Overcoming Challenges and Improvements in Best-Worst Elicitation: Determining What Matters to Japanese Wheat Millers

Paul F. Burke[+], Jordan Louviere[+], Edward Wei[+], Gordon MacAulay[++], Ken Quail[++], and Richard Carson[+]

[+] Centre for the Study of Choice (CenSoC), University of Technology, Sydney
[++] Grain Growers Ltd, North Ryde.

Word count: 7929 words

Abstract

Knowing how to produce what types of wheat with what characteristics and in what quantities is a key challenge for producer countries like Australia to successfully export wheat to various markets that consume it. Both producers and consumers would benefit by better matching what is produced to what the market(s) prefer and are willing to pay to have produced. Analysis of decision-maker choices is difficult as there are only a small number of millers in any one country that make buying decisions. Moreover, the buyers tend to use an extensive list of quality characteristics to inform their purchases. This research provides details of some of the insights that have been gained into this decision making context using best-worst scaling (BWS), as a choice-based measurement and modelling approach. The survey instrument was administered using CAPI in personal interviews with Japanese flour millers.

A small number of flour millers in Japan supply the entire government regulated market with products like Udon and Ramen noodles. Face-to-face interviews were conducted with 14 individuals from four different companies that account for about 74 per cent of wheat flour production in Japan. These individuals play various roles in wheat buying, production, distribution and marketing, such as production managers, quality control specialists and new product and scientific development managers. Based on the literature and pilot discussions with wheat buyers, a list was compiled of 31 factors (attributes) that could be considered by the individuals who influence wheat buying decisions. These included technical attributes (e.g., viscograph peak height; farinograph dough stability, etc) as well as attributes common in most business-to-business trade settings that are often cited as important in many agricultural trade contexts (e.g., price; country of origin; uniformity of shipment, etc).

Because there was limited prior research to guide the choice of factors (attributes) to design a survey to measure and model wheat choices, a BWS (case 1 or object case) choice task was designed and implemented to prioritise the 31 factors. A customised computer-assisted personal interview (CAPI) platform programmed in Adobe Flex was used. The Flex language and Actionscript was compiled to run under Adobe Air on a desktop computer. Descriptions of each factor (a glossary) and all survey questions were translated into Japanese (with Katakana and Hiragana true type font characters), and then back-translated into English to ensure that they were correctly translated from English to Japanese. Factors were assigned to choice sets using a balanced incomplete block design (BIBD), such that each appeared the same number of times and co-appeared with every other factor the same number of times. The CAPI survey had each of the 31 choice sets of six factors one-at-a-time, and survey respondents chose, respectively the most and least important factor in each set.

Operationally, each respondent evaluated all 31 BWS choice sets. The CAPI program shaded out chosen options so that respondents could make subsequent best-worst choices from a smaller list. Responses were recorded in real time to a local CSV file and each individual respondents' BWS scores were matched against each of their three 'top-of-mind' important characteristics (from the 31), in real time, to cross-check the accuracy of the BWS task and enhance respondent engagement. The cross-check also provided opportunities for researchers to immediately raise questions with respondents if there were inaccuracies.

The interviewers received generally favorable comments about the survey from the respondents, such as its uniqueness and the thoughtfulness required to answer the questions. The BWS results provided a ratio scale of importance, with each factor located at some point on that scale. The result was then used to design and implement a subsequent discrete choice experiment to model the choices of each flour miller participant. In this application BWS proved to be a valuable first stage in developing reliable and accurate models of miller's preferences and trade-offs between attributes. With an improved knowledge of the decision maker's priorities, plant breeding and planting choices can be better aligned to miller's preferences for varieties and types of wheat, thus enhancing the international competitiveness of Australian wheat.

Overcoming Challenges and Improvements in Best-Worst Elicitation: Determining What Matters to Japanese Wheat Millers

Introduction

Many different ways to identify and prioritise potential attributes for use in discrete choice experiments (DCEs) and surveys have been proposed and applied, such as Louviere's (1988) six question approach, or Timmermans' use of Kelly's Personal Construct Theory (see, Timmermans et al. 1982a;1982b). If properly designed, DCEs will reveal systematic relationships between the choices among the options and the variation in the levels of each attribute, thereby providing insights into the theoretically latent underlying preferences. In essence, changes in factors that matter more to each individual or segments in the sample or the sample as a whole should be associated with greater changes in choice. There are many applications of this DCE approach in several literatures, such as marketing, transport and environmental economics (e.g., Louviere, Hensher and Swait 2000; Auger et al. 2003; Verma, Louviere and Burke 2006).

In many applications deciding how many and which factors (attributes) to include in a choice study or experiment can be difficult. Unfortunately, however, omission of important factors can have dire consequences both statistically and substantively. Statistically, omitted variable bias is an obvious consequence. Thus, there is a trade-off between adding one or more factors with potentially greater task complexity due to more attributes to process and/or the requirements of experimental designs that may result in many more choice sets. The latter can occur despite the availability of sophisticated design technologies, such as alternative-specific and optimal designs (e.g., Louviere and Woodworth 1983; Kanninen 2002; Burgess and Street 2003; Street and Burgess 2007). From a modelling perspective, each new factor is associated with new interactions (typically unobserved) and associated parameters. The variation explained by them is often assumed to be negligible and literature dating back to the 1980s shows that incorrect additive models will almost always appear to fit choice data well (Dawes and Corrigan 1974; Louviere, Hensher and Swait 2000; Street and Burgess 2007).

From an experimental bias perspective, adding more factors to describe options that appear in a DCE can impact the complexity of the experimental task (Swait and Adamowicz 2001; DeShazo and Ferno 2002). In turn, this may induce behavioural outcomes, such that experimental participants may use heuristics and make spontaneous inferences to manage the added complexity (e.g., Kardes et al. 2004). Alternatively, systematically varying which attributes are included or excluded, can have some effects on preferences, but can also impact the consistency and randomness with which choices are made (e.g., Louviere, Islam and Burke, 2007).

