The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1751-1062.htm

Applying best-worst scaling to wine marketing

Eli Cohen
Department of Hotel and Tourism Management, Guilford Glazer School of Business and Management, Ben-Gurion University of the Negev, Beer Sheva, Israel; and School of Marketing, University of South Australia, Adelaide, Australia

Abstract

Purpose – Most marketing researchers use rating scales to understand consumer preferences. These have a range of problems, which can be greatly ameliorated by the use of a new technique, best-worst scaling (BWS). The purpose of this paper is to demonstrate the BWS method by an empirical example, which demonstrates the steps to design and analyze a BW study.

Design/methodology/approach – A brief critique of ratings and rankings is presented. Then the basic concept of BWS is described, followed by how to use the BW method to explore how Australian and Israeli consumers choose wine in a retail store. The paper demonstrates the design of the questionnaire as well as the steps to analyze and present the results.

Findings – The BWS approach can be easily implemented for research in wine business especially for multicultural comparisons as it avoids scale confounds. After transformation of the best and worst scores of each respondent for each attribute, the data can be analyzed directly using various statistical methods and can be expressed as choice probabilities.

Research limitations/implications – The advantage of BWS is its ability to compare attributes using B−W and B/W scores. The BW method provides a better discrimination of the attributes analyzed.

Practical implications – The simplicity of the analysis and graphical presentation makes a significant contribution to practitioners as the B−W counts and probabilities of attributes are easy to obtain and understand.

Originality/value – This paper presents BWS method in a form that researchers and practitioners can use and adopt for research and market surveys. The paper presents an empirical example using BWS method to determine the importance of wine cues while consumers are choosing wine in a retail store.

Keywords Market research, Wines, Consumer behaviour

Paper type Research paper

Introduction

Consumers in different countries evaluate and purchase goods based on product attributes, which may be particular to each culture. Researchers try to find out what product attributes most influence consumers' choice. Some examples of food products that researchers are trying to study consumers' preferences for are organic products, genetically modified food, eco-friendly products, low alcohol wines and more. Managers and marketers are trying to measure and forecast the importance and

The author thanks Professor Jordan Louviere for his assistance and support in the implementation of the method; Professor Larry Lockshin for his support, advice and encouragement to implement the best-worst scaling methodology to wine marketing and other research topics; and Dr Simone Muller for her significant input to the first version of the paper, and to the reviewers for their valuable comments. The author acknowledges the School of Marketing at the University of South Australia and the staff of the School for their trust and support. The study was funded in part by the Australian Grape and Wine Research Development (GWRDC).

International Journal of Wine Business Research
Vol. 21 No. 1, 2009
pp. 8-23

1751-1062
DOI 10.1108/17511060910948008

preferences for product attributes to better fit products to consumers' expectations and demands. Wine provides an example of a complex set of products and cues for the marketer to analyze. Among these cues are extrinsic cues such as brand, price, label and region and intrinsic cues such as taste, aroma and alcohol content. As intrinsic attributes are difficult to assess without tasting the wine, most research on wine buying behavior is based on extrinsic cues. Lockshin and Hall (2003) reviewed over 75 articles concerning consumer behavior for wine and noted many of the studies used simple surveys with rating scales to measure consumer preference for various wine attributes. Furthermore, there were conflicting findings in the rank order of the attributes for importance, though previously having tasted the wine, the price, the origin, the grape variety and the brand name of the wine were all mentioned frequently. The authors concluded that the best means to advance understanding of which attributes and combinations led consumers to purchase a particular wine was to use either choice-based experiments or analysis of actual consumer purchases.

Measuring actual consumer purchases has a major weakness as it only allows analysis of what consumers have purchased, but new attributes or combinations cannot be tested. Another disadvantage is that there is usually not enough information about the consumers to allow segmentation. Discrete choice experiments are a powerful and useful method in marketing (Louviere and Woodworth, 1983; Louviere *et al.*, 2000). The attributes are presented in various combinations, called product concepts, and the subject is forced to make trade-off between the concepts. The method allows new attributes and combinations to be tested for preference. One of the disadvantages of discrete choice is that the design and analysis is complex and use sophisticated computer programs. Another, and perhaps more serious limitation to discrete choice models is the difficulty of interpreting the data including the inability to compare utilities across different experiments (Louviere *et al.*, 2000). Hence, most researchers still tend to use simple rating scales for surveys, which seem to be easy to administer and analyze. The subjects are asked to rate their preferences for each attribute on a given scale such as 1-5, 1-7 or any other scale. Users of this method generally assume that the rating scales are interval scales and hence it is straightforward to analyze the data and draw conclusions using simple statistical procedures, such as comparing means of the attributes using *t*-tests and analysis of variance. Although the rating tasks are easy for respondents to complete, they weakly discriminate among the attributes (Hein *et al.*, 2008), due to several issues that increase the variance or noise in the results.

Finn and Louviere (1992) suggested the "best-worst" (BW) method to overcome some of the limitations of scale-based surveys. The respondent has to choose the most preferred item (called "most" or "best") and the least preferred item ("least" or "worst") in a set of items. Then, the researcher is able to obtain a full ranking of the items analyzed. Practical details in applying the method and an example using wine cues will be presented in this paper.

The next sections of the papers are structured as follows. First, commonly used scaling methods are described with their advantages and limitations. The following section describes the BW method and the design of the choice sets for the survey used as the case study. The method is demonstrated using wine attributes that influence consumers in purchasing wine in the supermarket or in a wine store. Two different countries are compared to illustrate the strengths of best worst scaling (BWS). The last section presents conclusions, advantages and limitations of the BW method.

