ANZAM Seminar

November 2021

Using Best-Worst Scaling to measure all sorts of things

Rating Scales are commonly used in business-related research – some examples

Please provide your response to each of the following questions about the service you received from and your attitudes toward your current mobile phone provider. Please circle the number that best reflects your view.

	Stron Disag	O J				Strongly Agree		
I would like to continue using my mobile phone provider's services after my existing contract expires	1	2	3	4	5	6	7	
My decision to use my mobile phone provider was a wise one	1	2	3	4	5	6	7	
The service provided by my mobile phone provider is outstanding	1	2	3	4	5	6	7	

	1	2	3	4	5	6	7	
Insignificant	0	0	0	0	0	0	0	Significant
Does not matter to you	0	0	0	0	0	0	0	Matters to you
Unimportant	0	0	0	0	0	0	0	Important
Worthless	0	0	0	0	0	0	0	Valuable
Means nothing to you	o	c	o	o	o	o	0	Means a lot to you

Ratings scales have advantages

Easy to use

Don't force discrimination

Many items can be used

Negative values are allowed

Reasonable statistical properties

Similar ordering to ranking

But ratings scales also have disadvantages

Especially response style issues

Social desirability biases

Extreme response biases

Acquiescence biases (i.e. most items are seen as important)

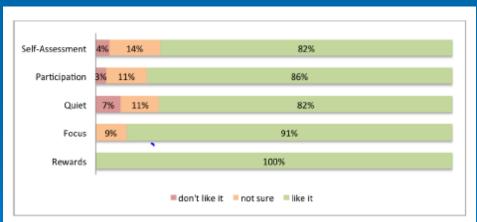


Figure 3. Histogram of the students' responses in the usability study for self-assessment, participation, quiet, focus and rewards features

So, we get:

Skewed data (generally negatively skewed)

Often unexpectedly high correlations

Different correlational structure than is found with ranked data

Another scaling option with real advantages in **some (many)** contexts is **Best-Worst Scaling** A short history:

Jordan Louviere invented BWS at Alberta in 1988

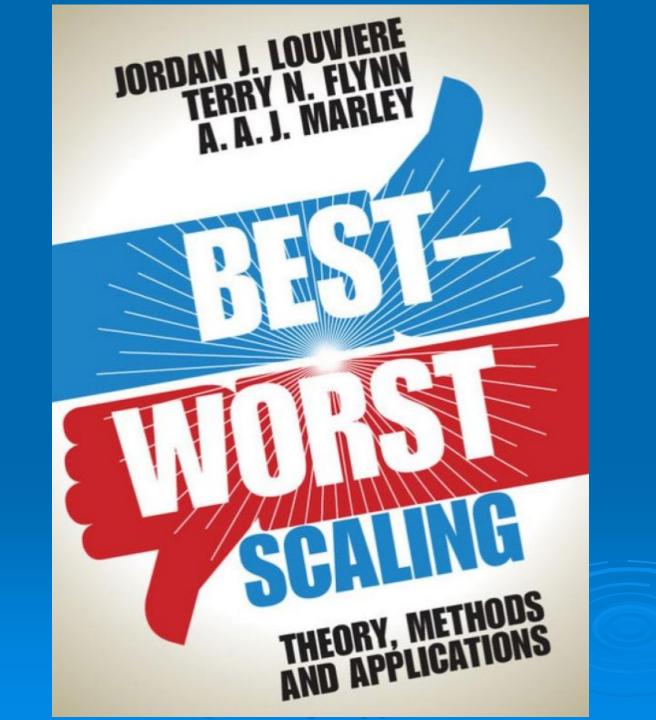
Finn & Louviere (1992) published the first important paper

Louviere & Swait (in a chapter in Bagozzi's *Advanced Methods of Marketing Research* in 1994) extended BWS to conjoint & discrete choice applications

Cohen won several "best paper" awards using BWS in the early 2000s

Marley & Louviere (2005) proved BWS's measurement & model properties

Many applications under way and the book published in 2015



There are 3 BW Cases – most researchers use only Case 1, which we will look at here

