

Conclusion

The hypothesis that Australian workers would rate the activities that describe their jobs in a similar way to U.S. workers was not supported. This is surprising given the cultural similarity between the two countries and cautions against the direct application of O*NET in Australia in the absence of supportive evidence. Admittedly, only a tiny section of O*NET was investigated in this study but this actually increases the concern because differences were found even on such a small scale.

The low interrater agreement on the two occupations where this could be investigated mirrored the literature on interrater agreement in job analysis which generally reports mixed results – some workers agree on how their job was described whereas others do not (Dierdorff & Wilson, 2003). Indeed, if Australian workers cannot agree amongst themselves on how to assess the activities involved in their work, little wonder that they disagree with U.S. norms on these factors.

There are many potential pitfalls in collecting job analytic information such as problems involving social pressure and cognitive biases (Morgeson & Campion, 1997) and this study is likely to have included many of them. There are also a number of specific limitations unique to the current study. First, the sample size is very small, especially in relation to the large number of occupations in O*NET. Thus the current study cannot be viewed as representative of O*NET or the world of work more generally and the results reported here must be seen as preliminary. They call for caution in the direct application of O*NET to Australia and highlight the need for further research on this topic. Related to the small sample size is the question of power; however, given that significant effects were observed, they are likely to be of sizable magnitude and therefore should not be ignored.

There was also a suggestion that some participants were not utilizing the scale anchors on the level ratings correctly. For example, one incumbent, describing their occupation as "business owner", rated the level of their interaction with the general public as 6, which was equivalent to "performing a monologue on national TV" according to the scale anchor. These problems would have been alleviated if each occupation were represented by more than one participant with gross inaccuracies or distortions being averaged out. An even better way to alleviate inflated ratings would be to have all questionnaires administered and scored by a trained experimenter in a structured interview format, but this was beyond the resources available for the study and is not the preferred method of using the questionnaires, even for the O*NET developers.

The O*NET database provides a broad and rich source of job information with many potential uses. The current study looked only at the applicability of GWAs

to the Australian context, but there are many other features of O*NET which need to be investigated including assessments of vocational interests and ability. The ideal study for future research would be to sample a large number of participants from many occupations and assess them on as many O*NET dimensions as possible. Nevertheless, we can conclude that, in the absence of adequate validation, at this stage O*NET should be used in Australia with caution.

References

- Australian Bureau of Statistics (1997). *ASCO: Australian Standard Classification of Occupations, Second Edition*. Canberra: Australian Bureau of Statistics.
- Bredt, J. C. (2001). An occupational view of the Australian labour force: Patterns of job growth and decline. *International Journal of Manpower*, 22, 414-424.
- Brough, P., & Smith, J. (2003). Job analysis. In M. O'Driscoll, P. Taylor, & T. Kalliath (Eds.), *Organisational Psychology in Australia and New Zealand*. Singapore: Oxford University Press.
- Dierdorff, E. C., & Wilson, M. A. (2003). A meta-analysis of job analysis reliability. *Journal of Applied Psychology*, 88, 635-646.
- Gudykunst, W. B., & Kim, Y. Y. (1992). *Communicating With Strangers: An Approach to Intercultural Communication, 3rd Edition*. New York: McGraw Hill.
- Hall, E. T. (1976). *Beyond Culture*. New York: Anchor Books.
- Harris, R. J. (1985). *A Primer of Multivariate Statistics*. Academic Press, Orlando, Florida.
- Hays, W. L. (1981). *Statistics (3rd Edition)*. New York: Holt, Rinehart and Winston.
- Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills: Sage.
- Morgeson, F. P., & Campion, M. A. (1997). Social and cognitive sources of potential inaccuracy in job analysis. *Journal of Applied Psychology*, 82, 627-655.
- Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., Campion, M. A., Mayfield, M. S., Morgeson, F. P., & Pearlman, K. (2001). Understanding Work Using the Occupational Information Network (O*NET): Implications for Practice and Research. *Personnel Psychology*, 54, 451-492.
- Singelis, T. M., & Brown W. J. (1995). Culture, self, and collectivist communication: Linking culture to individual behavior. *Human Communication Research*, 21, 354-89.

