. First, the middle response option was introduced by Likert to provide respondents with a means of self-reporting their neutral feelings to a set of extreme positive/negative items. Researchers infer that a respondent has a near neutral standing on the latent attribute when that respondent endorses the middle response option across several extreme scale items. The ideal point response process, however, suggests that a respondent's neutral standing on the latent attribute need not be inferred by the response options that he or she selects. Instead, a respondent's neutrality is inferred by examining the degree to which a respondent endorses the neutral *items* that are present in a scale. In short, the processes underlying ideal point responding suggests that standing on the attribute of interest is inferred “...based on the opinions that [respondents] accept or reject” (p. xii, Thurstone and Chave 1929); respondents self-report their neutrality by endorsing (i.e., accepting) neutral items and rejecting the extreme positive and extreme negative items.

Second, a particular objective response option has two corresponding subjective options with which it is associated (i.e., from above and from below). If respondents are following the ideal point response process, individuals with extreme positive and extreme negative standing on the latent attribute continuum will disagree with neutral items for different reasons (i.e., the items are not positive or negative enough, respectively). The same neutral items will be endorsed by individuals with neutral standing because the location of the item (i.e., neutral) matches the standing of the individual (i.e., neutral). Based on the difference between the individual's and item's locations, extreme positive items will be endorsed by extreme positive individuals and rejected by extreme negative and neutral individuals. For the same reason, extreme negative items will be endorsed by extreme negative individuals and rejected by extreme positive and neutral individuals. It makes little conceptual sense, according to the tenets of ideal point responding, for respondents to say they “*neither agree nor disagree from above*” or “*neither agree nor disagree from below*” an item. Indeed, it is hard to envision where on the latent attribute continuum an item and individual would need to fall such that the difference between the item's and respondent's location makes the respondent select the middle response option to that item while still self-reporting their standing on the attribute. In short, the ideal point response process does not allow for observed responses that are not explicit agreement or disagreement.

Indeed, if, according to the ideal point response process, respondents are expected to accept or reject items based on their standing on the attribute, items to which the middle response is a viable option may actually represent an irrelevant item (Thurstone 1928; Thurstone and Chave 1929). More specifically, if respondents use the middle option when responding to an ideal point scale item, it could suggest that explicit acceptance or rejection of the item is related to factors other than the construct at hand—that is, the item would be irrelevant to the construct being measured if the item cannot be accepted or rejected to convey attribute standing (Thurstone 1928). In fact, the results of the response scale studies cited above suggest that the neutral response might be used for many other reasons than conveying attribute standing lending support to this idea. It is possible, for example, that the middle response option may simply serve as a dumping ground for responses when a respondent has failed to interpret or understand the item (e.g., Kulas et al. 2008).

In sum, based on the above rationale, we submit that middle response options in ideal point scales are both unnecessary and at odds with the principles that underlie ideal point responding. Therefore, we posit that an even-numbered response option scale (i.e., one lacking a neutral response option) is the most appropriate choice for ideal point measures.

The Current Study

To investigate the claim that an even-numbered response option scale is more appropriate for ideal point scales, we administered the same ideal point scale under three quasi-experimental conditions. Individuals in each condition received a four-, five-, or six-option response scale, respectively. Responses from each condition were then compared with regard to one main criterion, model-data fit. We argue that an odd-numbered response scale does not match the principles of ideal point responding; therefore, individuals who received an odd-numbered response scale would not follow the ideal point response process. Stated differently, because individuals who received the even-numbered scales would follow the theoretical ideal point response process more closely than the individuals who received the odd-numbered scale, an ideal point IRT model would better fit even-numbered response scale data than odd-numbered response scale data.

Hypothesis Model-data fit will be better in the even-numbered response scale conditions compared to the odd-numbered response scale condition.

Note that we limit our discussion to the model-data fit criterion because, if a model does not fit the data (as is expected for the odd-numbered scale condition), interpretation of parameters (and information derived from said parameters) for this condition would be inappropriate (Chernyshenko et al. 2001). Similar to how it is inappropriate to interpret the conclusions from structural equation models that do not fit the observed data (e.g., Bollen 1989), it would be inappropriate to interpret parameters from an IRT model that did not fit the data (Chernyshenko et al. 2001). Therefore, we looked to confirm the mismatch between response scale and response process by showing that using an odd-numbered response scales resulted in misfit that would render subsequent interpretation of the IRT analysis inappropriate thereby providing strong support for the use of even-numbered response scales for ideal point measurement.

Moreover, we note that the demonstration conducted here utilized a scale measuring a construct that may be of little interest to most organizational researchers, namely attitudes toward abortion. That is, we analyzed responses to Roberts et al. (2000) attitudes toward abortion scale. This scale was selected because it has been shown to consistently fit the ideal point response process, and the scale contains items known to be located in the middle of the attitude continuum (e.g., Roberts et al. 2000, 2001, 2010). The goal of this study was to aid researchers and practitioners in the development of ideal point scales by outlining the role of the response scale—an all too often overlooked aspect of scale construction.

To facilitate this, we utilized scale data to which the ideal point model has shown consistently good fit. Although past research has shown that ideal point IRT models fit response data from workplace attitudes (e.g., Carter and Dalal 2010), personality (e.g., Stark et al. 2006), and vocational interest (e.g., Tay et al. 2009) better than dominance IRT models, there is little consistency in the number and nature of neutral items. Moreover, some (e.g., Chernyshenko et al. 2007) have tried to create ideal point scales measuring constructs of interest to organizational researchers (e.g., orderliness). Again, however, the number and nature of neutral items in these scales have not been systematically replicated. Utilizing one of these scales in this study would confound lack of fit due to the scale with lack of fit due to the nature of the response scale offered. Indeed, as Chernyshenko et al. (2001) noted, "...the issue of determining the generality required of an IRT model would be more complex because it becomes difficult to separate problems due to 'bad' items from problems due to inadequate IRT models" (p. 527). Although the items in the above noted scales are not "bad," ideal point IRT models have not been fit to response data from these scales as frequently as the Roberts et al. (2000) attitudes toward abortion scale. We believe utilizing a scale to which ideal point IRT models have consistently shown good fit addressed this potential confound.

