

Innovations in Measurement: Visual Analog Scales and Retrospective Pretest Self-Report Designs

Innovaciones en medición: Escalas Visuales Análogas y Diseños pre-test de Autoinforme Retrospectivo

Brittany K. Gorrall¹

Jacob D. Curtis²

Todd D. Little³

Pavel Panko⁴

Texas Tech University, United States

Abstract. The authors discuss limitations of two popular measurement procedures, the Likert scale and conventional pretest-posttest self-report design. Both techniques have limits and yet are often combined, leading to restricted fidelity for measuring change. The authors go on to discuss two innovations in measurement that provide researchers with greater assessment fidelity: Visual analog scales and the retrospective pretest design. Moreover, when used in combination, these innovations measurement techniques provide dramatic increases in the power to detect and quantify change.

Keywords. Attitudes, polytomous items, MRG, college students.

Resumen. Se discuten las limitaciones de dos procedimientos de medición extremadamente populares; la escala Likert y los diseños pre-post tradicionales. Ambos métodos tienen limitaciones; estos métodos son comúnmente combinados obteniendo como resultado una fidelidad limitada para la medición del cambio. Se discute sobre dos innovaciones en medición que proveen al investigador una mayor fidelidad: escalas visuales análogas y el diseño pre-test retrospectivo. Además, cuando se usan de manera combinada, estas innovaciones en medición proveen notables incrementos en el poder de detección y la cuantificación de cambios.

Palabras clave. Actitudes, ítems politémicos, MRG, estudiantes universitarios.

¹Brittany K. Gorrall. Institute for Measurement, Methodology, Analysis & Policy (IMMAP), Texas Tech University, United States. Postal Address: Department TTU-Education, Texas Tech University - National Wind Institute, 1009 Canton, Ave, Room Number 211, Lubbock, TX 79409, United States. Email: britt.gorrall@ttu.edu

²Jacob D. Curtis. Texas Tech University, United States. Email: jacob.curtis@ttu.edu

³Todd Daniel Little. Institute for Measurement, Methodology, Analysis, and Policy (IMMAP), Texas Tech University, United States. Email: yhat@ttu.edu

⁴Pavel Panko. Institute for Measurement, Methodology, Analysis & Policy (IMMAP), Texas Tech University, United States. Email: pavel.panko@ttu.edu



Introduction

The Likert scale and conventional pretest-posttest self-reports are two popular methods utilized by researchers to measure respondents' experiences. Renis Likert introduced the response scale in 1932, and it is still currently the most popular response format (Hodge & Gillespie, 2003). Likert scales make the respondent specify his or her experience by selecting one graduated category from the many category options, with Likert scales usually containing four to eleven categories (Flynn, 2004).

The categories are anchored by verbal descriptors (e.g., agree, disagree). In a similar vein, the conventional pretest-posttest self-report measurement design is by far one of the most popular methods to collect and analyze change data (Sprangers, 1989). After an individual participated in a treatment or experience, it is important to estimate the "true change" which occurred in the individual (Cronbach & Furby, 1970). However, both the Likert scale and the conventional pretest-posttest self-report have limitations.

The Likert scale's artificial categories are not sufficient to capture a continuous phenomenon (Joyce, Zutshi, Hrubes, & Mason, 1975). The conventional pretest-posttest self-report is limited in its assessment of change, and does not capture the full change story of the individual (Aiken and West, 1990). Two ways to improve upon the Likert scale and conventional pretest-posttest self-report measurement designs are through visual analog scales (VAS) and retrospective pretest-posttest self-report designs. Hayes and Patterson (1921) were the first to use and describe the VAS (Couper, Tourangeau, Conrad, Singer, 2006). The VAS is a horizontal continuous scale with verbal anchors on the extremes of a continuous line. Respondents mark a point on the continuum that best represents how they feel. The distance between the marked point and the origin of the line is measured to quantify the magnitude

of the response. In Figure 1, the respondent indicated feeling about 32% of the worst pain possible.

Reasons to use the VAS as opposed to the Likert scale.

