Brief Report

Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings

John T. Kulasa a,*, Alicia A. Stachowski b

a Department of Psychology, Whitney 303, Saint Cloud State University, St. Cloud, MN 56301, USA
b George Mason University, 4400 University Drive, MS 3F5, Fairfax, VA 22030, USA

Abstract

Although typically scored as indicating moderate or neutral trait standing, personality assessment respondents endorse the Likert-scale middle response category for a variety of reasons. Through the application of a cognitive processing model and an item characteristic orientation, middle category endorsements were found to exhibit a relatively high response latency, an “it depends” connotation, and a strong, negative relationship with item clarity. These general associations stress the importance of retaining unambiguous items for trait identification but also offer a tool to the personality assessment researcher – investigating the number of elicited middle category endorsements to identify trait indicators in possible need of contextualization.

1. Introduction

From an assessment developer’s perspective, an ideal respondent is inclined to endorse the Likert-type scale’s middle category only if his or her true construct standing is between the opposing positive and negative options of the response scale continuum (e.g., the individual is moderately Extraverted). When middle category endorsements are included in scale construction, the scored item-response is implicitly interpreted as reflecting such moderate trait standing. Unfortunately, respondents do not only use this category to indicate moderate construct standing (e.g., Hofacker, 1984; Klopfer & Madden, 1980; Kulas, Stachowski, & Haynes, 2008).

1.1. Respondent intentions with middle category endorsement

Shaw and Wright (1967) posited three possible respondent intentions with middle category endorsements. Individuals may select this category when (1) they have no attitude or opinion, (2) they are “balanced” in terms of evaluation, or (3) they have not clearly defined their attitude or opinion. Several investigations note that respondents sometimes use the middle category to indicate indifference or a lack of caring (DuBois & Burns, 1975; Nowlis, Kahn, & Dhar, 2002). Some endorse the middle category out of a reluctance to reveal a valenced answer to a personal question (Tourangeau, Smith, & Rasinski, 1997) or when their interpretation of an item is unclear (e.g., Goldberg, 1981). Collectively, these investigations point toward the middle response category being at least occasionally utilized as a “dumping ground” for not applicable, uncertain, indifferent or ambivalent response orientations.

1.2. Cognitive demands and middle category endorsement

Responding to personality items involves comparing an item to one’s self-schemata (Markus, 1977; Rogers, 1974). Questions that tap into a well-defined piece of self-knowledge are easily responded to and result in quicker decisions (and consequently faster response times; e.g., Hanley, 1965). As such, difficult items elicit greater response latencies than do relatively easier items (Yang, O’Neill, & Kramer, 2002). This general item-response effect has been looked at only sparingly with specific response categories. Temple and Geisinger (1990) investigated response time differences across personality item options of “true”, “false”, and “cannot say”. “Cannot say” responses exhibited longer response times than did either true or false endorsements, suggesting greater processing load associated with this response.

1.3. Item-antecedents of middle category endorsement

Tourangeau and Rasinski’s (1988) question-answering process-model posits that questionnaire respondents proceed through four steps: (1) understanding and interpreting the question, (2) retrieving information from memory, (3) consolidating retrieved information to form a judgment, and (4) reporting the judgment (i.e., choosing an item-response). Items that present a more “difficult” response situation through interfering with one or more of these steps (such as increasing cognitive demands or engendering an
unwillingness to divulge personal information) would be expected to result in more ‘invalid’ response orientations, including the “dumping ground” use of the middle response category. We therefore investigate two item characteristics associated with increased cognitive demands: item clarity (e.g., Step One; Goldberg, 1981) and the need for respondent self-awareness (e.g., Step Three; Weemts & Onwuegbuzie, 2001) and one item characteristic related to willingness to reveal a valenced response (intrusiveness; e.g., Step Four; Tourangeau et al., 1997).

1.4. Summary and hypotheses

The question-answering and general cognitive information processing frameworks predict that individuals may use personality response scale options for different reasons, under different circumstances, and with different associated amounts of cognitive effort. Regarding endorsement of the middle response category, increased cognitive load should result in longer response times (i.e., a longer amount of time devoted to making a decision) because of: (1) the multitude of reasons people choose to endorse the middle response category, and (2) the resulting ambiguity associated with the option. With a more difficult task (i.e., unclear, intrusive, or introspective question), the “neither agree nor disagree” option may be seen as an out for cognitively difficult choices. In addition to these process and item-antecedent associations, a collection of potential meanings, including the scoring-consistent meaning of “average” was culled from previous research to investigate whether or not middle category endorsement can be predominantly associated with a preferred meaning.

**Hypothesis 1.** Relative to other response choices, “neither agree nor disagree” responses will exhibit longer response latencies.

