Article

International Journal of Market Research

International Journal of Market Research
2019, Vol. 61(1) 394–407

DOI: 10.1177/1470785319831227
journals.sagepub.com/home/mre
SAGE

A comparison of scale attributes between interval-valued and semantic differential scales

Christos Themistocleous, Anastasios Pagiaslis, Andrew Smith and Christian Wagner
University of Nottingham, UK

Abstract

This article presents the results of an exploratory study comparing interval-valued scales (IVSs) and semantic differential scales (SDSs). The article investigates consumer perceptions regarding specific scale attributes and utilizes a controlled, between-subjects, experimental pen-and-paper design to assess the preferences of respondents when using the IVSs and SDSs. The rationale of this comparison lies with the fact that the newly introduced IVS has a built-in mechanism that allows the direct capture of respondent uncertainty toward the asked question, a feature that is absent from the SDS and other widely used, single-point capturing scales in marketing research such as the Likert and Stapel. Results show that overall consumer preferences of the IVS and SDS are equal, although "speed of use" results favor the IVS. The consistency of respondent evaluations regarding the two scales may indicate their interchangeability in marketing research and opens up pathways for future exploration of IVSs for the accumulation of more reliable and robust results. The main contribution of the article is the introduction of a novel IVS, within the context of marketing, for collecting respondent answers while also directly capturing respondent uncertainty. Furthermore, this article adds to the discussion of consumer perceptions and preferences regarding different scales, scale development, and optimal rating scales that may lessen ambiguity for survey respondents and researchers.

Keywords

experiments, interval-valued scale, scale attributes, semantic differential scale, survey

Introduction

Scale development, scale appropriateness, and scale utilization are crucial topics in the marketing research literature (Bendixen & Yurova, 2012; Cramphorn, 2012; Hanson & Rethans, 1980; Hartley & Betts, 2010; Hawkins, Albaum, & Best, 1974; Preston & Colman, 2000; Revilla, 2015;

Corresponding author:
Christos Themistocleous, Nottingham University Business School, University of Nottingham, Jubilee Campus, Nottingham NG8 1BB, UK.
Email: christos.themistocleous@nottingham.ac.uk

Rossiter, 2002, 2011). Marketing research commonly utilizes single-point capturing scales such as Likert (1934) scale and semantic differential scale (SDS; Osgood, 1952). Single-point capturing scales base their measurement properties on a varied number of response categories (points) linked to a question or statement, where the respondent chooses a single-point, with additional examples being the Stapel and visual analog scales.

Single-point capturing scales offer valuable information regarding respondents' perceptions on a specific topic. However, single-point capturing scales are limited in capturing uncertainty of respondent answers. Efforts to expand single-point capturing scales, for example, with phrase completion Likert scales (Hodge & Gillespie, 2003) or two-staged itemized rating scales (Albaum, 1997) have failed to address this built-in limitation of single-point capturing scale (Li, 2013).

Uncertainty in measurement traces to Zadeh (1965) who specifies that many objects, or constructs, encountered in the real world cannot be precisely defined as members or non-members in classes or sets and that our language incorporates imprecise, uncertain, and fuzzy information. Zadeh discusses memberships in the class of animals, which clearly includes dogs with a membership of 1 (belongs to class) and excludes objects such as rocks with a membership of 0 (does not belong to class). But objects such as a starfish or bacteria have an ambiguous and uncertain status in the class of animals, which is not exactly 1, and not exactly 0 either. Zadeh (1965; see also Chang & Yeh, 2002; Li, 2013; Smithson, 1987) sets the stage for the birth of the term *fuzzy sets* as a method that can capture information in which uncertainty is present with varying degrees and from a variety of sources.

Occasionally, the choice of a midpoint within a single-point capturing scale (if there is a midpoint) is erroneously associated with the respondent uncertainty. Nevertheless, when tracing back the research of Likert and Osgood, a midpoint response (i.e., fourth response category in a 7-point Likert-type scale or SDS) captures neutrality toward the level of agreement with a particular statement and not respondent uncertainty.

The capturing of respondents' uncertainty requires the development of more suitable scales such as the interval-valued scale (IVS) which provides the respondent with a choice of an interval when providing a response by positioning an ellipse on a straight line with polar adjectives on its two ends. Wagner, Miller, Garibaldi, Anderson, and Havens (2015) first introduced the IVS in surveys, providing respondents with the ability of recording their answer on a specific question. Their research notes that interval-valued survey responses offer richer and more complex information compared with single-point capturing scales through the quantification of respondent uncertainty, and the data collected can be comprehensively modeled through algorithms such as the interval agreement approach (IAA) to provide researchers with more insights regarding respondent perceptions as well as the (un)certainty of their responses which ultimately enhances test–retest validity and reliability.

The focus point of the IVS is linked to the fact that uncertainty and ambiguity is the norm in respondent assessments in a variety of domains and contexts, but this information is not encoded in a typical single-point capturing scale, such as the SDS and Likert scales, with the respondent likely to score a neutral rather than an ambiguous response if a midpoint category on the scale is provided. The IVSs with the inclusion of an extra, built-in dimension in their measurement provide researchers with direct insights on how uncertain a respondent might be in regard to a specific question. Nevertheless, capturing interval-valued data necessitates the widening of the spectrum of employable analysis techniques to utilize the additional information obtained (see further Ashtiani, Haghighirad, Makui, & Montazer, 2009; Wagner, Miller, & Garibaldi, 2014).

