# EVIDENCE-BASED SURVEY DESIGN: THE USE OF CONTINUOUS RATING SCALES IN SURVEYS

Seung Youn (Yonnie) Chyung | Ieva Swanson | Katherine Roberts | Andrea Hankinson

When practitioners and researchers develop structured surveys, they may use Likert-type discrete rating scales or continuous rating scales. When administering surveys via the web, it is important to assess the value of using continuous rating scales such as visual analog scales (VASs) or sliders. Our close examination of the literature on the effectiveness of the two types of rating scales showed both benefits and drawbacks. Many studies recommended against using sliders due to functional difficulties that cause low response rates.

## INTRODUCTION

In a structured survey questionnaire, survey respondents answer each survey item using the response scale provided. A popular response scale used in the social sciences is the Likert scale, which Rensis Likert, an American social psychologist, developed in the 1930s (Likert, 1932). The Likert scale uses a symmetrical (bipolar) format with disagreement options on one side, agreement options on the other side, and a neutral option in the middle (e.g., *Strongly Disagree*, *Disagree*, *Neutral*, *Agree*, *Strongly Agree*).

Survey developers can use other Likert-type scales, measuring the degrees of other attitudinal aspects such as being dissatisfied/satisfied or useless/useful, with or without a neutral option. Likert-type scales can also be unipolar, measuring the degree of a variable in one direction (e.g., *Never*, *Sometimes*, *About Half the Time*, *Most of the Time*, *Always*), and partially labeled, providing numerical points with only the two ending anchors (e.g., *Strongly Disagree 1 2 3 4 5 Strongly Agree*). Whether they are bipolar, unipolar, or fully or partially labeled, Likert-type scales are *discrete rating scales,* since they provide a discrete number of options from which to choose. The number of options may vary from 2 to 11 (Svensson, 2000).

An alternative to discrete rating scales are *continuous rating scales* that allow respondents to indicate a response anywhere within a given range. For example, while the 6-point pain measurement scale shown in Figure 1 is a discrete rating scale, the partially labeled scale with a horizontal line shown in Figure 2 is a continuous rating scale that allows respondents to indicate their response by marking a place on the horizontal line.

These different types of response scales give you different types of data. The 6-point discrete rating scale results in a collection of the six *ordered values* (e.g., *Mild* or *Moderate*). With the continuous rating scale, you get precise *numbers*—for example, 26 or 26.8 when a 0-to-100 range is used. Such different types of data lead to the use of different statistical analysis procedures. You would analyze the ordinal-type data obtained from bipolar or unipolar Likert-type discrete rating scales with non-parametric statistical procedures, although you may decide to use parametric statistical analyses on ordinal-type data if you find the data to be approximately normally distributed. On the other hand, continuous rating scales typically generate interval data, much like the data you would find from using a ruler, which increases the possibility of the data being normally distributed and allows the use of a wider range of statistical procedures. Thus, as a survey developer, you have to decide whether to use discrete or continuous rating scales in your structured survey questionnaires.

Although there are other issues that survey developers need to consider when developing survey instruments—

Performance Improvement, vol. 57, no. 5, May/June 2018


Published online in Wiley Online Library (wileyonlinelibrary.com) • DOI: 10.1002/pfi.21763

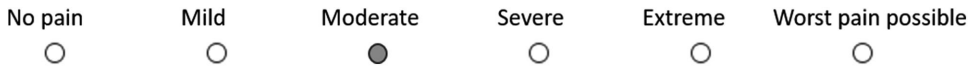


FIGURE 1. A DISCRETE RATING SCALE FOR MEASURING PAIN LEVELS


FIGURE 2. A CONTINUOUS RATING SCALE FOR MEASURING PAIN LEVELS, WITH TWO ENDING ANCHORS

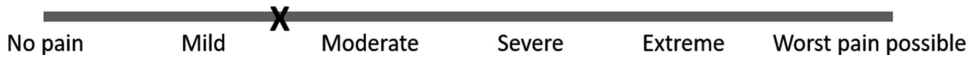


FIGURE 3. A CONTINUOUS RATING SCALE FOR MEASURING PAIN LEVELS, WITH SEVERAL DESCRIPTORS

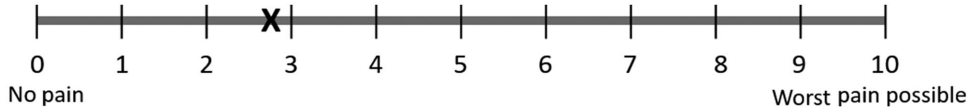


FIGURE 4. A NUMERICAL RATING SCALE FOR MEASURING PAIN LEVELS

such as whether to use a midpoint, whether to include negatively worded items, or whether to use an ascending or descending order of anchors—each of these topics is worth a separate article. Teams of researchers from the Organizational Performance and Workplace Learning department at Boise State University have been reviewing research articles and developing evidence-based recommendations that address these topics on developing survey instruments. For example, see Chyung, Roberts, Swanson, and Hankinson (2017) on the use of a midpoint on the Likert scale, and Chyung, Barkin, and Shamsy (2018) on the use of positively and negatively worded items in surveys.