Scaling methods

One typical form of a rating scale is the Likert-type scale where the subjects are asked to tick their rating for each attribute. Sometimes each response category is labeled and sometimes only the endpoints are indicated. Researchers in marketing often use adjectival descriptors to label the scale categories (e.g. "important", "not important", "good" or "fair"). One issue is that survey respondents do not use ratings the same way across respondents, and hence, the meaning respondents associate to the categories influences the perceived distance between categories (Crask and Fox, 1987). The distance between four and five for one person may be different that another's. Hence, treating these adjectival descriptors or categories as equal interval scales may generate different conclusions than if they are treated as ordinal scales. Furthermore, people may limit their responses to certain parts of a rating scale (Couch and Keniston, 1960; Bachman and O'Malley, 1984) and different parts of the scale are used more often than others by different cultures.

An example of rating food products is presented by Yeh *et al.* (1998) who compared the use of a nine-point hedonic scale to evaluate various foods between American, Korean, Chinese and Thai consumers living in the USA. They concluded that Chinese, Korean and Thai respondents use the nine-point hedonic scale differently from American respondents, irrespective of residency in the USA or length of stay. Moreover, there were no significant differences in food preferences for Thai and Korea consumers residing in the USA or their native countries. The results of surveys of different populations are subject to a range of biases resulting in scores or ratings, which are too similar or too difficult to interpret (Cohen, 2003; Cohen and Neira, 2003; Cohen and Orme, 2004; Finn and Louviere, 1992). There is empirical evidence showing that residents of different countries differ significantly in their responses (see for example, Baumgartner and Steenkamp, 2001; Chen *et al.*, 1995; Dolnicar and Grün, 2007). Cohen (2003) also claimed that segmentation studies in international markets show differences may be due to differences in scale usage rather than to real differences in consumers' preferences. As a result, the conclusions of international studies based on rating scales may be biased.

Attribute importance measured by rating scales is usually not measured relative to other attributes. Furthermore, some individuals truly might like (or dislike) nearly every attribute or combination of them and they rate them as "important" (or "not important"). Such responses do not provide adequate discrimination to help managers identify real priorities (Finn and Louviere, 1992). The relative importance of each attributes is then derived based on the averages determined across all respondents. So, it is not possible to draw reliable conclusions concerning the importance of issues or attributes as there is no possibility for respondents to made trade-offs between the attributes.

Another method to evaluate the relative importance of attributes is ranking. The method requires respondents to rank attributes in terms of a specific characteristic, for example wine attributes in terms of importance, or wine taste in terms of preference. The task is relatively easy for respondents to complete if the number of attributes is small. As the number of attributes increases the task becomes exhausting for respondents. There are ways to rank many attributes, but the task becomes overly complicated (see for example Weller and Romney, 1988).

The ranking task could be simplified by using paired comparisons (developed by Thurstone, 1927), which is probably the easiest and most reliable method for ranking; "even a child who is unable to understand a rating scale could perform a series of

paired comparisons reliably" (Cohen and Orme, 2004). Respondents are asked to choose which is "more" important (the other one is the "less" important) of two items that presented. Assuming n items, the number of possible pairs is $n(n-1)/2$. The disadvantage of the paired comparison is that the number of pairs required to be judged rises rapidly as the number of items increases. For example, for ten items we need 45 pairs to be compared, with 13 items we need 78 pairs and for 16 items the number of pairs increases to 120.

One way to reduce the number of subsets is to arrange the items in subsets of three or four items each and ask the subject to order, in terms of importance, the items in each subset. As the number of items in each subset increases, the number of subsets decreases. For example, if we want to compare 13 attributes and we use subsets of four items in each subset, we need only 13 subsets if we apply a balanced incomplete block design (BIBD). Instead of ranking four items in each choice set, one can choose the most preferred item ("best") and the least preferred item ("worst"). Thus, the "BW" method is an extension of paired comparisons. BW method models the cognitive process by which respondents identify the two items with the most and the least of a characteristic from designed subsets of three or more items. The method has several advantages that overcome the limitations of other methods of measurements such as rating-based methods.

BW scaling

The BW approach, also known as maximum difference scaling, was developed by Louviere and Woodworth (1990) and first published in 1992 (Finn and Louviere, 1992). The method as noted above is an extension of Thurstone's (1927) random utility theory for paired comparisons. As respondents can only choose one most and one least preferred item in each choice set, they are necessarily required to make trade-offs between items or benefits (Cohen, 2003). This overcomes the issue of many items having similar importance weights. Furthermore, BW avoids the problem of rating bias, as there is only one way to choose the "most" and the "least" preferred item, independent of the cultural background of the respondent (Cohen and Markowitz, 2002; Lusk and Briggeman, 2009). Hence, the method is powerful in conducting cross cultural studies in consumer behavior.

BWS methodology has been recently used in different areas such as social sciences, food and health care (for example Auger *et al.*, 2007; Cohen and Neira, 2003; Lee *et al.*, 2008; Lusk and Briggeman, 2009). Flynn *et al.* (2007) present an application of the Best-worst approach to health care to understand whether waiting time is more important than quality of care. Auger *et al.* (2004) tested country differences related to attitudes of individuals with respect to social and ethical issues such as human rights, child labor, animal rights and recyclable material. Consumers' preferences for minced pork patties were studied by Jaeger *et al.* (2008). The BW method also has been used to evaluate the importance of food values such as naturalness, taste, safety, origin, environmental impact and other factors (Lusk and Briggeman, 2009). However, there are only limited studies in wine marketing that used BW method (Goodman *et al.*, 2005, 2008).

Designing BW surveys

Marley and Louviere (2005) showed that the total choices over all subsets of the implied pairs are consistent with the multinomial logit model. An approximation of the model is achieved by calculating the differences of the total best and total worst frequencies for each item. Thus, as long as the experimental design is balanced, simply adding of

the number of times an item is chosen as worst and subtracting that from the total number of times it is chosen as best provides a scale that is about 95 per cent as accurate as using multinomial logit to model the same data (Auger *et al.*, 2004).

BIBDs are suggested to organize the items to be analyzed in choice sets. One advantage of BIBD is that large numbers of items can be studied in order to get the full ranking of all items in a relatively small number of subsets. The BIB designs control the number of times each pair is compared and by increasing the number of times each item compared with other item, the total number of subsets is increased and/or the number of items in each subset is increased. The simplest design is the one that each item appears only once with each other. Comparing each item with each other item more frequently increases the internal validity of the survey, but makes it longer and more repetitive for the respondent.