The VAS allows for finer distinctions than the Likert scale, providing a greater amount of information to the researcher (Aitken, 1969; Rausch & Zehetleitner, 2014). With computer aided technology, respondents can freely specify the exact position of their response which can subsequently be quantified up to the pixel. With Likert scales, respondents are limited to only certain categories on the continuum (Flynn, 2004). The increased sensitivity has been reported by researchers. When Joyce, Zutshi, Hrubes, and Mason (1975) measured pain on a four-point Likert scale and a VAS, the researchers concluded the VAS displayed more sensitivity. Neely and Borg (1995) reached the same conclusion when measuring perception of color change using both the Likert and VAS response formats. Compared to the Likert scale, the visual analog scale allows for true interval-level data while the Likert scale only allows for ordinal-level data.

The categories of Likert scales cannot be assumed to be equally spaced. For example, the distance on the continuum between 'somewhat agree' to 'strongly agree' might not be the same as the distance between 'somewhat disagree' and 'undecided' (Hodge & Gillespie, 2003; Svensson, 2001). The lack of precision of measurement means the data from Likert scales are at the ordinal-level, so it is not technically appropriate to sum or average such responses, even though researchers often do so. (Contrasted against the Likert scale, the visual analog scale provides a smoothed continuum for respondents (e.g., a score of 100 is twice the value of a 50 and is at the interval level). Also, some respondents find the VAS more intuitive (Little & McPhail, 1973; Aitken, 1969). For example, Zeally and Aitken (1969) concluded the VAS was the simplest method for their hospital patients.

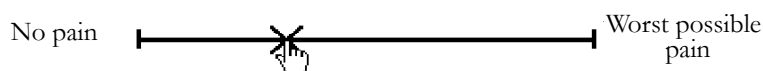


Figure 1. An example of a visual analogue scale

Other researchers also reported that children preferred the VAS (Abu-Saad, Kroonen, & Halfens, 1990; Shields, Palermo, Powers, Grewe, & Smith, 2003).

The biggest limitation of the VAS are related to the reliability of the scoring protocol (Bijur, Silver, & Gallagher, 2001). However, due to the recent advancements in technological capabilities, these limitations are negated by modern survey software. Historically, the VAS has not been widely used due to the practical limitations related to scoring (Couper, Tourangeau, Conrad, & Singer, 2006; Kersten et al., 2012). Researchers had to first measure the distance from the respondent's mark to the origin of the line with a ruler. Next, researchers needed to transcribe the distance for the data-entry; not only was this task time-consuming but the process was prone to error (Haefeli, & Elfering, 2006). Today, software can record the location of the click of the mouse automatically with no extra effort by the researcher.

Despite the aforementioned advantages of the VAS, researchers have not always reported positive results. Even though it is more precise, some researchers who have compared the VAS and the Likert scale reported that respondents could not distinguish more than seven to eleven different categories while using VAS (e.g., Thomeé et al., 1995; Munchi 2011). Also, some respondents (e.g., children and the elderly) prefer the Likert scale. (Couper, Tourangeau, Conrad, & Singer, 2006; Joyce, Zutshi, Hrubes, & Mason, 1975; Paediatr, 2004). The elderly described the difficulty in using VAS as requiring them to convert their responses using mathematical logic. Additionally, elderly participants reported that the VAS required more time to respond compared against the Likert scale (Joyce, Zutshi, Hrubes, & Mason, 1975). Although these criticisms have been identified in the literature, the documented benefits of the VAS scaling outweigh the potential problems. Furthermore, additional researcher training on the use of the VAS protocol can eliminate the potential scaling problems.

Reasons to use the retrospective pretest-posttest as opposed to conventional self-report methods

The second suggestion to improve the psychometric properties of traditional response formats is through

retrospective pre-test designs. Conventional pretest-posttest self-report measurements are frequently used to evaluate a treatment, intervention, or change in experience. Participants are given a questionnaire prior to receiving treatment, and are presented with the same questionnaire at the conclusion of the treatment.

The conventional pretest-posttest self-reports are contextually helpful in examining implicit attitudes or sentiments that individuals might not feel comfortable or able to articulate (Cohen, 2014). Yet, despite the frequent usage of conventional pretest posttest self-report measures, conventional self-report designs have methodological limitations.

