A NEW MODEL FOR THE FUSION OF MAXDIFF SCALING AND RATINGS DATA

JAY MAGIDSON[1]
STATISTICAL INNOVATIONS INC.
DAVE THOMAS
SYNOVATE
JEROEN K. VERMUNT
TILBURG UNIVERSITY

ABSTRACT

A property of MaxDiff (Maximum Difference Scaling) is that only relative judgments are made regarding the items, which implies that items are placed on separate relative scales for each segment. One can directly compare item preference for a given segment, but comparisons between respondents in different segments may be problematic. In this paper we show that when stated ratings are available to supplement the MaxDiff data, both sets of responses can be analyzed simultaneously in a fused model, thus converting to an absolute (ratio) scale. This allows individual worth coefficients to be compared directly between respondents on a common calibrated scale.

Our approach employs latent class methods, so that respondents who show similar preferences are classified into the same segment. The fused model also includes a continuous latent variable to account for individual differences in scale usage for the ratings, and scale factors for both the ratings and MaxDiff portions of the model to account for respondents who exhibit more or less amounts of uncertainty in their responses. The Latent GOLD Choice program is used to analyze a real MaxDiff dataset and show the differences resulting from the inclusion of ratings data.

INTRODUCTION

Ratings attempt to ascertain measures of *absolute* importance which allow respondents (or segments) to be compared directly to each other with respect to their preferences for each attribute. However, when measured in the usual way with a Likert scale, they suffer in the following respects:

1. Lack of discrimination – many respondents rate *all* attributes as important, and with more than 5 attributes, *all* respondents necessarily rate *some* attributes as *equally* important (on a 5-point scale).
2. Confounded with scale usage – ratings may not be directly interpretable as measures of preference because ratings elicited from a respondent are affected by that

[1] The authors gratefully acknowledge the assistance Michael Patterson from Probit Research Inc. for the MaxDiff design used and Mike MacNulty from Roche Diagnostics Corp. for providing the data and allowing it to be freely distributed via Statistical Innovations' website www.statisticalinnovations.com.

respondent's scale usage. Some respondents tend to avoid extreme ratings, while others prefer the extremes, etc.

Because of these limitations, the use of ratings alone has been avoided in favor of MaxDiff and other choice designs.

MaxDiff scaling and other discrete-choice modeling approaches have their own limitations, since they estimate worth (part-worth) coefficients based on *relative* as opposed to *absolute* judgments. This allows options for a given subject (segment) to be ranked on relative importance, but does not allow subjects (segments) to be compared with each other according to their preferences for any particular attribute.

For example, on a *relative* basis Mary may judge attribute D to be more important than C. However, in *absolute* terms she does not consider *either* to be very important (see Figure A). On the other hand, Jim may consider C to be more important than D, and consider both C and D to be very important. Given only their *relative* judgments, it may be tempting, but it is not valid, to infer that Mary considers D to be more important than does Jim (Bacon et al. 2007, 2008).

