Journal of Behavioral and Experimental Economics 106 (2023) 102049

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

An empirical comparison of conjoint and best-worst scaling case III methods

Haotian Cheng [a,*], Tong Zhang [b], Dayton M. Lambert [c], Ryan Feuz [d]

[a] School of Natural Resources and the Environment, University of Arizona, Tucson, AZ 85721, USA
[b] National School of Agricultural Institution and Development, South China Agricultural University, Guangdong, China
[c] Department of Agricultural Economics, Oklahoma State University, Stillwater, OK, 74078, USA
[d] Department of Applied Economics, Utah State University, Logan, UT 84322, USA

ARTICLE INFO

JEL Codes:
C81
D12
Q13

Keywords:
Best-worst survey
Conjoint survey
Biobased product

ABSTRACT

The Best-Worst Scaling (BWS) case III method, also called the BWS 'multi-profile case.' has been widely used to characterize survey respondent preferences for market goods. The BWS method is similar to conjoint analysis methods in that respondents select from a set of hypothetical item profiles with different attribute levels. Unlike conjoint methods, which allow respondents to select their best/most preferred profile, the BWS case III method asks respondents to select 'best' and 'worst' profiles in each choice set. This study compares consumer willingness to pay (WTP) estimates from conjoint and BWS case III survey formats. Data on consumer preferences for single-use eating-ware products made from biobased materials were collected. Results suggest that for the most preferred attribute levels, WTPs estimates are similar in magnitude and consistent for signs across methods. For least-preferred attributes, WTP estimates from the conjoint method are higher than those of the BWS method. However, the BWS WTP estimates have smaller confidence intervals.

1. Introduction

Revealed and stated preference methods are used to estimate consumer willingness to pay (WTP) for goods and services. Revealed preferences cannot be estimated for hypothetical or new products because purchasing data for these goods is unavailable (Samuelson, 1938, 1948). Researchers, therefore, apply stated preference methods, such as conjoint analysis, to determine potential demand for hypothetical goods, new products, or non-market goods (Lusk et al., 2006, 2008; Chen et al., 2010; Dauda & Lee, 2015; Verma & Chandra, 2017; Cheng et al., 2021; Hu et al., 2022). Two empirical approaches that elicit stated preferences include conjoint and best-worst scaling methods.

Conjoint analysis requires respondents to select one profile (a 'most preferred' choice) amongst several profiles, with options differentiated by unique combinations of attributes, each with two or more levels denoting quality or quantity. Respondents' answers presumably maximize their expected utility (Sen, 1994). Conjoint methods can be adapted to studying preferences in various market contexts. A few examples of conjoint studies highlight the method's flexibility across different situations. Lusk et al. (2006) used conjoint analysis to estimate consumer WTP for pork produced without subtherapeutic antibiotics. Lusk et al. (2008) used a non-hypothetical conjoint ranking approach to estimate consumer preferences for pasture-raised steak. Chen et al. (2010) used conjoint analysis to evaluate attributes of online shopping websites that increase consumer purchasing intentions. Dauda and Lee (2015) used conjoint analysis to estimate consumer preferences for online banking services. Verma and Chandra (2017) used conjoint analysis to identify consumer preferences for hotel accommodations. Cheng et al. (2021) estimated consumer WTP using conjoint methods for biobased products. Hu et al. (2022) conducted a conjoint study to compare consumer WTP values for dried noodles.

Best-worst scaling (BWS) is an alternative method for eliciting stated preferences. Finn and Louviere (1992) extended Richardson's (1938) max-diff preference ranking method to what is now called BWS. Empirical studies using BWS have increased rapidly since 2010 as an alternative method for determining stated consumer preferences (621 studies were identified by a Scopus search for the keywords "best--worst-scaling" since 2010). BWS evolved to handle three kinds of cases (Louviere et al., 2015). In BWS case I, the 'object case,' respondents assess the degree of importance of a set of subjects on a rating scale (Finn & Louviere 1992; see Zhang et al. (2015) and Pokhilenko et al. (2020) for additional reviews). In BWS case II (BWS-3), the 'profile case,'

* Corresponding author.
E-mail address: haotiancheng@arizona.edu (H. Cheng).

https://doi.org/10.1016/j.socec.2023.102049
Received 3 February 2023; Received in revised form 10 May 2023; Accepted 2 June 2023
Available online 4 June 2023
2214-8043/

respondents state their 'best' and 'worst' choices that correspond with attribute levels in a profile (Flynn, 2010; Soekhai et al., 2021; Himmler et al., 2021). BWS-3, or 'multi-profile cases,' is closely aligned with classical discrete choice experiments. Respondents are offered a sequence of choice sets from which selections are made (Louviere, Hensher & Swait, 2000). BWS-3 is most similar to conjoint analysis methods because respondents are asked to select item profiles among a set of profiles. However, unlike conjoint analysis wherein respondents are asked only to select a 'most preferred' item profile, the BWS-3 method tasks respondents with selecting 'best' and 'worst' profiles in a set. BWS-3 has also been used in business and health studies. For example, Mühlbacher et al. (2020) used the BWS-3 method to determine the most patient-relevant endpoints regarding the effects, risks, and administration of hemophilia A treatments. Yoo and Doiron (2013) used the BWS-3 method to estimate consumer preferences for nursing jobs. Thong et al. (2018) used the BWS III methods to estimate consumer preferences for beer.