This article is one of a series of articles on evidence-based survey design, focusing on the use of continuous rating scales. The purpose of this article is to provide (a) an overview of continuous rating scales as an alternative to discrete rating scales, and (b) research-based evidence on the benefits and drawbacks of using continuous rating scales.

## CONTINUOUS RATING SCALES AS AN ALTERNATIVE TO LIKERT-TYPE SCALES

### Development of Variations of Visual Analog Scales

A type of continuous (thus, *analog*) measurement instrument is the *visual analog scale* (VAS). When first introduced as a graphic rating scale (GRS) in the early 1920s (Hayes & Patterson, 1921; Patterson, 1922), the scale had several descriptors along the horizontal line as shown in Figure 3. By the 1960s, the term VAS began appearing in the literature (Wewers & Lowe, 1990). The VAS does not have descriptors along the line. With a paper-based VAS or a ruler-type VAS device, survey respondents indicate their rating by placing a mark on a horizontal line (often 100 mm in length) between two ending anchors, as shown in Figure 2. The VAS has been used frequently as a self-assessment tool in the medical field to measure subjective perceptions of symptoms, such as pain, tinnitus, or annoyance (e.g., Adamchic, Langguth, Hauptmann, & Tass, 2012; Bourdel et al., 2015).

Another variation of VAS is a numerical response scale (NRS) that presents numerical marks instead of verbal descriptors between the two ending points, as shown in Figure 4. We should note a caveat: an NRS is a continuous rating scale if the respondents are allowed to mark anywhere on a numerical continuum. An NRS can have the appearance of a continuous rating scale but is actually a discrete rating scale if survey respondents are asked to select one of the numerical anchors (0 to 10) provided on the scale (e.g., Bourdel et al., 2015).

There is similar confusion in the use of the term GRS. Despite the fact that the GRS was originally introduced as a continuous rating scale (see Patterson, 1922), a simple web search will reveal that the same term is used to describe discrete rating scales with several descriptive levels (e.g., *Exceeded expectations consistently*, *Exceeded expectations many times*, etc.) or discrete rating scales with graphically represented options or emoticons (e.g., ☺😐☹).

There are practical concerns related to using VASs on paper, as it is not as easy to accurately read the data marked by survey respondents as compared with using Likert-type discrete rating scales (Friedman & Friedman, 1986; Gardner, Cummings, Dunham, & Pierce, 1998; Reips &

Funke, 2008). When using paper-based VASs, researchers would use a ruler or stencil to determine exactly where the respondents' marks fall on the horizontal line and convert the marks to numerical scores (e.g., Briesch et al., 2013; Patterson, 1922), which is time-consuming. However, these drawbacks can be eliminated when you use web-based surveys, as the web survey system can digitally and automatically record the marking (Reips & Funke, 2008).

### Continuous Rating Scales on the Web

VASs used on web-based surveys allow survey respondents to *point and click* to indicate their response on a horizontal line. Some web-based survey systems such as Qualtrics™ offer bar scales that are similar to web-based VASs in that survey respondents *point and click* to indicate their response in a bar graph (see Figure 5).

Web-based survey tools also provide slider scales (or *sliders*) that are similar to VASs. Sliders are frequently used in social science research and market research (Funke, 2016). Similar to the traditional VAS 100-mm range, sliders can be designed to measure input using a 0-to-100-point range and even to a couple of decimal places (e.g., 56.23). With sliders, the survey respondents *drag and drop* the marker, as opposed to *pointing and clicking* when using web-based VASs without a marker. The marker can be placed at the low end or in the middle of the sliders (see Figure 6 and Figure 7). You may make a few numerical anchors visible or invisible to the survey respondents. You may hide numerical anchors from respondents to help them focus on their perception rather than paying attention to the numbers (e.g., Lin, Manuel, McFatter, & Cech, 2016).

Although sliders are often used as continuous rating scales, they can be designed to function as discrete rating scales (e.g., 10- or 11-point scales), similar to discrete NRSs, by making the marker snap to grid between grid lines (e.g., Thompson, Thomas, Manning, & Hogg, 2016). Alternatively, when sliders are set to capture input within a 0-to-100 range *without decimal points*, it is unclear whether they are a true continuous rating scale or a 101-point discrete rating scale. Some researchers refer to a 101-point slider as a quasi-continuous rating scale (e.g., Hadjiiski et al., 2007).