Therefore, we selected an attitude toward abortion scale to which ideal point IRT models have consistently shown good fit and have a consistent number of neutral items. To reiterate, however, we believe that the construct under investigation is less important than the conclusions derived which extend to any construct of typical behavior (Drasgow et al. 2010). Indeed, the theoretical argument in favor of even-numbered response option scales holds for any scale thought to follow an ideal point response process. Again, as the goal of this study was to advise researchers and practitioners on one aspect of ideal point scale construction (i.e., the type of response scale to offer), we believe the results of this study provide researchers and practitioners with some much needed guidance regarding the development of new ideal point measures of organizational constructs to leverage the benefits of ideal point measurement.

Methods

Participants and Measures

Three nonrandom samples ($N = 800, 797, 668$), collected during sequential academic semesters, from the same undergraduate student population were administered a scale which assessed attitudes toward abortion, as presented in

Roberts et al. (2000). Nineteen of the 20 items were administered to the three samples—one item was discarded due to a typographical error. The three scales differed only with respect to the number of response options provided. One group received a four-point response scale anchored as follows: *strongly disagree, disagree, agree, strongly agree*. Another group received a five-point response scale anchored as follows: *strongly disagree, disagree, neither agree nor disagree, agree, strongly agree*. Finally, the last group received a six-point response scale anchored as follows: *strongly disagree, disagree, slightly disagree, slightly agree, agree, strongly agree*. We note that two-option, *disagree, agree*, and three-option, *disagree, neither agree nor disagree, agree*, conditions were not collected in light of potential estimation issues in the IRT software used to estimate the ideal point IRT model that can occur with these types of response data (see Roberts and Thompson 2008).

In developing this scale, Roberts et al. (2000) wrote an initial set of 50 items to assess attitudes toward abortion. These items were written to cover the whole attitude continuum. After pilot testing and item analyses, a final set of 20 items was retained. An example anti-abortion item is "Abortion is unacceptable under any circumstances." An example pro-abortion item is "Abortion should be legal under any circumstances." Finally, an example moderate item is "There are some clear situations where abortion should be legal, but it should not be permitted in all situations." Roberts et al. demonstrated that responses to these items follow the ideal point response process (scale items are presented on Roberts et al. 2000, p. 20).

Model Parameter and Fit Estimation

To obtain item and person parameters, the Generalized Graded Unfolding Model (GGUM)—an ideal point IRT model—developed by Roberts et al. (2000) was fit to the data. The parameters were estimated using the GGUM2004 software (Roberts et al. 2006). Item parameters were estimated using the marginal maximum likelihood technique and person parameters (i.e., person scores) were estimated using an expected a posteriori method.

Model-data fit was assessed using the MODFIT 2.0 computer program (Stark 2007). This program computes model-data fit plots and χ^2 error values for item singles, doubles, and triples. Drasgow et al. (1995) noted that the adjusted (to $N = 3,000$) χ^2/df ratios less than 3.00 indicate good model fit. The adjustment is to insure the fit indices are comparable across sample sizes; therefore the χ^2/df ratios are adjusted to an $N = 3,000$ (see also, Tay et al. 2011). Recently, however, Tay et al. (2011) suggested that this criterion is inappropriate for item doubles and triples in studies with $N < 3,000$, and have suggested "reexamining" this criterion (p. 293). Therefore, to explore fit we inspected

item singles for adjusted χ^2/df ratios less than 3.00, but compared adjusted χ^2/df ratios for item doubles and triples across conditions such that smaller values were considered better fitting (e.g., Tay et al. 2011).

Results

Assessment of Unidimensionality

The GGUM is a unidimensional unfolding IRT model. To insure the attitudes to abortion scale items were essentially unidimensional, we conducted a principle components analysis (PCA). Roberts et al. (2000, p. 18) suggested that an item can be considered unidimensional if its communality derived from two principle components is greater than or equal to 0.30. Therefore, we forced a two-factor PCA for the four-, five-, and six-options conditions to determine if the communalities were greater than or equal to 0.30; results suggested that all items could be considered unidimensional in all conditions.

Assessment of Fit

Recall that IRT relates item and respondent characteristics to the probability of selecting a particular response alternative. This probabilistic relation is captured in the option response functions (ORFs) which depict the probability of any response alternative being selected given a certain level of the attribute (i.e., θ). Model-data fit assessment asks to what extent predicted responses, estimated from an IRT model (e.g., the GGUM), matches observed responses (Chernyshenko et al. 2001). To this end, we explored adjusted χ^2/df ratios for item singles, doubles, and triples. Whereas item singles represented how well the model was able to predict responses to a single item, doubles and triples assessed how well the model was able to predict *patterns* of responses. Moreover, misfit at the singles level would suggest that the ORFs are not adequately modeling the response process; however, misfit in the doubles and triples would suggest potential violations of the local independence (i.e., the assumption that the only factor impacting item responses is the latent trait).

Table 1 presents the adjusted χ^2/df ratios—a measure of the divergence between modeled and observed item responses—for each sample when the GGUM was fit to the responses. As discussed earlier, Drasgow et al. (1995) noted that adjusted χ^2/df ratios less than 3.00 suggest acceptable fit (applied to item singles only). As Table 1 shows, the model did *not* fit the data for the five-option scale, but did fit the data for both the four- and six-options scales. Indeed, the results of the item singles analysis

suggest that the respondents did not follow the ideal point response process in the odd-numbered condition, but did for the even-numbered conditions.¹ Moreover, the superior fit in the six-option condition was not simply due to more parameters being estimated than in the five-option condition as evidenced by the fit of the four-option condition being superior to the five-option condition as well. Given that item single adjusted χ^2/df ratios represent how well the estimated ORFs capture the observed data, these results suggest that the ORFs for the five-options condition were not an accurate probabilistic account for the responses. Although this could imply that the GGUM is not a viable model for the responses to the attitude toward abortion scale items, the good model-data fit in the four- and six-option conditions suggests that the response scale offered in the five-option condition and the response process implied by the ideal point model did not match. That is, given that the exact same items were used in each of the three conditions, the poor fit of the five-option condition cannot be attributed to the items. Moreover, that both even-numbered response scale conditions were fit well suggests that sampling bias did not likely account for the fit differences. The most likely explanation for the misfit in the five-option condition, therefore, is the fact that probabilistic account of the response process modeled by the GGUM did not match the nature of the response scale.