Questions naturally arise as to which method to use and under what circumstances. Potoglou et al. (2011) compared conjoint and BWS case II methods using an in-person survey. Respondents were asked to answer both conjoint and BWS case II questions. Potoglou et al. found no significant difference between the methods regarding stated preferences. Van Dijk et al. (2016) conducted a similar study where respondents answered conjoint and BWS case II questions. Their study found no differences in preference rankings between the methods. However, Van Dijk et al. (2016) found that respondents rated the BWS choice sets more challenging to answer. These findings are supported by Severin et al. (2013)'s research. Using a split survey design, half of the respondents answered conjoint questions. In contrast, the other half responded to BWS case II questions. Both methods produced similar results. These studies suggest conjoint and BWS-3 formats should produce similar results with the BWS method associated with greater complexity (i.e., higher costs).

The literature, however, is sparse regarding conjoint and BWS-3 comparisons. Yangui (2019) found that WTP values were similar for conjoint and BWS-3 methods. Xie et al. (2013) asked participants to answer conjoint and BWS-3 questions using an in-person interview format. They found that the estimates' variance was smaller for the BWS study than for conjoint results. Xie et al. concluded that, for the same prediction accuracy and precision, BWS-3 had the advantage of reducing respondent burden.

Possible reasons why the variance of BWS-3 and conjoint WTP estimates differ may relate to menu dependence and preference imprecision. Sen (1994) concluded that the offered choice sets might introduce context-dependency, or what Sen called 'menu dependency, which may contaminate rational decision-making and alter revealed preference ordering. Menu dependency may therefore cause inconsistencies in preference rankings and, in some empirical contexts, increase the variance around willingness to pay estimates. Another explanation could be related to preference imprecision (Maltz & Rachmilevitch, 2021; Chernev, Böckenholt, & Goodman, 2015; Bayrak & Hey, 2020). Preference imprecision is a consumer choice concept that refers to the degree of uncertainty or vagueness a consumer has regarding preferences for a product or service. If a consumer is uninformed about a product, then this uncertainty may introduce inconsistencies in choice rankings and increase the variance of WTP estimates. A related concept that could cause divergence between BWS and conjoint WTP estimates and variance is inattention bias (Malone & Lusk, 2019). Inattentive respondents may provide inconsistent preferences rankings if they are not paying attention to survey questions or information (Cheng et al., 2021). Some researchers suggest the BWS format may reduce intention bias (Erdem & Rigby, 2013; Massey et al., 2015; Shoji et al., 2021).

The research objective is to compare the BWS-3 and conjoint methods for eliciting consumer preferences through an online survey. The survey used a split treatment design with a conjoint group and BWS-3 group. The product is single-use eating ware (SUEW) made from biobased materials. This study randomly assigns respondents a conjoint or BWS-3 survey to compare WTP estimates from both groups. The study adds to the literature by comparing the conjoint and BWS-3 methods by contrasting the methods using an online survey rather than a survey administered in one location.

1.1. Data

Survey data are from Cheng et al. (2021). The study evaluated consumer preferences for single-used eating ware (SUEW) products made from biobased materials. The two surveys implemented in the study are identical, except for the structures of the choice experiment sections (conjoint or BWS-3). The data was collected using online conjoint and BWS-3 surveys launched by Qualtrics®, in October and November 2019. Respondents were 18 or older and were drawn randomly from a sampling frame of US households who self-identified as SUEW consumers. Qualtrics sent invitation emails to individuals by computer or cellphone. Respondents who completed the survey were compensated with coupons. There were $n = 1010$ and $n = 1000$ for the BWS and conjoint groups, respectively. These sample sizes were based on a 5% margin of error with a 95% confidence interval. Three separate surveys in the conjoint and BWS groups were divided into 'no information,' 'limited information,' and 'full information' treatments. The information treatments pertained to the level of information respondents received about the material composition of the SUEW product. Only the sample of 'full information' respondents from both the conjoint and BWS surveys were used here ($n = 335$ for the conjoint complete information group and $n = 345$ for the BWS complete information group). This sub-sample was used to avoid potential information effects. The average completion times for the conjoint and BWS-3 surveys were 15 and 22 min, respectively.

