

The Impact of Item Readability on the Endorsement of the Midpoint Response in Surveys

Pauline Velez and Steven D. Ashworth

San Francisco State University, San Francisco (CA) and Sempra Energy Utilities, San Diego (CA)

Researchers have debated including a midpoint response in questionnaire scales, and they have debated what the selection of that response option represents. Using data from employee opinion surveys of a large financial services organization based primarily in the United States, this study explored whether the endorsement of the midpoint response is related to an item's readability and its perceived clarity. The results showed the endorsement of the midpoint response was significantly correlated with the Flesch-Kincaid Grade Level index, average letters per word, and average number of syllables per word. The results also showed that as the perceived clarity of an item increases, the endorsement of the midpoint response decreases. These findings have important implications both to researchers and organizations that conduct surveys with their employees.

Keywords: Midpoint, neutral response, surveys, item readability.

Introduction

With the explosion of technology, surveys have become even more commonplace, allowing both researchers and organizations to gather data. However, the fundamentals of surveys (e.g., item construction, sampling, appropriate response scales) continue to be of paramount importance to the reliability and validity of the data collected. One of the fundamental issues is item readability. Item readability can be conceptualized from multiple standpoints.

On one hand, item readability can be considered from the perspective of language difficulty. Researchers have shown that the level of language difficulty in a survey may be too complex for certain sectors of the population (e.g., Terris 1949). In the 1950's, Rudolf Flesch introduced an objective formula to assess the level of language difficulty for written pieces. According to Flesch (1948), the more words in each sentence and the more syllables in each word, the less readable a passage. Immediately, Flesch's "ease of reading" formula became the best-known measure of language complexity or comprehension difficulty. Researchers have also applied the formula to survey items (Converse 1976) as a gauge for understanding the readability of one's survey items.

The practicality of using readability statistics to assess the level of language complexity on survey items seems apparent. Readability statistics would allow survey practitioners to assess the readability of their survey items. Further, researchers have found that when survey items are not understandable, the reliability and validity of survey responses becomes suspect. Klare (1950) found that questions using

language that is more difficult for the survey population had a higher frequency of "no opinion" responses. Razek and Cone (1981) found that material requiring more effort on the part of the reader to understand resulted in greater frustration and anger, which resulted in the reader feeling negative about the written piece. The potential impact of respondent frustration and anger on data quality is enormous. The goal of survey research is to gather information on how respondents feel about a particular construct. If survey data is influenced by general frustration and anger to the survey format, as opposed to the constructs of interest, then data quality becomes a bigger issue. When data quality is influenced by the respondent's reaction to the format of the survey, the data are no longer an accurate assessment of the respondent's thoughts or attitudes towards the construct of interest.

Another component of item readability is item clarity, or the extent to which the question or survey item is clear to the respondent. Coombs and Coombs (1976) examined whether item ambiguity or respondent uncertainty resulted in a higher amount of "no opinion" responses. The researchers concluded that certain characteristics of questions, such as question wording or content, increase the number of "no opinion" responses because they pose more cognitive difficulty. Therefore, if item ambiguity exists, it could be a confound in attitudinal measures, impacting the quality of one's data.

However, it is unclear what response option a respondent will select when an item is ambiguous and an attitudinal measure does not include a "no opinion" option. There are a number of possibilities. Respondents might leave the item unanswered. Or, respondents may select either a positive or negative response option. Or, if the attitudinal survey consists of a five-point Likert scale, with a midpoint response, then it seems reasonable that respondents who are confused by an item will select the midpoint response. In this instance, the use of the midpoint response signifies a lack of opinion due to respondent confusion, as opposed to a truly

Contact information: San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA (pvelez@sfsu.edu).

neutral stance. If a respondent were to choose the midpoint response in this situation, the data will contain systematic error. In other words, the choice of the midpoint response option no longer reflects one's neutral opinion; rather it reflects one's inability to understand the question. Such error in one's data impacts not only the reliability but the validity of one's findings. The issue of how an item's readability (i.e., language difficulty) influences the endorsement of the midpoint response has received limited attention (Scherpenzeel and Saris 1997; Saris et al. 2004).