In addition, the doubles and triples were uniformly lower for the four- and six-options condition than the five-option condition lending further support to the notion that the data from the five-option condition did not conform to the ideal point process despite the fact that the item content was constructed to follow the tenets of ideal point responding. Recall that the doubles and triples provide an index of how well the model can capture patterns of responses to groups of two or three items, respectively, and therein can be considered a stronger test of model-data fit (e.g., Chernyshenko et al. 2001). Moreover, simulations by Tay et al. (2011) showed that adjusted χ^2/df ratios values for doubles and triples can fluctuate with sample size, test length, and type of sample. Therefore, they recommended using the adjusted χ^2/df ratios for relative rather than absolute fit assessment. As can be seen, in the five-option condition, the model was unable to capture patterns of responses as well as in the four- or six-option conditions. Moreover, these values for doubles and triples are consistent with past studies (e.g., Stark et al. 2006; Tay et al. 2011). Based on this criterion, then, the probabilistic relation between item and person characteristics

¹ Differing sample sizes could account for the fit differences. To ensure that differing sample sizes did not substantially influence fit results, a random sample of $N = 668$ individuals from the four- and five-option conditions were taken and all analyses rerun with equal N . Results and conclusions did not differ; therefore, results with all individuals included are presented.

Table 1 Frequencies of adjusted (to $N = 3,000$) χ^2/df ratios for four-, five- and six-option response scales fit to the GGUM

χ^2/df ratios	<1	1 < 2	2 < 3	3 < 4	4 < 5	5 < 7	>7	Mean	SD
Four-option scale									
Singles	17	1	0	1	0	0	0	0.29	0.87
Doubles	0	0	0	2	0	2	17	16.57	9.60
Triples	0	0	0	0	0	1	8	13.82	3.23
Five-option scale									
Singles	4	2	0	3	2	3	5	4.55	3.58
Doubles	0	0	0	0	0	0	21	19.25	7.08
Triples	0	0	0	0	0	0	9	18.62	4.31
Six-option scale									
Singles	18	1	0	0	0	0	0	0.11	0.34
Doubles	0	0	1	2	5	4	9	7.46	3.82
Triples	0	0	1	0	2	3	3	6.92	3.80

captured in the ORFs was less able to model the patterns of responses in the five-option condition compared to the four- and six-option conditions. Based on the fit of item singles, doubles, and triples, we found support for the hypothesis that an odd-numbered response scale results in inferior model-data fit compared to the even-numbered response scale suggesting that using an odd-numbered response scale may not match with the ideal point response process.

To further demonstrate this, we inspected the response option fit plots for extreme and meaningfully neutral items. As Drasgow et al. (1995) noted, “Fit plots are primarily useful for discovering systematic misfit of a few aberrant response functions or a set of items over a particular θ [i.e., latent attribute continuum] range” (p. 163). Option fit plots show the overlay of observed and expected (as estimated by the model) probability of endorsing the particular response option across the range of the latent attribute continuum. Stated differently, option fit plots allow one to see where along the latent attribute continuum a particular response option is predicted to be used relative to when it was actually used. Large discrepancies between the observed and predicted usage of an option show that the model is unable to adequately reproduce the pattern of response option usage; that is, the particular response option is being used in a way that the model cannot adequately capture.

Option response fit plots for the middle response options to an extreme positive item, an extreme negative item, and two meaningfully neutral items are shown in Figs. 3, 4, 5, and 6, respectively. Roberts and Shim (2008) noted that item location estimates that are within the 10th and 90th percentiles of the distribution of attribute estimates can be considered as falling within the range of neutrality—that is, the items shown in Figs. 5 and 6 can be considered meaningfully neutral. These items were chosen because their item stems (see figure captions) best match typical extreme and neutral items. Moreover, these items show the

general trend in the data most clearly. Large divergences in conjunction with wide error bars are indicative of poor fit.

As can be seen in these examples, the *disagree/agree* and *slightly disagree/slightly agree* options (from the even-option response scales) were fit appreciably better than the *neither agree nor disagree* option (from the five-option response scale). Some aspects of these figures are worth noting. First, this misfit was seen primarily near 0 on the latent attribute continuum. That is, the middle response option was misfitting for individuals who would be characterized as having neutral attitudes toward abortion—the group for whom the middle category would be most appropriate in Likert/dominance scaling.

Second, as noted by a reviewer, the observed response functions are not symmetrical, especially for the middle response option. Recall that the observed response curves show the probability of using the response option along the latent attribute continuum. As seen in Figs. 3 and 4, for example, individuals with near 0 standing on the attitude continuum (i.e., neutral attitudes toward abortion) were the most likely to use the middle response option in actuality (i.e., the observed proportion of middle option responses). Observed rates of using the option, in general, decreased as attribute standing moved away from 0 in either direction. However, around an attribute standing of +1 for the item in Fig. 3, and -1 for the item in Fig. 4, the observed probability of using the middle response option increased slightly causing an asymmetry in the observed response functions. The +1 and -1 positions correspond to individuals with attitude standing in the direction of the item (i.e., slight pro-abortion and slight anti-abortion, respectively).

The asymmetry in observed response functions discussed above highlights the difficulty with interpreting middle response option usage under the assumptions of ideal point responding insofar as a second level of the trait appears to utilize the option. Stated differently, a mixture of classes of