From these, the aim of this article is to explore the IVS within a marketing research context and provide exploratory comparative results between the IVS and the widely used SDS in terms of respondent preferences of usability. If this new form of data capture proposed by the IVS is not amenable to respondents or if it is fatiguing, then the actual value of any analytical advantages may

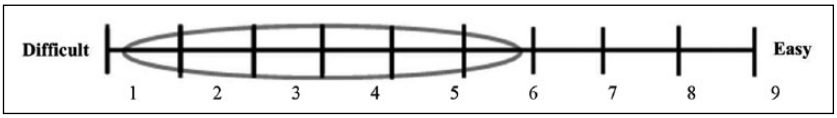

Figure 1. Graphical representation of the interval-valued scale.

be limited. This serves as the driving force of this research. The first part of the article introduces the concepts of the IVS and SDS and reviews the current analytical methods available in the literature while also reflecting on theory regarding specific scale attributes that are deemed to be essential for the measurement of satisfaction of respondents toward a measurement scale. The second part presents the between-subjects, experimental study that explores the comparability of scale attributes between IVSs and SDSs. The third and final part provides a discussion of the study results while also noting the article's contributions, limitations, and future research avenues.

IVSs

The original orthodoxy of scale development and design in marketing (Churchill, 1979) has come under substantive and influential criticism (e.g., Rossiter, 2002, 2011). The question of optimal length and structure of single-point capturing scales is an enduring question (see, for example, Cramphorn, 2012; Matell & Jacoby, 1971; Peterson, 1997). Rossiter's (2002) alternative paradigm of scale design and development "C-OAR-SE" focuses on improving the validity of marketing scales and the independence of scale performance from data-driven statistical techniques by privileging robust formulation of scale content. However, Rossiter (2002) notes that single-point capturing scales' answer categories need improvement and that alternative scales to the Likert and SDS need to be employed in marketing research.

The IVS has the potential to address shortcomings such as those highlighted by Rossiter regarding the increase of reliability, efficiency, and applicability of scales across different marketing contexts while also capturing respondent certainty with provided responses that can enhance test–retest reliability.

The present study utilizes IVS with a respondent-drawn ellipse on a line, and Figure 1 provides a visual representation of the scale. The line of the IVS represents the interval, and the ellipse represents the answer of the respondent on a question asking for the assessment of the level of difficulty.

The wider the ellipse, the more uncertain a respondent is regarding his or her answer to the relevant question (this method is readily explained to respondents). The closer the ellipse is toward the left-hand side, the more the respondent agrees with the assigned adjective of "difficult." The closer the ellipse is to the right-hand side, the more the respondent agrees with the "easy" adjective as an answer to the asked question. We note here that polar adjectives can change depending on what is being measured (i.e., good/bad, slow/fast, satisfied/dissatisfied). The width of the ellipse on the line represents the level of uncertainty. For example, when providing a respondent with a 10-point anchored line, they might not be certain as to whether their respective answer should be a 3 or a 4. In certain cases, respondents might even be more uncertain and could be assessing more than just two response categories. Furthermore, precise answers might lie in between response categories such as 4.5 and 3.2. Taking these into consideration, intervals of choice through the IVS allow the answer to lie in between response categories by including an interval that covers multiple (and not just a single) response categories. This way the IVS directly captures the uncertainty of the response through the width of the ellipse.

In the example found in Figure 1, the respondent has a clear tendency to rate the assessed topic as difficult, yet the width of the ellipse demonstrates a strong level of uncertainty as multiple response categories (2, 3, 4, 5, 6) are included in the interval that is captured through his or her drawing of the ellipse. More precisely, the respective ellipse of Figure 1 represents an interval of choice that starts and ends in-between response categories from 1.3 to 6.8.

As specified, occasionally, researchers support that single-point scales such as the SDS and Likert may capture uncertain responses by offering a middle-point response category; however, this approach does not clearly differentiate between actual uncertainty and a truly neutral response. Another limitation of these single-point capturing scales is that choice and level of expression for respondents is limited to a crisp number of response categories and choice in-between these is not provided. Initial studies by Miller, Wagner, Garibaldi, and Appleby (2012) noted that the IVS allows the precise identification of the degree of respondent uncertainty according to the width of the provided ellipse while providing a more granular understanding of respondent answers. It is also noted that the employment of IVSs may have the potential to reduce the required survey sample size when compared with the employment of single-point scales, without diminishing information returns or diminishing statistical validity and reliability (Miller et al., 2012).