Researchers use several methods to identify and choose attributes to include in DCEs, including qualitative research, expert judgement, and survey questions about attribute importance (typically using rating scales). Researchers also review previous academic or applied research or trade publications to determine attributes that appear to be suitable to include or exclude. Louviere, Hensher and Swait (2000) encourage and warn researchers to spend as much time as possible in advance to take advantage of "…theory, thinking, observation and just plain hard, empirical detective work" (p.122). Louviere (1988) used a six question open-ended qualitative approach for attribute identification and prioritisation to be then used to design and implement DCEs. This approach resulted in a consistent and systematic relationship between model predictions and real market choices (Louviere, Hensher and Swait, 2000, Chapter 13). Overall, however, there appears to be limited consensus or even a systematically consistent approach to attribute identification that can be observed in various literatures that use DCEs. Thus, it seems fair to say that this remains an open and important research question.

The purpose of this paper is to describe and discuss how to use BWS to reduce a potentially large list of relevant attributes to the ones that appear to matter most to the relevant choosers. Best-Worst Scaling (BWS) has been used by some researchers to reduce the number of factors to include in DCEs. For example, Burke et al. (2010) used a combination of qualitative research and BWS to reduce a list of 64 factors to a final list of nine overarching factors. These were then used in a DCE to study how variation in each overarching factor (which were modelled using 26 more specific set of attributes to describe them) affects museum choices. The context of the present application is prioritising the factors that matter most to millers when making the decision to purchase wheat from an imported source. A case study is described using BWS as an initial stage prior to designing and implementing several DCEs for a sample of Japanese and Korean millers. This paper is focussed on the results for a sample of Japanese millers already completed; the Korean work is currently underway. Also described and discussed are various challenges faced in this project, such as small sample sizes, typical of B2B applications, and a novel approach to eliciting data (CAPI). Empirically, the research reported below also provides insights into wheat characteristics and industry service factors that matter to Japanese wheat millers in their decisions to buy wheat and how important each characteristics or factor is in these decisions.

A Brief Overview of Best-Worst Scaling

BWS was first introduced by Jordan Louviere in a special session of the 1988 American Marketing Association's Advanced Research Techniques Forum (ART Forum) where Louviere demonstrated its use to an audience of research practitioners. This was followed by a working paper by Louviere and Woodworth (1990), and an application to measure citizen concerns about food safety in Alberta, Canada, by Finn and Louviere (1992). Since that time, BWS has been widely adopted by the marketing research community globally (2.9M Google hits by 28/03/13 for maximum difference scaling plus 2.2M hits for best-worst scaling). It has seen growing use by academics in several fields, such as marketing (e.g., Auger, Devinney and Louviere 2007; Louviere and Islam 2008; Jaeger et al. 2008; Cohen 2009), business logistics (Coltman et al. 2011), health economics (e.g., Flynn et al. 2007; Lancsar et al. 2007; Burge et al. 2011), retention of early career school teachers (Burke et al. 2013) and values and personality in social psychology (e.g., Lee et al. 2007; Lee et al. 2008), to name a few areas.

BWS is a relatively straightforward response elicitation method that asks people to choose the two items from a listing of several items that most and least match a given criterion. When Louviere first introduced BWS, he called it 'maximum difference scaling' (e.g., Cohen 2003), as the items chosen best and worst are the two items that are furthest apart on an assumed latent scale that underlies a respondents' preferences (Finn and Louviere 1992). This nomenclature has been widely adopted by research practitioners, especially in North America, but empirical tests over the past 20 plus years have consistently and repeatedly indicated that few people actually make these choices in a manner consistent with what a maximum difference choice process requires. Thus, Louviere and others changed the name of the elicitation procedure to "best-worst scaling". In a typical case 1 application, survey respondents are presented with several sets of items to be measured (i.e., “scaled”) and are asked to indicate the best and worst items in each set.

BWS case 1 typically can be used in almost every application in which rating scales have been widely used in the past; hence, the potential range of applications is very large. For example, Burge et al. (2011) asked respondents to nominate which aspect of social care (e.g., safety; cleanliness; employment) would impact their quality of life the most and least following a debilitating accident. Respondents evaluated lists of sentences that were both positive (e.g., “My home is as clean and comfortable as I want”) and negative (“Sometimes I don’t feel safe enough”) and simply chose the two sentences that, respectively, represented the most and least positive impacts on their quality of life.

Like traditional DCEs, BWS requires respondents to make trade-offs rather than evaluate the factors one-at-a-time in isolation. Thus, BWS provides significant advantages for prioritising factors compared with many traditional survey question formats that ask respondents to consider items one-at-a-time (e.g., rate each factor on a Likert scale ranging from low importance to high importance, as discussed by Louviere and Islam 2008). This use of rating scales is well-known to be susceptible to various types of biases due to differences in response styles. That is, different people can use the same numbers on a rating scale to indicate different things (Baumgartner and Steenkamp 2001). It is also cognitively attractive because respondents do not have to consider numbers or phrases to respond, nor do they need to decide how to allocate points or percentages or rank a lengthy list of items simultaneously (Louviere and Islam 2008).

Indeed, as Louviere and Woodworth (1983) noted, a properly designed choice experiment yields choice counts (frequencies of choices) that contain all the statistical information for estimating various choice models. This is not surprising, in so far as McFadden (1981) noted that choice data can be represented as a contingency (or crosstab) table that is sparse and incomplete. The use of and properties of counts in the statistical analysis of contingency table data is now well-established (e.g., Bishop, Fienberg and Holland 1975), and translates directly to DCEs, of which BWS case 1 is one type. Thus, simply counting the number of times each item is chosen best and worst allows for the calculation of “how much” a latent quantity of interest (e.g., factor importance) is associated with each choice option (item, sentence, thing, etc.). This simple method is robust if the DCE is properly designed and allows researchers to measure each item with minimal computational effort. As noted by Louviere (2013 forthcoming), simply calculating the square root of the ratio of most counts divided by least counts yields a measure of factor importance on a ratio scale (see also, Lancsar et al. 2007). Alternatively, it is possible to subtract least counts from best counts to obtain a difference scale centred at zero, where zero means an item that received as many best as worst counts. The formal mathematical properties of various BWS-related calculations are discussed by Marley and Louviere (2005).