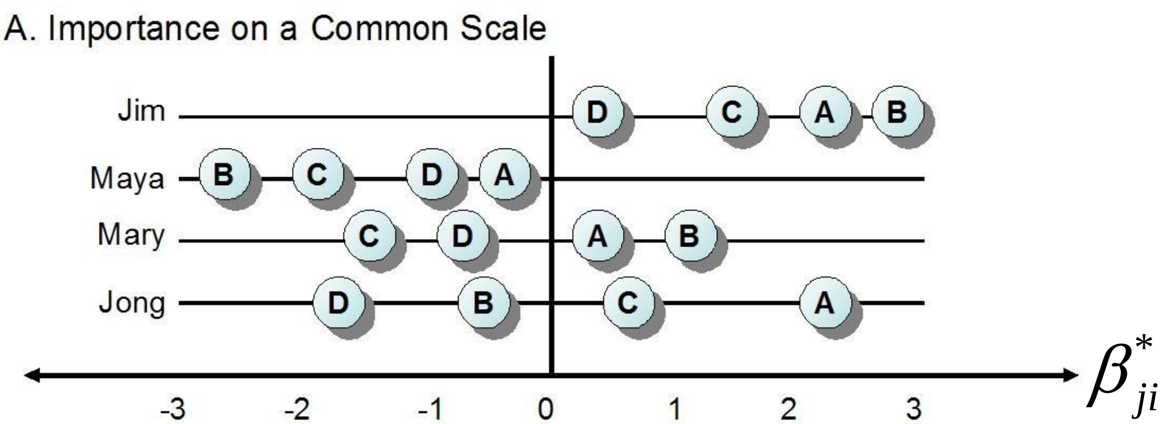


Figure A.
Comparison of Worths between 4 Respondents on a Common Absolute Scale

Approach

Our approach is to obtain maximum likelihood (ML) estimates for the segment sizes, worths, and other parameters that maximize the joint likelihood function based on both ratings and responses to MaxDiff tasks. A single nominal latent categorical variable is used for the segmentation. In order to equate the common information elicited in the ratings and MaxDiff tasks, a continuous latent variable is specified to account for individual differences in scale usage for the ratings (Magidson and Vermunt, 2007a), and scale factors are included in both the ratings and MaxDiff portions of the model which control for respondents who exhibit more or lesser amounts of uncertainty in their responses (Magidson and Vermunt, 2007b), as well as scale differences between rating and MaxDiff tasks.

A constant is added so that the calibrated worths correspond to expected log-odds of a higher, as opposed to a lower, attribute rating, thus providing a meaningful absolute metric for the worths. Hence, meaningful preference comparisons can be made between segments as well as between respondents.

Bacon, et al. (2007) utilized ratings data to uncover the zero point in individual MaxDiff worths using a somewhat similar data fusion methodology. Our approach differs from that of Bacon/Lenk in that it a) yields a segmentation, b) utilizes a simpler model, and c) it is implemented using commercially available software. Specifically, all models used here are developed using the syntax version of Latent GOLD Choice, release 4.5 (Vermunt and Magidson, 2008).

Our general approach extends directly to the use of ratings data in conjunction with *any* discrete choice model, including those where a "None" option is included in the design to provide an additional *absolute* alternative. Future research is planned that address such extensions.

We begin by discussing some important psychometric properties of ratings and choice data, and show that the lack of a common origin for choice data requires the use of identifying restrictions. In particular, we employ a simple example to show that the interpretation of the resulting MaxDiff worths may be very tricky and proper interpretation depends on the particular restrictions used (e.g., dummy or effect coding). We then present the results of a case study where a fused model is estimated and compared to results from a comparable MaxDiff model developed without use of ratings.

SOME PSYCHOMETRIC PROPERTIES OF RATINGS AND CHOICE DATA

Worth coefficients obtained with a choice/ranking/MaxDiff task for a particular latent class segment provide an *interval* scale, which is a *relative* scale in which the absolute zero point is unknown and/or unspecified. In a traditional latent class (LC) MaxDiff analysis where K segments (classes) are obtained, each segment k is associated with its own *separate* interval scale. The unknown zero points may be located at different positions along each of these scales.

For attribute j, denote the worth coefficient for segment-level *k* and individual level *i* as follows:

β_{jk} = worth ("Importance") for attribute j with respect to class k
j = 1,2,…,J items (attributes)
k = 1,2,…,K respondent segments (latent classes)

β_{jk} = individual worth coefficient for attribute j with respect to respondent i
i = 1,2,…,N respondents

Because the absolute zero points are unknown, the worth coefficients are not (uniquely) identifiable without imposing some restrictions. The most common identifying restrictions for LC choice analyses are effect and dummy coding, while for hierarchical Bayesian (HB) methods dummy coding is typically used[2]. As a simple hypothetical example, consider K = 3 segments

[2] Although this paper deals primarily with latent class (LC) analyses of MaxDiff data, identifying restrictions are also required in HB analyses. In traditional HB analyses of MaxDiff data, dummy coding restrictions are employed, in which individual worths for a particular (reference) attribute are taken to be 0 for all individuals. After obtaining such individual worths for each attribute, often the worths are 'normalized' by a) subtracting the mean worth across all attributes j=1,2,…,J for each individual i from that respondent's (non-normalized) worths, or b) converting the worths for each respondent to ranks for that respondent. Such normalized worths are similar to the use of 'effect' coding.

and J = 5 attributes (A, B, C, D and E), where each segment has the same importance ordering A > B > C > D > E.