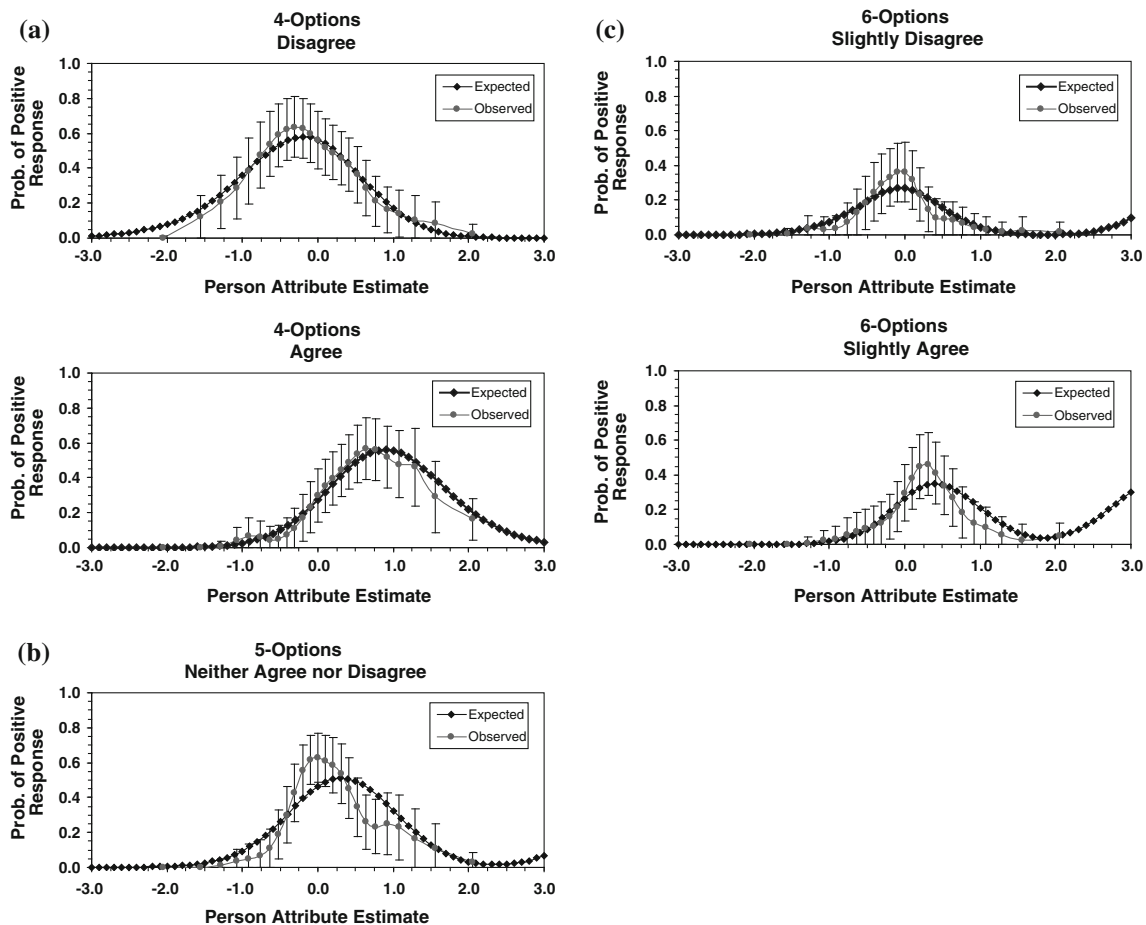


Fig. 3 Response fit plot for the middle options to an example positive dominance item (item number 19). **a** Four-point scale disagree/agree options, **b** five-point scale middle option, and **c** six-point scale slightly disagree/agree options. Shows the correspondence between observed

people who utilized this response option for different reasons may be present. As noted previously, Hernández et al. (2004) and Carter et al. (2011) showed that mixtures of individuals tend to use the middle response option on dominance scales for reasons other than inferring neutral standing on the attribute. For example, Carter et al. (2011) found that a latent class that relied heavily on the ? response to the JDI showed lower Trust in Management than those in latent classes that tended not to use the ? option. A similar situation may have occurred with this ideal point scale.

Incidentally, this issue can explain the large jump in doubles and triples for this condition. Recall that doubles and triples can indicate violations of local independence. This mixture of classes suggests that, in addition to one's attitudes toward abortion, there may have been some additional factor contributing to responses. The impact of this second factor appears to have been more pronounced for the five-option condition than the four- or six-option conditions. Moreover, these results highlight the difficulty of understanding what an observed response of the *neither*

and expected probabilities of selecting the response options based on the respondent's standing on the latent attribute continuum. Item stem: *abortion should legal under any circumstances*

agree nor disagree means when trying to assess one's attitude standing with an ideal point scales.

Finally, as can be seen, the model underpredicted the usage of the middle responses at some points along the attribute continuum and overpredicted the usage for others. Indeed, panel b in Figs. 3 and 4 show that the GGUM modeled the *neither agree nor disagree* option to be used most frequently by individuals with attitude standing slightly in the direction of the item (slightly pro-/anti-abortion, respectively); however, individuals with somewhat more neutral attitudes toward abortions were the ones who actually used the option the most (as indicated by the observed function). This pattern of misfit between the observed and modeled matched the *slightly agree* option in the six-option condition, though to a lesser extent. This was the general pattern across the items, and confirms that the middle option is not an appropriate match with the tenets of ideal point responding insofar as the model appeared to treat the middle response option as slight agreement. That is, it appears that the model treats anything that is not