The Measurement of Attitudes

The concept of an attitude is one of the most distinctive and indispensable concepts in the field of psychology (Zimbardo et al. 1990). The measurement of attitudes can be traced through relevant literature to as early as the 1920's (Thurstone and Chave 1929). One of the most common methods used to measure attitudes has been the standard, written questionnaire. The purpose of this study is to investigate the relationship between the endorsement of the midpoint response and the readability of the survey item.

The midpoint of any scale can represent a variety of opinions. For example, the midpoint can represent an "undecided", "don't know", "no opinion" or "neutral" response. Determining whether to offer a midpoint should be based on the theoretical construct the researcher is attempting to measure. In addition, one's decision to offer a midpoint could vary depending upon whether one is trying to measure a general attitude or a specific attitude.

Many arguments have been made both for and against offering a midpoint alternative. One of the main arguments in favor of offering a midpoint on a scale is that there are individuals who favor the middle position, and forcing them to choose a polar alternative will contribute to some form of random or systematic error (Schuman and Presser 1981). As early as 1944, Rugg and Cantril (1944) argued that offering a midpoint allowed for additional gradation of opinions. The inclusion of the midpoint gives the respondent more options. In addition, some researchers believe that offering a midpoint may make respondents more comfortable when selecting a response option (Nunnally 1967).

One of the main arguments against offering a midpoint is that it increases the amount of error in survey data. To begin with, the inclusion of a midpoint increases the probability of response styles (Klopfer and Madden 1980). For example, individuals who do not want to participate in the survey may respond to each item using the midpoint. In this instance, the individual differences in the responses to the survey items may not truly reflect the attitude in question. Similar to the concept of central tendency (Cascio 1991), the data is reflecting a response style of the individual. Lastly, many researchers who are against offering a midpoint believe that people do lean toward one direction, and that the midpoint attracts those individuals who do not feel strongly about the issue at hand (Schuman and Presser 1981). For a contrasting point of view, see Scherpenzeel and Saris (1997), who show that the inclusion of an explicit midpoint has a

moderate effect on validity.

The decision about whether to include a midpoint should be made during the design phase, with a thorough understanding of the consequences of including a midpoint upon the quality of one's data. If a researcher decides to offer a midpoint, an accurate understanding of the underlying concepts used by respondents who select the midpoint on attitudinal surveys is necessary. A great number of studies were conducted prior to 1950 on the middle alternative response. Research on the subject declined considerably from 1950 to the mid 1980's. In the last twenty years, a number of studies have begun to examine the midpoint response and respondents who select the middle alternative (Bishop 1987; Converse 1976; Coombs and Coombs 1976; Klopfer and Madden 1980; Presser and Schuman 1980; Schuman and Presser 1981). Research has been conducted to test whether certain item or respondent characteristics increases the number of individuals who select a certain response option (Coombs and Coombs 1976; Klopfer and Madden 1980; Schuman and Presser 1981). Klopfer and Madden (1980) found that individuals who are ambivalent about an attitude are more likely to select the midpoint. Bogart (1967) illustrated that items that are sensitive in nature, multidimensional, or unclear may affect the response option that is chosen. In addition, Bogart found that items that are unclear or confusing might result in more respondents selecting the midpoint. Thus, research has found that there are a variety of reasons an individual might select the midpoint including item characteristics (e.g., items are sensitive, multidimensional or unclear) and individual respondent characteristics (e.g., individual ambivalence, anxiety about responding). Some of these reasons are sources of systematic, controllable error in one's data, which further highlights the importance of training in survey item writing and the importance of considering whether a midpoint response is needed given the impact on data quality. One issue that intersects both item characteristics and individual respondent characteristics is item readability. Thus far, there is little research on the relationship between item readability and the endorsement of the midpoint response. This study addresses this gap in the literature.

The Present Study

The current study examines attitudinal items on an organization's annual employee opinion survey. An employee opinion survey was utilized in this study since the vast majority of respondents should have opinions on various facets of their job. Thus, it would be expected that this type of survey would elicit less midpoint responses from respondents. The employee opinion survey measures employee attitudes on a variety of workplace issues such as pay, benefits, job satisfaction, career opportunities, and company image.