**Hypothesis 2.** Unclear, personally intrusive, and introspective items will elicit more “neither agree nor disagree” responses than will more clear, less intrusive, and less introspective items.

**Research Question 1.** When presented with a limited number of alternative options (not applicable, uncertain, average, and it depends), do respondents who endorse the middle category favor one alternative over the others?

2. Method

2.1. Participants

One hundred and two undergraduate psychology students at a large Midwestern University participated for class extra credit. Seventy-six were female and 26 were male students.

2.2. Materials

The International Personality Item Pool (IPIP) is an internet-based item bank which at the time of investigation consisted of 2036 trait indicators. The computer program E-prime v1.1 (Schneider, Eschman, & Zuccolotto, 2002) was used to record response categories and latencies.

2.3. Procedure

One hundred items were selected from the IPIP item bank, with roughly a third selected based on subjective researcher assessment of intrusiveness, lack of clarity, and need for self-awareness. These items were presented to 51 undergraduate students who did not participate in the focal study, but who were asked to rate each item along a 5-point scale in terms of clarity (n = 16; $X = 3.86, s = .46$), need for self-awareness (n = 17; $X = 3.08, s = .48$), or intrusiveness (n = 18; $X = 2.25, s = .55$).

2.3.1. Focal study protocol

The general study procedure can be described as a within-subjects test–retest (1-week) design. One assessment administration offered typical response options of “strongly disagree”, “disagree”, “neither agree nor disagree”, “agree”, or “strongly agree” (this administration is referred to as the traditional form). Because response latency was of interest, the response options were located atypically – equidistant from the item prompt (in a ‘semi-circle’ presented above item prompts). A second assessment offered response options of “it depends”, “uncertain”, “average”, or “not applicable” (referred to as the forced-choice form). Here, response options were again placed at an equal distance around the center of the screen.

For both computer-administered forms, participants first completed 10 practice trials, and then responded to one of the 100-item assessments. The two assessments were counterbalanced for order of presentation. Each participant was instructed to respond “as quickly but accurately” as possible. Item administrations were preceded by a ready screen, asking the respondent to indicate when they were prepared to see and respond to the next question. Item presentations began with the cursor set at a location in the middle of the computer screen and response options located 165 pixels distant from the center start point.

3. Results

Fig. 1 presents the average reaction time (in milliseconds) for each traditional form option response to the 100-items across 102 individuals (Hypothesis 1). One thousand and 70 responses (1.7%) were made outside of the scor-able space (i.e., not within the specified response area). A one-way ANOVA revealed an omnibus response time difference across response option categories ($F = 36.55, p < .05$). Using Tukey’s least significant difference (LSD) criterion ($\alpha = .05$), all valid response options (strongly disagree, disagree, agree, and strongly agree) had quicker mean response times than did the “neither agree nor disagree” option. Applying the same Tukey’s LSD criterion, the “strongly agree” option was, on average, the most quickly endorsed response option (including its comparison with missed responses).

Regarding Hypothesis 2’s predicted antecedents, mean scores were created for each item’s pretest ratings of clarity, self-awareness, and intrusiveness. Rating scale options were then tallied across respondents (e.g., how many strongly disagrees, disagrees, neither agree nor disagrees, agrees, and strongly agrees were elicited for each item), and average item reaction times were computed. Table 1 presents bivariate correlations among all variables investigated in these item-level analyses (exhibiting mixed support for Hypothesis 2; predicted association with item clarity, lack of association with intrusiveness, and an opposite-from hypothesized association with ‘need for self-awareness’).

Using this item-level data for supplementary investigation of Hypothesis 1, a hierarchical regression was conducted on the average item–response time variable using the number of characters in an item as a primary covariate ($R = .75, p < .05$), item clarity, self-awareness, and intrusiveness as secondary covariates ($R = .82$, $R^2 = .11, p < .05$), and the number of strongly disagree, disagree, neither agree nor disagree, agree, strongly agree, or missing responses elicited by each item as the ultimate predictors ($R = .87$,}$
the analyses show that endorsement of the middle response category was predictive of response latency, even after taking item length, clarity, intrusiveness, and need for self-awareness into consideration.

3.1. "Meaning" of middle category endorsement

Eighty-two individuals completed both the traditional and forced-choice forms – this number differs from the numbers of individuals completing each form because some individuals did not return for their second scheduled questionnaire administration. Raw cross-tabulations of item endorsements across administrations revealed a difference in preferred forced-choice alternatives for different traditional form responses (e.g., different

\[ \Delta R^2 = .08, p < .05 \]. Table 2 presents these regression results.\footnote{Although "neither agree nor disagree" was the only significant variable in the option response third step, the agree category was excluded from this step because of a linear dependency caused by the specification of all possible item responses (tolerance estimate = 0).} The analyses show that endorsement of the middle response category was predictive of response latency, even after taking item length, clarity, intrusiveness, and need for self-awareness into consideration.