At the last ICMC Jordan Louviere’s keynote address (2013, forthcoming) focused on using BWS elicitation tasks as a way to model individuals, and integrated this with Louviere et al. (2008b) to demonstrate that one could use BWS as a way to obtain more ranking information per choice set for each survey respondent. That is, it is possible to ask experimental participants to make a second choice in addition to the traditional “best” choice in DCEs, namely the choice of their least preferred option. In BWS, this is sometimes referred to as case 3 or ‘multi-profile case’ (see Marley and Pihlens 2012). As discussed by Louviere, et al (2008b) asking this second choice question provides either a complete or partial ranking of the choice options in each choice set, which in turn allows for expanding the choice data for each person to create more (pseudo) choice observations for estimation purposes, enhancing estimation efficiency, and in some cases, permitting identification and convergence for single individuals (see also, Luces and Suppes 1965; Chapman and Staelin 1982). More recently, Collins and Rose (2011) use this elicitation format in DCEs to compare competing models of the choice process, such as sequential choices of the next two best options. Our focus in this paper is on the use of case 1 (object case) BWS (Marley and Louviere 2005) as a way to measure the relative importance of factors that may underlie choices in a particular domain of interest, in this paper the context of interest being the buying of wheat. So, the choice alternatives in this BWS case 1 application are attributes that can potentially drive flour miller wheat choices in Japan.

Background to Empirical Setting: The Value of Australian Wheat by Japanese Millers

As previously noted, the focus of this paper is on quantifying the importance of factors for flour millers in Japan who buy imported wheat to make value-added products like Udon and Ramen noodles. Regarding Australian wheat, much of what Japan imports is Australian soft wheat, largely white, with a low moisture content level, making it suitable for many end-user applications, particularly Udon noodles. Japanese demand for food wheat has been found to be highly elastic across varieties and country of origin: Australian soft wheat competes with US hard and soft wheat, which appear to be substitutes in the Japanese domestic market (Koo, Mao and Sakurai 2001). Substitution of US and Australian wheat also has been observed in other markets, such as South Korea, where cross-price elasticity results suggest US and Australian wheat act as close substitutes (Park 2010).

The decision of individual millers to purchase wheat is a complex one, potentially involving many factors complicated by differences in market regulations (Koo, Mao and Sakurai 2001). Wheat is a commodity with multiple characteristics, including dimensions like colour (includes whiteness and stability), texture (includes firmness and elasticity), and protein content. Unfortunately, the end-use value of these factors to millers who make purchase decisions is poorly understood. Specifically, while there are some obvious factors that millers use to value wheat (e.g., price), the relative value of the factors and how millers trade them off in evaluating competing wheat options is poorly understood. Researchers have largely concentrated on using import data to examine substitutions across different classes of wheat (e.g., soft, hard) and country of origin (e.g., Henning 1986; Alston et al. 1990, Koo, Mao and Sakurai 2001) rather than directly studying individual miller decisions. Further limitations arise due to most studies only considering a very small number of characteristics in any one study relative to a larger number potentially used by millers to make buying decisions.

A key research challenge in studying wheat buying decisions is that only a small number of milling companies in Japan import wheat and only a few employees are involved in recommending and purchasing this commodity (Lee, Lerohl and Unterschultz 2000). Moreover, to date few researchers have had access to data on decisions that can be directly linked to an individual employee of a milling company. Instead, researchers have largely concerned themselves with trying to model market effects and inferring the decision rules from the statistical results (e.g., Ahmadi-Esfahani and Stanmore 1994; Stiegert and Blanc 1997; Wilson 1989). For example, Wilson (1989) used a hedonic pricing model to conclude that data from many markets indicate millers are willing-to-pay a significant premium for wheat with higher protein levels, with less preference for hard over soft wheat. These types of models require high market volatilities over time and space to estimate factor effects while controlling correctly for autocorrelation due to the temporal nature of wheat supply and demand (Wilson 1989).

Very few studies could be found that tried to directly study the decisions of millers. One example is Lee, Lerohl and Unterschultz (2000), who studied American miller choices using a DCE. They found that bushel weight, price, country of origin and amylase (inversely related to falling number) were significant predictors in the demand for Canadian durum, a type of wheat that is mainly used in pasta production. Impacts of other characteristics (e.g., protein and grade) were sign-consistent with prior work, but insignificant. Gallardo et al. (2009) used a traditional conjoint analysis approach to study decisions about hard red winter wheat; they found that Mexican millers were willing to pay significantly more for quality characteristics as indicated by test weight, protein content, falling number and dough strength/extensibility. Kim (2000) also used conjoint analysis to study preferences of Japanese and Korean millers for various characteristics of wheat and flour used in noodle making. He studied six subsets of choices comprising two types of quality characteristics (intrinsic and purchase contract) by three wheat types (hard, semi-hard and medium). Intrinsic wheat quality characteristics were ash content, falling number, test weight and price; the purchase contracts included protein, country of origin, dockage and price.

There has been little recent research on how Japanese buyers evaluate Australian wheat, and most academic papers appeared in the 1990s (e.g., Ahmadi-Esfahani and Stanmore 1994; Ahmadi-Esfahani and Stanmore 1997; Stiegert and Blanc 1997). Unfortunately, since that time, many changes have occurred in the global marketplace (e.g., government regulation; changes in market power, and internationalisation of the handling and storage system), along with changes in competitive export quality and changes in consumers tastes and demands (Koo, Mao and Sakurai 2001). The export market is susceptible to environmental and weather effects with significant falls in the market share of Australian wheat during the drought period of 2006–2007 (Park 2010). Reliable and accurate models of miller trade-offs and preferences can lead to plant breeding and planting insights that better align to miller preferences for the varieties and types of wheat produced, thus enhancing export competitiveness. In fact, this knowledge can have a non-trivial impact on the Australian economy as wheat is the most important export crop, as well as being second-most important export commodity, representing $5.3B in the year ending June 2008. In 2007–08, just over half of Australia's wheat was exported (ABS 2011, p. 493). Thus, there is an opportunity to use DCEs and associated choice models to gain insights into how millers in Japan and elsewhere choose wheat import options. Towards this end, we proposed and applied BWS as a way to prioritise characteristics that millers may evaluate and trade off in choosing wheat import options.

Research Approach

Sampling Strategy and Sample

The research objective of this study was to gain insights into decision rules used by agents involved in purchases of Australian wheat in Japan. Access to Japanese milling companies was provided via personal email and phone contact to known associates by our partner researchers in Grain Growers Limited (see acknowledgements), and the timing of each visit was facilitated by representatives from Austrade. Grain Growers is the largest grains industry organisation in Australia, with a mission to promote the development of a sustainable, viable and efficient Australian grains industry. It undertakes scientific research on various aspects of grain growing and processing, and provides training, technical services and other information to the grains industry. It has over 17,000 Australian members.

