

Running head: *Neutral or Unsure*

Neutral or Unsure: Is there a Difference?

Christine E. DeMars

T. Dary Erwin

James Madison University

(2005, August). Poster presented at the annual meeting of the American Psychological Association, Washington, DC.

Abstract

University students responded to a survey measuring identity development using a 4-point Likert-type scale with two additional options: *neutral* and *unsure*. The level of identity development of students who chose *neutral* was compared to the level of identity development of students who chose *unsure* on the same item. On average, these two groups of students had similar scores. *Neutral* and *unsure* did not seem to be used to indicate different levels of the construct of interest. Often these two categories were used as a middle response, but on one scale they were used as a moderately high response.

Neutral or Unsure: Is there a Difference?

Why do respondents choose options such as *neutral* or *no opinion* or *unsure* on surveys or psychological instruments? Are such response options interchangeable? Taken literally, *neutral* would seem to indicate a middle level of endorsement, while *no opinion* or *unsure* would seem to indicate a lack of an opinion or a lack of interest in the topic. The purpose of this study was to examine differences between respondents who chose *neutral* and those who chose *unsure* on an instrument designed to measure identity development in college students.

The *neutral* response category falls under the broader classification of middle response options. Another example of a middle response option would be *about right* in a question where the options were *too much*, *not enough*, or *about right*. These options are not necessarily placed in the middle of the list of options, but they are interpreted to mean, and scored as, middle-level opinions. *Unsure*, *no opinion*, *cannot decide*, and *don't know* would seem to be a different class of options; they do not necessarily indicate a middle position but may be used when the respondent has no opinion or lacks enough information to form an opinion. One could reasonably argue these terms are not interchangeable, but they are all distinct from middle responses.

Respondents might choose either a *neutral* or *unsure* response because they do not want to exert the cognitive effort to form an opinion. Krosnick (1999) termed this satisficing because the respondent picks the first acceptable option rather than trying to decide on the most appropriate option. This tendency should be more frequent for respondents who are less motivated or questions that are less salient and thus require more effort. For example, Shoemaker, Eichholz, and Skewes (2002) showed that the proportion of *don't know* responses was correlated with the degree of cognitive effort required by opinion survey items. When

respondents use a satisficing strategy, middle responses and *unsure/no opinion* might be used interchangeably.

Baumgartner and Steenkamp (2001) proposed respondents might choose a middle category such as *neutral* "due to evasiveness . . . indecision . . . or indifference" (p. 145). None of these reasons would indicate a truly middle position and might be more equivalent to *unsure*. Liao (1995) suggested providing both a *neutral* and a *don't know* category. He noted there is an important distinction between not having an opinion and having an ambivalent opinion, and suggested that if only one of these categories were available it might be used for both types of responses. Presser and Schuman (1980) found that when some survey respondents were explicitly offered a neutral category and others were not, fewer respondents spontaneously responded *don't know* on the survey form where *neutral* was included. This finding suggests that some of those who responded *don't know* on the form without *neutral* would have substituted *neutral* had it been an explicit choice. Bishop, Hippler, Schwarz, and Strack (1988) included a middle response option in one survey form and a *no opinion* option in another form (neither survey form included both options). The proportions of respondents selecting these options were not equal, so apparently at least some respondents do make a distinction between the two categories even when only one of the two is explicitly presented.

If middle responses and unsure-type responses are included in a survey, or accepted if offered in an interview, than the researchers may be interested in how these responses are related to the construct being measured. For example, middle responses would seem like they should indicate middle levels of the construct. Unsure-type responses might also indicate middle level, or they might indicate extreme levels of the construct. In the area of identity development, for example, students might choose *unsure* if they had not yet reached a point at which the question

was meaningful; then *unsure* would indicate a low stage of identity development in which these students were not ready to struggle with the concept.