Including a 'Midpoint' Option in Responses to an Organisational Climate Survey: Does it Affect Scores, Data Quality or Respondent Satisfaction?

Louise P. Parkes (louise.parkes@voiceproject.com.au)

Voice Project, Department of Psychology,
Macquarie University, Sydney, NSW 2109 Australia

Peter H. Langford (peter.langford@voiceproject.com.au)

Voice Project, Department of Psychology,
Macquarie University, Sydney, NSW 2109 Australia

Abstract

The present study contrasted responses to 5-point (with midpoint) and 4-point (with no midpoint) Likert-type scales. Using responses from 321 participants completing an organisational climate survey, the study explicitly examined respondent preferences for response scale format, as well as examining psychometric qualities of the climate survey using the different formats. Results showed that respondents prefer scales that include a middle-point, and that preferences are affected by the experience of using the midpoint. While there were no differences found between the comparable mean scores and distributions for the two response scale formats, organisations need to be aware that commonly reported statistics, such as an 'agreement' score, vary substantially depending on the response scale used. There were no differences found between reliability and validity coefficients for the variables measured in the climate survey. However, in regression analyses the 5-point scale allowed greater discrimination between predictor variables, suggesting it captured greater independent variance.

Psychologists use Likert-type rating scales as one of their primary tools for measuring attitudes, judgements, personality and so on. However, there is some ambivalence about the optimum design of these scales, with a complex and disparate literature following 80 years of research (e.g., Symonds, 1924). Are such details really important, or are the effects trivial? Characteristics of questionnaire design have been found to account for at least two-thirds of the variation in measurement quality (Andrews, 1984) and a large proportion of the variance in reliability (Churchill & Peter, 1984).

The majority of research over the decades has focused on the effects of various response options on reliability and validity, with many studies using computer simulations. As such, the "psychological" process of people's cognitive and emotional response to various scales has often been overlooked. In the context of organisational surveys, employees' experience when responding to the survey contributes to ongoing

organisational relationships, and may affect the quality and quantity of responses. Enormous effort is invested, in both hours and dollars, to manage the perceptions, interest and motivation of employees participating in organisational surveys. Despite this, assumptions about respondent experiences and preferences have largely been based on anecdotal evidence and intuition.

One contentious issue is whether to allow respondents to have a "neutral" response, or whether to force a choice between a positive and negative judgement. In addition to examining the effects on scale qualities such as reliability and validity, this study directly investigated respondent preferences.

Why Include a Midpoint?

In general, respondents may use the midpoint as a valid rating point, an invalid response (opting out), or somewhere in-between – a 'satisficing' response.

As a valid rating point, the midpoint adds finer discrimination to a scale, and adds to data quality by extending the true variance captured. The middle-category represents not only a particular mid-'point', but a range of positions seen to lie between the polar opposites (a neutral 'region', Komorita, 1963).

Sometimes, especially when a survey does not include the option of 'not applicable' or 'don't know', respondents who truly cannot answer a question (don't understand the question, are not familiar with the object of the item, etc) may choose the middle-response. Such responding reduces the validity of the midpoint.

Finally, the midpoint may be used as a means to 'satisfice' because respondents don't care enough about the issue to invest the mental effort to decide either way (Krosnick, 1991). In an organisational climate survey, employees might be highly motivated to optimise responses – desires for self-expression, emotional catharsis, commitment to help the organisation make better decisions, etc, might all play a part. Nevertheless, factors of time-pressure, familiarity with the issue, cynicism or lack of faith in management's interest in the results, may lead to respondents using various cognitive heuristics or short cuts to save effort. There

are several options a respondent might take advantage of when satisficing on a Likert scale: if the midpoint is unavailable, respondents might simply agree with assertions (acquiescence), or choose a 'don't know' or 'not applicable' option.

The Effect of Excluding a Midpoint

So what happens when the neutral point is excluded, and respondents are forced to choose sides? Results have typically been investigated in terms of the impact on three different areas: actual responses – the mean, medians and response distributions; the quality of data obtained – discriminability/validity and reliability; and more recently, respondent satisfaction.