With effect coding, the identifying restrictions are: $\sum_j \beta_{jk} = 0 \qquad k = 1, 2, ..., K$

while for dummy coding with j=5 (E) as the reference category, the corresponding restrictions are: $\beta_{5k} = 0 \qquad k = 1, 2, ..., K$

Table 1 below provides worths resulting from the use of effect coding on a hypothetical example, where for each segment k=1,2,3 a zero worth corresponds to the average importance for that segment.

Table 1.
Example of Worths Resulting from Effect Coding

Attribute	Segment1	Segment2	Segment3
A	0.62	0.77	1.21
B	0.47	0.47	0.47
C	0.06	-0.12	-0.09
D	-0.37	-0.26	-0.50
E	-0.79	-0.85	-1.10
sum =	0	0	0

Table 2 provides worths resulting from the use of dummy coding for the same example, where for each segment k=1,2,3 a zero worth corresponds to the importance of E, the reference attribute.

Table 2.
Example of Worths Resulting from Dummy Coding with Reference Attribute

Attribute	Segment1	Segment2	Segment3
A	1.41	1.62	2.31
B	1.26	1.32	1.58
C	0.84	0.73	1.02
D	0.42	0.59	0.61
E	0	0	0

To see that the worth coefficients are not unique, it is easy to verify that the choice probability for any attribute j_0 obtained from effect coding, ($P_{j0.k}$), and dummy coding with attribute j as the reference ($P^*_{j0.k}$), are identical; that is,

$$P_{j_0.k} \equiv \exp(\beta_{j_0k}) / \sum_{j=1}^{5} \exp(\beta_{jk})$$

$$\begin{aligned} P^*_{j_0.k} &\equiv \exp(\beta^*_{j_0k}) / \sum_{j=1}^{5} \exp(\beta^*_{jk}) = \exp(\beta_{j_0k} - \beta_{j'k}) / \sum_{j=1}^{5} \exp(\beta_{jk} - \beta_{j'k}) \\ &= \exp(-\beta_{j'k}) \exp(\beta_{j_0k}) / \exp(-\beta_{j'k}) \sum_{j=1}^{5} \exp(\beta_{jk}) \\ &= \exp(\beta_{j_0k}) / \sum_{j=1}^{5} \exp(\beta_{jk}) \end{aligned}$$

As suggested earlier, the true (but unknown) zero may be at a different place for each segment along the interval scale for that segment. Thus, a comparison of worths between 2 segments for a given attribute cannot be used to determine whether segment #1 prefers that attribute more or less than segment #2. In this sense, it is not appropriate to compare worths between segments. Such comparisons can result in seemingly contradictory conclusions suggested by different identifying restrictions applied to the same worths as shown in Table 3 below:

Table 3.
Comparison of Worths Resulting from Different Identifying Restrictions

Attribute	Segment1	Segment2	Segment3
A	0	0	0
B	-0.15	-0.30	-0.74
C	-0.56	-0.89	-1.30
D	-0.99	-1.03	-1.71
E	-1.41	-1.62	-2.31

Dummy Coding with j=1 as reference:
Faulty Conclusion: Segment 3 believes **B is less important** than the other segments (relies on the mistaken assumption that the importance of A is identical for all segments)

Attribute	Segment1	Segment2	Segment3
A	0.62	0.77	1.21
B	0.47	0.47	0.47
C	0.06	-0.12	-0.09
D	-0.37	-0.26	-0.50
E	-0.79	-0.85	-1.10

Effects Coding:
Faulty Conclusion: Segment 3 believes **B is as important** as the other segments (relies on the mistaken assumption that the average importance for the 5 attributes is identical for all segments)

Attribute	Segment1	Segment2	Segment3
A	1.41	1.62	2.31
B	1.26	1.32	1.58
C	0.84	0.73	1.02
D	0.42	0.59	0.61
E	0	0	0

Dummy Coding with j=5 as reference:
Faulty Conclusion: Segment 3 believes **B is more important** than the other segments (relies on the mistaken assumption that the importance of E is identical for all segments)