As Louviere et al. (2013) note:

In case 1 (the object case), people choose the best and worst (on some subjective scale) from a set of objects

In case 2 (the profile case), people evaluate several profiles of objects described by combinations of attributes/features dictated by an underlying design; they "see" the profiles one at a time and choose the best and worst feature/attribute levels within each presented profile

In case 3, people choose the best and the worst profiles (choice alternatives) from various choice sets determined by an appropriate underlying design

The Louviere et al. (2013) IJRM article is a nice paper on Case 1



Contents lists available at SciVerse ScienceDirect

Intern. J. of Research in Marketing







Jordan Louviere a,*, Ian Lings b, Towhidul Islam c, Siegfried Gudergan d, Terry Flynn a

So there is quite a lot of useful background information in many places for those who are interested in BWS approaches

While there are some complicated ways to determine scores from BWS data

Many people use the differences in frequency counts to compute the score (e.g. Finn & Louviere 1992)

This formula is very simple:

 $Score_j = (\# best_j - \# worst_j)/(\# times option j appears)$

Very simple and easy to interpret, as it ranges from -1 to +1

1 implies always best

-1 implies always worst

0 implies never best or worst or same best and worst

Marley & Louviere (2005) have shown this score is not very biased – which means it can generally be used safely in our type of research

An alternative is a square root ratio scale that Marley and Louviere showed has "ratio-type properties" – but this score is harder to compute (although I have written a program to do this)

There are also couple of R packages that compute these Case 1 scores in a variety of ways – both simple and complex

bwsTools: An R Package for Case 1 Best-Worst Scaling

Mark H. White II

National Coalition of Independent Scholars

Package 'RcmdrPlugin.BWS1'

December 17, 2020

Type Package

Title R Commander Plug-in for Case 1 (Object Case) Best-Worst Scaling

Version 0.1-4

Date 2020-12-17

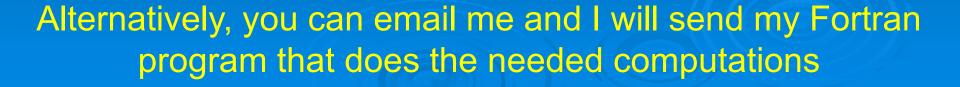
Author Hideo Aizaki

Maintainer Hideo Aizaki <azk-r@spa.n1fty.com>

Depends crossdes, support.BWS (>= 0.4-1), support.CEs, survival

Imports Remdr

Description Adds menu items to the R Commander for implementing case 1 (object case) best-worst scaling (BWS1) from designing choice sets to measuring preferences for items. BWS1 is a question-based survey method that constructs various combinations of items (choice sets) using the experimental designs, asks respondents to select the best and worst items in each choice set, and then measures preferences for the items by analyzing the responses. For details on BWS1, refer to Louviere et al. (2015) <doi:10.1017/CBO9781107337855>.



This ratio-scaled score is discussed in our 2008 paper

Journal of Personality Assessment, 90(4), 335–347, 2008 Copyright © Taylor & Francis Group, LLC ISSN: 0022-3891 print / 1532-7752 online DOI: 10.1080/00223890802107925



ARTICLES

The Best-Worst Scaling Approach: An Alternative to Schwartz's Values Survey

JULIE ANNE LEE, 1 GEOFFREY SOUTAR, 1 AND JORDAN LOUVIERE2

¹Business School, University of Western Australia, Perth, Australia ²School of Marketing, University of Technology, Sydney, Australia

Thus, the method outlined in Step 2 is a reasonable first approximation to the desired ratio scale. More formally, we have for value type j:

SVBWS score
$$V_j = \frac{1}{S} \sum_{s=1}^{S} \sqrt{\frac{Best \ v_j}{Worst \ v_j}}$$
. (4)

We can see the two scores high correlations by looking at the correlations between them