The most discussed and problematic limitation associated with conventional pretest-posttest self-report designs is response-shift bias. Response-shift bias can be defined as an instance when an individual changed his or her perception or understanding of their initial functioning in response to a treatment (Howard, 1980). However, after participating in a program, participants have gathered more knowledge about the construct and are therefore more suited to give an accurate report. Yet, the response on the posttest will be answered from a different frame of reference compared against the pretest reference frame (Drennan & Hyde, 2008; Howard, Dailey, & Gulanick, 1979; Howard, 1980). The internal axes of reference for the pretest and the posttest differed within the individual when they completed the self-report measures. Therefore, it is understandable why comparisons between the two self-reports are not appropriate (Bray, Maxwell, & Howard, 1984).

The concept of examining response shift bias as a latent variable meta-construct of three interrelated processes has been proposed by researchers (Schwartz and Schwartz, 1999; Sprangers, Carey, & Reed, 2004). The processes of the response shift bias meta-construct are categorized as recalibration, reprioritization, and reconceptualization. Recalibration refers to an individual's changes his or her internal standards; an indication of this response shift effect occurring would be when the individual changes his or her meaning of the target construct.

When an individual has a change in his or her values or priorities of the target construct, this change suggested a reprioritization response shift has occurred. an individual's change in his or her definition of the target construct. To test the degree of the response shift bias, particularly the three specific interrelated constructs of response shift bias, a latent variable modeling approaches such as factor analysis, confirmatory factor analysis, and longitudinal structural equation modeling have been proposed to measure response shift bias and its effect on the target construct of interest. For further discussion of these methods, please refer to Oort, Nieuwkerk, and Sprangers (2001), Schwartz et al. (2004), and Oort (2005).

To overcome some of the limitations imposed by conventional pretest-posttest self-report measures, the retrospective pretest-posttest design should be utilized. The retrospective pretest-posttest is specifically designed to control for response shift bias by accurately depicting change in an individual after a treatment or intervention (Howard, Dailey & Gulanick, 1979). To illustrate an instance of this design, imagine a group of individuals has participated in a certain type of treatment, however, they do not fill out a survey prior to the start of the program. Once the program has concluded, participants completed a posttest survey and a retrospective pretest.

The retrospective pretest allowed for the individuals to consciously reflect back to their state prior to the start of the program, and determine whether they have undergone some change in knowledge or attitude (Cohen, 2014; Davis, 2002; Howard, 1980). The posttest and retrospective pretest are taken at the same point in time and are responded to using the same internal frame of reference for both self-reports. With the posttest and retrospective pretest being answered from the same point of reference, comparisons of change can be determined from self-reports. As Hoogstraten (1985) and others have noted, the key benefit of the retrospective pretest-posttest designs over conventional pretest-posttest self-reports is the elimination of the validity threat due to response-shift bias. Additionally, individuals tend not to overestimate

or under-report their emotions and knowledge of their behaviors on retrospective self-report measures. Breetvelt and Van Dam (1991) conducted a study on a sample of cancer patients and demonstrated that there is an underreporting with measures of emotional behavior or attitudes. Because of this underreporting bias, the researchers advocated using the retrospective self-report design to control for response-shift bias. Retrospective pretest designs are convenient to assess change of knowledge or attitude in an individual. Lastly, retrospective self-reports are extremely flexible since questions can be formulated to actually reflect the program content as it evolves over time in the program (Pratt, McGuigan & Katzev, 2000).

For further discussion of the retrospective methods, and application refer to Howard et al. (1979), as well as Howard (1980), Nakonezny, Rodgers, and Nussbaum (2003), and Nakonezny and Rodgers (2005).

We have discussed why retrospective pretest self-report measures are an effective method of capturing change in response scales, especially after an individual has participated in a program or treatment. Yet, retrospective designs do have a few limitations that need to be mentioned. Drennan and Hyde (2008) noted how retrospective pretest self-report measures could suffer from social desirability and impression management. Demand characteristics and memory related problems have also been shown to influence the recall process in retrospective pretest self-reports (Pratt et al., 2000). Howard, Millham, Slaten and O'Donnell (1981) investigated social desirability and demonstrated that retrospective pretest-posttest self-reports actually diminish the effects of social desirability in participant responses.

