

# ***The use of midpoint on Likert Scale: The implications for educational research***

## 中間選項在李克特量表中的應用： 給教育研究的啟示

**Kwok Kuen TSANG**

*PhD Student, Faculty of Education, The University of Hong Kong*

### **Abstract**

This article is going to review the debate of the use of midpoints on Likert scale in order to understand the implications for scale construction. In this review, the major issues around this debate are identified. The first is methodological. The debate around this issue is the concern about whether the midpoints affect the reliability and validity of measurements. The second issue is epistemological. It concerns about how researchers exactly know the meaning of the responses into midpoints that they intend to measure. To some extent, the debate based on the methodological issue illustrates that both using and not using of midpoints are acceptable, because the midpoints may not really affect the reliability and validity. Therefore, the epistemological issue while designing the rating scale of a measurement is the focus of this article. The implications of the debate for educational research are also discussed in this article.

### **Keywords**

midpoint, Likert scale, scale construction

### **摘要**

本文回顧有關在李克特量表應用中間選項的爭論，希望著此了解中間選項對建構量表的作

用。本文將指出，關於中間選項的爭論主要圍繞兩個主題。一是方法論的，這一主題所關心的是中間選項對量表的信度和效度的影響；二是認識論的，這主題所關心的是研究者如何得知受試者對中間選項的理解跟他們所想的一樣。在某程度來說，關於方法論上的爭論普遍認同中間選項不一定對量表的信度和效度有影響，所以從應用和不應用中間選項均可接受。因此，在建構量表時，我們應更多地考慮認識論上的問題。另外，本文還討論了中間選項的爭論對教育研究的啟示。

## 關鍵詞

中間選項，李克特量表，量表建構

In educational research, Likert scale is commonly used to measure different kinds of variables, such as teacher stress and burnout (e.g. Chan, Chen, & Chong, 2010; Dworkin, 2002), self-efficacy (e.g. Brouwers, Tomic, & Stijnen, 2002; Cheung, 2006), school and teacher effectiveness (e.g. Bangert, 2006; Kyriakides, Campbell, & Christofidou, 2002; Kyriakides, Demetriou, & Charalambous, 2006; Reezigt & Creemers, 2005; Reynolds, 2001), school organization (e.g. Bowen, Ware, Rose, & Powers, 2007; Firestone, 1984; Firestone & Herriott, 1982; Herriott & Firestone, 1984) school climate and culture (e.g. Cavanagh & Dellar, 1996; Cavanagh & Waugh, 2004; Moos, 1987; Pang, 1998; Wagner, 2006), and the like. The reason is that Likert scale empowers educational researchers to effectively operationalize the variables and then identify their relationships in order to improve our educational system.

However, some researchers claim that the use of midpoints on Likert scale may affect research reliability and validity, but some other researchers disagree. It is necessary for education researchers to learn this debate about whether midpoint opinions are included in a scale, because the debate not only directly points to the problem of research quality but also the validity of research implications and recommendations to educational system. In this article, therefore, we will first review the debate in social research in general. On the basis of this review, then, we will suggest the implications of the debate for educational research.

## **Debate of the use of midpoints on Likert scale**

### **Methodological issue**

One concern among researchers about having midpoints on a Likert scale is the effects of the midpoints to the reliability and validity of measurements. Generally, the supporters of midpoint opinions claim that the midpoints can increase the reliability of measurement. For example, Courtenay and Weidemann (1985) assess the effects of midpoint answers (“don’t know”) to the Palmore’s Facts on Aging quizzes (FAQ) and conclude that the midpoint answers tend to enhance the reliability of FAQ. Another study conducted by Adelson and McCoach (2010) present similar findings. In that study, Adelson and McCoach compared the response pattern of elementary students who responded a mathematics attitudes instrument with a 4-point Likert scale with another group of elementary students who responded the same instrument but the scaling had an additional neutral point. The study shows that the scale including a neutral midpoint might be more appropriate for elementary students than the 4-point scale, because the reliability of the 5-point scale was statistically and significantly higher than the reliability of the 4-point scale.

On the other hand, the adversaries argue that the high reliability may be resulted from response set (Cronbach, 1950), especially the tendency to choose the midpoint options. Weems and Onwuegbuzie (2001) conduct three studies to show that there was a high rate of midpoint choices among their samples. This to some extent implies response set to the midpoints exist. Different from the findings found by the supporters of midpoints, the response set in Weems and Onwuegbuzie’s studies seems to attenuate the reliability rather than enhance it (Weems & Onwuegbuzie, 2001). In this sense, midpoints are not necessary to benefit the internal consistence of measurements.