Previous attempts to estimate or "measure" respondent uncertainty (i.e., attempts to quantify respondent uncertainty) are not a new concept (Ashtiani et al., 2009; Soll & Klayman, 2004; Teigen & Jorgensen, 2005; Wang, Wu, Wang, Zhang, & Chen, 2014); however, certain previous studies attempting to estimate the magnitude of uncertainty of a particular quantity using single-point capturing scales usually employ fixed levels of (un)certainty. For example, Soll and Klayman (2004, p. 304) employed statements such as "I am 80% sure that this happened between ____ and ___" to assess respondent uncertainty. This strand of research in psychology has highlighted the dangers of incorporating overconfidence in respondents' imprecise answers. Speirs-Bridge et al. (2010), for example, note that simple changes in question format can substantially reduce overconfidence in elicited responses. To this end, the IVS allows respondents to answer a question by providing an interval-valued response that has a built-in function to capture uncertainty within the initial answer and without having to include follow-up questions that ask the respondent of how sure he or she is with their answer, an attribute that reduces significantly response overconfidence (Speirs-Bridge et al., 2010).

As IVSs do not pre-assign any fixed level of (un)certainty and establish a direct method of estimating uncertainty through the utilization of an interval, they allow more freedom to respondents when expressing their answers compared with single-point capturing scales such as Likert and SDS (Wagner et al., 2014). Their argument finds support on the fact that single-point scales have predetermined response categories that limit respondent responses to a set range of discrete numerical values, usually seven categories (e.g., a 7-point Likert-type scale) or in certain cases more (e.g., as in the present study, a 10-point SDS). In IVS, the appropriate quantification and analysis of the interval width allows for clear identification of the uncertainty that respondents' answers may contain (Miller et al., 2012).

Analytical methods for IVSs

Obtaining location and central tendency statistics to allow analysis and understanding of IVS responses requires particular types of modeling, which is linked to the previous statement that the employment of IVSs and the direct capture of respondent uncertainty require the widening of data analysis techniques. To this end, initial analyses of data obtained by continuous scales include Liu and Mendel's (2008) interval approach (IA)—a modeling technique for raw interval-valued data designed to build type-2 fuzzy sets for analysis purposes, which requires a pre-processing stage.

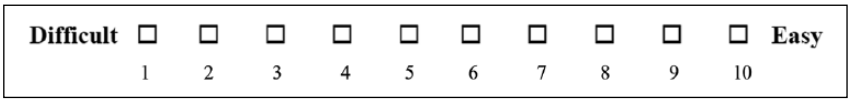

Figure 2. Graphical representation of the 10-point semantic differential scale (IVS).
IVS: interval-valued scale.

Focusing directly on data obtained by the IVS, Wagner et al. (2015) propose the IAA, an alternative method for modeling IVS data that does not require a pre-processing stage or outlier removal, thus allowing the inclusion of all responses, minimizing loss of information, and avoiding as much as possible assumptions about the distribution of the data. What they indicate from this is that the IAA allows researchers to model both intra- and inter-source (respondent) uncertainty allowing for (a) *a detailed assessment of test–retest consistency between responses*, (b) *better questionnaire design as uncertainty may derive from the question itself or the concept under examination* (something that can be captured and addressed appropriately from a simple pilot study), and (c) *attribution of specific degree of uncertainty to particular respondent characteristics* (e.g., people above the age of 70 years might be more uncertain with questions about the newest Apple products providing the researcher with an additional parameter for analysis and enhancement of findings).

The employability of IVS and the utilization of appropriate modeling and analysis techniques is a relatively new area compared with the well-established modeling and analysis techniques of discrete single-point scales in marketing research. Therefore, transition to full use of IVSs requires direct comparisons between the two different types of scales as well as further development of the available approaches for statistical analysis. This article focuses on the former.

SDSs and scale attributes

SDSs have found multiple applications in the marketing literature (e.g., see Barnett, 2004; Hartley & Betts, 2010; Rusell, 2010; Tsarenko & Tojib, 2009). Charles Osgood (1952) developed SDSs to provide quantifiable measurements for the meaning of language. SDSs are set up by using two descriptive polar adjectives (easy/difficult, fast/slow, invasive/non-invasive) at each end of the scale with two or more points in between depending on the concept measured. Figure 2 provides a visual depiction of a 10-point SDS again measuring the level of difficulty, with participants being asked to choose a single point to represent their response.

The optimum number of response categories that maximizes information acquisition in SDSs as well as Likert was, and still is, a subject of debate in the marketing and psychology literature. Initial comparisons by Matell and Jacoby (1972) with differing scale response categories (between 2 and 19) show that a small number of response categories as low as two are sufficient in research practice. This suggestion is in alignment with Green and Rao's (1970) argument who further note that six or seven response categories optimize obtainable information while response category increase provides marginal information gain.

However, Loken, Pirie, Virnig, Hinkle, and Salmon (1987) and Hancock and Klockars (1991) criticize the aforementioned assertion, suggesting that scales with more than seven response categories perform comparatively better in terms of validity than scales with less categories. More recently, Preston and Colman (2000) note that when examining how respondents perceive scales with differing response categories (from 2 to 101), the 10-point scale has the highest overall ranking in terms of respondents' preferences regarding ease of use, speed of use, and adequate expression of the respondents' perceptions.