A simple BIB design may be derived from a Latin Square design, which is used in agriculture and in industrial experiments when the items of interests have more than two levels and it is known that there are no interactions between the items. A Latin Square design for n items is organized by n rows and n columns, where each column and each row have all the items in different positions. Each row is considered as a block or a choice set. Latin Square designs seem to be a complete block design as each row contains all the attributes and the design is balanced as each attribute appears exactly the same frequency in all choice sets (Weller and Romney, 1988). If k columns ($k < n$) are omitted from a $n \times n$ Latin Square, the result is a Youden design that has $n - k$ attributes in each row, namely, an incomplete block design. The design is balanced, as each attribute appears exactly the same number of times across all rows or choice sets. More complicated Youden designs could be found in combinatorial books (e.g. Box *et al.*, 1978; Raghavarao and Padgett, 2005).

Researchers desiring to conduct BW studies should use a BIB design from a well-known source. The first step is to decide how many total attributes will be compared, and then to search for designs using that number or attributes. Researchers have to trade off the number if items per choice set vs the number of choice sets. Practice with BW designs seems to indicate that four to six items per choice is optimal for most respondents and most tasks. If the item labels are long, such as positioning statements (Remaud and Lockshin, 2009), then fewer than six items should be used. Respondents can typically undertake up to 20 choice sets, though anecdotal evidence suggests that boredom sets in after about 10-12 choice sets.

Example of a BW study

The next sections describe a BWS study. As noted above, the study starts with a literature review to identify the key attributes to be compared in the study. Then the issues of choosing a design are presented, followed by the data collection. The analysis is presented along with the interpretation of the results in several formats.

Attributes that influence consumers while purchasing wine in a retail store

There are many factors that influence consumers' choice for wine. Consumers are trying to reduce risk while buying a bottle of wine. The assessment of the quality of the wine is complicated and based on extrinsic cues such as price, brand, region, medal, variety and intrinsic cues such as taste, bouquet and other sensory characteristics of the wine. The taste of the wine was the most concerning risk for consumers (Mitchell and Greatorex, 1989) and it was found that the taste of the wine plays a dominant role for wine consumers (Koewn and Casey, 1995; Thompson and Vourvachis, 1995). In

most cases, tasting the wine is not possible during the process of selecting wine in a retail store, hence, most consumers try to choose the wine based on extrinsic cues or having previously tasting the wine.

The brand is one of the most important cues that consumers assess while purchasing a product or service (Gordon, 2002) and it is considered as the sum of the images that consumers have in their mind about a particular company. Another cue that commonly used in selecting wine is the origin of the wine which was found to be an important factor in the choice process of wine. In Australia it was the third important factor that influencing consumers in selecting wine (Batt and Dean, 2000). Other studies suggested that the region and the origin of the wine is an important cue in wine choice and has a major impact on wine purchase (Angulo *et al.*, 2000; Felzensztein and Dinnie, 2005; Felzensztein *et al.*, 2004; Gluckman, 1990; Orth *et al.*, 2005; Perrouty *et al.*, 2006; Rasmussen and Lockshin, 1999; Skuras and Vakrou, 2002). There are more factors that influence consumers in purchasing wine and related to price, such as grape variety, recommendation, reading about the wine, medal and front label (Combris *et al.*, 1997, 2000).

Based on the literature and wine experts, 13 attributes were chosen to represent the range of choice cues that consumers might use in wine selection. These attributes are presented in Table I. We applied a BIB design to allocate the attributes in the choice sets. After consideration, we decided that 13 choice sets were as many as the respondents could handle, because of the length of the rest of the questionnaire. The BIB design for v attributes is denoted as (b, r, k, λ) where b is the number of choice sets (blocks), r is the repetition per level, k is the number of items in each choice set (block size) and λ is the pair frequency. For example, the design noted as 13,4,4,1 for 13 attributes has 13 choice sets, each attribute appears 4 times across all choice sets, each choice set contains four attributes, and each attribute appears once with each other. The design is presented in Table II.

The design in Table II contains 13 choice sets, each choice set was presented in the questionnaire in a separate table. The 13 tables contain four attributes in each table. An example of one choice set (table) is presented in Table III and it represents choice set number 9 in Table II and contains attributes 9, 10, 12 and 5. In the choice sets that are presented to the subjects, the number of the attribute in the choice set is substituted with its description as in Table I.

	Attribute
1	Promotional display in-store
2	Grape variety
3	Origin of the wine
4	Information on the shelf
5	Alcohol level below 13%
6	Matching food
7	Information on back label
8	Medal/award
9	An attractive front label
10	Brand name
11	Someone recommended it
12	I read about it in a guide
13	Tasted the wine previously

Table I.
Wine attributes that consumers consider when purchasing wine

$v = 13$ Choice set no.	13,4,4,1 Attribute no.			
1	1	2	4	10
2	2	3	5	11
3	3	4	6	12
4	4	5	7	13
5	5	6	8	1
6	6	7	9	2
7	7	8	10	3
8	8	9	11	4
9	9	10	12	5
10	10	11	13	6
11	11	12	1	7
12	12	13	2	8
13	13	1	3	9

Table II.
Balanced incomplete blocks designs for 13 attributes

Remember the last time you purchased a bottle of wine to have with friends for dinner. For each of the following tables, tick the ONE reason that MOST influenced your choice and the ONE that LEAST

Least/worst	Attribute	Most/best
☐	1 An attractive front label	☐
☐	2 Brand name	☐
☐	3 I read about it	☐
☐	4 Alcohol level below 13%	☐

Table III.
An example of a BW choice set as presented to respondents

As the decision of purchasing wine depends on the consumption situation (Hall and Lockshin, 2000) the subjects were asked to choose the "best" (most important) and "worst" (least important) reason why the subject purchased a bottle of wine, related to the situation described in Table III. The situation was defined to standardize the situation and to avoid confusion with special situations where people's criteria might vary. Respondents find it easy to choose the best and worst from three to six items in a choice set (Cohen and Orme, 2004). More than five items in a choice set seems to be exhausting for respondents. As the number of attributes increases the number of choice sets increases and the questionnaire becomes long. One solution is to split the choice sets and to present only part of the choice sets to the respondents. This allows the use of large sets of attributes and this is useful if the assumption is that the respondents are relatively homogeneous. The disadvantage is that we obtain only partial information from each subject and hence segmentation of the sample is limited.