If attitudes reflect a valid neutral point or are indifferent, then responses should shift to match the underlying population distribution, as respondents are forced to make the effort to decide which end of the neutral region is closest to their true response. In this case the mean results are unaffected, as found in much past research (Biggers, 1977; Guy and Norvell, 1977; Mercer & Durham, 2001; Schuman & Presser, 1981). Alternately, satisficing respondents may simply move to another effort-reducing heuristic and such as acquiescence, resulting in an elevated mean score (Cronbach, 1950; Ray, 1983; Ray and Pratt, 1979; Schuman and Presser, 1981). Similarly, the 'don't know' / not applicable' option may be chosen instead.

If the neutral point represents the addition of a useful scale category, then scale reliability and validity should improve with inclusion of a midpoint, since validity generally tends to increase (and residual error decreases) as the number of answer scale categories increases (e.g., Andrews, 1984). However, random measurement error also increases with the number of responses, minimising reliability and validity coefficients. At the same time, correlated measurement error, such as method variance, inflates reliability and validity estimates. Cronbach (1950) remarked that it is an "open question" whether a finer scale of judgement gives more valid scores, or simply more response-set error.

Little research has focussed on the 'emotive' or 'affective' component of respondents' experience. It has generally been inferred that respondents prefer a midpoint option, for reasons of greater choice and for the option of satisficing. Preston and Colman (2000) measured respondent preferences to response scale stimuli, including both odd and even point scales, ranging from 2 to 11 points. It appeared that respondents were in a quandary, rating the smaller point scales (2, 3, & 4 points) as quicker to use, but not allowing them to express their feelings adequately. The 5-point scale was actually rated most favourably on "ease of use", and was rated significantly higher than the 4-point scale on "allowed you to express your

feelings adequately". These results suggest that respondents prefer a 5-point scale with a midpoint, to a 4-point scale.

The Present Study and Hypotheses

The present study contrasted responses to 5-point (with midpoint) and 4-point (without midpoint) Likert scales. In addition to examining the psychometric effects on response distributions and data quality, a chief aim of the study was to explicitly examine the effects of excluding the neutral option on respondent preferences and behaviour.

Overall, the evidence appears to suggest that when forced to choose, responses will reflect the underlying distribution, or a refusal to choose ('Not Applicable'). Increased acquiescence on the 4-point scale due to satisficing should be minimal.

Hypothesis 1 There will be no difference in comparable mean scores or response distributions for responses on 4 or 5-point scales.

Hypothesis 2 Respondents using a 4-point scale will be more likely to use the 'not applicable' option than those using a 5-point scale.

Reliability is measured here by Cronbach's alpha coefficient, and validity by the ability of scale scores to predict outcome variables in regression analyses. For reasons outlined above, the 5-point scale should capture greater true variance, but may also increase random measurement error.

Hypothesis 3 Responses on the 5-point scale will have reliability and validity coefficients equivalent to, or higher than, the 4-point scale.

It was expected that respondents would explicitly confirm assumptions about their preference for a response format that includes a midpoint.

Hypothesis 4 Respondents will explicitly indicate a preference for a 5-point scale, with a midpoint, over a 4-point scale.

Method

Participants

Students attending a Bachelor of Business Administration course completed the survey online as part of their course assessment. They responded to the survey online, through the Internet. There were 321 respondents to the survey, randomly allocated to a response form using either a 5-point scale (n=177) or a 4-point scale (n=144).

Survey

The Voice Climate Survey is designed to assess employee's perceptions of their work environment. The

survey comprised 31 scales assessing an organisation's performance on a wide range of HR and general management practices and 3 scales indicating employee engagement (a total of 130 items). Respondents rated each item by checking a response on a Likert scale with labels of: *Strongly Disagree* ("SA"), *Tend to Disagree* ("D"), *Mixed Feelings* ("M"), *Tend to Agree* ("A") and *Strongly Agree* ("SA"). Each point on the scale was labelled to maximise reliability, and the legend showing both the full anchor descriptions and the abbreviations was visible on each page. For respondents in the 4-point group, the Mixed Feelings option was not included. A "Not Applicable" category ("NA") was also included, and respondents were advised to use this option "When you don't feel as though a question is appropriate for you, don't have an opinion, or don't know the answer".