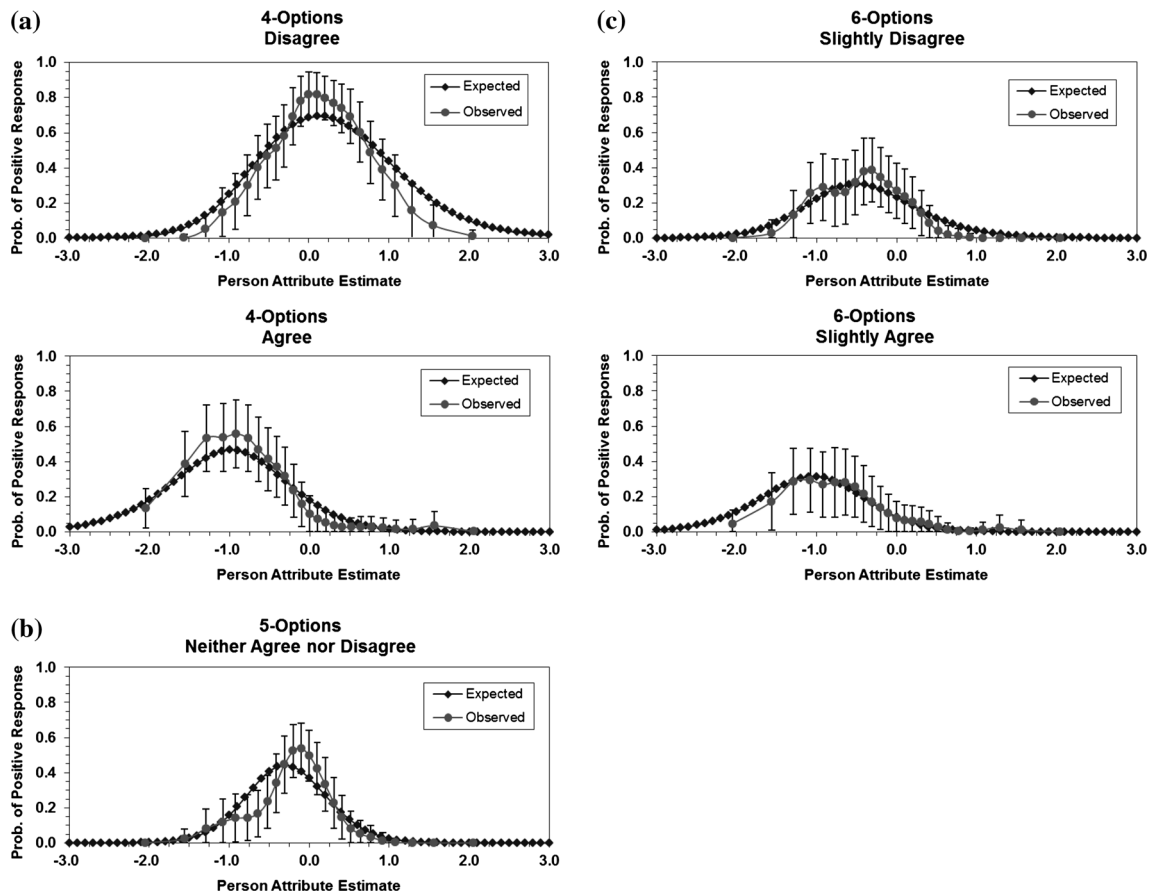


Fig. 4 Response fit plot for the middle options to an example negative dominance item (item number 1). **a** Four-point scale disagree/agree options, **b** five-point scale middle option, and **c** six-point scale slightly disagree/agree options. Shows the correspondence

between observed and expected probabilities of selecting the response options based on the respondent’s standing on the latent attribute continuum. Item stem: *abortion is unacceptable under any circumstances*

explicit disagreement with the item as slight agreement. Recall that the tenets of ideal point responding suggest that the respondent should, to some extent, endorse or reject an item—it seems that the mismatch between the response scale and response model led to the middle option being treated as slight agreement for extreme items.

As pointed out by a reviewer, the over/underprediction of the middle option may have also impacted the modeling of the neighboring response options. Indeed, although not depicted, where the model overpredicted and underpredicted the use of the middle responses option, the model underpredicted and overpredicted, respectively, the usage of the *disagree* option in the five-option condition. These results suggest that the inclusion of the middle option on the response scale not only resulted in that option being incorrectly modeled, but also impaired the modeling of the neighboring response options. Finally, we note that these figures are examples of the general trend in the data;² that

is, for none of the 19 items did the middle, *neither agree nor disagree* response fit better than the middle options for the even-numbered response option scales. As we noted above, the inappropriateness of the middle response option holds for all items on an ideal point scale. That is, although it may appear that middle response options are more confusing or problematic for neutral items, we argued that the middle option is inappropriate for all items on an ideal point scale. Indeed, if the middle response option was more problematic for neutral items, we would have seen the extreme positive and negative items fitting better than the meaningfully neutral items. Instead, misfit for each group of items was generally the same.

Discussion

Here, we have argued that the principles underlying ideal point responding support using an even-numbered response option scale; that is, a response scale that excludes a middle

² All figures are available from the corresponding author.

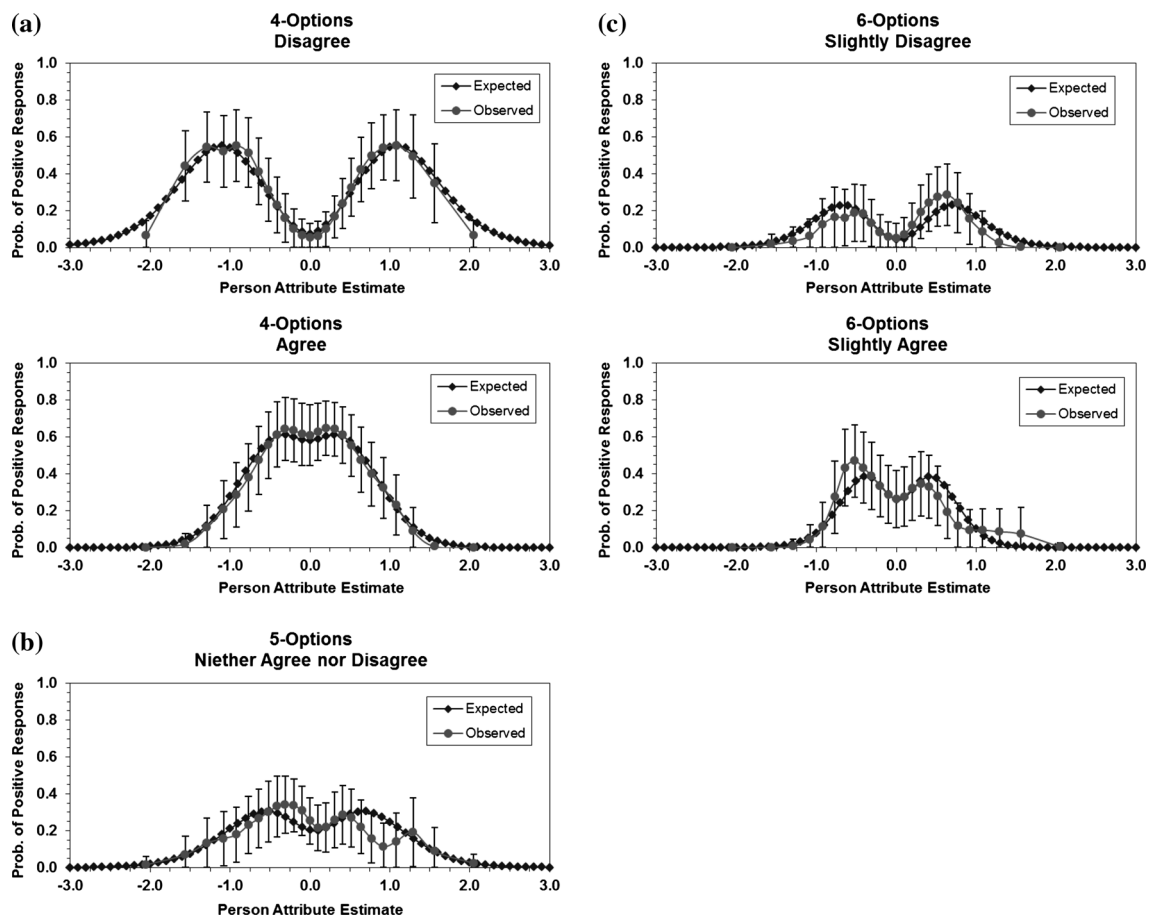


Fig. 5 Response fit plot for the middle options to an example meaningfully neutral item (item number 11). **a** Four-point scale disagree/agree options, **b** five-point scale middle option, and **c** six-point scale slightly disagree/agree options. Shows the correspondence

between observed and expected probabilities of selecting the response options based on the respondent's standing on the latent attribute continuum. Item stem: *I cannot whole-heartedly support either side of the abortion debate*

response option. As noted earlier, the prevailing view with respect to the ideal point response process is that each objective response option is associated with two subjective positions from which the respondent could be answering (i.e., from above or below the item's location). It is difficult to conceive of a location along the attribute continuum wherein an individual could respond *neutral*, *neither agree nor disagree*, or *?* to an item from above or below that item. Indeed, it is possible that offering such a response option may only serve as a catch-all category for respondents who are unsure how to respond to an item—this situation has been observed in dominance measurement where a middle option is necessary (e.g., Kulas et al. 2008).