Individual Worth Coefficients

From the example illustrated in Table 3 for *segment-level* worths, it is straightforward to show that the interpretation of *individual-level* worth coefficients is very tricky, whether such coefficients are obtained using LC or HB. Regarding LC, individual coefficients can be obtained from segment-level worths using the posterior membership probabilities for individual i, as weights, where the posterior probabilities $p_i = (p_{1.i,} p_{2.i,} p_{3.i})$ are obtained from LC analysis.

$$\beta_{ji} = \sum_k p_{k.i}\beta_{jk} \qquad (1)$$

For simplicity, suppose Mary has posteriors (1,0,0), Fred (0,1,0) and Jane (0,0,1). As seen above regarding the segment-level comparisons, seemingly contradictory conclusions can be obtained from dummy vs. effects coding. Table 4 suggests that Mary, Fred and Jane all have similar preferences for attribute B under effect coding. However, the preferences appear to be different when dummy coding is used with attribute A as reference.

Table 4.
Individual Coefficients Resulting from Dummy Vs. Effect Coding

Attribute	Mary	Fred	Jane
A	0	0	0
B	-0.15	-0.30	-0.74
C	-0.56	-0.89	-1.30
D	-0.99	-1.03	-1.71
E	-1.41	-1.62	-2.31
Attribute	Mary	Fred	Jane
A	0.62	0.77	1.21
B	0.47	0.47	0.47
C	0.06	-0.12	-0.09
D	-0.37	-0.26	-0.50
E	-0.79	-0.85	-1.10

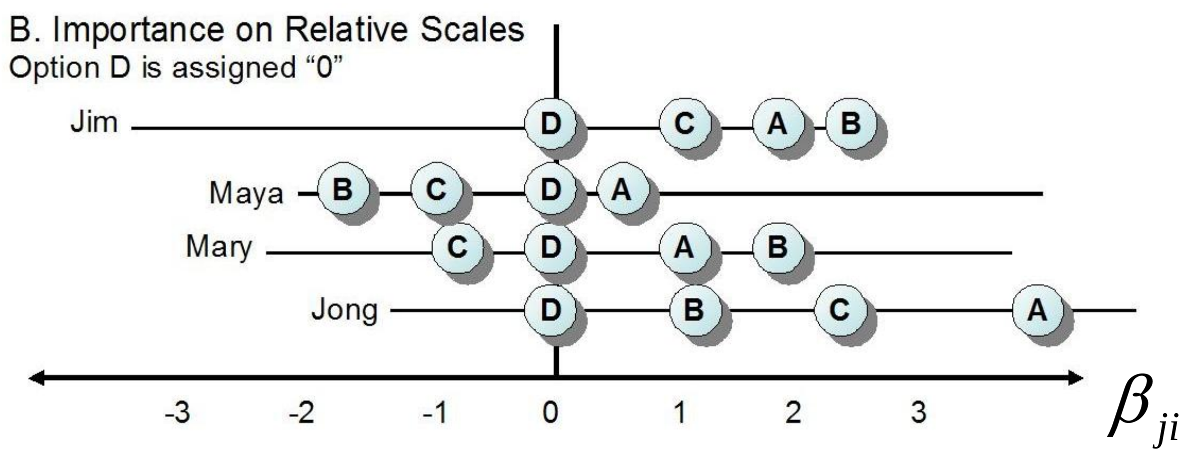
Identifying restriction = Dummy Coding (with A as reference):
Individual i’s worth for attribute j refers to
his/her inferred importance (preference) for attribute j
relative to his/her inferred importance for attribute A

Identifying restriction = Effects Coding:
Individual i’s worth for attribute j refers to his/her inferred importance (preference) for attribute j
relative to the average inferred importance (preference) for all of the attributes

This apparent discrepancy is resolved when we realize that under dummy coding with A as reference, the worths for individual i measure the importance of each attribute relative to the importance of attribute A for individual i. Under effect coding, the worths measure the importance of each attribute relative to the average attribute importance for that individual.

We conclude this section by revisiting our earlier example where worths were displayed in an *absolute* scale in Figure A with K=4 respondents. Figure B (below) displays the worths on a *relative* scale corresponding to dummy coding with D as reference.