In the following section, we present research evidence of the benefits and drawbacks of using continuous rating scales such as VASs or sliders in comparison with the use of Likert-type discrete rating scales. In doing so, we will define the two types of rating scales using the following criteria:

- Discrete rating scales record the respondents' choice among a discrete number of options, usually no more than 11.
- Continuous rating scales record the respondents' mark on a continuum with a minimum of 100 points (units).

We will also describe the rating scale appearance (paper- or web-based), which may influence the scale's effectiveness. In describing the research we reviewed, we characterized the type of scale used in the research as a VAS or a slider, based on the descriptions provided in the research reports. However, many research articles did not present a figure of the scale used; therefore, it is unclear whether VASs tested in web-based studies used a *point-and-click* method without a marker (like conventional paper-based VASs) or a *drag-and-drop* method with a marker (like sliders).

## RESEARCH FINDINGS ON BENEFITS AND DRAWBACKS OF USING CONTINUOUS RATING SCALES

### Benefits of Using Continuous Rating Scales

One of the benefits of using continuous rating scales lies in their usability. In a web-based study by Funke and Reips (2012) with 268 students at the University of Kassel in Germany, the researchers compared the quality of data collected from a set of partially labeled 5-point discrete rating scales (using radio buttons) and a set of 250-point continuous rating VASs (using a point-and-click method). After analyzing not only the sensitivity of data but also click frequency and completion speed, they found that respondents modified their ratings twice as often on the VASs as on the 5-point scales, without causing any negative impact on the mean scores, the response time, and the number of non-responses. However, it is unclear whether the increased frequency of response adjustment is a benefit of using VASs (i.e., the survey type) or of using the web (i.e., the delivery medium). Nonetheless, the researchers saw that being able to make precise decisions within the same amount of time as a benefit of using web-based VASs.

While Funke and Reips (2012) found no significant difference in response time between the use of 5-point scales and VASs in a web survey, a study conducted in Finland by Voutilainen, Pitkäaho, Kvist, and Vehviläinen-Julkunen (2015) showed positive outcomes of using a paper-based VAS. When patients (n = 150) responded to a 5-point Likert scale and a VAS on an 80-mm line to express their patient satisfaction before hospital discharge, their response time with a VAS was 28% (or three and a half minutes) faster. The researchers also recognized that the use of VASs

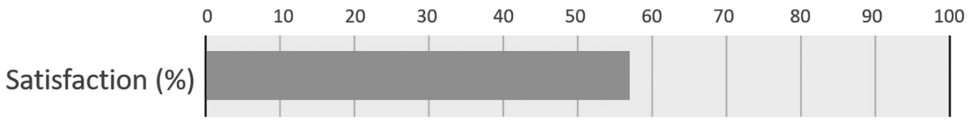


FIGURE 5. AN EXAMPLE OF BAR SCALES

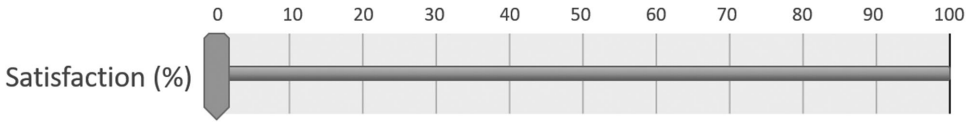


FIGURE 6. AN EXAMPLE OF SLIDER SCALES WITH NUMERICAL ANCHORS AND WITH A MARKER PLACED AT THE LOW END AS THE STARTING POINT

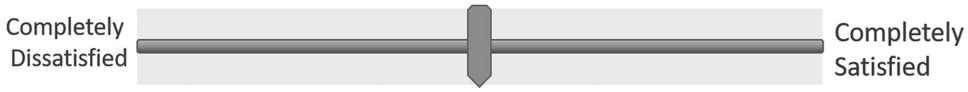


FIGURE 7. AN EXAMPLE OF SLIDER SCALES WITHOUT NUMERICAL ANCHORS AND WITH A MARKER PLACED IN THE MIDDLE AS THE STARTING POINT

was less vulnerable to age-related bias and helped avoid the ceiling effect that may occur when using a discrete rating scale with only a few options (i.e., a lot of data fall onto the top anchor of the scale resulting in little variance in data) compared with using 5-point Likert scales. For these reasons, the researchers concluded that there is strong support for the use of VASs in patient satisfaction surveys.

Another benefit of using continuous rating scales is their psychometric properties that produce interval data, allowing researchers to apply a wider range of statistical procedures than nominal- or ordinal-type scales allow. Evidence that data obtained from VASs are interval scale data is found in Reips and Funke's (2008) study involving 405 students at the University of Kassel. The researchers used web-based VASs created with the web-based software, VAS Generator, to compare the types of data obtained from six conditions made up by using two mathematical formats (percentages such as 20% or 50% and ratios such as 1/5 or 1/2) and three different lengths of VAS (50, 200, and 800 pixels). They found that regardless of the formats and lengths of VASs, the collected data showed roughly equal segments on the VASs, which indicates strongly that data collected with VASs are interval data.