The following attributes were used to develop SUEW product profiles: a) product degradability (3 levels; not degradable, degradable in 6 months (*Degrade6*), degradable in 24 months (*Degrade24*)); b) origin (2 levels; made in the US, or made elsewhere (*Origin*)); c) product content certification (2 levels; no or yes, (*Label*)); d) the material source (3 levels; plastic, paper (*Paper*), or wheat straw (*Wheat*)); and e) price for a 25-count of 10-inch size SUEW plates (6 levels; $2.27, $3.82, $5.36, $6.91, $8.45, and $10.00) (Table 1). Price points were determined from a review of 20 SUEW products. The highest price was $10.00 for a 25-count package of 10-inch plates. The lowest price for the same quantity and plate size was $2.27. These lower and upper bound prices were used to determine the other four price levels, which were uniformly distributed between the lower and upper price bounds.

In the conjoint choice experiment, respondents viewed a screen with five product profiles differing in attribute levels and a 'none of the above' (opt-out) option. The opt-out option was coded as an alternative-specific constant (ASC; Adamowicz et al., 1998). Participants were asked to select the most preferred alternative (Fig. 1).

In the BWS choice experiment, respondents viewed a screen with five product profiles that differed in attribute levels (Fig. 2). Respondents were asked to select the most and least attractive alternatives. Once selected, these alternatives were removed from the choice set, leaving three alternatives to rank 'most' or 'least' preferred in a second round. Completing the second round resulted in the complete ranking of all alternatives in each choice. One limitation of this study is that, for the BWS-3 question, no ASC option was included. For this reason, we split the conjoint sample into two groups; first, an ASC group including 119 respondents who selected the ASC at least once across all choice sets; and second, a non-ASC group that included 216 respondents who never selected the ASC option in any of the choice sets. This procedure allows for directly comparing the WTP results estimated for the conjoint non-ASC group with those estimated for the BWS-3 group while holding constant ASC effects.

Two trap questions were included in the survey. The purpose of the trap questions was to control respondent attentiveness. One trap question appeared before and the other after the choice experiment section.

Table 1
Choice experiment levels and attributes.

Attribute	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Degradability	Not degradable	6 months	24 months			
Content certification	No	Yes				
Material	Plastic	Paper	Wheat straw			
Origin	Made in the US	Made elsewhere				
Price ($/25 count)	$2.27	$3.82	$5.63	$6.91	$8.45	$10.00

Attribute	Plate A	Plate B	Plate C	Plate D	Plate E	None. I would not buy any of these plates
Made in the U.S.	No	Yes	Yes	Yes	No	
Source	Wheat Straw	Paper	Plastic	Wheat Straw	Plastic	
Degradable	No	Yes, 2 years	No	Yes, 2 years	Yes, 6 months	
USDA Certificated Bio-based	Yes	No	Yes	No	Yes	
Price for 25 Plates	$5.36	$3.82	$2.27	$10.00	$8.45	
Which Plate do you Prefer ➔	O	O	O	O	O	O

Fig. 1. Conjoint Question Example.

	Plate A	Plate B	Plate C	Plate D	Plate E
Made in the U.S.	No	Yes	Yes	Yes	No
Source	Wheat straw	Paper	Plastic	Wheat straw	Plastic
Biodegradable	No	Yes, 2 years	No	Yes, 2 years	Yes, 6 months
USDA Certificated Bio-based	Yes	No	Yes	No	Yes
Price for 25 plates	$5.36	$3.82	$2.27	$10.00	$8.45

Which plate do you prefer most?		Which plate do you least prefer?
○	Plate A	○
○	Plate B	○
○	Plate C	○
○	Plate D	○
○	Plate E	○

Fig. 2. Best-Worst Question Example.

Respondents who incorrectly answered a trap question were given a second chance to revise their answers. Respondents incorrectly answering the question on the second try were coded as inattentive (= ‘1’, ‘0’ otherwise). The inattentive dummy variable was interacted with each product attribute and price to control for respondent inattention.

2. Methods and procedures

Conditional logistic (CL) regression (McFadden, 1974) was used to estimate WTP for the conjoint group. BWS data are usually analyzed using rank-ordered logistic (ROL) regression (Scarpa et al., 2011). Both CL regression and ROL regressions were used to estimate WTP for the BWS-3 group to hold constant the influence of regression procedures.

We use McFadden (1974)’s random utility model (RUM) to estimate WTP. The linear indirect utility function is:

$$v_{ijt} = \mathbf{x}_{ijt}\beta - \beta_m \cdot p_{jt} + \in_{ijt} \tag{1}$$

where v_{ijt} is the utility derived by individual i on choice occasion (‘task’) t having selected choice alternative j; $\mathbf{x}_{ijt}$ is a vector of attributes for alternative $j = 1,\ldots J$; β includes the j attribute coefficients (β_j); $\beta_m$is the marginal utility of income (i.e., the coefficient on price, p_{jt}); and $\in_{ijt}$ is an unobserved random disturbance term from the extreme value distribution with an expected value of zero and variance $\sigma_{\in}^2$.