Accordingly, the study in the present article adopts Preston and Colman's (2000) suggestion and employs 10 response categories for the SDS while also anchoring the line of the IVS with 10 reference points for respondents to position their response by drawing an ellipse (see Figure 1 above). The latter design decision is particularly motivated by the aim to make both SDS and IVS as similar as possible in structure, thus facilitating direct comparison.

Preston and Colman's (2000) work also focused on the examination of respondents' preferences in regard to each single-point capturing scale included in their survey. Building upon the work of Jones (1968) which served as the first example for the measurement of respondent preferences toward single-point scales between 2 and 7 response categories, Preston and Colman (2000) focused on six scale attributes for the measurement of respondent preferences. They were (a) *ease of use*, (b) *speed of use*, (c) *ability to precisely record desired answers*, (d) *adequate expression of exact thoughts and feelings*, (e) *certainty/uncertainty with personal answers*, and (f) *overall satisfaction* with each scale. This scale adopts Preston and Colman's scale attributes for the comparison of respondent preferences between the IVS and SDS. The importance of this examination lies with the fact that if, for example, the newly introduced IVS does not allow adequate expression of thoughts and feelings, requires significantly more time, and is found to be more difficult to complete compared with the widely used SDS in marketing research, it may lead to demotivation, frustration, and by extension reduction to the quality of responses by the users of the scale. Therefore, exploration of respondent preferences toward the IVS serves as a fundamental step for the application of the IVS within marketing research.

Data and methods

The study is based on a controlled, between-subjects, laboratory experiment based on a quota sample of 122 UK adults encompassing a variety of ages, ethnic origins, and scale familiarity making it a suitable sample for comparative variance analyses (Harrell & Frank, 2001). The pen-and-paper format within the controlled laboratory environment ensured the maximization of internal validity and isolated the respondents from any external distractions during completion. The pen-and-paper format is in alignment with Gil and Rodriguez's (2012) as well Yusoff and Janor's (2014) examinations of interval-capturing scales, with the latter supporting that an online approach may produce differing results dependent on screen resolution and characteristics (e.g., color, refresh rate) that may influence the perceived length of intervals making results from different screen resolutions incomparable. The questionnaire was printed and handed out to the different respondents and, after a briefing process, was self-completed (sample of the questionnaire can be found in Supplemental Appendix). Sixty-one males and 61 females were recruited with a mean age of 32.4 years (SD = 13.132) and ranging from 19 to 66 years; Table 1 below summarizes the demographics of participants based on age, gender, educational background, and ethnicity.

A questionnaire with 18 questions was constructed to measure the sensitivity of respondents to the capture of private/sensitive information. The theme of this research is linked back to customer data accumulation procedures for profiling purposes, ultimately resulting to more effective and efficient targeting of customers for marketing campaigns. The questions were adopted from the research of Acquisti, John, and Loewenstein (2012) and asked the participants to rate a particular question as to the degree of invasiveness they felt the question involved (e.g., Question 1 reads, How invasive/intrusive do you feel this question is: How many times do you go to a restaurant in an average week?).

Two versions of the questionnaire were designed for this set of 18 questions—one utilizing the IVS and the other the SDS—leading to the two conditions of the between-subjects experiment (Table 2 summarizes the 18 questions). The rest of the questionnaire was common in both versions and designed to measure familiarity with surveys and survey scales, respondent perceptions of scale attributes, and demographics. Due to the fact that the IVS is relatively uncommon compared with single-point scales, the questionnaire included detailed instructions on how to complete the IVS.

Table 1: Summary of participant demographics.

	IVS			SDS			Total ($N = 122$)		
Participants	Male	Female		Male	Female		Male		Female
	30	31		31	30		61		61
Age	*M*	*SD*		*M*	*SD*		*M*		*SD*
	30.85	10.813		33.98	15.028		32.42		13.132
Education	Secondary	Grad.	Postgrad.	Secondary	Grad.	Postgrad.	Secondary	Grad.	Postgrad.
	4	25	32	15	25	21	19	50	53
Ethnic origin	White	Southeast Asia		White	Southeast Asia		White		Southeast Asia
	42	19		48	13		90		32

IVS: interval-valued scale; SDS: semantic differential scale.

More precisely, the following statement was included along with examples for visual aid (found in the Supplemental Appendix):

> This questionnaire follows an interval-based measurement meaning that you mark your answer with an ellipse on the given interval. Here is a brief explanation on how this Interval Valued Scale works. The closer the ellipse on the right-hand side of the scale the more sensitive the asked question is . . .

As respondents' perception of intrusiveness was ostensibly the subject of the study, in an attempt to limit order-of-presentation effects of questions which could prejudice the perceptions of respondents, the researchers created multiple versions of the IVS and SDS questionnaires with differential question sequences as proposed by Acquisti et al. (2012). Before assembling the questionnaire, a small pilot study with 10 PhD students and five academic members of staff was conducted to evaluate the assumed invasiveness of each question. After this, the permutated questionnaire versions were created using a block randomization method whereby questions were presented in a block of three questions of increasing invasiveness as per both the literature (Acquisti et al., 2012) and the pilot study. The blocks were then randomized across different questionnaires, creating four different versions of IVS and four different versions of SDS randomized questionnaires. We note that the real objective of the survey in respect to this article was the isolated assessment of the scales themselves and the comparison of the data acquired through the deployment of the two techniques, rather than the actual information on question sensitivity.