Even in discrete rating scales, using increased numbers of categories is shown to have benefits. In Leung's (2011) paper-based study of 1,217 Macau students in grades 7–12, using a Chinese version of the Rosenberg Self-Esteem Scale, the researcher compared data collected by 4-, 5-, 6-, and 11-point Likert scales. He found that the data obtained from an 11-point scale are closer to a normal distribution than the data obtained from Likert-type scales with six points or fewer. Also in Dawes' (2002) study involving a telephone survey of 751 Australian taxicab users and face-to-face interviews with 301 undergraduates using 5- and 11-point Likert scales, the researcher found that using 11-point scales resulted in more variance in the data. Therefore, 11-point Likert scales are more useful than 5-point Likert scales if one wishes to conduct statistical analyses (such as regression analysis) that require sufficient variance in data.

With the benefits reviewed in these studies, continuous rating scales such as VASs and sliders are frequently used in healthcare and the social sciences—for example, in tinnitus, endometriosis pain treatment, and orthopedics studies (Adamchic et al., 2012; Bourdel et al., 2015; Wall et al., 2017); in measuring happiness and patient satisfaction (Studer, 2012; Voutilainen et al., 2015); and in linguistics (Munson, Schellinger, & Edwards, 2017). Researchers in these studies found paper- or web-based VASs or sliders to be reliable measurement instruments, and recommended them over 5-, 7-, or 10-point discrete rating scales.

## No Clear Advantages of or Drawbacks in Using VASs or Sliders

Despite the benefits discussed previously, other research revealed evidence that calls into question clear advantages in using continuous rating scales over Likert-type scales. Several studies revealed no superiority in psychometric properties and indicated drawbacks of using paper-based continuous rating scales in terms of functionality and user preference. For example, Hilbert et al. (2016) completed a study in Germany with 866 university students using a paper-based survey to test three conditions: (1) a Yes-or-No dichotomous scale, (2) a 5-point Likert-type scale, and

(3) a 100-mm VAS. While Cronbach's alpha, a measure of internal consistency, increased from the Yes-or-No scale, to the 5-point scale, to the VAS, the Omega that compensates for the shortcomings of Cronbach's alpha (Peters, 2014) did not vary. The researchers thus concluded there were no significant differences among the three types of scales in terms of reliability. However, they indicated that the Yes-or-No scale and the 5-point scale were easier to use as compared with the VAS.

The web-based VASs or sliders also have some drawbacks. In a web-based study by Couper, Tourangeau, Conrad, and Singer (2006), 1,290 respondents provided their answers using a VAS with a marker (i.e., functioning as a slider), a Likert-type scale using radio buttons, or a numeric text-entry method. The researchers found no significant difference in variances due to the types of response methods. Briesch et al.'s (2013) study also found negligible variance attributed to the scale type (continuous vs. discrete) when it was paper-based. Couper et al. (2006), however, found problems with their slider. Missing data rates were on average twice as high with the slider as with the radio buttons and the numeric input method. The slider also suffered higher rates of non-completion and longer completion times compared to the two other response methods. These drawbacks of sliders are in contrast to the outcomes of using the paper-based VASs providing no difference or a quicker response time compared with the Likert-type scales reviewed in the previous section (Funke & Reips, 2012; Voutilainen et al., 2015).

Similarly, in Cook, Heath, Thompson, and Thompson's (2001) study, 4,407 respondents completed a web-based survey using two formats: a 9-point Likert-type scale and a slider scale that scored data on a 1-to-100 point range. The researchers found that both the Likert-like radio-button scale and the slider were reliable and psychometrically acceptable data-gathering tools. However, they also found that it took 71.2 seconds longer for respondents to complete the survey with the slider, as compared with using the radio-button scale.

Also in Maloshonok and Terentev's (2016) study, the researchers compared three types of user interfaces used in the web surveys of massive open online courses (MOOCs): a set of radio buttons, a slider, and a text-input box. About 3,000 online students were asked to select, indicate, or enter the number of hours by using one of the three scales. The researchers found that slider users frequently overestimated their numbers and used the *Don't know* button more often. Their conclusion was that radio buttons are easier to use and less challenging than a slider and recommended using traditional radio buttons for rating such questions.

Some web-based survey systems offer sliders with emoticons (e.g., as you move the marker on the slider, a thermometer next to the slider shows changes in temperature). It is tempting to use emoticon sliders, hoping to increase respondent engagement. However, these graphic sliders have similar drawbacks. In Derham's (2011) Australian web-based marketing research, respondents enjoyed using emoticon sliders more than using numbered and worded scales. Despite the engagement and enjoyment factors, the cognitive difficulties that respondents tend to encounter with sliders still applied. The emoticon sliders resulted in lower quality data due to higher levels of unanswered questions.