The CL regression framework for analyzing RUM specifications assumes that respondents select a 'best' alternative among a set of alternatives. The probability that a respondent selects alternative *j* as 'best' is (McFadden, 1974):

$$\Pr[j \mid \mathbf{x}_{it}] = \prod_{t=1}\prod_{i=1}\prod_{k=1}^{J-1}\frac{\exp(\mathbf{x}_{ijt}\beta)}{\sum_{k=1}^{J}\exp(\mathbf{x}_{ijt}\beta)} \tag{2}$$

ROL regression for the RUM specification assumes that respondents rank choice alternatives from best to worst. Let $r_{ijt} = (r_{i1t}, \ldots, r_{iJt})$ indicate an individual's choice ranking in descending order of preference. The probability an individual orders $\mathbf{r}$ in any particular ranking is (Scarpa et al., 2011):

$$\Pr[v_{it}(r_{i1t}) > v_{it}(r_{i2t}) > \cdots v_{it}(r_{iJt})] = \prod_{t=1}\prod_{i=1}\prod_{h=1}^{J-1}\frac{\exp(\mathbf{x}_{ijt}(r_h)\beta)}{\sum_{m=h}^{J}\exp(\mathbf{x}_{ijt}(r_m)\beta)} \tag{3}$$

where $\mathbf{x}_{ijt}(r_h)$ includes an alternative's attributes that receive rank h in the ordered set.

2.1. Willingness to pay

Marginal WTP is calculated as (Habb & McConnell, 2002):

$$\frac{\partial \widehat{WTP}_j}{\partial x_j} = \frac{\widehat{\beta}_j}{\widehat{\beta}_m} \tag{4}$$

where $\widehat{\beta}_j$ is the estimated parameter for attribute j and $\widehat{\beta}_m$ is the parameter estimate on price. The Krinsky and Robb parametric bootstrapping method was used to estimate the 95 percent confidence intervals for each attribute and each model's marginal WTP (Krinsky & Robb, 1986). WTP confidence intervals from the CL and ROL models were compared to determine statistically significant differences between methods (Schenker & Gentleman, 2001). The complete combinatorial method developed by Poe et al. (2005) was used to test for statistical differences among attribute WTP values across the models.

2.2. Estimation

The CL and ROL models were estimated using Stata's *clogit* and *rologit* commands (Stata 15.1, StataCorp LLC, College Station, TX, 2017). Both routines estimate the model parameters using simulated maximum likelihood (SML) (Cappellari & Jenkins, 2006).

3. Results

Table 2 presents the SML estimates of the CL methods for the conjoint ASC group, the conjoint non-ASC group, and the CL and the ROL methods for the BWS group. For attentive respondents, coefficients on attributes among the four models are significant at the 5% level. Only the Price and Degrade6 attributes are significant at the 5% level among the four models for inattentive respondents. For the BWS group, the confidence intervals for each attribute level are smaller than the conjoint group, consistent with Xie et al. (2013)'s findings.

All models produce similar attribute rankings (Table 3). *Degradability* has the average highest WTP premium in each model estimated, followed by *Source* and *Label*. The attribute exhibiting the lowest premium is *Origin*. This premium ranking is similar to the rankings found by Cheng et al. (2021). The non-ASC group WTP is slightly higher than the ASC group. The attribute premiums among the ASC and BWS groups (CL and ROL models) are similar, except for those for the *Label* and *Origin* attribute levels.

Table 4 presents the statistical comparison from Poe's procedure of the WTP values estimated by both groups. The attributes with the highest premiums, *Degradability* and *Source*, are not statistically different across methods. However, the WTP for the attribute with the lowest premium, *Origin* was statistically different among all four comparison groups (conjoint non-ASC group, conjoint ASC group, ROL-BWS, and CL-BWS). The WTP for the attribute with the second lowest premium, *Label*, is also statistically different when comparing the ASC group, ROL-BWS, CL-BWS, and ROL-BWS.

The results show that estimated WTPs are similar and consistent across methods for the most preferred attributes. In contrast, for least-preferred attributes, WTP estimates from the conjoint method are higher with larger confidence intervals than the BWS method. We do not believe menu dependence explains the differences in the WTP estimates. Each question's options (levels, orders, and the number of alternatives) are identical within the conjoint and BWS surveys. These findings could be explained as resulting from fatigue and preference imprecision (Bayrak & Hey, 2020; Wang, Yuan & Beck, 2022). Participant fatigue is

Table 2
Conjoint and BWS groups estimates.