The Japanese milling companies were asked to host a visit by two to three representatives of Grain Growers to hold discussions (common in interactions between Grain Growers and wheat buyers). The visiting party also included a member from the Centre for the Study of Choice, and a translator, provided by Austrade, to facilitate the technical aspects of a face-to-face computer assisted interview. Grain Growers requested an interview with as many employees involved in the buying decision for wheat noodle from Australia or other international suppliers as possible. Host companies and employees received no incentives to participate other than an expectation that the research findings would be discussed in subsequent visits and approved for dissemination throughout their own organisation.

Four of Japan's largest milling companies were visited; these companies account for around 74 per cent of the flour milling production in Japan. In each of the four meetings, three to five representatives from each company attended discussions and two to four completed the CAPI survey. In total, 14 interviews were conducted. An additional three responses were obtained online by employees absent at the time of data collection, bringing the total sample size to 17 employees. The individuals interviewed covered various roles in the decision making process, including managers and laboratory scientists. These roles were concerned with a mixture of issues relating to quality control, new product development and scientific work.

Developing the list of factors

The factors included in the BWS study were obtained from studying wheat shipment reports, previous literature, including specific reports published on the decision rules of Japanese buyers (e.g., Lee, Koo and Krause 1994) as well as prior choice modelling work on non-Japanese buyers (e.g., Lee, Lerohl and Unterschultz 2000) and a general understanding of the industry. This eventuated in a final list of 31 factors to be prioritised. They ranged from very technical factors (e.g., viscograph peak height; farinograph dough stability) to factors common to many business-to-business trade settings (e.g., price; country of origin; uniformity of shipment) that often are cited as important in many agricultural settings. As previously noted, the focus was on decisions by Japanese wheat buyers about imported wheat used to produce Udon noodles. Wheats used to produce Udon noodles tend to be creamier in colour with a soft but elastic texture, which differ from other noodle products, such as Ramen, Hokkien or instant noodles (Penfold 2010).

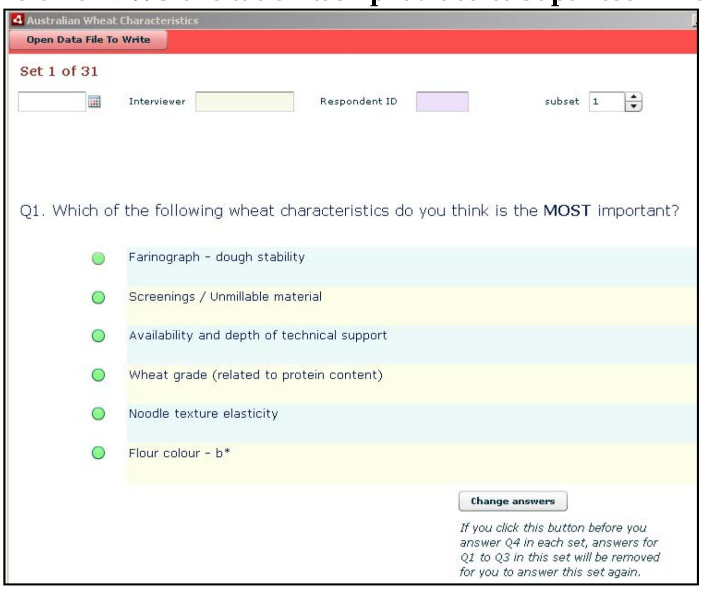
Experimental Design

The 31 factors were assigned to choice sets (comparison sets) using a Balanced Incomplete Block Design (BIBD). As noted by Louviere (2013 forthcoming) and Louviere et al. (2013, forthcoming), BIBDs are an obvious type of experimental design for case 1 BWS applications due to the fact that they produce: a) constant set sizes; b) each factor occurs across all choice sets the same number of times; and c) each factor co-occurs with every other factor the same number of times across all sets (Raghavarao and Padgett 2005; Street and Street 1987). In turn, "best and worst" elicitation tasks provide much more rank order information about choice options than only asking for "best. The particular type of BIBD that was selected for this application, known as a Youden design, has the benefit of controlling for order. That is, each factor appears in each of the six left-to-right positions across the choice sets (Raghavarao and Padgett 2005). BIBDs can be viewed as a systematic way of sampling from all possible ways of combining the 31 factors that produces sensible elicitation tasks that are consistent with random utility theory-based choice models. These properties are especially germane to this application because the total number of available respondents that can be sampled in this market is small. Thus, it is important to maximise the amount of choice data available, which led to asking each respondent to complete all 31 sets. This number of choice sets may seem large relative to papers published in the past 10-15 years in the choice modelling literature where many authors have limited the number of choices due to concerns about task complexity and respondent burden. The experience in CenSoC has been rather different, where it has been possible to routinely administer as many as 32 choice sets in case 3 BWS (traditional DCEs) applications. As noted by Louviere et al. (2011), the only impact of additional choice sets seems to be an increase in choice inconsistency (error variability), but as noted in that paper, this is drastically offset by the additional information that is obtained from the extra choice sets.