### Table 2
Hierarchical regression of response latency on item characteristics and number of elicited category responses.

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \beta )</th>
<th>( t )</th>
<th>( \Delta R^2 )</th>
<th>( \Delta F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td>.56</td>
<td>125.60</td>
</tr>
<tr>
<td>Number of item characters</td>
<td>.75</td>
<td>11.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td>.11</td>
<td>10.02</td>
</tr>
<tr>
<td>Self awareness</td>
<td>-.09</td>
<td>-1.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrusiveness</td>
<td>-.13</td>
<td>-1.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clarity</td>
<td>-.32</td>
<td>-4.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td>.08</td>
<td>5.97</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>-.14</td>
<td>-1.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>.25</td>
<td>2.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>.06</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>.03</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Missing&quot;</td>
<td>.09</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Overall model \( R^2 = .75, F_{10, 90} = 29.98, p < .05. Test of regression weights were two-tailed. Bold indicates \( p < .05. \)
likelyhoods of forced-choice endorsement across the five traditional form response categories; $X^2_{5} = 4139.45, p < .05$. This effect is highlighted in **Fig. 2**, where proportions of forced-choice form endorsements are organized by traditional form category response.

The figure reveals a tendency for “neither agree nor disagree” traditional form responses to be most commonly associated with an “it depends” response on the forced-choice form (with roughly a 50% probability). “It depends” becomes increasingly uncommon as the extreme agreement and disagreement options are approached. This addresses the research question – of the alternatives made available, is there a preference for the “it depends” alternative if “neither agree nor disagree” is selected on the traditional form.

4. Discussion

Personality assessment respondents may “neither agree nor disagree” with an item prompt because they are undecided, they do not understand the item, their response is conditional, or they have a neutral, moderate, or average construct standing. Given these options, however, there is a predominant general (across items and persons) tendency to endorse the middle category with an “it depends” response orientation, suggesting that conditional response interpretation of the category may be more common than the moderate-standing interpretation and use. This orientation is potentially useful, however, because it provides a directive for item revision: contextualization.

Our results suggest that it is not only less cognitively demanding to agree with a statement (relative to disagreeing; i.e., acquiescence), but that it is also less demanding to disagree than to endorse the middle response category (i.e., **Fig. 1**). This is consistent with the supposition that the category conveys multiple meanings and uses (obfuscation). It is also, however, consistent with the Grant, Button, and Noseworthy (1994) position that quicker extreme-point endorsement is related to more well-formulated, stable, or strong positions (i.e., response facilitation).

Although the current findings are taken as evidence of increased cognitive load associated with the interpretation and use of the middle response category, this alternative perspective is acknowledged. The regression analyses would seem to provide greater (item-level) support to our “middle category obfuscation” explanation of the **Fig. 1** effect, although we did not explicitly test the obfuscation versus facilitation perspectives – this would appear to be an interesting avenue for future research in the area of cognitive models of personality assessment.

**Kulas et al. (2008)** demonstrated that the Likert-type middle response category is at least sometimes used as a dumping ground for unsure or non-applicable responses. They further noted that the effect was pronounced when item content had been edited by replacing prompt verbs with less common synonyms. Collectively, the findings provide support for the notion of a cognitive load associated with the interpretation and use of the middle response category, this alternative perspective is acknowledged. The regression analyses would seem to provide greater (item-level) support to our “middle category obfuscation” explanation of the **Fig. 1** effect, although we did not explicitly test the obfuscation versus facilitation perspectives – this would appear to be an interesting avenue for future research in the area of cognitive models of personality assessment.

**Kulas et al. (2008)** demonstrated that the Likert-type middle response category is at least sometimes used as a dumping ground for unsure or non-applicable responses. They further noted that the effect was pronounced when item content had been edited by replacing prompt verbs with less common synonyms. Collectively, the findings provide support for the notion of a difficult survey task (operationalized as an unclear item) having an effect on response orientations, including increased middle category endorsement. Tourangeau and Rasinski’s (1988) question-answering process-model suggests that this middle response effect is most likely attributable to the first stage of the question-answering process (item interpretation).

The opposite-from hypothesized relationship between a perceived “need for self-awareness” and middle category use may reflect poor construct specification. Our pretest, in retrospect, did not do a good job differentiating the amount of future introspection/cognitive work required versus the amount of introspection/cognitive work that had already been devoted to a concept/item. Future investigations into item characteristic antecedents of response choice may be more successful pursuing related constructs such as “frequency of trait expression”, which is consistent with the Grant et al. (1994) perspective but also targets the second step of the Tourangeau and Rasinski (1988) model.