	Conjoint Sample				BWS Sample			
	ASC Group		Non-ASC Group		CL-BWS		ROL-BWS	
Attribute	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Price	−0.2453	0.0342***	−0.1583	0.0193***	−0.2382	0.0168***	−0.1898	0.0136***
ASC	1.0875	0.2605***	−18.0634	0.1673***				
Paper	0.5654	0.1256***	0.4588	0.0808***	0.5244	0.0659***	0.4359	0.0384***
Wheat	0.5368	0.1248***	0.4597	0.0897***	0.5064	0.0680***	0.4304	0.0410***
Dagrade6	1.0876	0.1410***	0.7891	0.0858***	0.8223	0.0678***	0.7716	0.0578***
Degrade24	0.7121	0.1343***	0.5513	0.0780***	0.5033	0.0604***	0.5207	0.0421***
Label	0.5395	0.0967***	0.4411	0.0548***	0.4127	0.0442***	0.1814	0.0257***
Origin	0.5208	0.0971***	0.3903	0.0589***	0.3184	0.0421***	0.1259	0.0202***
Inattention:								
d × *Price*	0.1662	0.0531***	0.1642	0.0384***	0.1911	0.0271***	0.1337	0.0203***
d × *ASC*	0.4770	0.4550	0.2214	0.3520				
d × *Paper*	0.1648	0.2604	−0.1242	0.1673	0.1392	0.1291	−0.2215	0.0675***
d × *Wheat*	0.1572	0.2597	−0.1999	0.1701	0.0891	0.1336	−0.2442	0.0638***
d × *Degrade6*	−0.9920	0.2527***	−0.7430	0.1358***	−0.5062	0.1145***	−0.5926	0.0777***
d × *Degrade24*	−0.7235	0.2685***	−0.6322	0.1335***	−0.3421	0.1114***	−0.4452	0.0694***
d × *Label*	−0.2791	0.1817	−0.1747	0.1047*	−0.1673	0.0839***	−0.0839	0.0516
d × *Origin*	−0.4212	0.1682**	−0.2059	0.0995*	−0.1865	0.0887***	−0.0314	0.0489
N	119		216		345		345	
Log likelihood	−2129		−3751		−5618		−18,150	

Table 3
Confidence intervals for the conjoint and BWS groups.

	Conjoint Sample				BWS Sample			
	ASC Group		Non-ASC Group		CL-BWS		ROL-BWS	
Attribute	Estimate	WTP 95% Confidence Interval	Estimate	WTP 95% Confidence Interval	Estimate	WTP 95% Confidence Interval	Estimate	WTP 95% Confidence Interval
Price	–		–		–		–	
ASC	4.43	[0.17,14.01]	−114.11	[−225.10,−76.40]				
Paper	2.30	[0.11,5.99]	2.90	[0.62,6.55]	2.20	[1.45,3.44]	2.30	[1.03,3.79]
Wheat	2.19	[0.23,6.17]	2.90	[0.20,7.33]	2.13	[1.30,3.48]	2.27	[0.82,3.86]
Dagrade6	4.43	[2.13,9.69]	4.98	[2.65,9.45]	3.45	[2.79,6.14]	4.07	[2.22,5.58]
Degrade24	2.90	[0.79,6.61]	3.48	[1.55,7.07]	2.11	[1.89,3.98]	2.74	[1.17,3.71]
Label	2.20	[0.62,5.31]	2.79	[1.30,6.11]	1.73	[0.43,1.82]	0.96	[0.92,3.15]
Origin	2.12	[0.68,4.71]	2.47	[1.01,4.95]	1.34	[0.18,1.22]	0.66	[0.53,2.20]
Inattention:								
d × *Price*	0.68	[−0.42,1.46]	1.04	[−0.19,2.32]	0.80	[0.37,1.06]	0.70	[0.44,1.20]
d × *ASC*	1.94	[−5.81.9.75]	1.40	[−8.98,11.79]				
d × *Paper*	0.67	[−4.26,6.07]	−0.78	[−5.89,4.10]	0.58	[−2.76,0.30]	−1.17	[−2.32,3.16]
d × *Wheat*	0.64	[−4.47,6.22]	−1.26	[−7.14,3.62]	0.37	[−2.64,0.20]	−1.29	[−1.74,2.72]
d × *Degrade6*	−4.04	[−10.64,−0.03]	−4.69	[−10.28,−1.49]	−2.13	[−5.21,−1.28]	−3.12	[−4.37,−0.05]
d × *Degrade24*	−2.95	[−8.92,1.37]	−3.99	[−9.42,−0.74]	−1.44	[−4.01,−0.77]	−2.35	[−3.33,0.73]
d × *Label*	−1.14	[5.59,2.71]	−1.10	[−4.76,2.14]	−0.70	[−1.82,0.55]	−0.44	[−2.68,0.62]
d × *Origin*	−1.72	[−5.36,1.32]	−1.30	[−4.46,1.15]	−0.78	[−1.30,1.03]	−0.17	[−2.65,0.69]
N	119		216		345		345	

Table 4
P-value comparing WTP between different groups.