		COMSQ	VALSQ	NETSQ	TRANSSQ	PERSSQ
COMBW	Pearson Correlation	.977**	.049	371**	302**	.152*
	Sig. (2-tailed)	.000	.485	.000	.000	.030
	Ν	203	203	203	203	203
VALBW	Pearson Correlation	.035	.971**	208**	033	014
	Sig. (2-tailed)	.618	.000	.003	.641	.845
	N	203	203	203	203	203
NETBW	Pearson Correlation	303**	109	.938**	.046	089
	Sig. (2-tailed)	.000	.121	.000	.518	.205
	N	203	203	203	203	203
TRANSBW	Pearson Correlation	210 ^{**}	.035	.037	.933**	096
	Sig. (2-tailed)	.003	.618	.604	.000	.173
	N	203	203	203	203	203
PERSBW	Pearson Correlation	.171	039	171*	209**	.970**
	Sig. (2-tailed)	.015	.585	.015	.003	.000
	N	203	203	203	203	203
INTERBW	Pearson Correlation	- 21g**	- 135	113	126	- 0/11

So - BWS produces a unidimensional interval or, potentially, a ratio-type scale from nominal level choice data – which is great for doing all sorts of analysis

BW data are easily obtained – here is an example

In this section, we will ask you to pick the most and least important values that guide your life. While more than one may be important or unimportant, please choose the MOST and the LEAST important to YOU as a guiding principle in YOUR life. There are 11 sets of statements in this section

> For more information hold your mouse pointer over any word in each set. Kandomized

Which is the MOST and LEAST important factor to you as a guiding principle in YOUR life?

Most Important		Least Important
0	Equality, a world at peace, social justice.	0
•	Clean, national security, social order.	0
0	Successful, capable, ambitious.	0
0	Devout, accepting portion in life, humble.	0
0	Protecting the environment, a world of beauty, unity with nature.	0
0	Helpful, honest, forgiving.	•

So, let's look at the differences when we use ratings and BWS to look at positioning – a key strategy issue (and discussed in detail in THE BOOK)

Here, we are looking at the importance professional service managers attribute to a number of positioning issues in their attempts to achieve their organisations' objectives

Interviews with professional service providers identified 10 market positioning strategies. Professional service firms seek to be seen as:

- A provider of value to their clients
- A quality communicator through databases
- A strong service quality provider
- An organisation with a strong, positive brand
- An organisation committed to clients
- A developer of networks among its clients
- A service innovator
- A transactional service provider
- An organisation with strong relationships with its clients
- An organisation that has quality interactions with its clients

Ratings measures were obtained using a seven-point scale ranging from 1 ('not very important') to 7 ('extremely important') in achieving objectives

For the BWS task, the strategy types were divided into subsets such that respondents saw each item six times and each pair of items four times

Respondents were asked to choose the strategy that was most important and the strategy that was least important to their achieving their goals for each of these subsets

The two approaches suggested similar importances

Spearman's rank correlation between the two orders was 0.88 - supporting this suggestion

Service quality provision and commitment to clients were the most important approaches

Taking a transactional approach and using databases were the least important approaches

The best-worst approach did not alter the relative importance of the positioning approaches suggested by the ratings data

The mean scores for the strategies were all significantly higher in the ratings data case than for the BWS data and seven of the ten strategies were significantly negatively skewed, which is suggestive of the "endpiling" that can result from acquiescence response bias

On the other hand, the BWS data had only one significant skew (quality interactions with clients) and that was positive

The BWS approach seems to control for acquiescence response bias

Of more interest was the pattern of the relationships between the various positioning approaches, as this aspect is most likely to be impacted by the biases that can affect ratings scales

All of the 45 correlations from the ratings data were positive, ranging from 0.14 (perceived value and developing networks) to 0.75 (quality interactions and strong relationships)

All were significant at the 5% level and 41 of the 45 were significant at least at the 1% level