Data collection and analysis

Data was collected in Australia during 2006 as part of a broader international study (see Goodman, 2009). The subjects were recruited in a combination of convenience sampling outside retail wine stores and from an online panel. The respondents were asked to choose the "best" (most important) attribute and the "worst" (least important)

attribute while considering purchasing a bottle of wine in a retail store. The questionnaire is designed to include 13 choice sets (tables) based on the design in Table II and a range of questions concerning the habits of drinking wine and demographic data as planned by the researchers. The paper questionnaire was presented to the respondents explaining the purpose of the study and how to fill the tables (choice sets) in the questionnaire. We included a worked example of a correctly filled in choice set. Because in the past we found that subjects do not fill the tables properly, such as choosing more than one most or least, or missing an answer in one or more table(s). Such a questionnaire is not valid for the survey as missing data leads to an unbalanced design and hence it should be omitted. One option to minimize missing values or ticking best or worst twice is using an on-line survey, which also has the advantage of presenting the choice sets at random. Furthermore, the respondent is not able to continue to the next choice set without filling in the displayed choice set properly. Another advantage is that the respondent can be asked to continue the task of choosing the next best and worst with the other items in the choice set, and hence a full ranking of the items in each choice set may be obtained. Randomization of the choice sets in a paper questionnaire is more complex. Recent unpublished research by Rungie (2008) indicates that there is a small effect of the order of the choice sets. This is typically done by using several versions of the questionnaire. This should be in mind when transforming the items in the choice sets to the original labels. Another simple approach to randomization is to randomize the sequence of the items and then allocate them in the choice sets based on the original design.

Data analysis

The first step before the analysis of the data is transforming the item numbers in each choice set to the original items numbers as in Table I. One possibility is using an Excel© spreadsheet that also is used for calculations and graphical presentations. After transforming the best and the worst in each choice set to the original item numbers, the best minus worst (B−W) for each item is calculated and then for each respondent we have 13 new B−W variables, one for each item. As each item appears four times in this design (Table II), each attribute could be chosen four times as best, as the maximum, and none as worst or vice versa, i.e. four times as worst and none as best. Consequently, the B−W scores for each attribute and individual can range from +4 to −4. Frequencies beyond this range indicate error(s) in the data.

The ranking of the attributes for all the subjects in the survey is obtained by ordering the BW score of each attribute after subtracting the number of times the attribute was least important (worst) from the number of times it was most important (best) in all choice sets (B−W score). Positive values of best minus Worst means that the given attribute was chosen more frequently as "best" than "worst" and negative scores should be determined. The average B−W scores is calculated by dividing the totals of B−W scores by the number of respondents and the frequency that each attribute appears in the design of the choice sets, i.e. four in our example.

The B−W scores and the "average B−W scores" of the wine attributes for Australian consumers when they consider purchasing wine in a retail store are presented in Table IV. We can easily see which factors are most important, which are similar and which have the least importance by far in influencing wine choice in Australia. A simple way of graphical presentation is plotting the B−W average scores vs the attributes as depicted in Figure 1. It is obvious that the most important attribute is "tasted the wine previously" for Australian consumers. The next important

Table IV.
Importance of wine attributes by Australian consumers ($n = 305$, ranked by B–W score)

No.	Attribute	Total best	Total worst	B–W score	Average B–W score
13	Tasted the wine previously	838	44	794	0.651
11	Someone recommended it	497	115	382	0.313
2	Grape variety	468	189	279	0.229
3	Origin of the wine	366	178	188	0.154
10	Brand name	401	218	183	0.150
8	Medal/award	364	217	147	0.120
12	I read about it	246	246	0	0.000
6	Matching food	269	348	−79	−0.065
7	Information on back label	194	275	−81	−0.066
4	Information on the shelf	99	364	−265	−0.217
9	An attractive front label	119	430	−311	−0.255
1	Promotional display in-store	85	512	−427	−0.350
5	Alcohol level below 13%	19	829	−810	−0.664

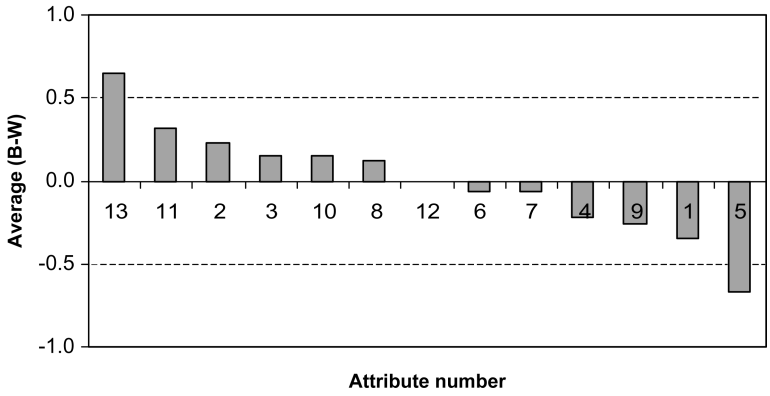

Figure 1.
BW average scores of wine attributes that influence consumer retail wine purchasing in Australia ($n = 305$)

attributes that most influence Australian consumers are "recommendation" and "grape variety".