Elicitation Task and CAPI Interface Development

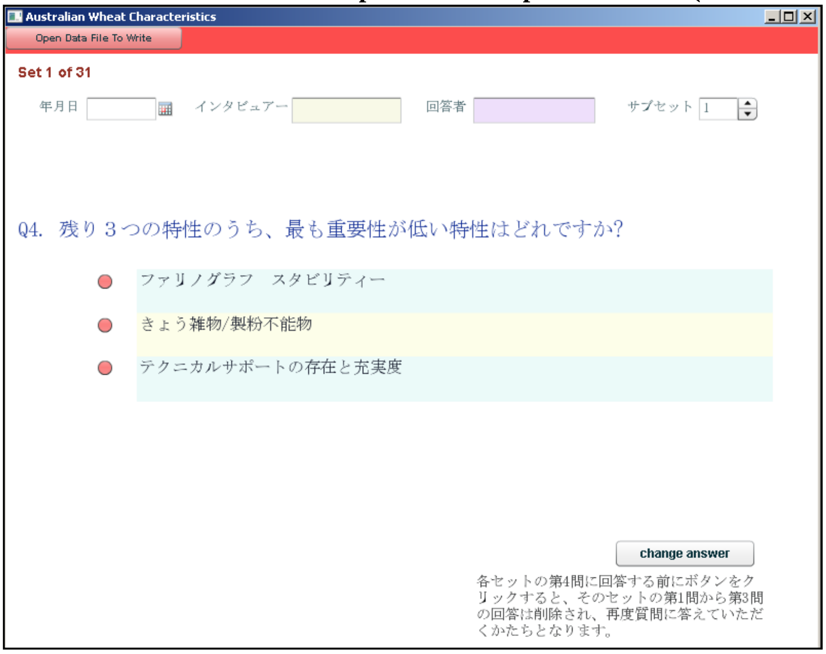
Because members of the research team are not fluent Japanese speakers, it was important to design and implement an elicitation process that would guide respondents through the interview process with minimal reliance on an independent translator. The translator was provided by Austrade to assist the research team. It was also assumed that there would be no internet access, precluding use of commercial survey software to field the survey. As much as possible the elicitation process was automated, particularly with processes designed to make the task easier, such as shading out already chosen items so that respondent could focus only on the remaining items. These objectives were accomplished by developing a customised computer assisted personal interview (CAPI) platform using Adobe Flex (using the Flex language and Actionscript). The software was compiled to run under Adobe Air on a laptop computer. The resulting survey platform showed each participant the 31 comparison sets that each contained six factors; descriptions of each factor were given in Japanese (with Katakana and Hiragana true type font characters). An independent commercial translation company was responsible for forward and backward translation between English and Japanese. Each participant was sent the translated list of factors and descriptions (one week before the face-to-face interview) via email so that they would know in advance what would be asked. An example of the task (in English) is provided by Figure 1.

Figure 1: English version of BWS elicitation task provided to Japanese millers (most important)



As can be seen in Figure 1, respondents were asked to choose one most important wheat factor. Following that choice, the option chosen as most important was shaded out leaving five factors from which to choose the least important. When asked for the most important, the dots associated with each factor were green; when asked for the least important, the dots were red. Two rounds of most and least important choices were used, such that in round three the participant would see only the four remaining options, and was asked to choose the second-most important factor. Finally, participants were shown a fourth screen with the three remaining options associated with red dots, and they were asked to choose the second-least important factor from that set. One of the latter rounds of questions appears in Figure 2, which illustrates the Japanese version that participants experienced. After making these four choices, participants proceeded to the next set of six items to make four choices in the same manner previously described. Participants took about 12 minutes to complete the 31 comparison sets.

Figure 2: Japanese version BWS elicitation task provided to Japanese millers (second least important)



Results

Preliminary real-time responses

Each participant's choices were recorded for each participant immediately to a local comma-delimited file. The software provided an automated real-time analysis of each participant's "most" and "least" choices that were then matched against the three "top-of-mind" factors that each participant nominated as being the most important of the 31. This was done to give feedback to the participants and verify the accuracy of the BWS elicitation method. It enabled the research team to ask clarifying questions when inconsistencies occurred. The participants made favourable comments about the survey task and method of implementation, noting that it was a unique experience and required their complete attention to answer the questions.

Best-Worst Importance Scores

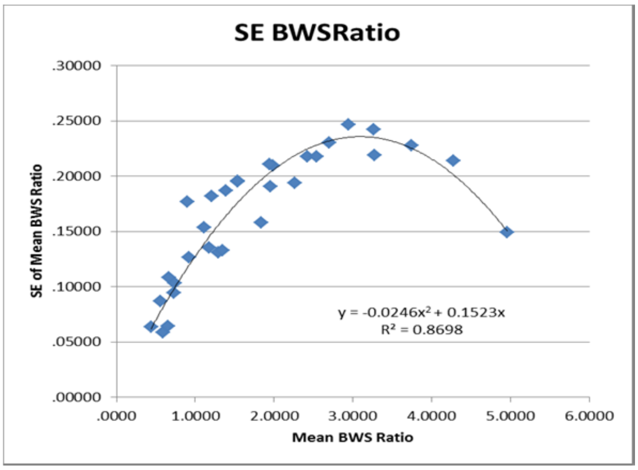
The case 1 BWS task yields measures of importance to wheat purchases for the 31 factors. The initial analysis was begun by simply calculating the ratio of the number of times a factor was chosen as "most" to the number of times it was chosen as "least" important in wheat purchases. Mean BWS scores and associated standard errors are in Table 1, sorted in order of most to least important. These provide a relative scale of "importance" and allow meaningful statements to be made about ratios or fractions of numbers. Thus "wheat grade" (23) is approximately 1/5 as important as "noodle texture elasticity" (1). Table 1 reveals that noodle elasticity was the most important factor, with the least important factor being availability and depth of technical support information associated with a purchase. The ratio of these two scores (4.9573/.4371) suggests that noodle texture elasticity is 11.3 times more important relative to technical support information.

Table 1: BWS ratio scores of importance of 31 wheat factors

Rank	Factor	Mean	Standard Error
1	Noodle texture elasticity	4.9573	.14892
2	Noodle texture firmness	4.2771	.21402
3	Protein content of wheat	3.7468	.22768
4	Noodle colour stability L*	3.2812	.21907
5	Uniformity of shipment	3.2653	.24239
6	Noodle colour stability b*	2.9465	.24681
7	Viscograph peak height	2.6973	.22999
8	Flour colour L*	2.5430	.21799
9	Wheat falling number	2.4243	.21776
10	Flour extraction rates	2.2656	.19372
11	Flour wet gluten	1.9876	.20969
12	Flour ash	1.9559	.19058
13	Price	1.9440	.21083
14	Flour colour a*	1.8350	.15806
15	Farinograph - dough stability	1.5420	.19512
16	Extensograph - maximum resistance BU	1.5360	.19540
17	Extensograph - extensibility cm	1.3849	.18691
18	Flour colour b*	1.3446	.13264
19	Test weight	1.2953	.13080
20	Data on specific parcel of wheat being shipped	1.2135	.18161
21	Farinograph - water absorption	1.1716	.13506
22	Moisture content of wheat	1.1165	.15341
23	Wheat grade	.9166	.12669
24	Country of origin	.8971	.17666
25	Availability and depth of quality assurance program	.7428	.10325
26	Screenings Unmillable material	.7314	.09423
27	Availability and depth of crop report	.6598	.10823
28	Grain hardness	.6498	.06416
29	Black point percentage	.5879	.05844
30	Grain colour	.5602	.08661
31	Availability and depth of technical support	.4371	.06371

The relationship between the BWS square root ratio importance measures and the associated standard errors are graphed in Figure 3. Figure 3 is not surprising in so far as the choice counts are multinomial outcomes that should exhibit an inverse-U shaped relationship with their corresponding means. So, Figure 3 simply confirms this expectation while showing that there was less variability in choices for importance measures at either extreme of the distribution.