Sliders are difficult to use and more challenging for respondents with lower-than-average education. Funke, Reips, and Thomas (2011) noted this issue in their study with 779 people who responded to a survey on health-related products, using slider or radio-button scales. In this study, although the slider did not show numeric values on the scale and had the appearance of a continuous rating scale, it was designed to produce seven discrete values. The researchers found that respondents using sliders experienced longer response times (independent of their education level) and that those with lower-than-average education experienced seven times more break-off rates. Thus, the researchers recommended using low-tech solutions with radio buttons instead of using slider scales, especially for respondents with low formal education.

Survey developers should also recognize respondents' increased access to mobile devices to complete survey requests. Funke's (2016) web-based study regarding the use of sliders (drag and drop), VASs (point and click), and Likert scales with radio buttons (point and click) noted that sliders were particularly difficult for those responding via mobile devices and caused longer response time. The researcher noted, however, that respondents could use VASs and radio buttons on mobile devices without difficulty. Thus, the researcher recommended using Likert scales with radio buttons or VASs that do not use an initial marker, in order to avoid the difficulties in dragging and dropping on mobile devices.

Finally, in any survey context, clear instructions can enhance the use of the survey instruments and improve their validity and reliability, which is the case when using web-based sliders. Roster, Lucianetti, and Albaum (2015) suggest that a survey designer include clear instructions regarding the use of sliders to mitigate missing data. They also recommend using a *Don't know* or *Prefer not to answer* option.

**TABLE 1** EVIDENCE-BASED RECOMMENDATIONS ON THE USE OF DISCRETE OR CONTINUOUS RATING SCALES IN STRUCTURED SURVEYS

| | DISCRETE RATING SCALES | CONTINUOUS RATING SCALES |
|---|---|---|
| Benefit | • Respondents find them easy to use.<br>• Likert-type scales generate higher completion rates, compared with sliders. | • They allow respondents to make precise decisions on their input.<br>• They help improve the instrument's reliability.<br>• They help increase the possibility of having normally distributed data, allowing the use of a wide range of statistical procedures. |
| Problem | • They likely generate ordinal-type data, allowing the use of a limited range of statistical analyses.<br>• They (e.g., 5-point scales) may be prone to a ceiling effect, resulting in little variance in data and causing difficulty during statistical analysis. | • When using paper-based VASs, it is time consuming to record data.<br>• When using sliders, respondents may have trouble in using them, spend more time to complete the survey, and produce more incomplete data (especially those with lower education levels).<br>• Sliders are difficult to use on mobile devices. |
| When | • It is appropriate to use them when presenting simple descriptive statistics such as frequency or mean values of the survey data. | • It is appropriate to use them not only to present simple descriptive statistics but also when performing statistical procedures that require interval type, parametric data, or when needing sufficient variance in data. |
| Strategies to Use | • Add *Don't know*, *Prefer not to respond*, or *Not applicable* options to collect valid data.<br>• Also see Chyung et al. (2017) for more information regarding the use of a midpoint in Likert-type discrete rating scales. | • Be aware that a slider can be designed as a discrete or continuous rating scale; setting it to snap from zero to 10 to 20, and so on, up to 100 makes it an 11-point discrete rating scale.<br>• Add *Don't know*, *Prefer not to respond*, or *Not applicable* options to collect valid data. |

**TABLE 2** RESEARCH EVIDENCE FOR THE USE OF CONTINUOUS RATING SCALES

| FOCUS | AUTHORS (YEAR) | SURVEY MEDIA USED | RECOMMENDATIONS BASED ON RESEARCH FINDINGS |
|---|---|---|---|
| **Continuous rating scales have benefits.** | Adamchic et al. (2012) | Unclear report on the media used | • VASs are reliable instruments for measuring subjective perceptions, such as chronic tinnitus. |
| | Bourdel et al. (2015) | Literature review | • VASs are the tool, most frequently reported in endometriosis pain treatment related publications. VASs are valid, reliable, and precise. |
| | Funke and Reips (2012) | Web-based survey | • Respondents modified their ratings with a web-based VAS twice as often as with a 5-point scale using radio buttons, although there were no differences in response times and data whether to use a VAS or a 5-point scale.<br>• The use of a VAS is recommended as it also allows detection of small differences and use of many statistical procedures. |

(*Continued*)