To further isolate effects, familiarity with different measurement scales between the two groups was measured via six binary/dichotomous (yes/no) questions. For example, the first familiarity question showed the respondents a 2-point Likert-type scale and asked respondents whether they were familiar with the scale. Questions were coded 0/1 and then summed thereby creating an index of familiarity that ranged from 0 to 6. This was to ensure consistency of participants' familiarity with different measurement scales between the two groups.

Respondent preferences of scale attributes were measured using the same question structure as Preston and Colman (2000). Six statements measured respondent preferences regarding each scale's (a) ease of use, (b) speed of use, (c) ability to precisely record desired answers, (d) adequate expression of exact thoughts and feelings, (e) certainty/uncertainty with personal answers, and (f) overall satisfaction.

Table 2: Summary of the 18 privacy-related questions.

Questions
1. How many sexual partners have you had since you became sexually active?
2. What is the amount of your household savings?
3. Have you ever been arrested?
4. Have you ever looked at pornographic material?
5. Have you ever tried illegal drugs like marijuana?
6. Have you lied about your income to an official service?
7. Have you ever suffered from cancer of any kind?
8. Have you downloaded illegally obtained pirated songs and/or movies from the Internet?
9. Have you witnessed a serious crime and failed to report it or stop it?
10. Have you called in sick when you were not sick either in your workplace, university, etc.?
11. How much alcohol do you consume on average per week?
12. Have you lied about your age to someone you were attracted to?
13. Have you claimed to have education that you didn't actually have (either on your CV or in person)
14. For which stores do you currently have loyalty cards?
15. Have you knowingly wasted energy, for example by not switching off the lights for convenience at your workplace, school or university?
16. Which mobile carrier do you currently use?
17. How often do you travel abroad either for holidays or business?
18. How many times do you go to a restaurant in an average week?

The invasiveness of each question was measured with SDS versus IVS scales. SDS: semantic differential scale; IVS: interval-valued scale.

Responses were recorded using a 10-point scale anchored, respectively, with point 1 (on the left) always representing a negative evaluation and point 10 (on the right) a positive evaluation. For example, Statement 1 read as follows: "How easy did you find the use of the interval-valued scale when assessing the intrusiveness of the questions above? 1 = *difficult*, 10 = *easy*." The questionnaire was provided in person by the principal investigators to each respondent in a controlled environment and required on average 30 min to complete. Amazon vouchers were provided to all 122 participants as incentives.

Results

Initial tests of demographic consistency ensured comparability of results between the IVS and SDS samples (see Table 3 below). A *t*-test analysis checked for cross-sample consistency of age, and two cross-tabulation analyses for gender and ethnicity. In regard to age, no statistically significant differences were identified between the IVS ($M = 30.852$, $SD = 10.813$) and SDS ($M = 33.984$, $SD = 15.028$) samples with $t(120) = -1{,}321$, $p = .183$, Cohen's $d = -0.239$ (95% confidence interval [CI] = [–0.595, 0.117]). For gender, cross-tabulation showed no statistically significant differences with $\chi^2(1) = 0.033$, $p = .856$, Cohen's $w = 0.032$ (95% CI = [–0.176, 0.208]). Similarly, ethnicity was also found non-significant with $\chi^2(1) = 1.525$, $p = .217$, Cohen's $w = 0.212$ (95% CI = [–0.166, 0.383]).

To ensure that familiarity did not influence the results of the scale attribute comparison, an additional *t* test checked cross-sample consistency for familiarity. Familiarity is non-significant between IVS ($M = 4.721$, bias-corrected and accelerated [BCa] CI = [4.534, 4.896]); $SD = 0.686$, BCa CI = [0.560, 0.800]) and SDS ($M = 4.410$, BCa CI = [4.100, 4.697]; $SD = 1.131$, BCa CI =

Table 3. Between-groups demographic consistency check.

Age consistency between groups								
Demographic variable	Group	*n*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Age	Group A (IVS)	61	30.852	10.813	−1.321	120	.183	0.239
	Group B (SDS)	61	33.984	15.028				
Gender and ethnicity consistency between groups								
Demographic variable	Group	*n*		χ^2	*df*	*p*		Cohen's *w*
Gender	Group A (IVS)	61		0.033	1	.856		0.032
	Group B (SDS)	61						
Ethnicity	Group A (IVS)	61		1.525	1	.217		0.212
	Group B (SDS)	61						

IVS: interval-valued scale; SDS: semantic differential scale.

[0.831, 1.363]) with the two samples exhibiting unequal variances (Levene's $F = 7.751$, $p = .0065$) with $t(98.911) = 1.839$, $p = .069$ (bootstrapped $p = .086$), Cohen's $d = -0.333$ (95% CI = [–0.025, –0.690]).