Attributes	ASC Group vs. Non-ASC Group P-Value Comparing WTP	ASC Group vs. CL-BWS P-Value Comparing WTP	ASC Group vs. ROL-BWS P-Value Comparing WTP	CL-BWS vs. ROL-BWS P-Value Comparing WTP
Paper	0.7462	0.4416	0.4977	0.5935
Wheat	0.7772	0.4641	0.5492	0.6295
Dagrade6	0.6974	0.1042	0.3239	0.8828
Degrade24	0.7574	0.1047	0.3991	0.9499
Label	0.8014	0.1813	0.0023	0.0020
Origin	0.7116	0.0378	0.0001	0.0014
Inattention:				
d × *Price*	0.9088	0.7420	0.5563	0.2151
d × *Paper*	0.1683	0.4702	0.0534	0.0039
d × *Wheat*	0.1037	0.4120	0.0432	0.0061
d × *Degrade6*	0.3312	0.9487	0.7819	0.0651
d × *Degrade24*	0.2352	0.8968	0.6954	0.0663
d × *Label*	0.5128	0.7034	0.8130	0.7181
d × *Origin*	0.6692	0.8849	0.9830	0.9132

associated with an increased cognitive burden imposed by the time and effort involved in taking a survey. The BWS method required more time (on average) for respondents to complete the survey than the conjoint method (22 min versus 15 min, respectively). This additional time may have induced fatigue effects. Past research evaluated various effects of fatigue on estimating stated preferences. Several studies demonstrated that WTP might decrease because of satisficing or straight-lining tendencies (Wang, Yuan & Beck, 2022; Layana & Lee, 2020). However, when using BWS case II data, Monte Carlo simulations have shown that the complete ranked BWS method improves the accuracy of estimates compared to incomplete ranked methods (Cheng, Feuz & Lambert, 2023), lending support for the choice of complete ranked methods. Preference imprecision could also explain the larger confidence intervals of the conjoint WTP estimates. In this study, respondents might have experienced greater preference imprecision in the conjoint survey due to soliciting incomplete ranked preferences, as preference incompleteness can be associated with preference imprecision (Bayrak & Hey, 2020). When faced with a conjoint task that does not require a complete ranking of preferences, consumers may experience preference imprecision for lower-ranked attributes due to the difficulty in distinguishing between minor differences in attribute levels and the potential for respondents to focus more on the most important attributes. Preference imprecision could result in larger confidence intervals across all WTP estimates compared to the complete-ranked BWS method.

4. Conclusions

This research note compared WTP estimates for extrinsic attributes of single-use eating ware using conjoint analysis and BWS-3 data collection methods. The attributes evaluated included biomaterial source, product degradability, USDA biobased certification, and manufacture origin. Findings suggest that conjoint analysis and BWS-3 estimations perform similarly for the most preferred attributes. However, there were differences in the WTP estimates for least-preferred attributes across the various methods. The WTP confidence intervals from BWS-3 were smaller than estimates from the conjoint data. However, increased precision must be balanced with the extra complexity (Dijk et al., 2016) and the time required to complete the BWS-3 survey.

The results contribute to the existing literature comparing the BWS and conjoint methods. Specifically, these results expand on those of Xie et al. (2013) and Yangui et al. (2019) by reducing the potential for sample bias by conducting an online survey rather than a sample from one location, more general and scientific sampling; and using a split sample design to help avoid information bias. Response fatigue was also reduced by splitting respondents into two groups that received either the conjoint or the BWS-3 format.

One limitation of this study is that for the BWS group, there was no ASC option in each question. Maltz and Rachmilevitch (2021) concluded that consumers might prefer an outside option when choices are complex. Thus, one direction of future research to address the limitation would be to add an alternative specific constant (ASC) option to the BWS choice and then compare the ASC groups' WTP estimates from the conjoint and the BWS methods. Another direction for future research is related to salience theory (Bordalo, Gennaioli & Shleifer, 2012). Bordalo, Gennaioli and Shleifer (2012) concluded that decision-makers exaggerate the probability of extreme events if they are aware of this possibility. Thus, a future research question could be to compare the subgroups who are educated with the information related to least-preferred attributes with a sub-group of participants who are unexposed to this information.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by the Sparks Chair in Agricultural Sciences & Natureal Resources, Oklahoma State University.