There were not strong distinctions between the strategies, supporting the earlier suggestion that ratings data suffer from acquiescence response bias, as this leads to "a tendency (for scales) to correlate positively" (Diamantopoulos et al., 2006)

On the other hand, the BWS correlations were positive and negative, ranging from -0.31 (strong relationships and a transactional approach) to 0.14 (databases and networking)

Just over half of the correlations (24 of the 45) were significant, but only 10 were significant beyond the 1% level

BWS seems to have overcome the acquiescence bias problem evident with the ratings data and BWS also led to a logical set of relationships

If response biases exist, the underlying structure produced is often less complex than it is in reality, as an underlying response bias factor explains a considerable proportion of the variance

A factor analysis of the ratings data found two factors with eigenvalues greater than one that, together, explained 63% of the variance, with the first factor explaining 43% of the 63% (i.e. 68% of the explained variance was in the first factor)

This suggests a single strong underlying factor and the presence of acquiescence bias

A factor analysis of the BWS data found four components with eigenvalues greater than one that, together, explained 59% of the variance in that data set

In this case, the first factor explained only 17% of the variance, while the other three factors had almost equal impacts (16%, 13% and 13% respectively)

Most of the strategies did not load highly onto the first component and high loadings were spread across the four factors – no acquiescence bias was evident

Acquiescence biases may hide complex structures that provide additional insights we lose when ratings scales are used – we may not be explaining anything, as we really only have a response effect - this may explain a lot of past results

These interrelationships can also be examined by "mapping" the strategies

	Transactional approach
	O _උ Database user Develop networks
Strong relationships O Quality interactions	Strong brand O O hno vativeness
Commit to clients Service quality	
Perceived value C	
55	

	Service quality O	Perceived value O	
Strong relationships			Innovativeness O Quality interactions O
Strong brand O			Turna shira da susa da
Commit to clients O			Transactional approach O
		Develop networks Database user O	

Ratings Data

BWS Data

Cliff's (1966) GOF measure between the two maps was only 0.45, which doesn't suggest congruency. Further, the correlation between the inter-point distances (Green & Rao, 1972) was only 0.20, supporting this suggestion

Clearly there are real differences here – which has implications for a lot of things we might have measured in the past

Another issue concerns subgroup analysis or segmentation – let's use personal values to look at this

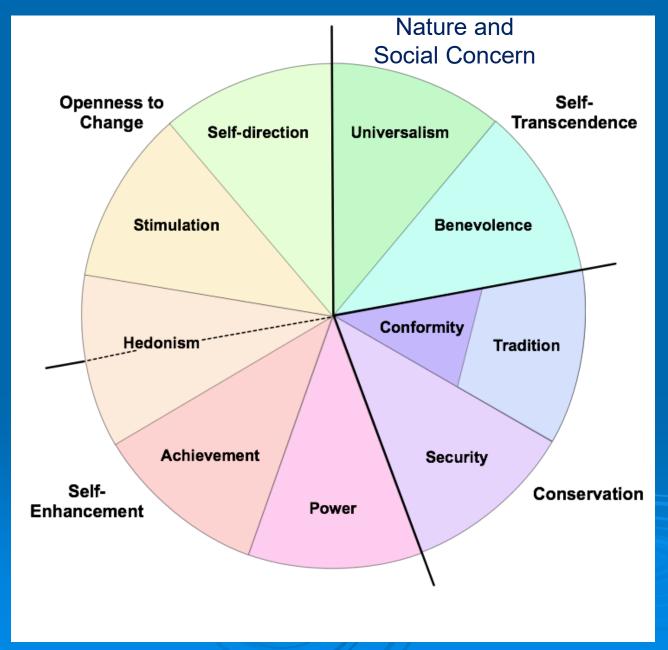
A lot of people have explored personal values - but few have examined subgroups

This may be due to measurement issues, rather than a lack of clear, reasonable subgroups

BWS may provide an answer here as well

- especially as values can be in opposition –
 so BWS is likely to work better than ratings data
- We will use Schwartz's model to look at this