More Australian consumers consider the "origin of the wine" as important (B−W > 0) while "information on the shelf" for example is less important for most consumers (B−W < 0), and "alcohol level below 13 per cent" is the least important for Australian consumers. The attributes in the middle of the scale (such as "medal", "I read about it" or "matching food") were either not often chosen as best or worst, or were chosen as best the about the same frequency as worst. It is important to understand the nature of the BW variation and to look for underlying segments in the analyzed sample. However, this issue is beyond the scope of this article and is discussed more completely by Mueller and Rungie (2009).

The question whether "grape variety" is significantly different from the "origin of the wine" or between any two attributes could be answered applying one way analysis of variance and a *post hoc* test, such as Tukey's B across all possible pairs of attribute means. We can see the statistically similar groups of attributes that emerge within the data (Table V). It is obvious that "grape variety", "origin of the wine", "brand name" and "medal/award" are not significantly different, while "tasted the wine previously" is significantly more important ($p < 0.05$) from "someone recommended it". Our

Attribute		B–W score	Average B–W score	Comparison of means*						
13	Tasted the wine previously	794	0.651	×						
11	Someone recommended it	382	0.313		×					
2	Grape variety	279	0.229		×	×				
3	Origin of the wine	188	0.154			×				
10	Brand name	183	0.150			×				
8	Medal/award	147	0.120			×				
12	I read about it	0	0.000				×			
6	Matching food	−79	−0.065				×			
7	Information on back label	−81	−0.066				×			
4	Information on the shelf	−265	−0.217					×		
9	An attractive front label	−311	−0.255					×	×	
1	Promotional display in-store	−427	−0.350						×	
5	Alcohol level below 13%	−810	−0.664							×

Note: *Tukey B $p < 0.05$; items with ×s in the same column or row are not significantly different

Table V.
Mean comparison of attributes for Australian data ($n = 305$)

experience shows that the BW method provides better discrimination compared to rating scale method. This is in agreement with Cohen and Neira (2003) and Hein *et al.* (2008), who found greater discrimination when the BW method was used compared to traditional scaling methods.

Another way to compare attribute importance is to derive ratio scores by taking the square root after dividing the total best (B) scores by the total worst (W) scores for each person (adding 0.5 to each W avoids dividing by zero). The resulting coefficient measures the choice probability compared to the most important item benchmark of 100 per cent (Auger *et al.*, 2007; Flynn *et al.*, 2007; Lee *et al.*, 2008; Marley and Louviere, 2005). The square root of (B/W) for all attributes (sqrt(B/W)) is scaled by a factor such that the most important attribute with the highest sqrt(B/W) becomes 100. All attributes can then be compared to each other by their relative sqrt(B/W) ratio. The result is interpreted as X per cent (e.g. 60 per cent) as likely to be chosen best as the most important.

The relative importance for each attribute for Australian wine consumers is presented in Table VI. The most important attribute for Australian consumers is "tasted the wine previously" and it is denoted as 100. All other attributes are related to this attribute and could be interpreted as relative to the most important attribute or to each other. The relative importance (Table VI) of "someone recommended it" is considered to be 47.8 and "matching food" is only 20.2. We conclude that "tasted the wine previously" is about twice more important as "someone recommended it" and four times as important as "matching food". In other notation, the probability of choosing "matching food" as important is about 20.2 per cent comparing to the probability of choosing "tasted the wine previously" as important (100 per cent), while choosing wine in a retail store.

As the individual B−W scores are unbiased and do not suffer scale confounds, it is straightforward to compare Australian consumers with another culture. A similar survey with the same design as in Table II was performed in Israel during 2006. The B−W scores were calculated for each respondent as well as the total B−W, the average B−W scores and the relative importance, and they are presented in Table VI. It is clear that the most important attribute for Israeli consumers is "tasted the wine previously" as for Australian consumers.

No.	Attribute	Australia $n = 305$ B–W score	Australia Average B–W score	Australia SQRT (B/W)	Australia Relative importance	Israel $n = 184$ B–W score	Israel Average B–W score	Israel SQRT (B/W)	Israel Relative importance
13	Tasted the wine previously	794	0.651	4.34	100.0	559	0.760	6.29	100.0
11	Someone recommended it	382	0.313	2.07	47.8	134	0.182	1.41	22.4
2	Grape variety	279	0.229	1.57	36.2	125	0.170	1.44	22.9
3	Origin of the wine	188	0.154	1.43	33.0	−157	−0.213	0.57	9.0
10	Brand name	183	0.150	1.35	31.2	133	0.181	1.45	23.1
8	Medal/award	147	0.120	1.29	29.8	−26	−0.035	0.93	14.7
12	I read about it	0	0.000	1.00	23.0	122	0.166	1.46	23.2
6	Matching food	−79	−0.065	0.88	20.2	169	0.230	1.58	25.1
7	Information on back label	−81	−0.066	0.84	19.3	−127	−0.173	0.58	9.2
9	An attractive front label	−311	−0.255	0.53	12.1	−222	−0.302	0.45	7.1
4	Information on the shelf	−265	−0.217	0.52	12.0	−236	−0.321	0.38	6.1
1	Promotional display in-store	−427	−0.350	0.41	9.4	−134	−0.182	0.68	10.8
5	Alcohol level below 13%	−810	−0.664	0.15	3.5	−340	−0.462	0.31	4.9

Table VI.
Importance of wine attributes by Australian and Israeli consumers (ranked by Australian B–W scores)

The relative importance, "matching food", "I read about it", "grape variety" and "someone recommended it" for Israeli consumers is almost the same, far away from "tasting the wine previously". The probabilities of choosing these four attributes as important are almost equal and are about 25 per cent. Comparing consumers' preferences in the two nations, the relative importance of "information on the shelf" in Australia is 12.0 while in Israel is 6.1 (Table VI). Although most consumers in Australia and in Israel consider "information on the shelf" is of less importance (negative B−W), we may conclude that the probability of choosing "information on the shelf" by Australian consumers about twice as by Israeli consumers.

Figure 2 illustrates the comparison of the relative importance of the items for consumers in the two countries. While tasting the wine has the highest probability to be chosen by both Israeli and Australian consumers, and is recorded as 100, all other attributes are presented in bars relative to "tasting the wine previously". It can be seen that "I read about it" (attribute #12) is of the same relative importance to both Australian and Israeli consumers, while "origin of the wine" is of much higher importance to Australian consumers.