References

Adamowicz, W., Boxall, P., Williams, M., & Louviere, J. (1998). Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *American Journal of Agricultural Economics, 80*(1), 64–75. https://doi.org/10.2307/3180269

Bayrak, O., & Hey, J. D. (2020). Understanding preference imprecision. *Journal of Economic Surveys, 34*(1), 154–174. https://doi.org/10.1111/joes.12343

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics, 127*(3), 124–1285. https://doi.org/10.1093/qje/qjs018

Cappellari, L., & Jenkins, S. P. (2006). Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *Stata Journal, 6*(2), 156–189. https://doi.org/10.1177/1536867x0600600202

Chen, Y. H., Hsu, I. C., & Lin, C. C. (2010). Website attributes that increase consumer purchase intention: A conjoint analysis. *Journal of Business Research, 63*(9–10), 1007–1014. https://doi.org/10.1016/j.jbusres.2009.01.023

Cheng, H., Lambert, M. D., DeLong, K. L., & Jensen, K. L. (2021). Inattention, availability bias, and attribute premium estimation for a biobased product. *Agricultural Economics, 53*(2), 274–288. https://doi.org/10.1111/agec.12679

Cheng, H., Feuz, R., & Lambert, D. M. (2023). A comparison of best-worst scaling marginal and rank methods. *Applied Economics Letters*. https://doi.org/10.1080/13504851.2023.2187019. In press.

Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology, 25*(2), 333–358. https://doi.org/10.1016/j.jcps.2014.08.002

Erdem, S., & Rigby, D. (2013). Investigating heterogeneity in the characterization of risks using best worst scaling. *Risk Analysis, 33*(9), 1728–1748. https://doi.org/10.1111/risa.12012

Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: The case of food safety. *Journal of Public Policy & Marketing, 11*(2), 12–25. https://doi.org/10.1177/074391569201100202

Flynn, T. N. (2010). Valuing citizen and patient preferences in health: Recent developments in three types of best–worst scaling. *Expert Review of Pharmacoeconomics & Outcomes Research, 10*(3), 259–267. https://doi.org/10.1586/erp.10.29

Habb, T. C., & McConnel, K. E. (2002). *Valuing environmental and natural resources*. Northampton, MA: Edward Elgar.

Himmler, S., Soekhai, V., van Exel, J., & Brouwer, W. (2021). What works better for preference elicitation among older people? Cognitive burden of discrete choice experiment and case 2 best-worst scaling in an online setting. *Journal of Choice Modelling, 38*, Article 100265. https://doi.org/10.1016/j.jocm.2020.100265

Hu, W., Sun, S., Penn, J., & Qing, P. (2022). Dummy and effects coding variables in discrete choice analysis. *American Journal of Agricultural Economics, 104*(5), 1770–1788. https://onlinelibrary.wiley.com/doi/10.1111/ajae.12311.

Krinsky, I., & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics, 68*(4), 715. https://doi.org/10.2307/1924536

Layana, M. C., & Lee, J. G. (2020). Respondent fatigue in estimates of the cost of white-collar crime: Implications from willingness-to-pay surveys. *Criminal Justice Policy Review, 31*(9), 1366–1389. https://doi.org/10.1177/0887403419900245

Louviere, J. J., Hensher, D. A., Swait, J. D., & Adamowicz, W. (2000). *Stated choice methods: Analysis and applications* (1st ed.). Cambridge University Press.

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling theory, methods and applications*. Cambridge University Press.

Lusk, J. L., Norwood, F. B., & Pruitt, J. R. (2006). Consumer demand for a ban on antibiotic drug use in pork production. *American Journal of Agricultural Economics, 88*(4), 1015–1033. https://doi.org/10.1111/j.1467-8276.2006.00913.x

Lusk, J. L., Fields, D., & Prevatt, W. (2008). An incentive compatible conjoint ranking mechanism. *American Journal of Agricultural Economics, 90*(2), 487–498. https://doi.org/10.1111/j.1467-8276.2007.01119.x

Malone, T., & Lusk, J. L. (2019). Releasing the trap: A method to reduce inattention bias in survey data with application to U.S. beer taxes. *Economic Inquiry, 57*(1), 584–599. https://doi.org/10.1111/ecin.12706

Maltz, A., & Rachmilevitch, S. (2021). A model of menu-dependent evaluations and comparison-aversion. *Journal of Behavioral and Experimental Economics, 91*, 1–8. https://doi.org/10.1016/j.socec.2020.101655

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142. Retrieved from https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf.

Massey, G. R., Wang, P. J., Waller, D. A., & Lanasier, E. V. (2015). Best–worst scaling: A new method for advertisement evaluation. *Journal of Marketing Communications, 21*(6), 425–449. https://doi.org/10.1080/13527266.2013.828769

Mühlbacher, A. C., Sadler, A., Lamprecht, B., & Juhnke, C. (2020). Patient preferences in the treatment of hemophilia A: A best–worst scaling case 3 analysis. *Value in Health, 23*(7), 862–869. https://doi.org/10.1016/j.jval.2020.02.013

Poe, G. L., Giraud, K. L., & Loomis, J. B. (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics, 87*(2), 353–365. https://doi.org/10.1111/j.1467-8276.2005.00727.x