Based on this reasoning, we expected that an ideal point IRT model would not be able to adequately model observed responses of the middle option (e.g., *neither agree nor disagree*) to ideal point scale items; the result would be poor model-data fit when responses are made on a scale containing a middle option. The current study found support for this hypothesis; the ideal point IRT

model showed good fit to responses from even-numbered response scales (i.e., four- and six-options), but not responses from an odd-numbered response scale. Moreover, inspection of the option fit plots (Figs. 3, 4, 5, 6) confirmed that the middle, *neither agree nor disagree* response option was being modeled poorly compared to the two options that differentiated the middle option for the four- and six-option conditions. This was true even though the five-option response scale had more parameters than the four-option response scale from which to model response behavior; in all, results suggested that the superior fit of the even-numbered response scales was neither due to modeling more parameters nor due to sample differences, but was likely due to the mismatch of the odd-numbered response scale and the ideal point response process.

It was noted earlier that it is possible that a middle response option on ideal point scales may serve as a “dumping ground” for responses in instances where an individual does not know how to respond. Although other

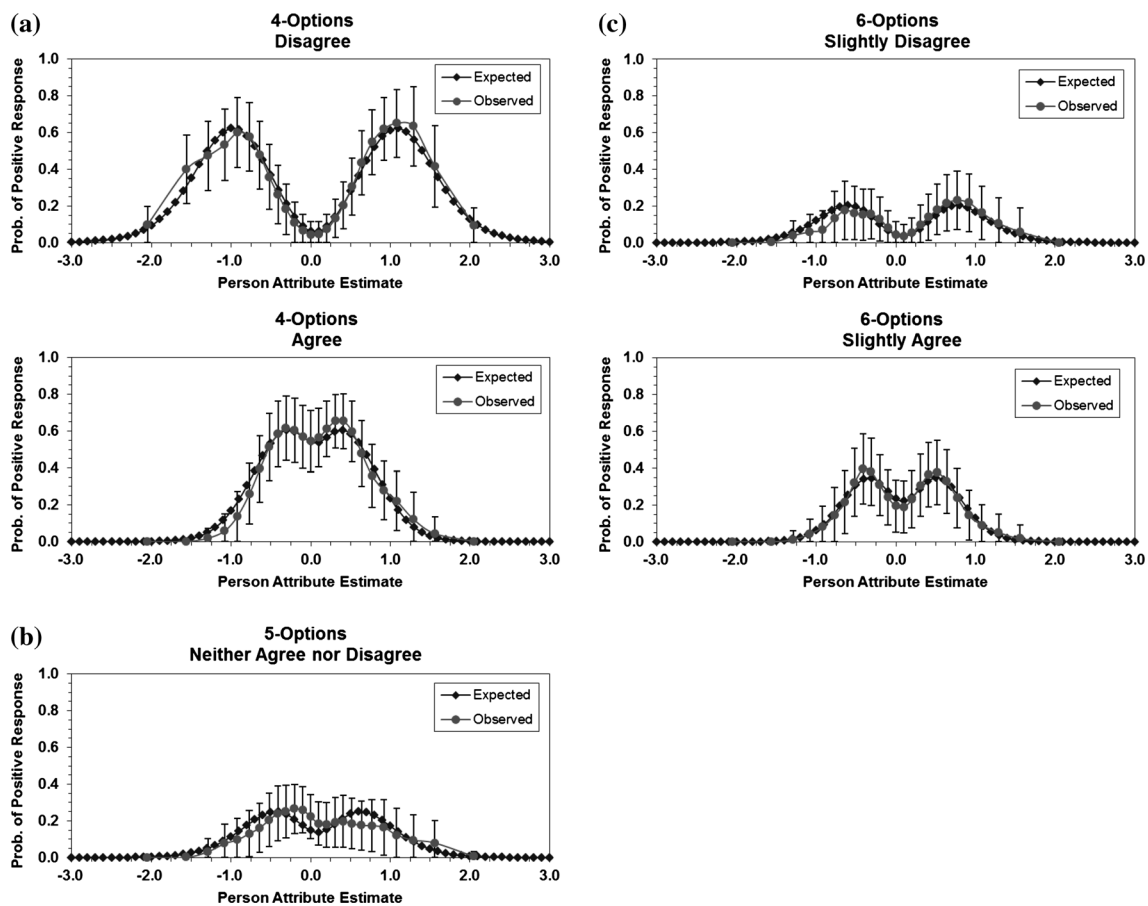


Fig. 6 Response fit plot for the middle options to an example meaningfully neutral item (item number 12). **a** Four-point scale disagree/agree options, **b** five-point scale middle option, and **c** six-point scale slightly disagree/agree options. Shows the correspondence

between observed and expected probabilities of selecting the response options based on the respondent’s standing on the latent attribute continuum. Item stem: *my feelings about abortion are very mixed*

explanations exist (e.g., attitudinal indifference, weak attitude strength, etc.), the lack of fit for this response option into the five-option condition lends some support to this notion. Stated differently, the poor fit of the *neither agree nor disagree* response option provided some support to the idea that this option is not being used in accordance with the assumptions of ideal point responding.

Although the specific reason(s) for the middle option misfitting will need to be further explored, the results of the study showed that including a middle response option on ideal point scales can actually result in modeling the wrong response process. Specifying the wrong response model can impair practical uses of IRT including misestimating person scores, and the inappropriate application of differential item functioning (DIF) and computer adaptive testing (CAT). Stated differently, assuming the wrong response model can result in invalid estimates of attribute standing (Roberts et al. 1999), incorrect conclusions from DIF analyses, and miscalibrated CAT sessions (e.g., Tay et al. 2011); these results suggest this is possible if a middle response option is offered on ideal point scales.