Figure 3: Relationship between BWS mean and standard error



As discussed by Louviere, et al (2013, forthcoming), there are several ways to calculate case 1 (and other cases) BWS measures. A simple measure discussed by Marley and Louviere (2005) is the difference in best and worst counts (i.e., best counts minus worst counts). Figure 4 shows that this measure is monotonically related to the square root ratio measures, and indeed it is approximately logarithmically related, and consistent with expectations. That is, the best minus worst score differences are linearly related to estimates from conditional logit models, whereas the square root ratios are proportional to best choice counts. So, the BWS differences should be logarithmically related to the square root ratios, which Figure 4 indicates is approximately the case.

Figure 4: Relationship between best-worst ratio scores and best minus worst

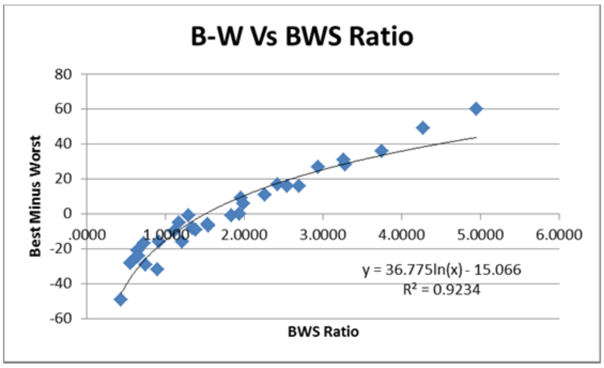
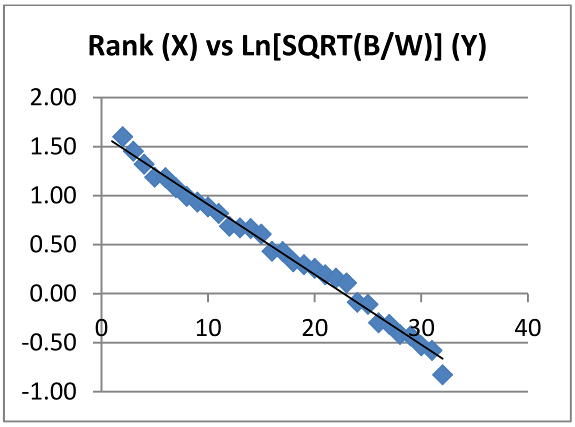


Figure 5 graphically displays the relationship between the rank position of each factor and the natural log of the square root of best/worst. The natural log is approximately linearly related to the rank positions of each factor. The regression results are shown in Table 2 below. The fit is very high ($r^2 = 0.99$), and the slope is very reliably estimated.

Figure 5: Relationship between importance ranking and Ln[SQRT(B/W)]

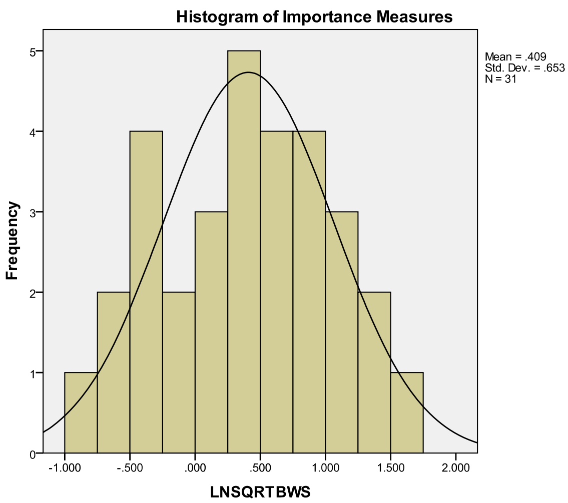


Histograms of the square root ratio measures are shown in Figure 6, along with means and standard deviations. The histogram suggests that differences of 2 x the standard deviation (approximately 1.3) are significant. Thus, the first ranked factor is significantly higher in importance than the third, by implication it is significantly more important than all lower ranked factors. The second ranked factor is significantly more important than the sixth ranked and all lower factors. The third ranked factor is significantly more important than the eight ranked factor and all below it, and so forth. In this way, it was possible to identify the 12 to 14 most important factors to use in the second (DCE) stage of the research (not reported here).

Table 2: Regression Results Relating Rank Order to Ln[SQRT(B/W)]

Effect	Estimate	StdErr	t-stat	P(t)
(Constant)	1.554	.024	65.257	.000
RankOrder	-.072	.001	-55.066	.000

Figure 6: Histogram of Importance Measures



Discussion and Conclusions

Case 1 (object case) of Best-Worst Scaling (BWS) was used to measure the importance of 31 factors that potentially could drive Japanese wheat purchases. The results will then be used as inputs to a second survey that was based on a discrete choice experiment (DCE) using the factors measured as most important by the process in this paper. It was found that Japanese millers who buy wheat to make Udon noodles care most about noodle texture elasticity and firmness. They also consider protein content (consistent with prior research, such as Stiegert & Blanc 1997; Saito, Saito, Kondo and Osanami 2009), colour stability (stability measured by L* and b* indicators) and differences in colour measures (i.e., differences in redness (a*), yellowness (b*) and whiteness (L*) of the flour or noodle), with flour colour important relative to noodle colour (i.e., the BWS measure for flour colour L* is 8th most important, with a* 14th and b* 18th). Shipment uniformity (whether the cargo received is consistent across containers and/or consistent across hatches within a ship) also is among the top five most important factors. Perhaps surprisingly for a commodity like wheat, price was not among the top 10 factors in importance, instead positioned in 13th place. This probably is due to the unique situation in Japan whereby the Japanese Food Agency determines import quotas, regulates purchases and determines prices (Koo, Mao and Sakurai 2001). In addition, the government subsidises domestic wheat production (Koo, Mao, Sakurai, 2001).