**TABLE 2** Continued

| FOCUS | AUTHORS (YEAR) | SURVEY MEDIA USED | RECOMMENDATIONS BASED ON RESEARCH FINDINGS |
|---|---|---|---|
| | Reips and Funke (2008) | Web-based survey | • Web-based VASs measure data on an interval scale, which allows the use of a wide range of statistical procedures when analyzing the data obtained from VASs. |
| | Studer (2012) | Web-based survey | • There were no differences in data whether to use a VAS or a 10-point Likert scale when measuring happiness.<br>• Using a VAS is still preferable to using a 10-point scale as it overcomes the ordinal measurement of happiness. |
| | Voutilainen et al. (2015) | Unclear report on the media used | • It was 28% quicker to complete a VAS than to complete a 5-point Likert scale; the use of a VAS helps avoid the ceiling effect compared to when using a Likert scale.<br>• The use of a VAS rather than a 5-point Likert scale is recommended for measuring patient satisfaction. |
| | Wall et al. (2017) | Web-based survey | • A continuous slider scale showed higher inter-rater reliability compared with a 7-point ordinal scale.<br>• It is recommended to use a continuous rating scale than an ordinal rating system for orthopedic measurement. |
| **Even having more response points in discrete rating scales shows some benefits.** | Dawes (2002) | In-person interviews and telephone surveys | • If it is simply to present descriptive statistics, either a 5-point Likert scale or an 11-point scale works, although an 11-point scale may be more easily interpreted.<br>• If it is to conduct a regression-type analysis, an 11-point scale can be more useful because it results in more variance in the data. |
| | Leung (2011) | Paper-based survey | • Compared with data obtained from 4-, 5-, and 6-point Likert scales, data obtained from an 11-point scale is the closest to a normal distribution; thus, an 11-point scale is recommended. |
| **Paper- or web-based VASs are not always functionally better. There are drawbacks in using sliders.** | Couper et al. (2006) | Web-based survey | • There is no significant difference in variances whether to use a slider, select a radio button (for a Likert-type scale), or enter text in a textbox.<br>• A slider takes longer to complete, has more missing data, and suffers higher non-completion rates. |

(*Continued*)

**TABLE 2** Continued

| FOCUS | AUTHORS (YEAR) | SURVEY MEDIA USED | RECOMMENDATIONS BASED ON RESEARCH FINDINGS |
|---|---|---|---|
| | Hilbert et al. (2016) | Paper-based survey | • Among a dichotomous scale, a Likert-type scale, and a VAS, there were no significant differences in terms of reliability and validity.<br>• For practical use, Likert scales are easy to understand and use, in contrast to VAS. |
| | Cook et al. (2001) | Web-based survey | • Nine-point Likert-type scales using radio buttons and unnumbered graphical sliders both produced reliable data.<br>• It takes somewhat longer for respondents to manipulate sliders than to use radio buttons. |
| | Derham (2011) | Web-based survey | • The use of sliders with emoticons was the least effective compared to the use of numbered and worded scales, so the use of sliders with emoticons is not recommended. |
| | Funke (2016) | Web-based survey | • Sliders (drag and drop) produced more incomplete and longer response time than web-based VASs (point and click) and Likert scales with radio buttons.<br>• Use a Likert scale with radio buttons or VASs (without an initial marker). |
| | Funke et al. (2011) | Web-based survey | • Seven-point slider scales produced about seven times more break-offs among respondents with low formal education and, regardless of education levels, significantly increased response time, compared with 7-point radio-button scales.<br>• Do not use slider scales; use low-tech solutions such as radio buttons. |
| | Maloshonok and Terentev (2016) | Web-based survey | • More respondents selected *Don't know* while using a slider than those who used radio buttons.<br>• The use of radio buttons is less challenging as compared with the use of slider and text box interfaces. |
| | Roster et al. (2015) | Web-based survey | • The use of a 5-point slider and 5-point radio buttons yielded no significant differences in response rates, completion time, and mean scores.<br>• When using sliders, provide clear instructions and offer *Don't know* or *Prefer not to answer* options to reduce missing data. |

[a]Some had an appearance of a continuous rating scale but produced discrete values.

*Based on the novelty effect of sliders, survey developers may feel tempted to replace conventional Likert-type discrete rating scales with sliders, without knowing their true effectiveness.*

## SUMMARY

Practitioners and researchers who develop their own survey instruments need to make many design-related decisions. In this article, we addressed one of the design issues: the pros and cons of using continuous rating scales such as VASs or sliders, compared with using Likert-type discrete-rating scales. In addressing this topic, we also needed to consider the impact of the delivery media used to administer the survey instruments. Traditionally, surveys have been administered on paper and some on ruler-type devices (e.g., pain-measuring VASs). However, due to the popularity of web-based survey systems and respondents' increased access to the web via desktop and mobile devices, practitioners and researchers now frequently administer their surveys via the web. Web-based survey systems commonly provide sliders as one of the response scale options. Based on the novelty effect of sliders, survey developers may feel tempted to replace conventional Likert-type discrete rating scales with sliders, without knowing their true effectiveness.