Pokhilenko, I., Janssen, L. M. M., Hiligsmann, M., Evers, S. M. A. A., Drost, R. M. W. A., Paulus, A. T. G., et al. (2020). The relative importance of education and criminal justice costs and benefits in economic evaluations: A best–worst scaling experiment. *PharmacoEconomics, 39*(1), 99–108. https://doi.org/10.1007/s40273-020-00966-8

Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., et al. (2011). Best–worst scaling vs. discrete choice experiments: An empirical comparison using social care data. *Social Science &Amp; Medicine, 72*(10), 1717–1727. https://doi.org/10.1016/j.socscimed.2011.03.027

Richardson, M. W. (1938). Multimensional psychophysics. *Psychological Bulletin, 35*, 659–660. https://ci.nii.ac.jp/naid/10009703610/.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica, 5*(17), 61–71. https://doi.org/10.2307/2548836

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica, 15*(60), 243–253. https://doi.org/10.2307/2549561

Scarpa, R., Notaro, S., Louviere, J., & Raffaelli, R. (2011). Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons. *American Journal of Agricultural Economics, 93*(3), 813–828. https://doi.org/10.1093/ajae/aaq174

Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician, 55*(3), 182–186. https://doi.org/10.1198/000313001317097960

Sen, A. (1994). The formulation of rational choice. *The American Economic Review, 84*(2), 385–390.

Severin, F., Schmidtke, J., Mühlbacher, A., & Rogowski, W. H. (2013). Eliciting preferences for priority setting in genetic testing: A pilot study comparing best-worst scaling and discrete-choice experiments. *European Journal of Human Genetics, 21*(11), 1202–1208. https://doi.org/10.1038/ejhg.2013.36

Soekhai, V., Donkers, B., Levitan, B., & de Bekker-Grob, E. (2021). Case 2 best-worst scaling: For good or for bad but not for both. *Journal of Choice Modelling, 41*, Article 100325. https://doi.org/10.1016/j.jocm.2021.100325

Shoji, Y., Kim, H., Kubo, T., Tsuge, T., Aikoh, T., & Kuriyama, K. (2021). Understanding preferences for pricing policies in Japan's National Parks using the best–worst scaling method. *Journal for Nature Conservation, 60*, Article 125954. https://doi.org/10.1016/j.jnc.2021.125954

Thong, N. T., Thanh, B. Q., Solgaard, H. S., & Yang, Y. (2018). The role of packaging format, alcohol level and brand in consumer's choice of beer: A best-worst scaling multi-profile approach. *Food Quality and Preference, 65*, 92–100. https://doi.org/10.1016/j.foodqual.2017.11.005

van Dijk, J. D., Groothuis-Oudshoorn, C. G., Marshall, D. A., & IJzerman, M. J. (2016). An empirical comparison of discrete choice experiment and best-worst scaling to estimate stakeholders' risk tolerance for hip replacement surgery. *Value in Health, 19*(4), 316–322. https://doi.org/10.1016/j.jval.2015.12.020

Verma, V. K., & Chandra, B. (2017). Sustainability and customers' hotel choice behaviour: A choice-based conjoint analysis approach. *Environment, Development and Sustainability, 20*(3), 1347–1363. https://doi.org/10.1007/s10668-017-9944-6

Wang, C. X., Yuan, H., & Beck, J. T. (2022). Too tired for a good deal: How customer fatigue shapes the performance of Pay-What-You-Want pricing. *Journal of Business Research, 144*, 987–996. https://doi.org/10.1016/j.jbusres.2022.02.014

Xie, F., Pullenayegum, E., Gaebel, K., Oppe, M., & Krabbe, P. F. M. (2013). Eliciting preferences to the EQ-5D-5 L health states: Discrete choice experiment or multiprofile case of best–worst scaling? *The European Journal of Health Economics, 15*(3), 281–288. https://doi.org/10.1007/s10198-013-0474-3

Yangui, A., Akaichi, F., Costa-Font, M., & Gil, J. M. (2019). Comparing results of ranking conjoint analyses, best–worst scaling and discrete choice experiments in a nonhypothetical context. *Australian Journal of Agricultural and Resource Economics, 63*(2), 221–246. https://doi.org/10.1111/1467-8489.12292

Yoo, H. I., & Doiron, D. (2013). The use of alternative preference elicitation methods in complex discrete choice experiments. *Journal of Health Economics, 32*(6), 1166–1179. https://doi.org/10.1016/j.jhealeco.2013.09.009

Yusuf Dauda, S., & Lee, J (2015). Technology adoption: A conjoint analysis of consumers' preference on future online banking services. *Information Systems, 53*, 1–15. https://doi.org/10.1016/j.is.2015.04.006

Zhang, J., Reed Johnson, F., Mohamed, A. F., & Hauber, A. B. (2015). Too many attributes: A test of the validity of combining discrete-choice and best–worst scaling data. *Journal of Choice Modelling, 15*, 1–13. https://doi.org/10.1016/j.jocm.2014.12.001