Several studies have looked at this issue. One method used by several researchers is to administer one form with a middle response or unsure-type category, and another form without this category. If the other responses are distributed among the remaining options (most studies have used only two additional options) in similar proportions on the two forms, then the researchers can conclude that the middle response or unsure-type category is used about equally throughout the attitude range. This seemed to be true in a survey asking for data about recently deceased relatives (Poe, Seeman, McLaughlin, Mehl, & Dietz, 1988) and in interviews covering social and political attitudes (Presser & Schuman, 1980). However, in Bishop's (1987) interviews, for about half the items the presence of a middle response changed the distribution of the proportions choosing the other categories. To explain why some questions might be more impacted than others, Madden and Klopfer (1978) suggested that for issues regarded by respondents as more important, more effort would be exerted in trying to respond to the items and *cannot decide* would only be used when the respondent genuinely felt unsure. On issues regarded as less important, Madden and Klopfer suggested respondents might choose *cannot decide* as the easiest response (similar to Krosnick's idea of satisficing). They administered two forms of surveys to the same college students, one week apart; one form had the choices *yes* and *no*, and the other added the choice *cannot decide*. One survey was about capital punishment, an area in which Madden and Klopfer believed these students held strong opinions, and the other was about Sunday observance, an area in which they expected students to hold less strong opinions. *Cannot decide* was substituted about equally for *yes* and *no* on the capital punishment survey, but more frequently for *no* on the Sunday observance survey.

Another method used to explore the construct levels of those who choose middle responses or unsure-type responses is the nominal response model, an item response theory (IRT) model which does not assume any particular order or spacing to the categories a-priori. Hanisch (1992) examined data from scales of work satisfaction and retirement satisfaction. The response options were *yes*, *no*, and *?*. She showed that respondents who chose *?* tended to have satisfaction levels in between those choosing *yes* and those choosing *no*, but closer to those choosing *no*.

The nominal response model was also used in a study of an earlier version of the identity measure to be explored in the current study. In that study (DeMars & Erwin, 2004), *neutral or unsure* was provided as a single response category. For one of the four identity scales, usage of *neutral or unsure* was most common for those with middle levels of identity development or remained constant throughout the development range. For two of the four identity scales, usage of *neutral or unsure* was most common for those with low levels of identity development. For the remaining scale, on one item *neutral or unsure* was used by those with middle levels and on the other items it was used by those with low levels. The current study extends this work by exploring whether *neutral* and *unsure* indicate different developmental levels. Respondents were given each of these options along with a 4-point Likert-type scale. The research questions were:

- 1) Do those choosing *neutral* and those choosing *unsure* have different levels of the construct measured by the survey questions (identity development)?

There is little empirical evidence on which to predict this outcome, but other researchers have suggested that logically *neutral* and *unsure* have different meanings though they may be used similarly if only one of these options is provided (Liao, 1995; Presser & Schuman, 1980).

Bishop, Hippler, Schwarz, and Strack's (1988) findings suggest that these responses are not used exactly interchangeably when only one is offered.

- 2) In terms of average scores on the construct, where does *neutral* fall in relationship to the other response categories? Where does *unsure* fall?

Based on the usual meaning of the term *neutral*, we expect the identity levels of those choosing *neutral* to lie between the identity levels of those choosing the two categories in the middle of the scale. Based on the findings of DeMars and Erwin (2004) with the combined *neutral or unsure* option, we expect the identity levels of those choosing unsure to fall towards the lower end of the identity range.

Method

Participants

Participants were 1388 university students. These students were required to participate in university program assessment activities. The identity measure was administered along with achievement tests of General Education course objectives. Students were informed through written materials and verbal instructions that the results would be used for program evaluation.

Instrument

The Erwin Identity Scale VI (EIS) measures identity development in college students. It has four scales: Affective Identity, Confidence, Conceptions about Body and Appearance (CBA), and Communal Identity. In a previous study where *neutral or unsure* was provided as a single option, on the Affective and Communal scales, and most items on the Confidence scale, *neutral or unsure* was used more frequently by those with lower levels of development. On the CBA scale, *neutral or unsure* was used about equally throughout the developmental range; it was used somewhat more frequently by those in the middle range of development.

On the EIS VI, both *neutral* and *unsure* were provided as separate response options. They were placed to the right of the primary response scale (*very true of me, somewhat true of me, not*

really true of me, and *not at all true of me*). Students indicated their responses on a scannable bubblesheet.