Intrusive items demonstrated no linear relationship with the frequency of elicited middle category endorsements. Intrusiveness is a construct that was not presumed to impact respondent fatigue or cognitive load, but one that was thought to contribute to a reluctance to provide an accurate/valenced response based on a willingness to divulge personal information. It is possible that the confidential data collection process, chosen items, and/or student sample used in the current study moderated the effect of intrusiveness on unwilling-

![Fig. 2. Likelihood of forced-choice response (it depends, average, uncertain, n/a) by traditional category endorsement (sa, a, n, d, sd).](image-url)
ness to endorse a valenced position. The most extreme item, “Think constantly about sex” only yielded an average intrusiveness rating of 3.67 on our 5-point scale (3 = “moderately intrusive”; 88% of rated items had average intrusiveness scores below three).3

4.1. Limitations

So as not to imply a continuum, the forced-choice response options were placed equidistant from one another around the question stem. The “it depends” and “average” options were most frequently endorsed, but these options always occupied the bottom two alternatives. Ideally, these locations would have been counterbalanced across participants. Additionally, the number of alternatives was limited to four – these were selected from prior investigations and theoretical orientations on the meaning of the middle response option. While there are other alternatives that could have been offered (e.g., “neutral”), a conundrum existed in identifying an optimal number of discriminating meanings that had previously been associated with the middle response category but also did not overload the cognitive capacity of the respondent.

Because we were interested in item antecedents of middle category endorsement, we chose items that were predicted to be less clear, more intrusive, and requiring a greater amount of self-awareness. Preexisting scales or underlying dimensions were not considered in the selection of these items. The item selection method compromised our ability to investigate the impact of middle category endorsements on scale level statistics (such as reliability or validity estimates). It would be advisable for future investigations in this area to consider a priori scale associations prior to item selection when investigating the intention, meaning, or (especially) impact of response scale interpretation and use.

Although presented as a general main effect, there was considerable cross-item variability in the strength of the “it depends” orientation. Example items that strongly conformed to this effect were: “Have excellent ideas” (95% of “neither agree nor disagree” responses were “it depends”) and “Remain calm under pressure” (82%). Example items that reversed this general pattern were “Don’t know why I’m angry” (0% of “neither agree nor disagree” responses were “it depends” upon retest) and “Experience longer periods of sadness or depression than other people seem to” (10%; although both of these items elicited a low overall number of “neither agree nor disagree” responses). The “it depends” middle response main effect, therefore, likely needs further exploration and refinement (focusing on modifying item characteristics that increase or decrease the “it depends” orientation).

4.2. Implications and future directions

Respondents prefer to have a middle option provided when they complete questionnaires and inventories (e.g., McDonald, 2004). Two broad practice-based recommendations therefore arise from this investigation. First, if the potential for “dumping ground” use of the middle response option is of concern, all administered items should be screened for clarity and/or an “it depends” response option should be presented to respondents. Second, assessment publishers and researchers who investigate and find high frequencies of middle category response may have a specific target for item revision – contextualization.

Ideally, middle category endorsements that convey a “dumping ground” orientation of cannot decide, it depends, or uncertain would be distinguished from “valid” middle category endorsements that convey orientations of equally true or false, neutral, moderate, or average. It is possible that person characteristics such as ability and motivation may prove fruitful in this pursuit. It is not completely surprising that different respondents may possess different response orientations - one respondent’s “agree” may be another respondent’s “strongly agree.” This, however, is a difference of internal calibration. More problematic is one respondent’s “it depends” and another respondent’s “moderate” or “average.” This represents a qualitative difference of category – rating dimension association, rather than a mere difference of quantitative calibration. Clearly response scales that elicit responses qualitatively different than others along an intended continuum (i.e., agreement) call into question the propriety of simple summed scale definitions.

In the absence of a predictive model of middle category meaning, the current study’s recommendations are first to generate clear, unambiguous, and easily understood item stems. This should limit the absolute number of “it depends”-oriented middle response endorsements. This of course is an old and continuing recommendation, but it is aimed in this context at a different criterion (i.e., limiting the number of conditional middle response endorsements that are elicited from items). Secondly, assessment researchers may find the results useful in the editing and revision of their assessments – items eliciting a large number of middle category endorsements should be considered for a specific form of editing: contextualization.

An extended report of this study can be obtained from John Kulas (jtkulas@stcloudstate.edu).

References


---

3 Although the range of intrusiveness ratings was constrained, note that intrusiveness did exhibit strong associations with valenced category endorsements.