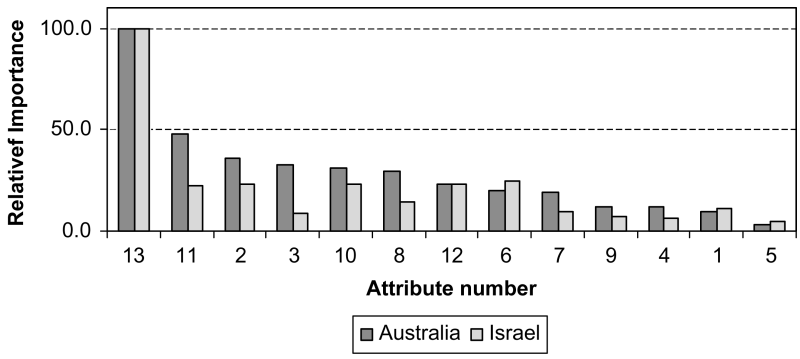

Figure 2.
Relative importance of wine attributes that influence consumer purchasing wine in Australia ($n = 305$) and in Israel ($n = 184$)

Significant differences (if exist) between consumers' perceptions in Australia and in Israel and within segments in both countries could be found applying simple statistical procedures using the B–W scores (Goodman *et al.*, 2008). As the B–W measures do not suffer scale bias, the data do not need any further transformation and many statistical procedures could be applied to the data, such as factor analysis, cluster analysis and other statistical procedures (see for example Cohen *et al.*, 2009; Mueller and Rungie, 2009).

Conclusions

Researchers have been trying to find what wine attributes most influence consumers' perception of wine quality. This is a complex task, as the perceived quality of the wine is based on many attributes that might influence consumers in the process of their choice. Consumers usually assess wine quality based on intrinsic cues such as alcohol level, wine style, taste and other sensory attributes, and extrinsic cues such as brand name, region, packaging and others. As most of the intrinsic cues can only be assessed during consumption, the ability of a consumer to evaluate the wine is limited without tasting the wine. Hence, consumers usually rely on extrinsic cues as indicators of quality. Among these cues are brand name, origin of the wine, region and wine awards. However, there are trade-offs among these cues for example, between brand name and region of the vineyard. The brand name is a combination of attributes that reflect the reputation of the company, the wine maker and the wine style, and the region reflects the influence of soil, climate, irrigation and other environmental factors. Understanding which attributes and/or combinations of them influence consumers in their choice of wine, choice-based surveys should be conducted. Yet, choice-based experiments are limited to extrinsic cues and they are relatively complex and required a relatively large number of respondents as the number of attributes to be analyzed increases.

One of the popular methods to measure consumers' preferences is using surveys where subjects are asked to rate or rank their preferences for each attribute on a given scale. The rating tasks are easy for respondents to complete and for researchers to analyze. Researchers assume that the rating scales are interval scales and hence it is straightforward to apply simple statistical analyzes, such as comparing means of the attributes. Yet, the attributes are rated independently and not compared with other attributes and hence, researchers cannot evaluate the relative importance of an attribute to the others as there are no trade-offs among the attributes. Furthermore, some consumers might like almost every attribute or consider most of the attributes as important. Such responses do not provide adequate discrimination and therefore, it is not possible to draw reliable conclusions concerning the importance of attributes. Studies in different cultures show that different cultures use different parts of the rating scales. Hence, the results of surveys of different populations are subject to a range of biases resulting in scores or ratings, which are too similar or too difficult to interpret. As a result, conclusions of international studies based on rating scales may be biased.

The BW method overcomes most of the limitations of rating and ranking methods. The respondents are provided choice sets with three or more items, and they are forced to choose the best/most important and the worst/least important item from each set. Unlike rating scales, there is no bias in the choice as there is only one option to choose something that is "most" or "least" important. The key issue for implementation is to design a series of choice sets that include all the items of interest and all possible

comparisons an equal number of times for each respondent. BIBDs are suggested to create the choice sets. The BIB designs control the number of times each pair is compared and by increasing the number of each item compared with other items, the total number of subsets is increased and/or the number of items in each subset is increased. One advantage of BIBD is that a large number of items can be studied and the full ranking of all items can be obtained with a relatively small number of subsets.

The BWS method is an approach that has much to offer to researchers in wine business and marketing as well as for marketers in general. The BW method has several advantages compared to other scaling methods. Respondents are asked to make trade-offs between items, which other rating methods do not use. BWS provides a better discrimination of the items in the study. The method allows the researcher to construct an individual-level ranking in a relatively easy structure, especially using on-line surveys where respondents are asked to choose the second and/or third set of best and worst in each subset. The BW scaling method provides the ranking of the items in the study and allows the researcher to measure the relative importance of each attribute to the other as a ratio scale of the probabilities of choosing each attribute. The method yields a score of BW for each attribute that could be analyzed using many statistical procedures without standardization of the data. The ratio of B/W of each attribute is in correlation with the probability of choosing the attribute as important and can be directly compared to other attributes.

The paper presents an empirical example using BWS method to determine the importance of wine cues while consumers are choosing wine in a retail store. The method can be implemented to other wine characteristics or to a wide range of marketing questions. Researchers can estimate the probability of choosing Chardonnay compared to Sauvignon Blanc for example, or any other wine variety. The question whether Chilean wine is preferred to French wine (or vice versa) by consumers in different countries could be answered and as well as what price consumers are willing to pay. What is the probability of choosing an imported wine compared to another local or imported wine? Is the brand name is more important than the region or country of origin? BWS lends itself to a wide range of comparisons. In this special issue, for example Remaud and Lockshin (2009) compare the salience of different positioning statements for a specific wine region. Such questions could be answered using BW methodology by using a consumer survey based on BW choice sets. Studies in different cultures can be easily compared as there is no bias in the choice as there is only one way to choose something as "best" and something as "worst".