Figure B.
Worths Corresponding to Dummy Coding with Item D as Reference



Suppose we know the absolute 0 for each individual (or segment), as determined from an analysis based on the additional ratings data. Then, the worths given in Figure B could be calibrated to provide the absolute scale as shown in Figure A, where for each of the N = 4 respondents, or segments, (Jim, Maya, Mary and Jong), the J = 4 attributes (A, B, C, and D) are positioned according to the calibrated score for that respondent. Since we now have a ratio scale (with a common zero point), it is appropriate to infer, for example, that C is more important to Jim than to Jong (i.e., 1.5 > 0.7).

Specifically, the worths displayed as separate interval scales for each segment (respondent) in Figure B may be calibrated to the ratio scale (Figure A) by adding the appropriate class-specific constants:

$$\beta^{*}_{jk} = \beta_{jk} + c_k \quad k = 1, 2, ..., K$$

For this example, the constants are: $c_1 = 0.2$, $c_2 = -0.8$, $c_3 = -0.8$, $c_4 = -1.6$.

CASE STUDY OF LAB MANAGERS

To illustrate the process of obtaining interval scales and transforming them to a common ratio scale, we utilize a MaxDiff Case Study of N = 305 Lab Managers. Each respondent was randomly assigned to 1 of 5 blocks, each being exposed to 16 different MaxDiff tasks. Each of the 16 x 5 = 80 choice sets contained 5 randomly selected items from J = 36 total attributes measured. From each choice set, respondents selected the Most and Least Important attribute.

Latent GOLD Choice (Vermunt and Magidson, 2008) models the traditional MaxDiff as a sequential choice process. The selection of the best option is equivalent to a first choice. The selection of the worst alternative is a (first) choice out of the remaining alternatives, where the choice probabilities are negatively related to the worths of these attributes. Scale weights of +1 and -1 are used to distinguish the most from the least important and thus obtain the appropriate likelihood function.

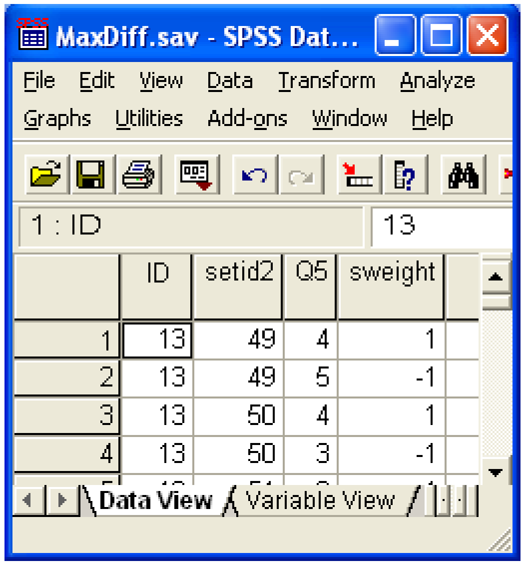
Table 5.
Cross-Tabulation of Most (+1) and Least (-1) Important Choices for MaxDiff Task #1

sweight * Q5 Crosstabulation

		Q5					
		1	2	3	4	5	Total
sweight	-1	8	39	12	3	17	79
		10.1%	49.4%	15.2%	3.8%	21.5%	100.0%
	1	40	7	20	6	6	79
		50.6%	8.9%	25.3%	7.6%	7.6%	100.0%
Total		48	46	32	9	23	158
		30.4%	29.1%	20.3%	5.7%	14.6%	100.0%

Figure C shows the LG Choice data file setup indicating that respondent #13 selected the 4th alternative (Q5 = 4) as most important in choice tasks #49 and #50, the 5th as least important in task #49 and the 3rd as least important in task #50 (setid2 = 49 and 50).

Figure C.
Response file for MaxDiff Model estimated using Latent GOLD Choice program.



Following the MaxDiff section, all respondents rated all 36 attributes on a 5-point importance scale with the end points labeled 1 “Not Important” and 5 “Extremely Important”.

We begin by analyzing the data without ratings and compare the results with these obtained from the data fusion model.