Nevertheless, some researchers argue that the use of reliability as a criterion to judge the merit of midpoints is inappropriate (Chang, 1994). As Cronbach (1950, p.22) already notes, “there is no merit in enhancing test reliability unless validity is enhanced at least proportionately.” In other words, validity should be a better criterion than reliability (Chang, 1994). Some studies evaluate the impacts of midpoints on measurement validity. However, the findings are also contradictory. For instance, some studies find that the construct validity may not be influenced by the midpoints (Adelson & McCoach, 2010; Kulas, Stachowski, & Haynes, 2008), but some researchers suggest the omission of the midpoints may impair the validity (Johns, 2005).

One possible reason explaining such contradictory findings is that the reliability and validity may be independent of the number of scale points, including the use or not use of midpoints, on Likert scale (Dawes, 2001a; Matell & Jacoby, 1971). Another possible explanation is that there are other factors mediating the relation of the use of midpoints to the measurement reliability and validity, such as respondents' response style (Clarke, 2001; Lee, Jones, Mineyama, & Zhang, 2002; Wong, Tam, & Fung, 1993) and reverse coding (Weems & Onwuegbuzie, 2001).

### **Epistemological issue**

Another concern about the use of midpoints on Likert scale is epistemological. This means whether and how researchers exactly know the meaning of the responses into midpoints that they intend to measure. Originally, the meaning of midpoints on Likert scale refers to neutral i.e., neither agree nor disagree (Raaijmakers, Hoof, Hart, Verbogt, & Wollebergh, 2000). Therefore, this kind of options is desirable because it avoids forcing respondents to choose agree or disagree options, that may evoke misleading conclusion, if they really hold neutral opinions towards the items. However, some scholars already note that midpoints may have many different meanings such as "neither agree nor disagree", "undecided", "don't know", and "no opinion" (Raaijmakers, et al., 2000). Thus, it is possible for respondents to interpret the midpoints in several different ways that may be totally different from the original or intended meaning, especially when the midpoints are not clearly defined (Kulas, et al., 2008).

Worcester and Burns (1975) conduct a very interesting experiment to investigate this issue. In the experiment, the subjects were randomly assigned into four groups. Each group of the subjects was required to answer three questions that were the same for each group except the rating scales (4-point or 5-point) and the scale option labels (e.g. "tend to agree", "agree", and "2" assumed to be the same meaning). First, all of the subjects were asked to give their answers by using a discrete verbal scale (the Likert scales being tested); then, they were asked to indicate the answers again on a continuous non-verbal scale (literally straight, blank line). After that, Worcester and Burns compared the answers between the verbal and non-verbal scales. They found that the midpoint selections could mean "neither agree nor disagree", "tend to agree", and "tend to disagree" among the subjects. Thus, they concluded that the similar or the same options may mean different things to different people.

Similar to Worcester and Burns, Kulas, et al. (2008, p.251) claim that midpoints

may be viewed by the respondents as a “dumping ground” for unsure or non-applicable responses, “if the respondent[s] did not view the middle response option as existing along the agreement continuum.” In other words, midpoints may not really represent the opinion of “neither agree nor disagree”.

To some extent, this argument gets supports from another line of studies that aims to learn the effects of the midpoints on survey results. For example, Garland (1991) asks his respondents to give opinions about the importance of product labeling with a Likert scale and he finds that more negative ratings were obtained when midpoints were removed from the scale. Dawes (2001b) conducts a similar study, in which the respondents were asked to identify their satisfaction towards their insurance company with a scale either with or without midpoints, and also finds the similar results. In addition to negative rating, some studies indicate that an increase in positive rating may occur when a scale does not include midpoints (Worcester & Burns, 1975). These effects of the denial of midpoints may be explained by that respondents may “use the midpoint to avoid reporting what they see as less socially acceptable answer” (Johns, 2010, p.7) in order to please the interviewers (Garland, 1991). If it is the case, the selection of midpoint may no longer imply neutrality. In other words, midpoints may be harmful to measurement validity.

However, the supporters of the use of midpoints provide confronting evidences. For instance, Raaijmakers, et al. (2000) argue that the midpoints are necessary. This is because the respondents, who do not have enough knowledge to response the items, might minimize unresponse rate by selecting the midpoint to indicate the sense of “undecided” or “don’t know”. In addition, Matell and Jacoby (1972) discover that a negative correlation between the number of scale options and the opportunity that midpoints become a dumping ground. This implies that midpoints may be more appropriate in a scale with more scale options.

## **Implication to educational research**

According to the literature reviewed above, it is obvious that there is still no conclusion whether the midpoints on Likert scale are desirable or not. Nevertheless, according to the methodological viewpoint, i.e. the issue about the impact of midpoints on measurement reliability and validity, both use and not use of midpoints are acceptable because the midpoints may not really affect the reliability and validity (Dawes, 2001a; Matell & Jacoby, 1971). Therefore, it is suggested that educational researchers should

take more consideration to the epistemological issue while designing the rating scale of a measurement.