The case 1 BWS approach also allowed separation of the importance of country of origin relative to other factors. This is advantageous because market data often contains inherent correlations between country of origin and various underlying characteristics (Wilson 1989). For example, Australia predominantly exports white wheat with low moisture content, posing challenges for separating price variations associated with colour variations.

The least important choices suggest opportunities to change the content and manner in which information is provided to buyers. Specifically, supplementary technical information that provides parcel-specific data (ranked 20th), quality assurance programs (25th), crop reports (27th) and technical support information (31st) all featured as relatively unimportant factors out of the 31 considered. This poses interesting issues as to whether these factors actually are unimportant or reflect historic behaviour by exporters that could be changed. The current study does not address these issues, but does hint at interesting future lines of research.

Of course, the case 1 BWS approach has some limitations. For example, the task involved 31 sets of six factors and four questions (most, least, second most, second least) were asked of each participant. No attempt was made to investigate whether respondents became fatigued during the task, and it may be important to pursue this in future research. By way of contrast, many participants reported a high level of engagement with what they saw as a unique way to elicit factor importance; yet, they also reported that the trade-offs were cognitively challenging to maintain during the 10 to 15 minutes when they answered the questions. As noted earlier in this paper, Louviere et al. (2011) conducted a systematic study of the impact of number of choice sets (as well as other factors) on response rates and choice consistency, and found that the primary impacts were on choice consistency, but that losses in reliability due to inconsistency were far outweighed by gains in choice information from extra sets.

Acknowledgements

This project was funded by the Australian Research Council (ARC) in conjunction with Grain Growers Ltd., as part of a Linkage Project (LP120100493). The views and opinions expressed are those of the authors alone. The authors would like to express their thanks to all those who participated in the research, particularly the employees of the Japanese milling companies that participated in the research along with the management of these companies that allowed the researchers to conduct the interviews. The researchers also acknowledge the assistance of members of Austrade who assisted in arranging visits to the milling companies and further procurement of responses from employees. We would also like to thank Gail Bradford, Gnana Bharathy and other members of the Centre for Study of Choice (CenSoC) for assisting in the research.

References

Ahmadi-Esfahani, F.Z. and Stanmore, R.G. 1994. Quality premiums for Australian wheat in the growing Asian markets. *Australian Journal of Agricultural Economics*, *38*(3), 237-250.

Ahmadi-Esfahani, F., and R. Stanmore. 1997. Export Demand for Attributes of Australian Wheat in Asia and the Middle East. *Journal of Food Policy*, *22*(2):145~154.

Auger, Pat, Paul Burke, Timothy M. Devinney, and Jordan J. Louviere. 2003. What Will Consumers Pay for Social Product Features? *Journal of Business Ethics, 42* (3), 281-304.

Auger, Pat, Timothy M. Devinney, and J. Louviere. 2007. To What Extent Do Consumer Ethical Beliefs Differ Across Countries? A Cross-Country Investigation Using Best-Worst Scaling Methodology. *Journal of Business Ethics* 70 (3): 299-326.

Australian Bureau of Statistics (ABS). 2011. Year Book Australia, 2009-10, 1301.0.

Baumgartner, H., and J-B.E.M. Steenkamp. 2001. Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research* 38: 143-156.

Bishop, Y. Fienberg, S. and Holland, P. 1975. Discrete multivariate analysis, Cambridge, Mass: MIT Press.

Burgess, Leonie and Deborah J. Street. 2003. Optimal Designs for 2k Choice Experiments. *Communications in Statistics: Theory & Methods*, 32 (11), 2185-.

Burke, Paul F., Burton, C., Huybers, T., Islam, T., Louviere, J.J. & Wise, C. 2010. The Scale-Adjusted Latent Class Model: Application to Museum Visitation, *Tourism Analysis*, 15(2), 147-165.

Burke, Paul F., Schuck, S., Aubusson, P., Buchanan, J., Louviere, J.J., & Prescott, A. 2013. Why do early career teachers choose to remain in the profession? The use of best–worst scaling to quantify key factors, *International Journal of Educational Research,* in press.

Burge, Peter, Dimitris Potoglou, Terry Flynn, John E. Brazier, and Ann Netten. 2011. Best-worst scaling: Consistency of preferences with discrete choice experiments and stability over time. International Choice Modelling Conference, Leeds.

Chapman, R.G., & Staelin, R., 1982. Exploiting rank order choice set data within the stochastic utility model. *Journal of Marketing Research* 19 (3), 288-301.

Cohen, S.H. 2003. Maximum difference scaling: Improved measures of importance and preference for segmentation. *Sawtooth Research Paper Series*, Sequim, WA: Sawtooth Software, Inc.

Cohen, E. 2009. Applying Best-Worst Scaling to Wine Marketing. *International Journal of Wine Business Research*, 21(1), 8-23.

Collins, Andrew T. and Rose, John M. 2011. Estimation of stochastic scale with best-worst data. *International Choice Modelling Conference*, Leeds.

Coltman, T. R., Devinney, T. M. and Keating, B. W. 2011. Best–Worst Scaling Approach to Predict Customer Choice for 3PL Services. *Journal of Business Logistics,* 32(2), 139–152.

DeShazo, J.R. and Fermo, G. 2002. Designing choice sets for stated preference methods: The effects of choice complexity on choice consistency. *Journal of Environmental Economics and Management*, 44 (1), 123-43.

Finn, A. and J.J. Louviere. 1992. Determining the appropriate response to evidence of public concern: The case of food safety. *Journal of Public Policy and Marketing* 11 (1): 12-25.

Flynn, T., Louviere, J.J., Peters, T., & Coast, J. 2007. Best-worst scaling: What it can do for health care research and how to do it*. Journal of Health Economics 26*, 171-189.

Gallardo, R.K., Lusk, J.B., Holcomb, R.B. Rayas-Duarate, P. 2009. Willingness-to-Pay for Attribute Level and Variability: The Case of Mexican Millers Demand for Hard Red Winter Wheat. *Journal of Agricultural and Applied Economics*, 41(3), 599-611.

Islam, T., Louviere, J. J., and Burke, P. F. 2007. Modeling the Effects of Including/Excluding Attributes in Choice Experiments on Systematic and Random Components. *International Journal of Research in Marketing,* 24 (4), 289-300.