Results from MaxDiff Model Development without Ratings:

In this section we present results from the estimation of MaxDiff models without use of the ratings. We began by estimating traditional LC MaxDiff (models without use of the ratings). Such models specify a single categorical latent variable with K classes to account for heterogeneity among K latent segments. We found that the 5- and 6-class models fit these data best (i.e., corresponding to 5 or 6 latent segments) according to the BIC[3] criterion (see Table 6).

Table 6.
Model Summary Statistics for Traditional LC MaxDiff Models
MD1: MaxDiff without Scale Factors

# Classes (K)	LL	BIC(LL)	Npar
1	-11944.36	24088.94	35
2	-11750.41	23906.96	71
3	-11612.85	23837.77	107
4	-11491.34	23800.69	143
5	-11383.80	23791.53	179
6	-11280.74	23791.34	215
7	-11188.79	23813.39	251

[3] BIC balances fit with parsimony by penalizing the log-likelihood (LL) for the number of parameters (Npar) in the model. The model(s) with the lowest BIC (highlighted) are selected as best.

However, these traditional models make the simplifying assumption that all respondents have the same error variance, which can yield misleading results (see Magidson and Vermunt, 2005, Louviere, et al., 2009). Since this assumption tends to be overly restrictive, we specify a 2nd latent factor with S latent scale classes to account for differential uncertainty in the models. That is, we allow K x S total latent classes to account for heterogeneity in the data, where the S classes are structured such that

$$\beta_{jks} = \lambda_s \beta_{jk}$$

where λ_s denotes the scale factor for respondents in sClass = s, s=1,2,…,S. Thus, $\lambda >$ 1 yields greater spread among the choice probabilities for the alternatives j=1,2,…,J, which reflects more certainty. For identification purposes, we restrict λ_1=1, and require that λ_s > 1 for s>1, so that the first sClass corresponds to the least certain class.

For these data we found that S = 2 scale factor levels[4] fit best, such models resulting in lower BIC values (Table 7) than the corresponding models with S = 1 (Table 6).

Table 7.
Model Summary Statistics for MaxDiff Models with S = 2 Scale Factors
MD2: MaxDiff with Scale Factors

# Classes	LL	BIC(LL)	Npar
2	-11709.50	23842.29	74
3	-11575.23	23785.41	111
4	-11457.39	23761.39	148
5	-11349.27	23756.79	185
6	-11248.05	23766.00	222

The 5 class model fits best, which yields a structured LC model with 2x5 = 10 joint latent classes (Magidson, and Vermunt, 2007b). The less certain sClass (s=1) consists of 61.3% of all respondents. A scale factor of 2.14 was estimated for sClass #2, the more certain class, compared to the scale factor of 1 for sClass #1. The 2 latent variables Class and sClass are highly correlated in this solution, as evident from the cross-tabulation shown in Table 8.

[4] The inclusion of a 3-level scale factor in the 5-class model resulted in a worse model fit (higher BIC).

Table 8.
Cross-tabulation of the Segments (Class) by the Scale Classes (sClass)

	sClass		
lamda =	1	2.14	
Class	1	2	Size
1	91.9%	8.1%	22.0%
2	95.0%	5.0%	24.6%
3	23.0%	77.0%	20.6%
4	70.6%	29.4%	13.3%
5	18.7%	81.3%	19.5%
Total	61.3%	38.7%	

Most of the cases in Classes 3 and 5 are in the more certain sClass (sClass #2), while the reverse is true for cases in the other Classes.

Regardless of the sClass, cases in the same class have similar preferences that differ from the preferences of cases in other Classes. For example, in MaxDiff choice task #1 (shown as set #1 in Table 9), cases in segment 1 (Class 1) selected alternatives #1 or #4 as most important regardless of their sClass. Similarly, Class 2 selected alternative #3, Class 3 selected alternative #4 or #1, Class 4 cases express a very strong preference for alternative #1 and Class 5 tends toward alternatives #4 and #5.