To some extent, it is hard for educational researchers to know exactly the meaning of the midpoint responses. For instance, how should we interpret why a teacher choose the midpoint to the following statement retrieved from Kyriakides, et al.'s (2002) questionnaire about teacher effectiveness: "Students' achievement in relation to teachers' objectives"? The teacher may select the midpoint for a variety of purposes, such as: he or she may want to express neutrality about the relationship between students' achievement and teachers' objectives; he or she may not know the relationship; or he or she may avoid to select "disagree" or "very disagree", even though this may be his or her true thought, because he or she may think disagreement about the statement is socially undesirable.

Nevertheless, it is argued that the inclusion of midpoints on a scale is necessary. This is because we cannot sure whether the meaning of "agree" or "disagree" response, for instance, really implies the respondents' agreement or disagreement towards the items. Some respondents may select these two options because there is no an option referring to "neutral", "undecided" or "don't know". In this sense, we may need to take a risk that we may make an inaccurate conclusion due to the scale without such midpoints. The inaccurate conclusion may affect the validity of the implications and recommendations to improve our educational system.

If this is right, educational researchers need to think how to minimize the disadvantage of the use of midpoints such as respondents' misinterpretation to the midpoint opinions, response set to midpoints, and social desirable responses through midpoints. One possible way is a careful use of option labels. For example, Worcester and Burns (1975) discover that the balance side point options (e.g. the point 2 and 4 of a 5-point Likert scale) that are labeled as "slightly agree/disagree", "fairly agree/disagree" or "quite agree/disagree" are more preferable than that labeled as "agree" and "disagree" only, because the adverbs tend to reduce the number of midpoint selections.

Another way is to define the midpoints as clear as possible. Some studies suggest that the use of midpoints as a dumping ground may be more likely to occur when the option labels are difficult to understand (Cummins & Gullone, 2000; Kulas & Stachowski, 2009). Therefore, for example, it is more desirable to refer midpoints to the label of "neither agree nor disagree" or "neutral" instead of number "3". Alternatively, educational researchers may add "non-applicable" or "N/A" options in a Likert scale (Kulas, et al., 2008). These

two approaches may be able to solve the problem of “untrue” middle response category endorsement (Kulas & Stachowski, 2009; Kulas, et al., 2008).

Finally, the limitations of the midpoints may be reduced by increasing scale sensitivity. To increase scale sensitivity means to increase number of scale options (Cummins & Gullone, 2000). Some researchers show the increase in scale sensitivity may decrease the midpoint selections (Matell & Jacoby, 1972). They suggest that the midpoint selections tend to more often occur on 3-point and 5-point scale, but less on 7-point and 19-point scale (Matell & Jacoby, 1972). In other words, the response set to midpoints and social desirable responses through midpoints may be minimized by increasing scale sensitivity.

## Conclusion

In this paper, the literature about the debate of the use of midpoints on Likert scale is reviewed. It finds that the debate seems to focus on two issues: methodological issue – the impact of midpoints on measurement reliability and validity – and epistemological issue – whether and how researchers know the meaning of midpoint responses that are the intended meaning designed by the researchers. After reviewing the arguments from both supporters and opponents, it is suggested that a scale with midpoints is appropriate for educational research because such an inclusion may not necessarily be harmful to the measurement reliability and validity, but also avoid forcing respondents to choose a direction. Nevertheless, it is still noted that there are some limitations of the use of midpoints, such as respondents’ misinterpretation to the midpoint opinions, response set to midpoints, and social desirable responses through midpoints. Nevertheless, the limitations may be minimized by the careful use of the option labels, the clear definition of the midpoints, the inclusion of “N/A” options in a Likert scale, and the increase in scale sensitivity.

## References

- Adelson, J. L., & McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point Likert-type scale. *Educational and Psychological Measurement, 70*(5), 796-807. doi: 10.1177/0013164410366694
- Bangert, A., W. (2006). The development of an instrument for assessing online teaching effectiveness. *Journal of Educational Computing Research, 35*(3), 227-244.
- Bowen, G. L., Ware, W. B., Rose, R. A., & Powers, J. D. (2007). Assessing the functioning of schools as learning organizations. *Children & School, 29*(4), 199-208.
- Brouwers, A., Tomic, W., & Stijnen, S. (2002). A confirmatory factor analysis of scores on the teacher efficacy scale. *Swiss Journal of Psychology, 61*(4), 211-219.
- Cavanagh, R. F., & Dellar, G. B. (1996). *The development of an instrument for investigating school culture*. Paper presented at the Annual Meeting of the American Educational Research Association.
- Cavanagh, R. F., & Waugh, R. F. (2004). Secondary school renewal: The effect of classroom learning culture on educational outcomes. *Learning Environments Research, 7*, 245-269.
- Chan, A. H. S., Chen, K., & Chong, E. Y. L. (2010). *Work stress of teachers from primary and secondary schools in Hong Kong*. Paper presented at the the International MultiConference of Engineers and Computer Scientists 2010 Vol III, Hong Kong.
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*(3), 205-215. doi: 10.1177/014662169401800302
- Cheung, H. Y. (2006). The measurement of teacher efficacy: Hong Kong primary in-service teachers. *Journal of Education for Teaching: International Research and Pedagogy, 32*(4), 435-451. doi: 10.1080/02607470600982134
- Clarke, I. I. (2001). Extreme response style in cross-cultural research. *International Marketing Review, 18*(3), 301-324.
- Courtenay, B. C., & Weidemann, C. (1985). The effects of a "don't know" response on Palmore's Facts on Aging quizzes. *The Gerontologist, 25*(2), 177-181.
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10*(1), 3-31. doi: 10.1177/001316445001000101
- Cummins, R. A., & Gullone, E. (2000). *Why we should not use 5-point Likert scales: The case for subjective quality of life measurement*. Paper presented at the Proceedings, Second International Conference on Quality of Life, Singapore: National University of Singapore.