Comparison of IVS versus SDS attributes

To test the comparability of IVS and SDS scale attributes, the researchers conducted multiple *t* tests reported in Table 4 below. The table includes bootstrapped CIs for all comparisons to strengthen the reporting of the results (DiCiccio & Efron, 1996). As evident from Table 3, most scale attributes were not found to be statistically significantly different between IVS and SDS samples, with the exception of speed of use. Findings from the analysis indicate that speed of use interestingly results in favor of IVS with $t(109.61) = 2.070$, $p = .04$ (bootstrapped $p = .046$), Cohen's $d = 0.375$ (95% CI = [–0.017, 0.733]). Explanations in regard to this finding are primarily linked to the work of Kashdan, Rose, and Fincham (2004) and Preston and Colman (2000) and are presented in the "Discussion" section.

Discussion

The present article examines respondent preferences toward IVSs and SDSs on a set of scale attributes. Respondent perceptions indicate strong and perhaps surprising satisfaction with the novel IVS compared with the more familiar SDS. Five out of the six examined scale attributes—namely, (a) *ease of use*, (b) *ability to precisely record desired answers*, (c) *adequate expression of exact thoughts and feelings*, (d) *certainty/uncertainty with personal answers*, and (e) *overall satisfaction*—were found to have no significant differences between the IVS and SDS. The consistency of respondent perceptions regarding the two scales may indicate their application in marketing research, a finding that is primarily useful for the IVS which has not been previously examined within a marketing context neither examined for respondent preferences nor compared with more

Table 4. *T*-test results comparing IVS versus SDS.

T tests of scale attribute mean differences between IVS and SDS samples, overall $N = 122$

Scale attribute	Sample	*M*	*M* BCa 95% CI[a]	*SD*	*SD* BCa 95% CI[a]	*M* difference	*M* difference BCa 95% CI[a]	*t*	*df*	*p*	Cohen's *d*[b]	95% CI
Ease of use[c]	IVS	7.377	[6.781, 8.029]	2.222	[1.873, 2.484]	0.672	[–0.288, 1.587]	1.457	113.504	.148	0.294	[–0.062, 0.651]
	SDS	6.705	[6.042, 7.368]	2.342	[2.568, 3.081]							
Speed of use[d]	IVS	7.492	[6.901, 8.043]	2.038	[1.665, 2.336]	0.918	[–0.035, 1.771]	2.070	109.614	.041*	0.375	[0.017, 0.733]
	SDS	6.574	[5.831, 7.259]	2.802	[2.468, 3.069]							
Precision allowed by each scale	IVS	6.967	[6.240, 7.633]	2.575	[2.100, 2.926]	−0.082	[–1.130, 0.864]	−0.172	120	.864	−0.031	[–0.386, 0.324]
	SDS	7.049	[6.375, 7.704]	2.692	[2.252, 3.061]							
Expression of exact feelings allowed by each scale	IVS	6.541	[5.810, 7.257]	2.592	[2.297, 2.825]	0.028	[–0.788, 1.266]	0.575	120	.566	0.104	[–0.251, 0.459]
	SDS	6.262	[5.557, 6.898]	2.756	[2.450, 3.015]							
Certainty/uncertainty with personal answers through each scale	IVS	6.443	[5.768, 7.108]	2.605	[2.271, 2.868]	0.541	[–0.368, 1.457]	1.116	120	.267	0.202	[–0.154, 0.558]
	SDS	5.902	[5.167, 6.649]	2.750	[2.428, 3.017]							
Overall satisfaction for each scale	IVS	7.361	[6.743, 8.028]	2.345	[1.872, 2.688]	0.426	[–0.499, 1.343]	0.968	120	.335	0.176	[–0.180, 0.531]
	SDS	6.934	[6.327, 7.511]	2.516	[2.154, 2.864]							

IVS: interval-valued scale; SDS: semantic differential scale; BCa: bias-corrected and accelerated; CI: confidence interval.

[a]Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples.

[b]Cohen's *d* calculated using Hedge's *g* modification as Cohen's $d = (M_2 - M_1) / SD_{pooled}$, where $SD_{pooled} = \sqrt{(SD_1^2 + SD_2^2) / 2}$. Cohen's *d* calculated using G*Power 3.1.9.2.

[c]Levene's test of equality of variances is significant; table reports *t* test statistics without assuming equal variances (Levene's $F = 9.847$, $p = .002$).

[d]Levene's test of equality of variances is significant; table reports *t* test statistics without assuming equal variances (Levene's $F = 17.070$, $p = .000$).

*Bootstrapped $p = .046$.

widely used scales such as the SDS. Furthermore, both scales appear to be acceptable to respondents and the instructions for use are readily assimilated.

Based on the weaker performance of the SDS in regard to the speed of use attribute, it may allude to the fact that individuals prefer SDSs with fewer response categories. As noted, Preston and Colman (2000, p. 10) showed that respondents rate single-point scales with the fewest response categories (2-point, 3-point, and 4-point) as the *quickest to use*. Nevertheless, this article employs a 10-point SDS because, according to Preston and Colman (2000), this number of categories maximizes *ease of use* and *adequate expression of participant perception*.