Table 9.
Predicted Choice Probabilities* for Items in MaxDiff Task #1 by Class and sClass

	Predicted Choice Probabilities for Set #1 (Most Important Alternative)										
	Class =	1		2		3		4		5	
	sClass =	1	2	1	2	1	2	1	2	1	2	Overall
Set 1(n=79)	Class Size	0.20	0.02	0.23	0.01	0.05	0.16	0.09	0.04	0.04	0.16	
Choice	Item #											
1	20	0.39	0.57	0.09	0.01	0.25	0.23	0.73	0.97	0.21	0.19	0.36
2	36	0.08	0.02	0.01	0.00	0.05	0.01	0.02	0.00	0.07	0.02	0.04
3	24	0.08	0.02	0.70	0.94	0.18	0.11	0.07	0.01	0.17	0.12	0.23
4	11	0.30	0.33	0.16	0.04	0.40	0.61	0.13	0.02	0.29	0.39	0.26
5	28	0.14	0.07	0.04	0.00	0.12	0.04	0.05	0.00	0.25	0.28	0.10

*** All results reported are maximum likelihood estimates as obtained from the syntax version of the Latent GOLD Choice program.**

The estimates for the worths for the 5-class solution with scale factors are given in Table 10.

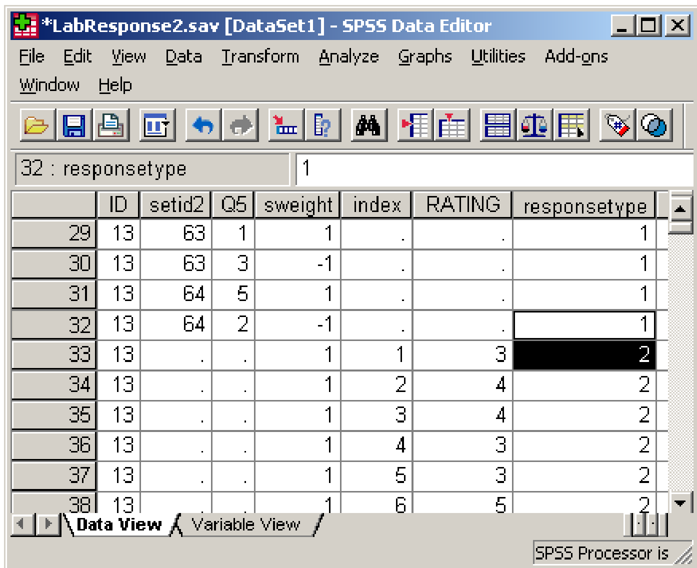
Table 10.
MaxDiff Worths for the 5-class Model with 2 Scale Factor Classes – Effect Coding (Worths shown are for s = 1). Alternatives selected most and least frequently are shaded.

	Class					sClass		
	1	2	3	4	5	1	2	
Item	22.0%	24.6%	20.6%	13.3%	19.5%	61.3%	38.7%	Average
2	1.91	2.36	1.41	2.34	1.03	1	2.143	1.80
1	1.82	2.60	1.04	1.99	.70			1.65
3	1.51	2.00	1.14	1.41	1.15			1.47
7	2.14	1.36	1.20	.89	.49			1.27
5	1.32	1.48	.84	.63	.65			1.04
4	.07	1.68	.69	2.52	.28			.96
6	1.39	1.11	.22	-.11	.44			.70
10	.67	1.13	.34	.62	.41			.66
8	.22	1.07	.99	.38	.33			.63
12	.35	1.05	.20	-.05	.79			.53
13	1.06	.25	.21	-.16	.60			.43
9	.48	.59	.11	.05	.16			.31
17	.91	.29	-.24	-.41	.17			.20
14	.67	-.10	.29	-.22	-.23			.11
11	-.05	.17	.22	.03	.08			.09
20	.20	-.46	-.25	1.76	-.26			.06
15	.35	-.18	-.12	-.34	.23			.01
25	.37	.03	-.20	-.95	-.26			-.13
23	-.73	-.73	.64	.41	.10			-.13
16	-.94	.72	-.28	.30	-.52			-.15
24	-1.37	1.61	-.60	-.55	-.47			-.19
22	.66	-.77	-.36	-.25	-.51			-.25
18	-.80	-.30	-.50	.63	-.07			-.28
19	.46	-1.08	.21	-.36	-1.00			-.36
21	-.28	-.31	-.41	-.92	-.26			-.40
27	-1.03	-.31	-.30	-.39	-.09			-.43
26	-1.22	-.38	-.73	.36	.10			-.44
30	-.82	-1.57	-.23	.29	-.28			-.63
29	-.20	-1.85	-.13	-1.06	-.04			-.67
32	-.95	-1.49	-.06	-1.29	.22			-.71
31	-1.70	-1.29	.01	-.06	-.67			-.83
28	-.81	-1.29	-1.01	-.94	-.08			-.84
34	-1.56	-1.05	-.74	-1.69	-.89			-1.15
33	-1.03	-2.02	-1.05	-1.42	-.53			-1.23
35	-1.71	-2.06	-.74	-1.78	-.49			-1.37
36	-1.38	-2.28	-1.82	-1.68	-1.30			-1.72