- Dawes, J. (2001a). Comparing data gathered using five point vs eleven point scales. *Bridging Marketing Theory and Practice 1*, 1-8.
- Dawes, J. (2001b). The impact of mentioning a scale mid-point in administering a customer satisfaction questionnaire via telephone. *Australasian Journal of Market Research*, 9(1), 11-18.
- Dworkin, A. G. (2002). Teacher burnout. In D. L. Levinson, P. W. Cookson & A. R. Sadovnik (Eds.), *Education and sociology: An encyclopedia* (pp. 659-664). New York: RoutledgeFalmer.
- Firestone, W. A. (1984). *The study of loose coupling: Problems, progress, and prospects*. Paper presented at the The Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Firestone, W. A., & Herriott, R. E. (1982). Two images of schools as organizations: An explication and illustrative empirical test. *Educational Administration Quarterly*, 18(2), 39-59. doi: 10.1177/0013161x82018002004
- Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, 2, 66-70.
- Herriott, R. E., & Firestone, W. A. (1984). Two images of schools as organizations: A refinement and elaboration. *Educational Administration Quarterly*, 20(4), 41-57. doi: 10.1177/0013161x84020004004
- Johns, R. (2005). One size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion and Parties*, 15(2), 237-264.
- Johns, R. (2010). Likert items and scales. Retrieved from Survey Question Bank website: <http://www.surveynet.ac.uk/sqb/datacollection/likertfactsheet.pdf>
- Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, 43(3), 489-493. doi: 10.1016/j.jrp.2008.12.005
- Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology*, 22(3), 251-259. doi: 10.1007/s10869-008-9064-2
- Kyriakides, L., Campbell, R. J., & Christofidou, E. (2002). Generating criteria for measuring teacher effectiveness through a self-evaluation approach: A complementary way of measuring teacher effectiveness. *School Effectiveness and School Improvement*, 13(3), 291-325.
- Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research*, 48(1), 1-20. doi: 10.1080/00131880500498297

- Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert scale. *Research in Nursing & Health*, 25(4), 295-306.
- Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, 31(3), 657-674. doi: 10.1177/001316447103100307
- Matell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert scale effects of testing time and scale properties? Effects of testing time and scale properties. *Journal of Applied Psychology*, 56(6), 506-509.
- Moos, R. H. (1987). *The Social Climate Scales: A User's Guide*. California: Consulting Psychologists Press.
- Pang, N. S. K. (1998). The binding forces that hold school organizations together. *Journal of Education Administration*, 36(4), 314-333.
- Raaijmakers, Q. A. W., Hoof, A. v., Hart, H. t., Verbogt, T. F. M. A., & Wollebergh, W., A. M. (2000). Adolescents' midpoint response on Likert-type scale items: Neutral or missing values? *International Journal of Public Opinion Research*, 12(2), 208-216.
- Reezigt, G. J., & Creemers, B. P. M. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement*, 16(4), 407-424.
- Reynolds, D. (2001). Beyond school effectiveness and school improvement? In A. Harris & N. Bennett (Eds.), *School effectiveness and school improvement: Alternative perspectives* (pp. 26-43). London: Continuum.
- Wagner, C. R. (2006). The school leader's tool for assessing and improving school culture. *Principal Leadership*, 7(4), 41-44.
- Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. *Measurement and Evaluation in Counseling and Development*, 34(3), 166-176.
- Wong, C. S., Tam, K. C., & Fung, M. Y. (1993). Differences between odd and even number of response scale: Some empirical evidence. *Chinese Journal of Psychology*, 35(2), 75-86.
- Worcester, R. M., & Burns, T. R. (1975). A statistical examination of the relative precision of verbal scales. *Journal of the Market Research Society*, 17(3), 181-197.