Jaeger, Sara R., Jørgensen, Anne S., Aaslyng, Anne S., Bredie, Wender L.P. 2008. Best–worst scaling: An introduction and initial comparison with monadic rating for preference elicitation with food products. *Food Quality and Preference, 19*(6), 579-588.

Kardes, F. R., Posavac, S. S., & Cronley, M. L. 2004. Consumer inference: A review of processes, bases and judgment context. *Journal of Consumer Psychology*, 14, 230–256.

Kanninen, B.J. 2002. Optimal designs for multinomial choice experiments. *Journal of Marketing Research, 39*(2), 214-227.

Kim, R.B., J. Unterschultz, and M. Veeman. 2000. Japanese Miller's preferences for wheat and flour: A stated Preference Analysis, *Western Agricultural Economic Association Annual Meeting*.

Koo, W.W., Mao, W. and Sakuarai, T. 2001. Wheat demand in Japanese flour milling industry: a production theory approach, *Agricultural Economics* 24, 167-178.

Lancsar, E., J.J. Louviere, and T.N. Flynn. 2007. Several methods to investigate relative attribute impact in stated preference experiments. *Social Science & Medicine,* 64(8), 1738-53.

Lee, J. A., Soutar, G.N., & Louviere, J. 2007. Measuring Values Using Best-Worst Scaling: The LOV Example. *Psychology & Marketing,* 24(12), 1043-1058.

Lee, J. A., G. Soutar, and J. Louviere. 2008. The Best-Worst Scaling Approach: An Alternative to Schwartz's Values Survey. *Journal of Personality Assessment* 90(4): 335-347.

Lee, J-H, Koo, W.W. & Krause, M.A. 1994. Japanese Wheat Import Demand, Agricultural Economics Report No. 317, Department of Agricultural Economics, North Dakota State University, Fargo, North Dakota.

Lee, M., Lerohl, M. and Unterschultz, J. 2000. Buyer preferences for durum wheat: a stated preference approach. *International Food and Agribusiness Management Review* 3, 353-366.

Louviere, J. 2012. Choice modelling: the state of the art and the state of practice. In S. Hess & A. Daly (Eds.), *Proceedings from the second International Choice Modelling Conference*. Edward Elgar.

Louviere, J., Carson, R., Burgess, L., Street, D., and Marley, A. A. J. 2011. Sequential Preference Questions: Factors Influencing Completion Rates Using An Online Panel. *CenSoC Working Paper No. 11-009*, Centre for the Study of Choice (CenSoC), University of Technology, Sydney.

Louviere, J. J., Hensher, D. A., & Swait, J. D. 2000. *Stated choice methods: Analysis and application*. Cambridge: Cambridge University Press.

Louviere, J., & Islam, T. 2008. A Comparison of Importance Weights and Willingness-To-Pay Measures Derived from Choice-Based Conjoint, Constant Sum Scales and Best-Worst Scaling. *Journal of Business Research, 61*(9), 903-911.

Louviere, J.J., Street, D.J. , Burgess, L. , Wasi, N. , Islam, T. , Marley, A.A.J. 2008. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modelling* 1(1), 128-163.

Louviere, J. J., & Woodworth, G. 1983. Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. *Journal of Marketing Research, 20*, 350-367.

Louviere, J.J., and G. Woodworth. 1990. Best-worst scaling: A model for largest difference judgments. *Working Paper, Faculty of Business, University of Alberta*.

Luce, D. and Suppes, P. 1965. Preference, utility, and subjective probability. In: Bush, R. R., Galanter, E. (Eds.), Handbook of Mathematical Psychology. John Wiley and Sons Inc., New York, pp. 249-410.

Marley, A.A. and J.J. Louviere. 2005. Some probabilistic models of best, worst, and best/worst choices. *Journal of Mathematical Psychology* 49 (6): 464-80.

Marley, A. A. J., & Pihlens, D. (2012). Models of best-worst choice and ranking among multi-attribute options (profiles). *Journal of Mathematical Psychology, 56*, 24-34.

McFadden, D. 1981. Econometric Models of Probabilistic Choice, in Structural Analysis of Discrete Data with Econometric Applications, in C. F. Manski and D. McFadden (eds.), MIT Press, C Cambridge, MA, MIT Press, 198-272.

Park, Jaehong. 2010. Essays on the International Grain Marketing and Demand System Analysis, PhD dissertation, Oklahoma State University.

Penfold K. 2010. Wheat export expansion by the packet. *Ground Cover*, 84.

Raghavarao, D., and L.V. Padgett. 2005. Block designs: Analysis, combinatorics and applications. Hackensack, NJ: World Scientific.

Saito, Yoko & Saito, Hisamitsu & Kondo, Takumi & Osanami, Fumio, 2009. Quality-oriented technical change in Japanese wheat breeding. *Research Policy*, 38(8), 1365-1375.

Stiegert, K. & Blanc, J.P. 1997. Japanese demand for wheat protein quantity and quality. *Journal of Agricultural and Resource Economics*, 22(1), pp. 104-119.

Street, D., & Burgess, L.B. 2007. The Construction of Optimal Stated Choice Experiments: Theory and Methods. Hoboken, New Jersey: John Wiley and Sons.

Street, A.P., & Street, D.J. 1987. Combinatorics of Experimental Design. Claredon: Oxford University Press.

Swait, J., & Adamowicz, W. 2001. Incorporating the effect of choice environment and complexity into random utility models. *Organizational Behavior and Human Decision Processes*, 86(2), 141−167.

Timmermans, Harry, van der Heijden, Rob, and Westerveld, Hans. 1982a. Cognition of urban retailing structures: A Dutch case study. *Tijdschrift voor Economische en Sociale Georgrafie*, 73, 2-12.

Timmermans, Harry, van der Heijden, Rob, and Westerveld, Hans. 1982b. The identificatrion of factors influencing destination choice: An application of the repertory grid methodology, . *Transportation*, 11, 189-203.

Verma, R., Louviere, J. J., and Burke, P.F. 2006. Using a Market-Utility-Based Approach to Designing Public Services: A Case Illustration from United States Forest Service. *Journal of Operations Management*, 24 (4), 407-416.

Wilson W. 1989. Differentiation and Implicit Prices in Export Wheat Markets. *Western Journal of Agricultural Economics*, 14(1): 67-77.