We reviewed many research studies that investigated the effectiveness of using continuous rating scales delivered on paper, using ruler-type devices, and on the web, as compared with using Likert-type discrete rating scales. Some research indicated benefits of using continuous rating scales, while other research did not show such benefits. There are, in fact, some drawbacks in using continuous rating scales, especially sliders in web-based surveys. Based on the research evidence, it is our conclusion that no one type is inherently better than the other; however, practitioners and researchers need to select the one that works better for the purpose of the instrument and should exercise caution about using sliders, given their known drawbacks. Concerning sliders, the consensus is to pay attention to the functional issues presented previously since the disadvantages they bring to surveys seem to outweigh perceivable advantages. In addition, practitioners and researchers should be aware that a VAS or slider that has the appearance of a continuous rating scale may be designed to produce discrete values. We encourage researchers to clearly describe the type of scales used in research as discrete or continuous rating scales.

*Based on the research evidence, it is our conclusion that no one type is inherently better than the other; however, practitioners and researchers need to select the one that works better for the purpose of the instrument and exercise caution about using sliders, given their known drawbacks.*

Table 1 presents a summary of the benefits and problems associated with using discrete and continuous rating scales, as well as strategies to employ when using each type of scale. Table 2 presents a summary of the research evidence used in generating these recommendations.

## References

Adamchic, I., Langguth, B., Hauptmann, C., & Tass, P.A. (2012). Psychometric evaluation of visual analog scale for the assessment of chronic tinnitus. *American Journal of Audiology*, *21*(2), 215–225. https://doi.org/10.1044/1059-0889(2012/12-0010)

Bourdel, N., Alves, J., Pickering, G., Ramilo, I., Roman, H., & Canis, M. (2015). Systematic review of endometriosis pain assessment: How to choose a scale? *Human Reproduction Update*, *21*(1), 136–152. https://doi.org/10.1093/humupd/dmu046

Briesch, A. M., Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2013). The influence of alternative scale formats on the generalizability of data obtained from direct behavior rating single-item scales (DBR-SIS). *Assessment for Effective Intervention*, *38*(2), 127–133. https://doi.org/10.1177/1534508412441966

Chyung, S.Y., Barkin, J., & Shamsy, J. (2018). Evidence-based survey design: The use of negatively-worded items in surveys. *Performance Improvement, 57*(2).

Chyung, S.Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, *56*(10).

Cook, C., Heath, F., Thompson, R.L., & Thompson, B. (2001). Score reliability in web- or internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, *61*(4), 697–706.

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*(2), 227–245. https://doi.org/10.1177/0894439305281503

Dawes, J. (2002). Five point vs. eleven point scales: Does it make a difference to data characteristics? *Australasian Journal of Market Research*, *10*(1), 39–47.

Derham, P.A.J. (2011). Using preferred, understood or effective scales? How scale presentations effect online survey data collection. *Australasian Journal of Market & Social Research*, *19*(2), 13–26.

Friedman, L.W., & Friedman, H.H. (1986). Comparison of itemised vs graphic rating scales: A validity approach. *Journal of the Market Research Society*, *28*(3), 285–289.

Funke, F. (2016). A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*, *34*(2), 244–254. https://doi.org/10.1177/0894439315575477

Funke, F., & Reips, U-D. (2012). Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods*, *24*(3), 310–327. https://doi.org/10.1177/1525822X12444061

Funke, F., Reips, U-D., & Thomas, R.K. (2011). Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review*, *29*(2), 221–231. https://doi.org/10.1177/0894439310376896

Gardner, D.G., Cummings, L.L., Dunham, R.B., & Pierce, J.L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, *58*(6), 898–915. https://doi.org/10.1177/0013164498058006003

Hadjiiski, L., Chan, H.-P., Sahiner, B., Helvie, M.A., & Roubidoux, M.A. (2007). Quasi-continuous and discrete confidence rating scales for observer performance studies: Effects on ROC analysis. *Academic Radiology*, *14*(1), 38–48. https://doi.org/10.1016/j.acra.2006.09.048

Hayes, M.H.S., & Patterson, D.G. (1921). Experimental development of graphic rating method. *Psychological Bulletin*, *18*(2), 98–99.

Hilbert, S., Küchenhoff, H., Sarubin, N., Nakagawa, T.T., & Bühner, M. (2016). The influence of the response format in a personality questionnaire: An analysis of a dichotomous, a Likert-type, and a visual analogue scale. *Testing, Psychometry, Methodology in Applied Psychology*, *23*(1), 3–24. https://doi.org/10.4473/TPM23.1.1

Leung, S-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research. 37*(4), 412–421. https://doi.org/10.1080/01488376.2011.580697

Likert, R. (1932). A technique for the measurement of attitudes. In R.S. Woodworth (Ed.), *Archives of Psychology* (Vol. *22*, No. 140, pp. 5–55). New York, NY: The Science Press.