The purpose of the present study is to examine whether the readability level of attitudinal survey items is related to the endorsement of the midpoint response. The hypotheses to be investigated are as follows:

Table 1: Sample listing of items with respective readability statistics

Item	Flesch – Kincaid Grade Level Index	Words per sentence (mean)	Letters per word (mean)	Syllables per word (mean)
How satisfied are you that the < company name > practices and procedures which you are required to follow on your job allow you to provide service to customers that exceeds their expectations?	16	30	5.13	1.66
When new procedures affecting your job are introduced, how satisfied are you with the explanation you receive of how these procedures work?	14	22	5.27	1.77
How satisfied are you with your involvement in decisions that affect your work?	8	13	5.07	1.53
I have enough information to do my job well.	5	9	3.88	1.44
I like the kind of work I do.	1	8	2.62	1.00

1. As the reading grade level of an item increases, the endorsement of the midpoint response increases.
2. As suggested by Coombs and Coombs (1976), as the perceived clarity of an item, based on expert rater ratings, decreases the endorsement of the midpoint response increases. In this instance, the perceived clarity of an item might represent issues over and above item readability. In other words, other factors besides the reading grade level of an item, might impact the perceived clarity of an item.

If a relationship between the endorsement of the midpoint response and an item's reading grade level does exist (hypothesis 1), it would suggest that researchers contribute to this potential source of error by wording items at a higher reading grade level than the respondents possess. Further, in order to minimize this type of error, researchers can assess the language difficulty of the items as part of the questionnaire construction phase and appropriately match the reading grade level of the survey to the population's reading grade level. Items with a higher reading grade level may need to be reworded to ensure the majority of respondents will understand the item. Finally, researchers should be concerned with the impact of this form of systematic error in terms of the reliability and validity of one's findings.

If a relationship were found between an item's clarity and the endorsement of the midpoint response (hypothesis 2), it would further substantiate the belief that respondents are more likely to select the midpoint response when the item is unclear or not understood. Thus, the selection of the midpoint response does not represent a neutral opinion; rather it represents respondent confusion in the form of systematic error. In addition, it would further support any analyses that have found a direct relationship between reading grade level and the endorsement of the midpoint response.

Method

Procedures

Items from the organization's annual employee opinion survey, from the years 1988 through 1991, were utilized.¹ Given that the unit of analysis of this study is an item, it was necessary to span four years to have a sufficient number of items for the analyses. Since many of the same items were used throughout the four years, this study only examined each item the first year it was used. One hundred and fifteen items were used.

The topic areas of the employee opinion survey include customer focused quality, training, performance management, communication, empowerment, work involvement, teamwork/cooperation, satisfaction with the job, job security, working conditions, workload, immediate manager or supervisor, workforce diversity, pay, benefits, recognition, company image, and leadership.

To assess the language complexity of a piece of written work, one must decide which readability statistics to use. When deciding between usage of the Flesch Reading Ease formula and the Flesch-Kincaid Grade Level index, the main factor the researcher must consider is the length of the writing that will be assessed. The Flesch Reading Ease formula is appropriate for longer pieces of writing since it examines the number of syllables per 100 words. For shorter items, such as attitudinal survey items, the Flesch-Kincaid Grade Level index is more appropriate since the formula examines the number of syllables per word regardless of the number of

¹ Data for this study were obtained from employees of a financial service organization in the United States. The respondents represented one business unit within the organization. The data from each year represents approximately 17,000 to 45,000 employees, including management.

words.

For the current study, all items were assessed on item readability based upon two measures. First, the items were analyzed in terms of readability statistics using the reference software, Grammatik IV (please see Table 1 for a sampling of items and the respective readability statistic). The readability statistics that were analyzed included the Flesch-Kincaid Grade Level index, the average number of words per sentence, the average letters per word, and the average number of syllables per word. As a second measure of item readability, the items were rated on a ten-point scale in terms of its clarity by a team of expert raters. The expert raters were psychologists who had been trained on survey research methodology and had years of experience in writing survey items and instruments. Clarity was defined as how well an item is written. The raters assessed whether the content of an item was clear or ambiguous. The clarity measure was chosen and developed to be conceptually different from the Flesch-Kincaid Grade Level Index, which is highly weighted toward the number of syllables per word. In order to assess the reliability of the ratings, intraclass correlations (Shrout and Fleiss 1979) were performed.