Despite the noted methodological limitations of self-reports, the retrospective pretest is a valuable strategy to control for response-shift bias and underestimation or overestimation of program effects. Schwartz et al. (2004) validated the retrospective pretest to be able to detect recall bias and recalibration bias. They compared the retrospective pretest results against covariance analytic approaches to see how the three different aspects of response shift bias interrelate. The results of the study were not redundant; they found

that individuals changed their internal standards over time. Additionally, this finding is compounded by the authors' suggestion of measuring response shift bias using the retrospective pretest-posttest coupled with latent variable modeling methods.

Likert scales and conventional pretest-posttest self-report measures are certainly the dominant methods used in measuring subjective experiences in contemporary social and behavioral sciences. Yet, these techniques are not sufficient in capturing the true experiences of an individual. To capture and measure change after a proposed treatment, researchers should consider incorporating both VAS and retrospective pretest-posttest self-reports in their research designs.

Research such as this has shown that retrospective pretest designs can overcome the limitations of conventional pretest-posttest self-reports, particularly the threat of response-shift bias. However, retrospective pretest self-reports are not welcomed in all research circles, with points of contention being related to philosophical objections (Howard, 1980). On the contrary, we suggest that researchers should consider implementing both VAS and retrospective pretest-posttest measurements in future research. We also encourage researchers to think innovatively in development measurement systems that are highly sensitive to changes in the level of a construct and to changes in the construct over time.

References

- Aiken, L.S., and West, S.G. (1990). Invalidity of true experiments, self-report pretest biases. *Evaluation Review*, 14, 374-390.
- Aitken, R.C. (1969). Measurement of feelings using visual analogue scales *Proceedings of the Royal Society of Medicine*, 62, 989-993.
- Abu-Saad, H. H., Kroonen, E., & Halfens, R. (1990). On the development of a multidimensional Dutch pain assessment tool for children. *Pain*, 43, 249-256
- Bijur, P.E., Silver, W., & Gallagher, J. (2001). Reliability of the visual analog scale for measurement of acute pain. *Academic Emergency Medicine*, 8, 1153-1157.
- Bray, J.H., Maxwell, S.E., & Howard, G.S. (1984). Methods of analysis with response-shift bias. *Educational Measurement and Psychological Measurement*, 44, 781-804.
- Breetvelt, I.S., & Van Dam, F.S.A.M. (1991). Underreporting by cancer patients: the case of response-shift. *Social Science Medicine*, 32, 981-987.
- Cohen, E.H. (2014). Self-assessing the benefits of educational tours. *Journal of Travel Research*, published online 11 Sept 2014, 1-9.
- Cronbach, L.J., & Furby, L. (1970). How we should measure "change"-or should we? *Psychological Bulletin*, 74, 68-80.
- Couper, M. P., Tourangeau, R., Conrad, F. G., & Singer, E. (2006). Evaluating the effectiveness of visual analog scales: A Web experiment. *Social Science Computer Review*, 24, 227-245.
- Davis, G.A. (2002). Using a retrospective pre-post questionnaire to determine program impact. Paper presented at Mid-Western Education Research Association.
- Drennan, J., & Hyde, A. (2008). Controlling response shift bias: The use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment & Evaluation in Higher Education*, 33, 699-709.
- Edmonson D.R. (2005). Likert Scale: A History. Retrieved from CHARM database.
- Finkelstein, J.A., Quaranto, B.R., & Schwartz, C.E. (2014). Threats to the internal validity of spinal surgery outcome assessment: recalibration response shift or implicit theories of change? *Applied Research Quality Life*, 9, 215-232.
- Flynn, D., van Schaik, P., & van Wersch, A. (2004). A comparison of multi-item Likert and visual analogue scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment*, 20, 49-58.
- Haefeli, M., & Elfering, A. (2006). Pain assessment. *European Spine Journal*, 15, 17-24.
- Hayes, M., & Patterson, D. (1921). Experimental development of the graphic rating method. *Psychological Bulletin*, 18, 98-99.