At the end of the survey the following question was included to directly assess respondent preferences regarding rating scales: "When answering these types of surveys, do you prefer a 5-point response scale with a mid-point (e.g., has a "Mixed Feelings" or "Neutral" option), or a 4-point response scale that doesn't include a mid-point?". There were also two open-ended questions: "From your experience in completing employee opinion surveys like these, (1) name one characteristic of the survey design that you particularly LIKE (either from your experience with this current survey or others)" and (2) "name one characteristic of the survey design that you particularly DISLIKE (either from your experience with this current survey or others)."

Results

Mean Response and Response Distributions

Looking first at the overall response distribution for those who completed the 5-point scale, that is, averaged over respondents and items, we note that the distribution is not normal, but has an overall positivity bias. Of the total responses, 58% were rated favourably, compared to 20% unfavourable, and 22% mixed. This use of the midpoint is consistent with previous research, with frequencies around 20% for a 5-point scale (Mattel & Jacoby, 1972).

The frequency of responses on the 5-point scale were recalculated, excluding responses to the neutral point, enabling the remaining four responses to be compared directly to the four-point scale (Presser & Schumann, 1980).

As seen in Figure 1, when excluding the midpoint, the distribution of responses across the options is very similar. There are *not* more favourable responses for the 4-point group (73% favourable versus 75% for the 5-point group).

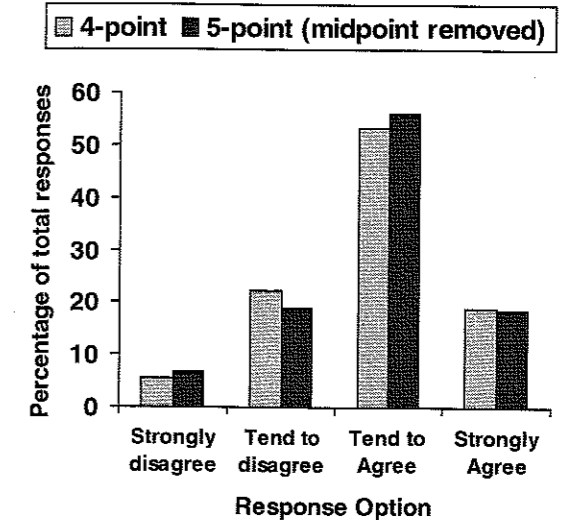


Figure 1: Frequency of responses on the 4-point versus 5-point scales.

In order to compare equivalent mean scores, ratings on the 4-point scale were rescaled on 4 equal intervals on a scale of 1 to 5 (i.e., 1, 2.33, 3.67, 5). The mean response across all items was not significantly different for the 4-point and 5-point scales ($M = 3.48$, $SD = .52$; $M = 3.47$, $SD = .48$, respectively).

Consistent with hypotheses respondents chose the 'not applicable' option more often when the midpoint was not available (for 2% of items on the 5-point scale, 6% of items for a 4-point scale, $p < .05$).

Reliability and Validity

There was negligible difference between the internal reliabilities for climate scales on the 4-point and 5-point response scales. The mean alpha for the 4-point scales was .80 (range from .69 to .95) and for the 5-point scales was .79 (range from .58 to .95).

In an attempt to measure the valid variance captured by each of the scale formats, a regression analysis was conducted using the climate scales to predict Employee Engagement. The 28 scales were collapsed into six higher-order dimensions (Langford, 2007), with the intention of increasing the stability of the β coefficients with a higher variable-to-respondent ratio. A slightly smaller sample of responses from the 5-point group was selected randomly ($n = 136$) so that it more closely matched the 4-point group in size and power ($n = 138$). Overall, climate dimensions measured on the 4-point scale accounted for similar variance in Employee Engagement ($R^2 = 44.5$), compared to the 5-point scale ($R^2 = 44.3$). However, analyses using the 5-point scale identified three, rather than two, independently significant predictors. That is, the dimensions had

greater discriminability when assessed on a 5-point scale than when measured using the 4-point scale.