There are several limitations to the method. First, it becomes complicated to analyze many attributes in a single survey. A full ranking of each individual becomes a difficult task and using only part of the items to different respondents limits the possibility of segmentation, although recent advances in latent clustering provide a degree of analysis with partial designs (Vermunt and Magidson, 2005). Furthermore, even with a relatively small number of items (10-15), respondents' perception is that the task is boring as there are many repeated items across all choice sets. The perception of respondents using the BW method is that the researcher is trying to judge their cognitive processes and whether they are choosing consistently the best and worst items. However, our experience suggests that it is relatively easy for respondents to answer 15 choice sets or less, using a paper questionnaire.

There are many advantages to BW scaling which can be beneficial in wine marketing research. As the number of items increase it is recommended to use an

online survey that limits the errors made by respondents while using a paper survey. This paper provides guidance and examples to the design and analysis of surveys using the BW method.

References

Angulo, A.M., Gil, J.M., Gracia, A. and Sanchez, M. (2000), "Hedonic prices for Spanish red quality wine", *British Food Journal*, Vol. 2 No. 7, pp. 481-93.

Auger, P., Devinney, T.M. and Louviere, J.J. (2004), "Consumer's social beliefs, an international investigation using best worst scaling methodology", working paper, University of Melbourne, Melbourne Business School, Melbourne.

Auger, P., Devinney, T.M. and Louviere, J.J. (2007), "Using best worst scaling methodology to investigate consumer ethical beliefs across countries", *Journal of Business Ethics*, Vol. 70, pp. 299-326.

Bachman, J.G. and O'Malley, P.M. (1984), "Yea-saying, nay-saying, and going to extremes: Black–White differences in response style", *The Public Opinion Quarterly*, Vol. 48 No. 2, pp. 491-509.

Batt, P.J. and Dean, A. (2000), "Factors influencing the consumer's decision", *Australia and New Zealand Wine Industry Journal Marketing Supplement*, Vol. 15 No. 4, pp. 34-41.

Baumgartner, H. and Steenkamp, J.-B.E.M. (2001), "Response styles in marketing research: a cross national investigation", *Journal of Marketing Research*, Vol. 38 No. 2, pp. 143-56.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York, NY.

Chen, C., Lee, S. and Stevenson, H.W. (1995), "Response style and cross-cultural comparison of rating scales among East Asian and North American students", *Psychological Science*, Vol. 6 No. 3, pp. 170-5.

Cohen, S.H. (2003), "Maximum difference scaling: improved measures of importance and preference for segmentation", *Sawtooth Software Conference Proceedings, Sequim, WA*.

Cohen, S.H. and Markowitz, P. (2002), "Renewing market segmentation: some new tools to correct old problems", *ESOMAR*, September. *ESOMAR 2002 Congress Proceedings, ESOMAR: Amsterdam*, pp. 595-612

Cohen, S.H. and Neira, L. (2003), "Measuring preference for product benefits across countries: overcoming scale usage bias with maximum difference scaling", *ESOMAR 2003 Latin America Conference Proceedings, Amsterdam*.

Cohen, S. and Orme, B. (2004), "What's your preference?", *Marketing Research*, Vol. 16, pp. 32-7.

Cohen, E., d'Hauteville, F. and Sirieix, L. (2009), "A cross cultural comparison of choice criteris for wine in restaurants", *International Journal of Wine Business Research*, Vol. 21 No. 1, pp. 50-63.

Combris, P., Lecocq, S. and Visser, M. (1997), "Estimation of a Hedonic price equation for Bordeaux wine: does quality matter?", *Economic Journal*, Vol. 107, pp. 390-402.

Combris, P., Lecocq, S. and Visser, M. (2000), "Estimation of a hedonic price equation for Burgundy wine", *Applied Economics*, Vol. 32 No. 8, pp. 961-7.

Couch, A. and Keniston, K. (1960), "Yeasayers and naysayers: agreeing response set as a personality variable", *Journal of Abnormal and Social Psychology*, Vol. 60, March, pp. 151-74.

Crask, M.R. and Fox, R.J. (1987), "An exploration of the interval properties of three commonly used marketing research studies: a magnitude estimation approach", *Journal of the Marketing Research Society*, Vol. 29 No. 3, pp. 317-39.

Dolnicar, S. and Grün, B. (2007), "Cross-cultural differences in survey response patterns", *International Marketing Review*, Vol. 24 No. 2, pp. 127-43.

Felzensztein, C. and Dinnie, K. (2005), "The effects of country of origin on UK consumers' perceptions of imported wines", *Journal of Food Products Marketing*, Vol. 11 No. 4, pp. 109-17.

Felzensztein, C., Hibbert, S. and Vong, G. (2004), "Is the country of origin the fifth element in the marketing mix of imported wine? A critical review of the literature", *Journal of Food Products Marketing*, Vol. 10 No. 4, pp. 73-84.

Finn, A. and Louviere, J.J. (1992), "Determining the appropriate response to evidence of public concerns: the case of food safety", *Journal of Public Policy and Marketing*, Vol. 11 No. 1, pp. 12-25.

Flynn, T.N., Louviere, J., Peters, T.J. and Coast, J. (2007), "Best worst scaling: what it can do for health care research and how to do it", *Journal of Health Economics*, Vol. 26, pp. 171-89.

Gluckman, R.L. (1990), "A consumer approach to branded wines", *International Journal of Wine Marketing*, Vol. 2 No. 1, pp. 27-46.

Goodman, S. (2009), "A multi-country comparison of consumer choice for wine in retail stores", *International Journal of Wine Business Research*, Vol. 21 No 1,

Goodman, S., Lockshin, L. and Cohen, E. (2005), "Best worst scaling: a simple method to determine drinks and wine style preferences", paper presented at the 2nd Annual International Wine Marketing Symposium, Sonoma State University, Rohnert Park, CA, 8-9 July.