Data from the computerized databases containing the results of the employee opinion surveys conducted between 1988 and 1991 were accessed to determine the percentage of midpoint responses. For those items that appear more than once in the four-year span, the percentage of midpoint response was based on the first year the item appeared on an employee opinion survey.

Results

The purpose of this study was to explore the relationship between item readability and the endorsement of the midpoint response. This section explores the relationships between an item's readability and an item's perceived clarity to the endorsement of the midpoint response.

The means and standard deviations for various readability statistics and the percent of midpoint responses for all 115 items are listed in Table 2.

Table 2: Means and standard deviations of readability statistics for survey items (n=115)

	Mean	S.D.
Percent of midpoint responses	19.00	.08
Flesch-Kincaid Grade Level Index	11.52	4.01
Average number of words per sentence	15.24	5.29
Average number letters per word	5.24	.96
Average number of syllables per word	1.77	.31

To test the hypothesis that as the reading grade level of an item increases, as measured by the readability statistics, the percent of midpoint responses increases, correlations were performed between the percent of midpoint responses and the Flesch-Kincaid Grade Level Index, the average number of syllables per word, the average number of words per sentence, and the average letters per word.

The correlation matrix is given in Table 3. Specifically, the percent of midpoint responses was significantly correlated with the Flesch-Kincaid Grade Level Index ($r(114) = .25, p < .01$), indicating that as the reading grade level of an item increases, the percent of midpoint responses increases. In addition, the percent of midpoint responses is significantly correlated with the average letters per word ($r(114) = .20, p < .05$) and the number of syllables per word ($r(114) = .22, p < .05$). A significant correlation was not found between the percent of midpoint responses and the average number of words per sentence ($r(114) = .13, p < .05$).

In order to investigate the hypothesis that as perceived clarity of an item decreases the percent of midpoint responses increases, a team of seven experts rated each item on a ten-point scale in terms of its clarity, with clarity being defined as how well an item is written. The reliability of the ratings was assessed using intraclass correlations. The reliability coefficient of the expert ratings proved to be acceptable (ICC (2,7) = .75).

The clarity of an item was positively and significantly correlated with the percent of midpoint responses ($r(114) = .26, p < .01$). Since an increase in the clarity rating indicates an item is less clear, a positive correlation symbolizes that as the clarity of an item decreases, the percent of midpoint responses increases.

Discussion

The purpose of this study was to determine if aspects of attitudinal survey items affect the endorsement of the midpoint response. Specifically, this study addressed the relationship between an item's readability and perceived clarity to the endorsement of the midpoint response.

The first hypothesis, that as the reading grade level of an item increases, the percent of midpoint responses associated with the item increases, was supported. Specifically, the higher the reading grade level of an item, the more midpoint responses it will have. In addition, the more letters per word and the more syllables per word, the more midpoint responses.

Since items with a high reading grade level increase the probability of the respondent not understanding the items, the respondent is more likely to select the midpoint response, thereby creating a response tendency that adds error to the data. Alternately, it is possible that when a respondent encounters a complex question, the respondent opts to select the midpoint response, rather than expend the necessary cognitive energy to interpret the question. In such cases, the researcher can no longer assume that the data in the midpoint response truly represents the neutral position of the respondents. Instead, some of the data may represent respondent confusion based on lack of understanding of the item or the respondent's unwillingness to expend a great deal of cognitive effort to interpret the survey question, resulting in systematic error in one's data.