Schwartz's Initial Values Model



To look at this issue, travellers and young adults in China and the USA were surveyed using

- 1. The traditional Schwartz Values Ratings
 Survey (SVS) for which raw scores and
 standardised (or Z) scores were computed
- 2. Lee, Soutar and Louviere's (2008) Schwartz Values Best Worst Survey (SVBWS)

The various data sets were clustered to see if there were any meaningful subgroups

The SVS (Z) data suggested a two cluster solution, the SVS raw data suggested a three cluster solution and the SVBWS data suggested a four cluster solution for both the **USA** and China

Discriminant analysis was used to clarify the cluster solutions for the two groups

The SVS (Z) scores produced only 2 clusters - which meant only one discriminant function could be estimated

- The single function explained most of the variation between the Chinese and American sub-groups
- which suggests there were meaningful differences between the groups

However, in both countries, the two groups attached more or less importance to all values

a common but not very useful outcome with this type of ratings data

The unstandardised SVS data suggested three clusters in both countries, allowing two discriminant functions to be estimated

However, 99% of the explained variance in China and 96% of the explained variance in the USA was due to the first function, suggesting only one function should be used

The discriminant analysis again showed the China and USA clusters were a function of respondents agreeing more or less to all of the values (with the third being a moderate group) – this result was no more useful than the standardised SVS outcome

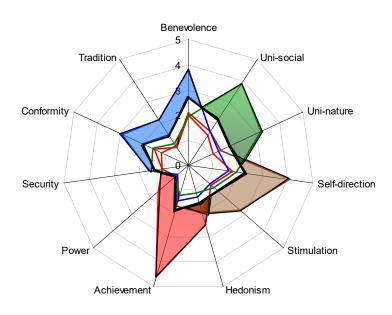
The SVBWS data, however, suggested four clusters in both countries, allowing three discriminant functions to be estimated

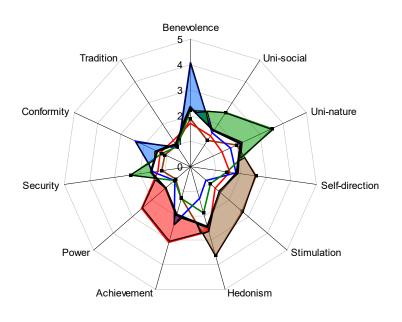
In both countries, all functions were significant and explained most of the inter-group variation

In contrast to the SVS data, the SVBWS discriminant analysis results found useful information about relevant sub-groups

US International Travelers

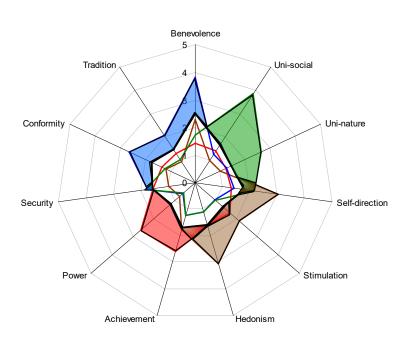
China International Travelers

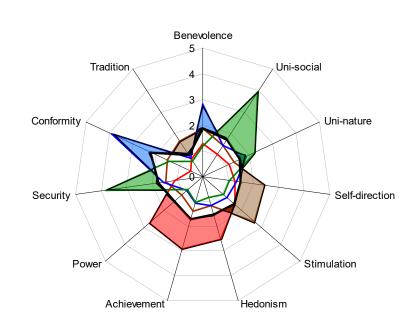




US Young Adults

China Young Adults





There were similarities in the values groups within and across the two countries, which would not have been obvious had ratings scales been used to measure values

I wonder how many "real subgroups" have been missed by using ratings scales in all sorts of contexts

Mueller and Rungie have suggested some useful ways to examine BWS data further

IJWBR 21,1

Is there more information in best-worst choice data?