Respondent Preferences

Of those in the 5-point group, 98% of respondents used the midpoint option at least once. Use of the midpoint ranged from not at all, to 81 of 130 items, $M = 28$, $SD = 15.6$.

When asked directly, the majority of respondents preferred a 5-point scale (15% preferred 4-point, 18% don't mind either way, and 66% preferred a 5-point scale). Interestingly, there was a significant difference between the groups that had completed surveys with either 4 or 5-point scale formats (see Figure 2). The 5-point group showed a stronger preference for the inclusion of a midpoint.

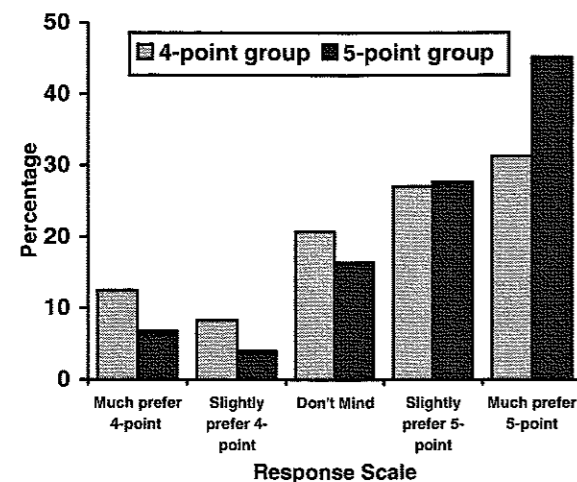


Figure 2: Scale preferences for respondents in the 4 and 5-point groups.

Analysis of open-ended responses revealed that when asked for a characteristic that they like, 20% of respondents referred to a characteristic of the response scale. Of these, the majority identified a preference for a multiple choice answer scale (68%), saying it was easy to use and efficient; 10% of respondents liked the number of options on their scale (i.e., they were extensive enough); and 14% liked having a midpoint.

When asked for a characteristic that they dislike, 17% of responses referred to characteristics of the response scale. Of these, the majority identified that they disliked not having an N/A option (33%) or a midpoint/neutral option (31%).

When comparing responses from the 4-point and 5-point groups, the starkest differences were over preferences for a midpoint. Reiterating the quantitative response above, of respondents who commented on the

response scale, 20% liked having a midpoint in the 5-point group, compared to 4% in the 4-point group. However, when identifying dislikes, 40% in the 4-point group disliked not having a midpoint, compared to 19% in the 5-point group.

Discussion

Confirming anecdotal evidence, our data showed that respondents clearly prefer scales that include a middle-point. Almost all respondents used the midpoint at least once if it was available, and two-thirds reported that they preferred using a rating scale with a 'mixed feelings' or 'neutral' option. Respondents in the 5-point response group professed a stronger preference for a midpoint option than those in the 4-point group, and in open-ended questions, the 5-point group was five times more likely to say they liked having a midpoint option than the 4-point group. The experience of using the midpoint appears to have made them more sensitive to this design characteristic (similar to results found by Guy & Norvell, 1977). At the same time, when asked what they disliked, respondents in the 4-point group were twice as likely to cite not having a midpoint option. It appears they may have experienced greater frustration with this scale characteristic. Effects like these suggest the same response scales should be used throughout an entire survey, and any possible alterations minimised, to reduce method bias. It is also a consideration for future research.

A comparison of the response distributions indicated that respondents appeared to use the midpoint as a neutral region, presumably leaning toward one of the polar alternatives. That is, when the midpoint was unavailable, the majority of respondents were able to choose a response on either side of the scale. Consistent with prior research, there were no differences found between the mean scores and distributions for the two scales, demonstrating that acquiescence did not increase with the 4-point scale. Nevertheless, a small percentage of respondents (4% more in the 4-point condition) could not, or would not, choose, and opted for the 'Not Applicable' option.

Despite respondent preferences, and consistent with previous research (Andrews, 1984; Churchill & Peter, 1984), the inclusion or exclusion of a midpoint on the Likert response scale did not appear to affect the quality of survey measurement. No significant differences were found between reliability and validity coefficients for each of the scale formats.