Goodman, S., Lockshin, L. and Cohen, E. (2008), "Examining market segments and influencers of choice for wine using the best worst choice method", *Revenue Internationale des Sciences Sociales – Special Wine Marketing Edition*, Vol. 8 No. 1, pp. 94-112.

Gordon, W. (2002), "Minding your brand manners", *Marketing Management*, Vol. 11 No. 5, September/October, pp. 18-20.

Hall, J. and Lockshin, L. (2000), "Using means-end chains for analysing occasions – not buyers", *Australasian Marketing Journal*, Vol. 8, pp. 45-54.

Hein, K.A., Jaeger, S.R., Carr, B.T. and Delahunty, C.M. (2008), "Comparison of five common acceptance and preference methods", *Food Quality and Preferences*, Vol. 19, pp. 651-61.

Jaeger, S.R., Jørgensen, A.S., Aaslyng, M.D. and Bredie, W.L.P. (2008), "Best-worst scaling: an introduction and initial comparison with monadic rating for preference elicitation with food products", *Food Quality and Preferences*, Vol. 19, pp. 579-588.

Koewn, C. and Casey, M. (1995), "Purchasing behaviour in the Northern Ireland wine market", *British Food Journal*, Vol. 97 No. 11, pp. 17-20.

Lee, J.A., Soutar, G. and Louviere, J. (2008), "The best worst scaling approach: an alternative to Schwartz's values survey", *Journal of Personality Assessment*, Vol. 90 No. 4, pp. 335-47.

Lockshin, L. and Hall, J. (2003), "Consumer purchasing behaviour for wine: what we know and where we are going", paper presented at the International Wine Marketing Colloquium, Adelaide, July, CD-ROM, available at: www.unisa.edu.au/winemarketing/conferences/proceedings.asp

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000), *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge.

Louviere, J.J. and Woodworth, G.G. (1983), "Design and Analysis of simulated consumer choice allocation experiments: an approach based on aggregate data", *Journal of Marketing Research*, Vol. 20, pp. 350-67.

Louviere, J.J. and Woodworth, G.G. (1990), "Best worst scaling: a model for largest difference judgments", working paper, Faculty of Business, University of Alberta, Edmonton.

Lusk, J.L. and Briggeman, B. (2009), "Food values", *American Journal of Agricultural Economics*, Vol. 91 No. 1, pp. 1-13.

Marley, A.A.J. and Louviere, J.J. (2005), "Some probabilistic models of best, worst, and best worst choices", *Journal of Mathematical Psychology, Journal of Mathematical Psychology*, Vol. 49, pp. 464-80.

Mitchell, V.W. and Greatorex, M. (1989), "Risk reducing strategies used in the purchase of wine in the UK", *European Journal of Marketing*, Vol. 23 No. 9, pp. 31-46.

Mueller, S. and Rungie, C. (2009), "Is there more information in best worst choice data? Using the attitude heterogeneity structure to identify consumer segments", *International Journal of Wine Business Research*, Vol. 21 No. 1, pp. 24-40.

Orth, U.R., Wolf-McGarry, M. and Dodd, T.H. (2005), "Dimensions of wine region equity and their impact on consumer preferences", *Journal of Product & Brand Management*, Vol. 14 No. 2, pp. 88-97.

Perrouty, J.P., d'Hauteville, F. and Lockshin, L. (2006), "The influence of wine attributes on region of origin equity: an analysis of the moderating effect of consumer's perceived expertise", *Agribusiness*, Vol. 22 No. 3, pp. 323-41.

Raghavarao, D. and Padgett, L.V. (2005), *Block Design: Analysis, Combinatorics and Applications*, World Scientific Publishing, NJ.

Rasmussen, M. and Lockshin, L. (1999), "Wine choice behaviour: preliminary research on the effect of regional branding", *The Australia and New Zealand Wine Industry Journal*, Vol. 11 No. 1, pp. 36-46.

Remaud, H. and Lockshin, L. (2009), "Building brand salience for commodity-based wine regions", *International Journal of Wine Business Research*, Vol. 21 No. 1, pp. 79-92.

Rungie, C. (2008), personal communication.

Skuras, D. and Vakrou, A. (2002), "Consumer's willingness to pay for origin labeled wine: a Greek case study", *British Food Journal*, Vol. 104 No. 11, pp. 898-912.

Thompson, K.E. and Vourvachis, A. (1995), "Social and attitudinal influences on the intention to drink wine", *International Journal of Wine Marketing*, Vol. 7 No. 2, pp. 35-45.

Thurstone, L.L. (1927), "A law comparative judgment", *Psychological Review*, Vol. 34, pp. 273-86.

Vermunt, J.K. and Magidson, J. (2005), *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Statistical Innovations Inc., Belmont MA.

Weller, S.W. and Romney, A.K. (1988), *Systematic Data Collection*, Sage, Newbury Park, CA.

Yeh, L.L., Kim, K.O., Chompreeda, P., Rimkeeree, H., Yau, N.J.N. and Lundahl, D.S. (1998), "Comparison in use of the 9-point hedonic scale between Americans, Chinese, Koreans, and Thai", *Food Quality and Preferences*, Vol. 9 No. 6, pp. 413-19.

Further reading

Cunningham, W.H., Cunningham, I.C.M. and Green, R.T. (1977), "The ipsative process to reduce response set bias", *The Public Opinion Quarterly*, Vol. 41 No. 3, pp. 379-84.

Louviere, J.J. and Islam, T. (2004), "A comparison of importance weights/measures derived from choice-based conjoint, constant sum scales and best worst scaling", CenSoC working paper 04-003, September.

About the author

Eli Cohen is Senior Lecturer at Ben Gurion University of the Negev in Israel, where he is teaching Food and Beverage Management, Food Science and Quality Management. He is also Adjunct Senior Lecturer at the School of Marketing, The University of South Australia. His research interests cover restaurant practice and menu analysis, consumer choice of wine in retail stores and in restaurants and wine tourism. Eli Cohen can be contacted at: elico@bgu.ac.il; eli.cohen@unisa.edu.au