24

Using the attitude heterogeneity structure to identify consumer segments

Simone Mueller and Cam Rungie

Ehrenberg-Bass Institute for Marketing Science, University of South Australia,

Adelaide, Australia

International Journal of Wine Business Research Vol. 21 No. 1, 2009 pp. 24-40 © Emerald Group Publishing Limited 1751-1062 DOI 10.1108/17511060910948017

Their table can be a very useful starting point

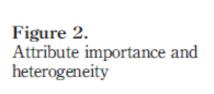
	Attribute	Best	Worst	Aggregated B-W	Mean of individual B-W	Stdev of individual B-W	Sqrt B/W	Sqrt stand
	I have had the wine before							
	and liked it	790	70	720	2.37	1.64	3.36	100
	I matched it to my food	521	156	365	1.20	1.89	1.83	54
	Suggested by another at							
	the table	434	161	273	0.90	1.87	1.64	49
	Try something different	333	174	159	0.52	1.57	1.38	41
	Region	354	277	77	0.25	2.16	1.13	34
	I had read about it, but							
	never tasted	265	202	63	0.21	1.61	1.15	34
	Waiter recommended	196	274	-78	-0.26	1.79	0.85	25
	Suggestion on the menu	197	279	-82	-0.27	1.43	0.84	25
Table I.	Varietal	164	275	-111	-0.37	1.68	0.77	23
Attribute importance on	Available by the glass	209	453	-244	-0.80	2.00	0.68	20
aggregated level and	Promotion card on the table	213	508	-295	-0.97	1.89	0.65	19
summary of individual	Available in half bottle (375 ml)	165	500	-335	-1.10	1.86	0.57	17
B-W scores $(n = 304)$	Alcohol level below 13%	111	623	-512	-1.68	1.75	0.42	13

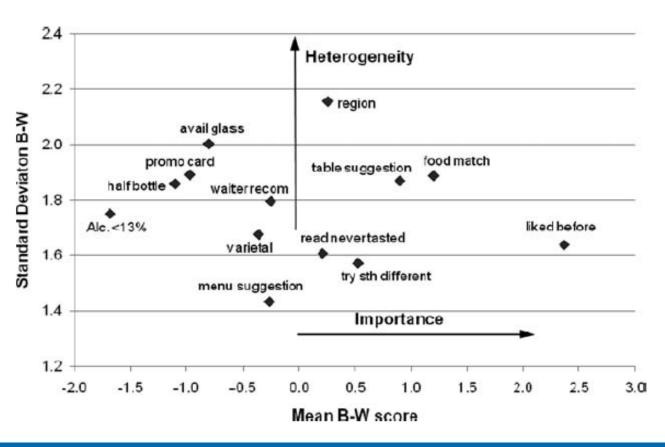
The sqrt stand is simply (sqrt B/W)/(max sqrt B/W)*100 – so lowest is $(0.42/3.36)*100 = 0.125*100 = 12.5 \sim 13$

There is a very useful diagram as well



30





Attributes in the north east are important and have heterogeneity – so of real interest to managers

Attributes in the north west are less important, but have heterogeneity, so they need to be considered

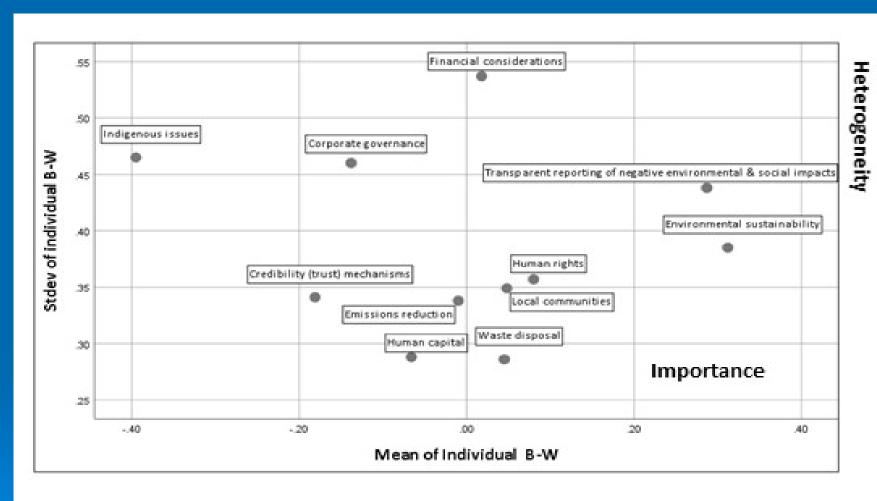
Attributes in the south have very consistent responses – little heterogeneity