- Hodge, D. R. & Gillespie, D. F. (2003). Phrase Completions: An alternative to Likert scales. *Social Work Research*, 27, 45-55.
- Hoogstraten, J. (1985). Influence of objective measures on self-reports in a retrospective pretest-posttest design. *The Journal of Experimental Education*, 53, 207-210.
- Howard, G.S. (1980). Response-shift bias: A problem in evaluating interventions with pre-post self-reports. *Evaluation Review*, 4, 93-106.
- Howard, G.S., Dailey, P.R., & Gulanick, N.A. (1979). The feasibility of informed pretests in attenuating response-shift bias. *Applied Psychological Measurement*, 3, 481-494.
- Howard, G.S., Millham, J., Slaten, S., & O'Donnell, L. (1981). Influence of subject response style effects on retrospective measures. *Applied Psychological Measurement*, 5, 89-100.
- Joyce, C.R.B., Zutshi, D.W., Hrubes, V., & Mason, R.M. (1975). Comparison of fixed interval and Visual Analogue Scales for rating chronic pain. *European Journal of Clinical Pharmacology*, 8, 415-420.
- Kievit, W., Hendriks, J., Stalmeier, P.F.M., van de Laar, M.A.F.J., Van Riel, P.L.C.M., and Adang, E.M. (2010). The relationship between change in subjective outcome and change in disease: A potential paradox. *Quality of Life Research*, 19, 985-994.
- Likert, R. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 140, 5-55.
- Little, J. C., & McPhail, N. I. (1973). Measures of depressive mood at monthly intervals. *British Journal of Psychiatry*, 122, 447-452.
- Munshi J. A method for constructing Likert scales. Sonoma State University. [Internet]. 1990. Retrieved from: <http://www.munshi.4t.com/papers/likert.html>
- Nakonezny, P.A., and Rodgers, J.L. (2003). An empirical evaluation of the retrospective pretest: Are there advantages to looking back? *Journal of Modern Applied Statistical Methods*, 4, 240-250.
- Nakonezny, P.A., Rodgers, J.L., and Nussbaum, J.F. (2003). The effect of later life parental divorce on adult-child/older-parent solidarity: A test of the buffering hypothesis. *Journal of Applied Social Psychology*, 33, 115-1178.
- Neely, G., & Borg, E. (1995). Properties of a category ratio scale (CR-10) and the Visual Analogue Scale (VAS): A comparison with magnitude estimation, line production, and category scaling (Tech. Rep. No. 791). Stockholm, Sweden: Stockholm University, Department of Psychology.
- Pratt, C.C., McGuigan, W.M., and Katzev, A.R. (2000). Measuring program outcomes: using retrospective pretest methodology. *American Journal of Evaluation*. 21, 341-349.
- Rausch, M. & Zehetleitner, M. (2014). A comparison between a visual analogue scale and a four point scale as measures of conscious experience of motion. *Consciousness and Cognition*, 28, 126-140.
- Shields, B. J., Palermo, T. M., Powers, J. D., Grewe, S. D., & Smith, G. A. (2003). Predictors of a child's ability to use a visual analogue scale. *Child: Care, Health and Development*, 29, 281-290.
- Sprangers, M. (1989). Subject bias and the retrospective pretest in retrospect. *Bulletin of the Psychometric Society*, 27, 11-14.
- Svensson, E. (2001). Construction of a single global scale for multi-item assessments of the same variable. *Statistical Medicine*, 20, 3831-3846.
- Thoméé R., Grimby G., Wright B.D., & Linacre J.M. (1995). Rasch analysis of Visual Analog Scale measurements before and after treatment of patellofemoral pain syndrome in women. *Scandinavian Journal of Rehabilitation Medicine*, 27, 145-151.
- Zealley A.K., & Aitken R.C.B. (1969). Measurement of mood. *Proceedings of the Royal Society of Medicine*, 62, 993-996.

Received: February 2nd 2016

Accepted: April 7th 2016