Another gauge used to assess readability is perceived clarity. The second hypothesis, that as perceived clarity of an item decreases, the percent of midpoint responses asso-

Table 3: Correlation matrix for percent of midpoint responses and various readability statistics

	Percent of midpoint responses	Flesch-Kincaid Grade Level	Words per sentence (mean)	Letters per word (mean)
Flesch-Kincaid Grade Level	.25**			
Words per sentence (mean)	.13	.42**		
Letters per word (mean)	.20*	.80**	.07	
Syllables per word (mean)	.22**	.89**	.00	.86**

* $p < 0.05$ ** $p < 0.01$

ciated with the item increases, was supported. Perceived clarity was defined as how well an item is written, and was based on ratings assigned by experts. The finding indicates that as the clarity of an item decreases, the percent of midpoint responses increases. This further substantiates the belief that respondents are more attracted to the midpoint response when the item is unclear or not understood.

Limitations

While this study does offer some interesting findings regarding item readability and the endorsement of the midpoint response, there are a number of limitations to this study. While this study did find a relationship between item readability and the endorsement of the midpoint response, the analyses were correlational in nature. As such, the issue of causality cannot be addressed. In addition, it is possible that the relationship between item readability and the endorsement of the midpoint response can be influenced by other unobserved variables. For example, the reading grade level of an item might be influenced by the conceptual clarity of an item or it might be influenced by personal tendencies to endorse the midpoint. This study does not address such unobserved variables and further research is needed to address these issues.

This study utilizes responses to employee opinion surveys. The choice to use employee opinion surveys was driven by the fact that most employees will have an opinion about various components of their job, thus it was anticipated that such surveys would not have a high level of endorsement of the midpoint response. On the other hand, it is unclear whether the results of this study would generalize to other types of surveys. It is quite possible that the motivation level of employees in this survey differs from respondents who take other types of surveys. Employees who completed this survey might have felt as if this survey was their opportunity to change their work environment by providing feedback to the organization. Future research is needed to explore the generalizability of these findings to other types of surveys with different populations.

Finally, this study does not address the impact of survey length on the endorsement of the midpoint response. It is possible that as the length of a survey increases, respondents are more likely to endorse the midpoint response as they progress through the survey such that items that appear later in the survey elicit a higher frequency of midpoint responses. In other words, as respondents get tired, they might

be more likely to select the easy answer (e.g., the middle alternative). Thus far, researchers have found that the position of an item in the questionnaire does impact the item's reliability and validity (e.g., Scherpenzeel and Saris 1997; Saris et al. 2004).

Implications of This Study

This study has provided additional insight to the age-old debate of whether to include a midpoint response. Particularly, the reading grade level of survey items is related to the percent of midpoint responses associated with items, and, ultimately, does create a certain amount of error. This is possibly the greatest contribution of this study considering the lack of literature that exists on the endorsement of the midpoint response as it relates to the readability of survey items.

From the perspective of readability, this study has illustrated that the interpretation of midpoint data must include some discussion of the readability of the survey, especially if the survey did not examine this issue during the design phase. This becomes more important to those researchers who do not analyze midpoint data. Decisions on whether to include the midpoint on surveys and to analyze midpoint data could skew the results. Further, it is possible that the inclusion of a midpoint could result in systematic error, impacting both the reliability and validity of one's data. Thus, examining the readability of one's survey before it is deployed would help to address potential concerns over the readability of the survey items and assist in revising potential confusing items.

The largest implication of this study suggests that surveys need to be assessed in terms of readability to ensure the questionnaire will not create a higher level of cognitive demand for the intended audience. It appears that the temptation exists to overestimate the degree to which one's audience is informed about the survey content and able to understand survey items, especially in organizational surveys. This study highlights the importance of considering readability when piloting and proofing questions, even when trained survey experts develop the questions. Although the specific organization used in this study does employ experts who review the questions on the employee opinion survey, some of the error could have been eliminated had the items been more readable. As more and more organizations move to surveying employees and customers, the implications of this study are magnified. Organizations need to consider readability in the piloting process of their questionnaires. Researchers and practitioners can easily assess the readability of their survey

items, using, for example, Microsoft Word software. By using the spelling and grammar options in Word, one can invoke the feature to show readability statistics whenever a spell check is done. Thus, exploring the readability of one's survey items has become much more user friendly.