A recent DBA project by Wendy Elliott is worth a look – she examined what retail investors want in mining company sustainability reports

Table 5.1: Attribute Importance (Aggregated) and Individual B-W Scores

Attribute	Best	Worst	Aggregated B-W (B-W Score)	Mean of B-W Score	Mean of B-W Score / no. attributes	StDev of Individual B-W	Sqrt B/W	Sqrt Stand
TR Negative environmental/social impacts	464	103	361	1.72	0.29	0.44	2.65	100
environmentar/social impacts	404	103	301	1./2	0.29	0.44	2.03	100
Environmental sustainability	444	51	393	1.87	0.31	0.38	2.61	98
Financial considerations	298	275	23	0.11	0.02	0.54	1.92	73
Human rights	215	114	101	0.48	0.08	0.36	1.72	65
Local communities	211	150	61	0.29	0.05	0.35	1.68	63
Waste disposal	150	93	57	0.27	0.05	0.29	1.49	56
Emissions reduction	144	157	(13)	(0.06)	(0.01)	0.34	1.43	54
Corporate governance	152	326	(174)	(0.83)	(0.14)	0.46	1.35	51
Human capital	93	176	(83)	(0.40)	(0.07)	0.29	1.23	46
Credibility (trust) mechanisms	92	320	(228)	(1.09)	(0.18)	0.34	1.13	43
Indigenous issues	47	545	(498)	(2.37)	(0.40)	0.46	0.82	31

We can also look at whether there is heterogeneity – using Mueller and Rungie's (2009) approach



What does this tell us?

The BWS approach also has several advantages when undertaking cross-country research, such as the values study discussed earlier

- 1. BWS produces scores that are equivalent across countries and do not need to be standardised prior to making comparisons
- 2. BWS has only has two verbal scale terms (most important and least important or some such), while rating scales often include multiple verbal scale terms

So BWS reduces translation issues, which can cause very significant problems

- 3. BWS does not use numbers eliminating problems when numbers have meanings in a country, such as four being an unlucky number in parts of China
- 4. BWS is relatively easy for respondents, as all they need to do is choose the most and least important from different sets of items
- 5. BWS measures generally take much less respondent time than the equivalent rating scale tasks that usually use multiple item scales, which can be important when budget constraints limit researchers' ability to collect data

The unique combination of advantages offered by the BWS approach makes it a very real alternative that should be considered when undertaking cross-country research

There are some implications

Ratings Scales

People endorse most things as important – so

Responses are often very skewed

High positive correlations, even between incompatible things – which means we may have lost our ability to see sensible differences

Are ratings data really interval level?

Best-Worst Scaling

Forces trade offs - so

Less skewed

Sensible positive & negative correlations

BW scores are at least interval level data

Helps solve some crosscultural issues

There seem to be good reasons to consider using a BW scale when collecting importance type data or "trade-off" type data, which is often the case in many research areas