Recommendations for Research and Practice

As a reviewer notes, questions may be raised about why ideal point measurement is better relative to the more straightforward dominance approach. The primary advantage of ideal point measurement is the more reliable and valid indexing of attribute standing. That is, because ideal point scales contain items that are located across the range of the attribute continuum, these scales provide more psychometric information (i.e., measurement precision) across the range of the attribute. Dominance scales, however, tend to have items located near one another resulting in measurement precision in a relatively narrow range of the attribute. Therefore, using ideal point scales should improve the precision with which attributes are measured (Chernyshenko et al. 2007). More reliable measurement will result in more valid test-score inferences.

As the goal of this paper was to provide researchers and practitioners with advice on one aspect of ideal point scale construction, we offer some recommendations based on these results. In general, a question is raised when one asks

what it means to be neutral to an ideal point item, suggesting that a middle response option should not be offered on ideal point scales. The results of this study shed additional light on this concern. When the items are extreme, a middle option may appear to be necessary. This is only true, however, if the respondents are assumed to follow a dominance response process wherein neutral standing on the attribute is inferred from selection of the middle response option. When administering a scale to which respondents are assumed to follow the ideal point response process, on the other hand, middle response options are inappropriate because respondents are assumed to endorse or reject each item (e.g., Roberts et al. 1999; Thurstone and Chave 1929). When coupled with items throughout the range of the latent attribute distribution, standing on the latent attribute distribution can be estimated based on the pattern of responding to all the items.

Based on the results of this study, then, we recommend that an even-numbered response scale that excludes a middle response option be used with ideal point scales—this should allow the ideal point response process to be followed more closely. More specifically, we suggest scale developers use a six-option response scale with ideal point scales for two main reasons: (1) the six-option response scale condition had the best model-data fit, and (2) the larger number of parameters will result in more psychometric precision (i.e., IRT information). For these reasons, then, we believe a six-option response scale is most appropriate for ideal point scales.

Here, we contend that a middle response option should be excluded when using ideal point measures; that is, researchers and practitioners creating ideal point scales should use an even-numbered response option scale. It is entirely likely, however, that given the lack of attention paid to the response scale on ideal point measures, readers may have already collected data using a middle anchor. We offer some practical guidance for such situations. First, the conceptual appropriateness of the middle option should be considered. One should consider, for example, whether there is a compelling reason for a person to select the middle option in response to items on ideal point scales.

The second consideration involves the empirical appropriateness of the middle option—that is, the degree to which an IRT model can accurately represent the way in which people use response options. Interestingly, ideal point IRT models can provide a more reasonable probabilistic account of the middle option for neutral items than dominance IRT models. More specifically, whereas the dominance model required a unimodal response function for the middle option, the ideal point model allows for a bimodal curve. This means that persons of different attribute levels may choose the option for different reasons (see Carter and Dalal 2010). In short, it is possible that some scale data collected

using odd-numbered response scales may still be sufficiently modeled by an ideal point IRT model. That is, it is possible for ideal point IRT models to fit responses to scales with an odd-number response options (e.g., Carter and Dalal 2010; though, likely not as well as if that data were collected using an even-numbered response scale). Indeed, this is possible when scales show few, if any, meaningfully neutral items, and/or if the scale was originally developed under dominance responding assumptions. This study utilized a scale that did not meet either of these criteria; therefore, the odd-numbered response data were not modeled well. In light of these findings, however, and in conjunction with the conceptual inappropriateness of a middle anchor, an even-numbered response scale should be used with ideal point scales whenever possible.

Limitations and Future Directions

In light of some evidence that ideal point response models can model middle response options (e.g., Carter and Dalal 2010), one may wonder to what extent these findings are likely to generalize to other scales beyond the one examined here. Importantly, the amount of misfit may be scale specific, but the logic of excluding the middle option holds for every ideal point scale. That is, whereas the theoretical rationale for an even-numbered response scale holds for all ideal point scales, the empirical consequences will depend on numerous factors. As just noted, this can include things like the number of meaningfully neutral items and/or whether or not the scale was originally developed under ideal point assumptions. An additional explanation may be attitude strength. That is, it is possible that attitudes that are held more strongly (e.g., attitudes toward abortion) are likely to result in worse modeling of the middle option than attitudes that are held less strongly (e.g., attitudes toward prisoners). For these strongly held attitudes, respondents are likely to discriminate well between items and response scale points. Indeed, recent work (Lake et al. 2013) has shown that when attitudes are held strongly, response latitudes are narrower suggesting better discrimination between response options when responding to dominance items. Although this work was conducted with dominance scales, a similar process may hold for ideal point scales such that more strongly held attitudes are likely to result in worse probabilistic accounts of the middle response option. That is, because respondents distinguish between response options better when attitudes are held more strongly, the inappropriateness of the middle option would be more apparent in these circumstances. Future research should fully explore the conditions under which ideal point IRT models will be able to model the middle response option including the role of response latitudes (Roberts et al. 2010).

In addition to exploring if and when the middle response option may be modeled by ideal point IRT models, future research should replicate and extend these findings by investigating other attitude (e.g., job satisfaction) as well as personality constructs to further confirm the arguments outlined here. Indeed, Drasgow et al. (2010) argued that all attitudes and personality constructs should follow the ideal point response process.

Here, we demonstrated the negative impact on fit as a result of including a middle option in the response scale. Future research should investigate other consequences from including a middle response option; especially when the scale fits the data. As noted above, there may be instances when ideal point IRT models fit responses from ideal point scales that used an odd-numbered response scale. When the odd- and even-numbered response scale data are fit well by the ideal point model, researchers can investigate the impact on test and item information, item location estimates, and even person attribute estimation, to name a few. These analyses were not possible here because the ideal point IRT model did not fit the five-option condition. When the model does not fit the data, "...IRT results may be suspect" (p. 524, Chernyshenko et al. 2001). In short, it would have been inappropriate to interpret the results of the five-option condition in light of the fact that the model did not fit. Therefore, investigations into more specific consequences of the including a middle response option were not possible here.