The present results corroborate that a 10-point SDS may indeed impact satisfaction in respect to speed of use. More importantly, this result may also be explained due to the fact that the IVS does not as such force participants to make a choice between discrete options but instead allows them to more freely and directly express their opinion. For example, two options may fit the participant's perception equally well, but SDS forces them to choose one. By allowing the respondent to include more than one response category in their answer through the width of the ellipse on the scale, the respondent's decision of which category(ies) to choose can be a quicker and easier one. Finally, this finding may also be linked to another stream of literature indicating that curiosity in respect to using a new scale (IVS in this case) can lead to increased cognitive stimulation and positive subjective perceptions, thus leading to the high rating of IVS (Kashdan et al., 2004). However, as such an effect is not apparent for any other properties, the latter is arguably more unlikely.

The main contribution of the present article is the exploration of IVS for capturing respondent data in a marketing context. The IVS is an alternative to traditional single-point scales. The direct comparison of scale attributes between two fundamentally different scales established a basis for exploring their complementarity and interchangeable nature in marketing research, respectively. The article compares respondent perceptions regarding predefined scale attributes between IVS and SDS and finds that IVS is not considered more difficult to use, instead actually considered quicker to complete. The fact that the IVS has the advantage of explicitly modeling response uncertainty without asking the participant to make an explicit judgment on their uncertainty in contrast to other approaches such as that by Soll and Klayman (2004) makes it a useful tool for inclusion in future questionnaire-based research. This potential provides researchers with new pathways to measuring responses, leading to a wider range of analytical capabilities and improved, data-driven insight, compared with the ones traditionally offered. It is expected that through the modeling of the uncertainty captured by IVS, researchers can extend the depth of their findings and draw more meaningful conclusions regarding behaviors, preferences, and perceptions around a particular topic.

As questions regarding optimized length and structure of rating scales are still prominent, the present article adds to the discussion (Hanson & Rethans, 1980; Matell & Jacoby, 1971; Preston & Colman, 2000; Rocereto, Puzakova, Anderson, & Hvolkiin, 2011; Rossiter, 2011) of scale development and optimal rating scales toward lessening ambiguity for survey respondents and users of research (Rossiter, 2002).

Overall, the results of the article suggest the IVS format holds promise, as an alternative scale, because respondents report the scale as quicker and because IVS has structural advantages compared with single-point scales. IVS offers a straightforward answer format that can be conveniently applied in pen-and-paper surveys without causing respondent fatigue or delay in completing a questionnaire. Such a feature of a rating scale is of particular importance to both marketing academics and practitioners.

Finally, the article acknowledges certain limitations, which give rise to future research suggestions. The present study offers an initial stepping-stone to increase the use of IVS in

academic and practitioner marketing research, taking advantage of the uncertainty-capturing abilities of the scale. Future research avenues can also include the comparison of respondent perceptions between the IVS and other single-point capturing scales (e.g., Stapel, visual analog). Future research should also expand data collection to an online environment and address potential effects of screen resolution and screen characteristics (e.g., colors, refresh rate; Yusoff & Janor, 2014). This can further reinforce the literature on scale preferences of respondents regarding single-points capturing and IVSs, setting a step further the applicability of IVSs in marketing research. At the time of writing, ongoing research is focusing on a detailed investigation of the cognitive and statistical properties (location, scale estimators, and psychometric properties) of IVSs and direct comparison with classical single-point scales examining the applicability and usefulness of the IVS.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

ORCID iD


Christos Themistocleous https://orcid.org/0000-0001-9074-7711

References

Acquisti, A., John, L. K., & Loewenstein, G. (2012). The impact of relative standards on the propensity to disclose. *Journal of Marketing Research*, *49*, 160–174.

Albaum, G. (1997). The Likert scale revisited: An alternate version. *Journal of the Market Research Society*, *39*, 331–332.

Ashtiani, B., Haghighirad, F., Makui, A., & Montazer, C. (2009). Extension of fuzzy TOPSIS method based on interval-valued fuzzy sets. *Applied Soft Computing*, *9*, 457–461.

Barnett, T. W. (2004). Consumer disclosure and disclosure avoidance: A motivational framework. *Journal of Consumer Psychology*, *14*, 41–51.

Bendixen, M., & Yurova, Y. (2012). How respondents use verbal and numeric rating scales: A case of rescaling. *International Journal of Marketing Research*, *54*, 261–282.

Chang, Y. H., & Yeh, C. H. (2002). A survey analysis of service quality for domestic airlines. *European Journal of Operational Research*, *139*, 166–177.

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*, 64–73.

Cramphorn, M. (2012). More scales than a fish? *International Journal of Marketing Research*, *54*, 739–749.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*, 189–212.

Gil, M. A., & Rodriguez, G. G. (2012). Fuzzy vs. Likert scale in statistics. *Combining Experimentation and Theory*, *271*, 407–420.

Green, P. E., & Rao, V. R. (1970). Rating scales and information recovery: How many scales and response categories to use? *The Journal of Marketing*, *34*, 33–39.

Hancock, G. R., & Klockars, A. J. (1991). The effect of scale manipulations on validity: Targeting frequency rating scales for anticipated performance levels. *Applied Ergonomics*, *22*, 147–154.