One interesting result was that regression analyses using the 5-point scale identified a larger number of independently significant predictors than with the 4-point scale. That is, after accounting for some common method variance, the 5-point scale appeared to allow greater discrimination between predictor variables. This

suggests that the 5-point scale may enable finer discrimination for ratings, extending the variance captured (Ramsay, 1973; Garner, 1960; Bendig, 1954).

The present results also have some important implications for the way survey results are reported. For ease of understanding when presenting results of employee surveys to management, it is common to report a 'percentage favourable' or 'agreement' statistic, to indicate item approval. If a scale with a midpoint has been used, respondents with 'mixed feelings' may not be included as favourable, but may still be appropriately included in the total denominator used to calculate the statistic. A 4-point scale forces respondents to choose a favourable or non-favourable stance. Because of the positivity bias of the data in this study, and many other similar studies, reports of the 'agreement' statistic will vary substantially depending on the response scale used (e.g., 58% on 5-point scale versus 73% for 4-point scale for the current study). While these appear quite disparate, means on all the scales (and overall) were not significantly different for respondents answering on 4 or 5-point scales. When respondents choosing the midpoint are excluded from the 'percentage favourable' calculations altogether, frequencies are more comparable (75% versus 73% respectively). Managers may be well advised to note differences in response scales or methods of reporting statistics before comparing reports.

Finally, the decision to include or exclude a midpoint option for respondents appears to rely most heavily on the purposes of the research. Certainly when choosing between a 4 and 5-point rating scale, the psychometric properties are both good. The response distribution remained the same, as did the mean response. Rather, the inclusion of the midpoint added another data point to the results, improved discriminability of the variables being measured, and was explicitly preferred by most respondents. In cases where the researcher is seeking to repeatedly survey the same group (as in annual organisational surveys), improving people's motivation to provide optimising responses is a strong argument for including the middle-point.

References

Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modelling approach. *Public Opinion Quarterly*, 48, 409-442.

Bendig, A. W. (1954). Transmitted information and the length of rating scales. *Journal of Experimental Psychology*, 47, 303-308.

Biggers, J. L. (1977). Use of a neutral point with the Dogmatism Scale. *Journal of Experimental Education*, 45, 56-58.

Churchill, G. A., Jr., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, 21, 360-375.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, 3-31.

Garner, W. R. (1960). Rating scales, discriminability and information transmission. *Psychological Review*, 67, 343-352.

Guy, R. F. & Norvell, M. (1977). The neutral point on a Likert Scale. *The Journal of Psychology*, 95, 199-204.

Harter, J. K. (1997). The psychometric utility of the midpoint on a Likert scale. *Dissertation Abstracts International*, 58, 1198.

Komorita, S. S. (1963). Attitude content, intensity, and the neutral point on a Likert scale. *Journal of Social Psychology*, 61, 327-334.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.

Langford, P. H. (2007). *Psychometrics of the Voice Climate Survey: Evidence for a lower and higher-order factor structure of work practices and outcomes*. Manuscript submitted for publication, Macquarie University.

Matell, M. S. & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, 56, 506-509.

Mercer, C. & Durham, T. W. (2001). The Hood Mysticism Scale: Does the presence of a neutral response-item affect response style. *Psychological Reports*, 88, 335-338.

Preston C. C. & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1-15.

Ramsay, J. O. (1973). The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika*, 38, 513-533.

Ray, J. J. (1983). Reviving the problem of acquiescence response bias. *Journal of Social Psychology*, 121, 81-96.

Ray, J. J. & Pratt, G. J. (1979). Is the influence of acquiescence on 'catchphrase' type attitude scale items not so mythical after all? *Australian Journal of Psychology*, 31, 73-78.

Schuman, H. & Presser, S. (1981). *Questions and Answers in Attitude Surveys*. New York: Academic Press.

Symonds, P. M. (1924). On the loss of reliability in ratings due to the coarseness of the scale. *Journal of Experimental Psychology*, 7, 456-461.