Future Research

Given the paucity of research on the endorsement of the midpoint response, further research is needed. This study provides some interesting initial insights; however it does not address the issue of causality nor does it provide insight into other possible unobserved variables that might influence the endorsement of the midpoint response. In addition, this study utilizes data from an English-language employee opinion survey within one organization in the United States; therefore it is unclear whether these results would generalize to other types of surveys with different populations. Additional research could shed light on these issues.

Although research does exist that examined the relationship of the level of education, sex, age, and race and the endorsement of the midpoint response, thus far, the results have been inconclusive. If further insight is desired, the possibility of replicating such studies still exists, but more substantive findings might emerge by examining segments of respondents. Segmentation analyses, popular in the market research field, could provide the interaction between education, sex, age, race and the endorsement of the midpoint response, distinguishing how much variation in midpoint responding is due to individual differences. Future research might also examine the relationship between item readability and other forms of responding. For example, what would the impact of readability be on a survey that did not contain a midpoint response category?

Conclusion

This study contributes to the literature in terms of both the midpoint response, as well as the readability of attitudinal survey items. Specifically, this study shows that items with a high reading grade level have more midpoint responses. Readability, from this standpoint, becomes an important issue in questionnaire design and this study suggests it behooves researchers and practitioners to make survey items as clear as possible. Further, researchers and practitioners should utilize the various tools at their disposal (e.g., pilot testing, assessing the reading grade level of their items) to ensure their items are clear before deploying their survey. As suggested earlier, commonly used word processing software packages such as Microsoft Word provide an option to assess the readability statistics of any piece of written work.

This study also provides some reinforcement that writing survey items is both an art and a science. With advances in technology and the ability to survey individuals online easily, this study confirms the importance of training in how to

design, write, and implement surveys, and highlights the implications of poorly designed surveys on data quality. This point cannot be understated. Research continues to show measurement error impacts the quality of one's data and researchers and practitioners must take proactive steps to ensure they minimize measurement error.

References

- Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51, 220-232.
- Bogart, L. (1967). No opinion, don't know, and maybe no answer. *Public Opinion Quarterly*, 31, 332-345.
- Cascio, W. F. (1991). *Applied psychology in personnel management*. New Jersey: Prentice Hall.
- Converse, J. M. (1976). Predicting no opinion in the polls. *Public Opinion Quarterly*, 40, 515-530.
- Coombs, C. H., & Coombs, L. C. (1976). "Don't know": Item ambiguity or respondent uncertainty? *Public Opinion Quarterly*, 40, 497-514.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Klare, G. R. (1950). Understandability and indefinite answers to public opinion questions. *International Journal of Opinion and Attitude Research*, 4, 91-96.
- Klopfert, F. J., & Madden, T. M. (1980). The middlemost choice on attitude items: Ambivalence, neutrality, or uncertainty? *Personality and Social Psychology Bulletin*, 6, 97-101.
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw Hill.
- Presser, S., & Schuman, H. (1980). The measurement of a middle position in attitude surveys. *Public Opinion Quarterly*, 44, 70-85.
- Razek, J. R., & Cone, R. E. (1981). Readability of business communication textbooks-an empirical study. *Journal of Business Communication*, 18, 33-40.
- Rugg, D., & Cantril, H. (1944). The wording of questions. In H. Cantril (Ed.), *Gauging public opinion*. Princeton, NJ: Princeton University Press.
- Saris, W. E., Veld, W. van der, & Gallhofer, I. (2004). Development and improvement of questionnaires using predictions of reliability and validity. In S. Presser (Ed.), *Methods for testing and evaluating survey questionnaires*. San Francisco, CA: John Wiley & Sons Inc.
- Scherpenzeel, A. C., & Saris, W. E. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods and Research*, 25, 341-353.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. San Diego, CA: Academic Press Inc.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Terris, F. (1949). Are poll questions too difficult? *Public Opinion Quarterly*, 13, 314-319.
- Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude*. Chicago: Univ.Chicago Press.
- Zimbardo, P. G., Ebbeson, E. B., & Maslach, C. (1990). *Influencing attitudes and changing behavior*. New York: McGraw-Hill.