In addition, future research can look to explicitly test alternative explanations for middle response option use on ideal point scales (e.g., dumping ground, respondent indifference, respondent confusion, etc.) to understand when and why respondents might utilize the neutral option on an ideal point scale. Moreover, more research and guidance is needed on item creation (e.g., Carter et al. 2010; Dalal et al. 2010). Although the response scale is important, a more complete understanding of how to create items that follow the ideal point process is needed to complement these results regarding the response scale. Finally, ideal point scales have been criticized for being difficult to observe-score (e.g., Dalal et al. 2010). Although the IRT models are valuable for scoring ideal point scales, they require large sample sizes that are prohibitive for most research and practice purposes. Therefore, along with how to create ideal point scales, a straightforward method for observe-scoring ideal point measures will help facilitate the widespread adoption of ideal point measurement in research and practice.

References

- Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology*, 49(2), 347–365.
- Andrich, D., & Styles, I. M. (1998). The structural relationship between attitude and behavior statements from the unfolding perspective. *Psychological Methods*, 3, 454–469.
- Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51, 220–232.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Carter, N. T., & Dalal, D. K. (2010). An ideal point account of the JDI Work satisfaction scale. *Personality and Individual Differences*, 49, 743–748.
- Carter, N. T., Dalal, D. K., Lake, C. J., Lin, B. C., & Zickar, M. J. (2011). Using mixed-model item response theory to analyze organizational survey responses: An illustration using the Job Descriptive Index. *Organizational Research Methods*, 14, 116–146.
- Carter, N. T., Griffith, R. P., Feitosa, J., Moukarzel, R., Kung, M.-C., Lawrence, A. D., & O'Connell, M. (2012, April). Predicting non-invariance across cultures using cultural uncertainty avoidance. In N. T. Carter, & A. D. Mead (Chairs), *Recent developments in personality measurement invariance: Time, culture, and forms. Symposium presented at the 27th Annual Meeting of the Society for Industrial and Organizational Psychology, San Diego, CA*.
- Carter, N. T., Lake, C. J., & Zickar, M. J. (2010). Toward understanding the psychology of unfolding. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 511–514.
- Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, 36, 523–562.
- Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Towards increasing the flexibility of personality measures. *Psychological Assessment*, 19, 88–106.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Dalal, D. K., Withrow, S., Gibby, R. E., & Zickar, M. J. (2010). Six questions that practitioners (might) have about ideal point response process items. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 498–501.
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 Years after Likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 465–476.
- Drasgow, F., Levin, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19, 143–166.
- Droba, D. D. (1930). *A scale for measuring attitude toward war*. Chicago: The University of Chicago Press.
- Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions* (2nd ed.). New York: Routledge.
- Hanisch, K. A. (1992). The Job Descriptive Index revisited: Questions about the question mark. *Journal of Applied Psychology*, 77, 377–382.
- Hernández, A., Drasgow, F., & González-Romá, V. (2004). Investigating the functioning of a middle category by means of a mixed-measurement model. *Journal of Applied Psychology*, 89, 687–699.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1, 104–121.

- Kalton, G., Roberts, J., & Holt, D. (1980). The effects of offering a middle response option with opinion questions. *The Statistician*, 29, 65–78.
- Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business Psychology*, 22, 251–259.
- Lake, C. J., Withrow, S., Zickar, M. J., Wood, N. L., Dalal, D. K., & Bochinski, J. (2013). Understanding the relation between attitude involvement and response latitude using item response theory. *Educational and Psychological Measurement*, 73, 690–712.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 5–55.
- Likert, R., Roslow, S., & Murphy, G. (1934). A simple and reliable method of scoring the Thurstone attitude scales. *Journal of Social Psychology*, 5, 228–238.
- Nowlis, S. M., Kahn, B. E., & Dhar, R. (2002). Coping with ambivalence: The effect of removing a neutral option on consumer attitude and preference judgments. *Journal of Consumer Research*, 29, 319–334.
- Oswald, F. L., & Schell, K. L. (2010). Developing and scaling personality measures: Thurstone was right—But so far, Likert was not wrong. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 481–484.
- Presser, S., & Schuman, H. (1980). The measurement of a middle position in attitude surveys. *Public Opinion Quarterly*, 44, 70–85.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A generalized item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, 24, 3–32.
- Roberts, J. S., Fang, H., Cui, W., & Wang, Y. (2006). GGUM2004: A Windows-based program for estimate parameters in generalized graded unfolding model. *Applied Psychological Measurement*, 30, 64–65.
- Roberts, J. S., & Laughlin, J. E. (1996). A unidimensional item response model for unfolding responses from a graded disagree–agree response scale. *Applied Psychological Measurement*, 20, 231–255.
- Roberts, J. S., Laughlin, J. E., & Wedell, D. H. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement*, 59, 211–233.
- Roberts, J. S., Lin, Y., & Laughlin, J. E. (2001). Computerized adaptive testing with the generalized graded unfolding model. *Applied Psychological Measurement*, 25, 177–196.
- Roberts, J. S., Rost, J., & Macready, G. (2010). MIXUM: An unfolding mixture model to explore the latitude of acceptance concept in attitude measurement. In S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 175–197). Washington, DC: American Psychological Association.
- Roberts, J. S., & Shim, H. S. (2008). *GGUM2004 technical reference manual (v1.1)*. Atlanta, GA: Georgia Polytechnic University.
- Roberts, J. S., & Thompson, V. M. (2008, June). *Accuracy of alternative parameter estimations methods with the Generalized Graded Unfolding Model*. In Paper presented at the 73rd Annual International Meeting of the Psychometrics Society, Durham, NH.
- Stark, S. (2007). *MODFIT 2.0: A computer program for model-data fit* [Author].
- Start, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91, 25–39.
- Tay, L., Ali, U. S., Drasgow, F., & Williams, B. (2011). Fitting IRT models to dichotomous and polytomous data: Assessing the relative model-data fit of ideal point and dominance models. *Applied Psychological Measurement*, 35(4), 280–295.
- Tay, L., Drasgow, F., Rounds, J., & Williams, B. A. (2009). Fitting measurement models to vocational interest data: Are dominance models ideal? *Journal of Applied Psychology*, 94(5), 1287.
- Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *The Journal of Educational Psychology*, 16, 433–451.
- Thurstone, L. L. (1928). Attitudes can be measured. *The American Journal of Sociology*, 33, 529–554.
- Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude: A psychophysical method and some experiments with a scale for measuring attitude toward the church*. Chicago: The University of Chicago Press.
- Zickar, M. J., & Broadfoot, A. A. (2009). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 37–60). New York: Routledge.