Results

Items were calibrated and responses were scored using a graded response model (Samejima, 1969). Both *neutral* and *unsure* were coded as *not-administered* because in the graded response model the item categories need to be in a pre-specified order and the placement of *neutral* and *unsure* was unknown. In item response theory models, including the graded response model used here, after the items are calibrated to the same scale, a respondent's score is the trait level at which his or her responses are most likely, given the locations of the items on the scale. If responses are not available for a few items, on average it will not affect the value of the trait level at which the likelihood peaks, though it will flatten the likelihood a bit and widen the confidence interval around the estimated trait level. If the data set had been considerably larger, the nominal response model could have been used and *neutral* and *unsure* could have been treated as valid categories instead of coded as *not administered* because the categories do not need to be in a pre-specified order under the nominal response model. Given the size of the data set, however, this was not a reasonable option.

After items were calibrated and respondents were assigned scores by maximum-likelihood estimation, items were selected for the analysis if at least 30 respondents selected *neutral* and at least another 30 selected *unsure*. Thirty-four items met this criterion, 17 Affective items, 2 CBA items, 9 Confidence items, and 6 Communal items. For these selected items, the mean identity score (Affective, Confidence, Communal, or CBA, as appropriate) was calculated for those choosing each of these options. The magnitude of the difference between the means was indexed by Hedge's g (the difference between the means divided by the pooled standard deviation; Hedges, 1982).

Across the 34 items, g ranged from -0.38 to 0.41, with a weighted mean of 0.03. Thus, on average the scores on the construct of interest did not differ for those selecting *neutral* compared to those selecting *unsure*. However, the range of g values appeared large, so perhaps these options were used differently on some of the individual items. Techniques from meta-analysis were used to assess whether the variance in g 's was larger than would be expected by chance if all items had a common effect size. The probability values may not be strictly accurate because the items were not independent experiments. A respondent could choose *neutral* or *unsure* to more than one item, so some respondents' scores were included in more than one of the comparisons. Across the 34 items, 909 students were included in at least one comparison; of these, 288 were included in at least five comparisons. So the probability values and confidence intervals should be interpreted somewhat cautiously.

The homogeneity statistic Q (Shadish & Haddock, 1994, p. 266) was used to check whether the variance in effect sizes was greater than expected given the null hypothesis of a shared common g , where $Q = \sum_{i=1}^k \frac{(g_i - \bar{g})^2}{v_i}$, where g_i is the effect size g for item i , v_i is the variance of g_i , \bar{g} is the weighted mean effect size with each g_i weighted by $1/v_i$, and k is the number of items. Under the null hypothesis, Q is distributed as χ^2 with $k - 1$ degrees of freedom. For these data, $Q(33) = 34.7$, $p = 0.61$. The variance in the effect sizes, then, was not inconsistent with what would be expected with a common effect size (though again it should be noted that using some overlapping samples for the g 's may affect the accuracy of the probability value). Thus it was meaningful to estimate a 95% confidence interval around the weighted mean g . This confidence interval was from -0.04 to 0.09 so the weighted mean effect size was not significantly different from zero (and was not meaningfully different from zero even if it had been significantly different).

These results indicate that students who choose *neutral* and those who choose *unsure* have similar scores on the construct of interest. The next focus of the research was how their scores compared to the scores of students choosing the other options. This was addressed by comparing the mean maximum likelihood scores (based on the response pattern to all items except those answered *neutral* or *unsure*) of those choosing each option. This procedure was in some ways similar to Sympson and Haladyna's (1988; Haladyna & Kramer, 2005) polyweighting system for determining how to score each option on a multiple-choice test, except that maximum likelihood scores were used in the present study.