Some examples show how widely used BWS really is

Cites	Per year	Rank	Authors	Title	Year	Publication
715	51.07	2	TN Flynn, JJ Louviere, TJ Peters, J Coast	Best-worst scaling: what it can do for health care research and how to do it	2007	Journal of health economics
691	57.58	409	JL Lusk, BC Briggeman	Food values	2009	American journal of agricultural
518	37.00	7	P Auger, TM Devinney, JJ Louviere	Using best-worst scaling methodology to investigate consumer ethical beliefs across countries	2007	Journal of business ethics
470	78.33	1	JJ Louviere, TN Flynn, AAJ Marley	Best-worst scaling: Theory, methods and applications	2015	
300	23.08	23	JJ Louviere, T Islam	A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant su	2008	Journal of Business Research
251	20.92	5	E Cohen	Applying best-worst scaling to wine marketing	2009	International journal of wine business research
249	19.15	6	JA Lee, G Soutar, J Louviere	The best-worst scaling approach: An alternative to Schwartz's values survey	2008	Journal of personality assessment
238	21.64	375	DMH Thomson, C Crocker, CG Marketo	Linking sensory characteristics to emotions: An example using dark chocolate	2010	Food Quality and Preference
224	74.67	426	S Mohammad	Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words	2018	Proceedings of the 56th Annual Meeting of the
223	20.27	14	TN Flynn	Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling	2010	Expert review of pharmacoeconomics & outcomes
218	7.27	291	JJ Louviere, GG Woodworth	Best-worst scaling: A model for the largest difference judgments	1991	
217	54.25	376	SM Mohammad	Word affect intensities	2017	arXiv preprint arXiv:1704.08798
215	26.88	4	J Louviere, I Lings, T Islam, S Gudergan	An introduction to the application of (case 1) best-worst scaling in marketing research	2013	International journal of
207	14.79	9	JA Lee, GN Soutar, J Louviere	Measuring values using best-worst scaling: The LOV example	2007	Psychology & Marketing
192	14.77	289	KA Hein, SR Jaeger, BT Carr, CM Delahunty	Comparison of five common acceptance and preference methods	2008	Food quality and preference
178	17.80	16	D Potoglou, P Burge, T Flynn, A Netten, J Malley	Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data	2011	Social science &
164	14.91	478	S Mueller, L Lockshin, JJ Louviere	What you see may not be what you get: Asking consumers what matters may not reflect what they choose	2010	Marketing Letters
160	16.00	413	S Dekhili, L Sirieix, E Cohen	How consumers choose olive oil: The importance of origin cues	2011	Food quality and preference
157	22.43	3	TN Flynn, AAJ Marley	Best-worst scaling: theory and methods	2014	Handbook of choice modelling
155	11.92	12	TN Flynn, JJ Louviere, TJ Peters	Estimating preferences for a dermatology consultation using best-worst scaling: comparison of various methods of analy	2008	BMC medical
151	18.88	294	E Lancsar, J Louviere, C Donaldson, G Currie	Best worst discrete choice experiments in health: methods and an application	2013	Social science &
140	11.67	15	L Casini, AM Corsi, S Goodman	Consumer preferences of wine in Italy applying best-worst scaling	2009	International Journal of Wine
140	10.77	21	SR Jaeger, AS Jørgensen, MD Aaslyng	Best-worst scaling: An introduction and initial comparison with monadic rating for preference elicitation with food prod	2008	Food Quality and
132	12.00	31	JJ Louviere, TN Flynn	Using best-worst scaling choice experiments to measure public perceptions and preferences for healthcare reform in Aus	2010	The Patient: Patient-Centered Outcomes Research
128	25.60	8	KL Cheung, BFM Wijnen, IL Hollin, EM Janssen	Using best-worst scaling to investigate preferences in health care	2016	
112	14.00	19	PF Burke, S Schuck, P Aubusson, J Buchanan	Why do early career teachers choose to remain in the profession? The use of best-worst scaling to quantify key factors	2013	International Journal of
108	21.60	11	AC Mühlbacher, A Kaczynski	Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview	2016	Health
107	6.69	32	S Goodman, L Lockshin, E Cohen	Best-worst scaling: a simple method to determine drinks and wine style preferences	2005	

44 papers with BWS in the title published were published in 2020 and 53 so far in 2021

Many researchers are seeing people's implicit "trade-offs" can be better measured using a BWS approach - We should seriously consider this approach before beginning a research project (including doctoral research)

Thank you for listening

Are there any questions?