Lin, H-C., Manuel, J., McFatter, R., & Cech, C. (2016). Changes in empathy-related cry responding as a function of time: A time course study of adult's responses to infant crying. *Infant Behavior and Development*, *42*, 45–49. https://doi.org/10.1016/j.infbeh.2015.10.010

Maloshonok, N., & Terentev, E. (2016). The impact of visual design and response formats on data quality in a web survey of MOOC students. *Computers in Human Behavior*, *62*, 506–515. https://doi.org/10.1016/j.chb.2016.04.025

Munson, B., Schellinger, S.K., & Edwards, J. (2017). Bias in the perception of phonetic detail in children's speech: A comparison of categorical and continuous rating scales. *Clinical Linguistics & Phonetics*, *31*(1), 56–79. https://doi.org/10.1080/02699206.2016.1233292

Patterson, D.G. (1922). The Scott Company graphic rating scale. *Journal of Personnel Research*, *1*(8–9), 361–376.

Peters, G.-J.Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, *16*(2), 56–69.

Reips, U-D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS generator. *Behavior Research Methods*, *40*(3), 699–704. https://doi.org/10.3758/BRM.40.3.699

Roster, C.A., Lucianetti, L., & Albaum, G. (2015). Exploring slider vs. categorical response formats in web-based surveys. *Journal of Research Practice*, *11*(1), Article D1. Retrieved from http://jrp.icaap.org/index.php/jrp/article/view/509/413

Studer, R. (2012). Does it matter how happiness is measured? Evidence from a randomized controlled experiment. *Journal of Economic & Social Measurement*, *37*(4), 317–336. https://doi.org/10.3233/JEM-120364

Svensson, E. (2000). Comparison of the quality of assessments using continuous and discrete ordinal rating scales. *Biometrical*

*Journal*, *42*(4), 417–434. https://doi.org/10.1002/1521-4036(200008)42:4

Thompson, J.D., Thomas, N.B., Manning, D.J., & Hogg, P. (2016). The impact of greyscale inversion for nodule detection in an anthropomorphic chest phantom: A free-response observer study. *British Journal of Radiology*, *89*(1064), 1–7. https://doi.org/10.1259/bjr.20160249

Voutilainen, A., Pitkäaho, T., Kvist, T., & Vehviläinen-Julkunen, K. (2015). How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *Journal of Advanced Nursing*, *72*(4), 946–957. https://doi.org/10.1111/jan.12875

Wall, E.J., Milewski, M.D., Carey, J.L., Shea K.G., Ganley, T.J., Polousky, J.D., & Zbojniewicz, A. (2017). The reliability of assessing radiographic healing of osteochondritis dissecans of the knee. *The American Journal of Sports Medicine*, *45*(6), 1370–1375. https://doi.org/10.1177/0363546517698933

Wewers, M.E., & Lowe, N.K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, *13*(4), 227–236. https://doi.org/10.1002/nur.4770130405

SEUNG YOUN (YONNIE) CHYUNG, EdD, is a professor in the Department of Organizational Performance and Workplace Learning in the College of Engineering at Boise State University (http://opwl.boisestate.edu/faculty-staff/faculty/yonnie-chyung/). She teaches graduate courses on Program Evaluation and Quantitative Research in Organizations. She runs a Workplace-Oriented Research Central (WORC) lab where teams of practitioners and researchers conduct research on various topics in the HPI context. She may be reached at ychyung@boisestate.edu

IEVA SWANSON, MS, is an instructional designer for Scottrade in St. Louis, MO. Since 2007, her professional and volunteer experience has included multimedia course design and development, instruction, facilitation, and learning management system administration. She volunteers as technology chair for ISPI Bay Area/Boise State Chapter (BABS). She earned her MS in Organizational Performance and Workplace Learning from Boise State University in 2016, and participates in Dr. Seung Youn (Yonnie) Chyung's Workplace-Oriented Research Central (WORC) lab. She may be reached at ievaswanson@u.boisestate.edu

KATHERINE ROBERTS, MS, is a senior instructional designer for WellMed Medical Management, Inc. in San Antonio, TX. She graduated with her MS in Organizational Performance and Workplace Learning from Boise State University in 2017. Her interests include interactive e-learning design, animation, ethics, performance improvement, and continued learning through Dr. Seung Youn (Yonnie) Chyung's Workplace-Oriented Research Central (WORC) lab. She may be reached at katherin.roberts@gmail.com

ANDREA HANKINSON, MA, MS, is the performance excellence consultant at the Public Service Commission, a department of the Government of Alberta, in Edmonton, Alberta, Canada. She holds an MS in Organizational Performance and Workplace Learning from Boise State University and an MA in English from Simon Fraser University. She has worked extensively in online and distance learning and is a freelance writer and instructor of writing. She may be reached at andrea.hankinson@gmail.com