For most of the 34 studied items, the scores of the combined *neutral* or *unsure* group fell in between those choosing *somewhat true of me* and those choosing *not really true of me*. For 5 of the 6 items on the Communal scale, however, the scores of the *neutral* or *unsure* group fell in between those choosing the highest category and those choosing the next-highest category (generally *not at all true of me* indicated the highest level of identity, but on reverse-scored items *very true of me* was the highest category). This pattern also occurred for 1 of the 17 Affective items, and 1 of the 2 CBA items. The opposite pattern occurred for another of the Affective items and 1 of the 9 Confidence items; the scores of those choosing *neutral* or *unsure* fell between the scores of those choosing the lowest category and those choosing the next-lowest category.

Discussion and Conclusions

Based on these results, students who chose *neutral* did not have scores that consistently differed from those who chose *unsure*. Therefore it seems reasonable to combine these categories for scoring the EIS or to obtain larger groups for research purposes. When these groups were combined, their scores generally fell in the middle of the response categories for the Affective and Confidence scales. However, their scores generally fell between categories 3 and 4 (highest

category) for the Communal scale. This contradicts the findings of a study with an earlier version of the EIS (DeMars & Erwin, 2004) in which those with low levels of identity were more likely to choose the category *neutral or unsure* on the Affective and Communal scales. Aside from some changes in the items and the fact that *neutral* and *unsure* were offered as separate categories in the present study, there were differences in the techniques used in the analysis. In the earlier study, scores on each scale were estimated with the nominal response model, treating *neutral or unsure* as one of the categories, while in the present study scores were estimated with the graded response model, treating both *neutral* and *unsure* as missing data. A bigger difference, though, was that in the earlier study the focus was on where usage of each category peaked -- essentially, the modal score. In the present study, the focus was on mean scores for the students who chose each category. Particularly when the response curve is fairly flat, as it tended to be for the *neutral or unsure*, the mean can be a considerable distance from the mode.

In conclusion, *neutral* and *unsure* seem to be used by those with similar scores on the EIS, on average. These categories appear to be used as a middle response on the Affective and Confidence scales, and as a somewhat high response on the Communal scale. From this, it seems that how these categories are used depends on the construct the scale measures and perhaps on the individual item, so no generalizations about scoring weights can be made for other instruments.

References

- Baumgartner, H., & Steenkamp, J-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143-156.
- Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51 (2), 220-232.
- Bishop, G. F., Hippler, H., J., Schwarz, N., & Strack, F. (1988). A comparison of response effects in self-administered surveys. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 321-340). New York: John Wiley & Sons.
- DeMars, C. E., & Erwin, T. D. (2004). Scoring *Neutral or Unsure* on an Identity Development Instrument for Higher Education. *Research in Higher Education*, 45, 83-95.
- Haladyna, T. M., & Kramer, G. (2005, April). *Poly-scoring of multiple-choice item responses in a high-stakes test*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal.
- Hanisch, K. (1992). The job descriptive index revisited: Questions about the question mark. *Journal of Applied Psychology*, 77 (3), 377-382.
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92, 490-499.
- Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology*, 50, 537-567.
- Liao, T. F. (1995). The nonrandom selection of don't knows in binary and ordinal responses: Corrections with the bivariate probit model with sample selection. *Quality and Quantity*, 29 (1), 87-110.
- Madden, T. M., & Klopfer, F. J. (1978). The "Cannot Decide" option in Thurstone-type attitude scales. *Educational and Psychological Measurement*, 38 (2), 259-264.

- Poe, G. S., Seeman, I., McLaughlin, J., Mehl, E., & Dietz, M. (1988). "Don't Know" boxes in factual questions in a mail questionnaire: Effects on level and quality of response. *Public Opinion Quarterly*, 52 (2), 212-222.
- Presser, S., & Schuman, H. (1980). The measurement of a middle position in attitude surveys. *Public Opinion Quarterly*, 46 (1), 70-85.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplements*, 17.
- Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper and L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 261-281). New York: Russell Sage Foundation.
- Shoemaker, P. J., Eichholz, M., & Skewes, E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research*, 14, 193-201.
- Sympson, J. B., & Haladyna, T. M. (1988, April). An evaluation of "polyweighting" in domain-referenced testing. Paper presented in C. E. Davis (Chair), *New developments in polychotomous item scoring and modeling*. Symposium conducted at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED294911)