Hanson, J. W., & Rethans, A. J. (1980). Developing interval scale values using the normalized rank method: A multiple context, multiple group methodology. *Advances in Consumer Research*, *7*, 672–674.

Harrell, J., & Frank, E. (2001). *Regression modelling strategies: With application to linear models, logistic regression and survival analysis*. New York, NY: Springer.

Hartley, J., & Betts, L. R. (2010). Four layouts and a finding: The effects of changes in the order of the verbal Labels and numerical values on Likert-type scales. *International Journal of Social Research Methodology*, *13*, 17–27.

Hawkins, D. I., Albaum, G., & Best, R. (1974). Stapel scale or semantic differential in marketing research? *Journal of Marketing Research*, *11*, 318–322.

Hodge, D. R., & Gillespie, D. (2003). Phrase completions: An alternative to Likert scales. *Social Work Research*, *27*, 45–54.

Jones, R. R. (1968). Differences in response consistency and subjects' preferences for three personality inventory response formats. In G. A. Mendelsohn (Ed.), *Proceedings of the 76th Annual Convention of the American Psychological Association* (pp. 247–248). Washington, DC: American Psychological Association.

Kashdan, T. B., Rose, P., & Fincham, F. D. (2004). Curiosity and exploration: Facilitating positive subjective experiences and personal growth opportunities. *Journal of Personality Assessment*, *82*, 291–305.

Li, Q. (2013). A novel Likert scale based on fuzzy sets theory. *Expert Systems with Applications*, *40*, 1609–1618.

Likert, R. (1934). A simple and reliable method of scoring the Thurstone attitude scales. *Journal of Social Psychology*, *5*, 228–238.

Liu, F., & Mendel, J. (2008). Encoding words into interval type-2 fuzzy sets using an interval approach. *IEEE Transactions on Fuzzy Systems*, *16*, 1503–1521.

Loken, B., Pirie, P., Virnig, K. A., Hinkle, R. L., & Salmon, C. T. (1987). The use of 0-10 scales in telephone surveys. *Journal of the Market Research Society*, *29*, 353–362.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study 1: Reliability and validity. *Educational and Psychological Measurement*, *31*, 657–674.

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

Miller, S., Wagner, C., Garibaldi, J., & Appleby, S. (2012). Constructing general type-2 fuzzy sets from interval-valued data. In J. Liu, C. Alippi, B. Bouchon-Meunier, G. W. Greenwood & H. A. Abbass (Eds.), *IEEE World Congress on Computational Intelligence* (pp. 1–8). New York, NY: IEEE.

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, *49*, 197–237.

Peterson, R. A. (1997). A quantitative analysis of rating-scale response variability. *Marketing Letters*, *8*, 9–21.

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

Revilla, M. (2015). Effect of using different labels for the scale in a web survey. *International Journal of Market Research*, *57*, 225–238.

Rocereto, J. F., Puzakova, M., Anderson, R. E., & Hvolkiin, K. (2011). The role of response formats on extreme response style: A case of Likert type vs semantic differential scale. *Advances in International Marketing*, *10*, 53–71.

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *Research in Marketing*, *19*, 305–335.

Rossiter, J. R. (2011). Marketing measurement revolution: The C-OAR-SE method and why it must replace psychometrics. *European Journal of Marketing*, *45*, 1561–1588.

Rusell, G. J. (2010). *Itemized rating scales (Likert, semantic differential and Stapel)* (1st ed.). Hoboken, NJ: John Wiley.

Smithson, M. (1987). *Fuzzy set analysis for behavioural and social sciences*. New York, NY: Springer-Verlag.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology Learning Memory and Cognition*, *30*, 299–314.

Speirs-Bridge, A., Fidler, F., Mcbride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgements of experts. *Society of Risk Analysis*, *30*, 512–523.

Teigen, K. H., & Jorgensen, M. (2005). When 90% confidence intervals are only 50% certain: On credibility of credible intervals. *Applied Cognitive Psychology*, *19*, 455–475.

Tsarenko, Y., & Tojib, D. R. (2009). Examining customer privacy concerns in dealings with financial institutions. *Journal of Consumer Marketing*, *26*, 468–476.

Wagner, C., Miller, S., & Garibaldi, J. (2014). Similarity based applications for data-driven concept and word models based on type-1 and type-2 fuzzy sets. In O. Cordon (Ed.), *IEEE International Conference on Fuzzy Sets* (pp. 1–9). New York, NY: IEEE.

Wagner, C., Miller, S., Garibaldi, J. M., Anderson, D. T., & Havens, T. C. (2015). From interval-valued data to general type-2 fuzzy sets. *IEEE Transactions on Fuzzy Systems*, *23*, 248–269.

Wang, J., Wu, J., Wang, J., Zhang, H., & Chen, X. (2014). Interval-valued hesitant fuzzy linguistic sets and their applications in multi-criteria decision-making problems. *Information Sciences*, *288*, 55–72.

Yusoff, R., & Janor, R. M. (2014). Generation of an interval metric scale to measure attitude. *SAGE Open, 4*, 1–6.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353.