ESTIMATION OF THE FUSED MODEL

The model results discussed thus far were all based solely on the MaxDiff tasks alone. In contrast, the fused model utilizes both the MaxDiff tasks (response type = 1) and the ratings (response type = 2) for all 36 attributes. Figure D below illustrates what the response file looks like for respondent #13 for MaxDiff tasks #63 and #64 and ratings for attributes 1 – 6.

Figure D.
Data File Containing Responses to the MaxDiff Tasks (responsetype = 1) Followed by the Ratings (responsetype=2)


*LabResponse2.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

32 : responsetype 1

	ID	setid2	Q5	sweight	index	RATING	responsetype
29	13	63	1	1	.	.	1
30	13	63	3	-1	.	.	1
31	13	64	5	1	.	.	1
32	13	64	2	-1	.	.	1
33	13	.	.	1	1	3	2
34	13	.	.	1	2	4	2
35	13	.	.	1	3	4	2
36	13	.	.	1	4	3	2
37	13	.	.	1	5	3	2
38	13	.	.	1	6	5	2

Data View Variable View

SPSS Processor is

As before, the parameters of the MaxDiff part of the model are the worths β_{jk} and scale factors λ_s, where $\lambda_1 = 1$ for identification. The worths vary across latent segments and the scale factors across scale classes. The ratings are modeled using the following cumulative logit model:

$$\log P(y_{ji} \geq d \mid k,s) / P(y_{ji} < d \mid k,s) = \lambda_s^{'}(\alpha_d + c_k + \beta_{jk} + \theta_i) \qquad (2)$$

Here, β_{jk} are the common worth parameters of the rating and MaxDiff models, c_k are the parameters determining the location of the classes (the key additional information obtained from the ratings), α_d is the threshold parameter corresponding to response category d, θ_i is the random effects for dealing with individual scale usage differences, and $\lambda_s^{'}$ are the scale factors for the rating part of the model.

Calibrated worths can be obtained by setting the random effects to zero and selecting a value for d. We select d=4, which allows the calibrated scale to be interpreted as the cumulative logit associated with levels 4-5 vs. 1-3. Thus, a value of 0 for a given attribute means that the probability of rating it above 3 equals .5. The formula for the calibrated worths is:

$$\beta_{jks}^{*} = \lambda_s'(\alpha_4 + c_k + \beta_{jk}) \qquad (3)$$

The fusion between the MaxDiff and Ratings models is provided by the common joint discrete latent factors Class and sClass, together with the equality restriction placed on worths estimated in the 2 models. More specifically, the latent variable "Class" provides a common segmentation – respondents in the same class have the same worths. The latent variable sClass distinguishes respondents who are more certain from those who are less certain in their responses to the MaxDiff and Rating tasks. For identification, the scale factor for the first sClass is set to one for the MaxDiff tasks, and corresponds to the least certain group with respect to their responses to the MaxDiff tasks. That is, the sClasses are ordered from low to high based on the estimated scale factor. The scale factors for the Rating tasks are unrestricted.

Detailed model specifications based on the Latent GOLD Choice program are provided in the appendix for all models.

The best fitting fused model again had five classes and two scale classes. However, unlike the 5-class MaxDiff model presented above, there was no significant correlation between the classes (Class) and the scale classes (sClass). Compare Table 11 with Table 8.

Table 11.
Cross-Tabulation of the Segments (Class) by the Scale Classes (sClass)

		sClass		
	s =	1	2	
MaxDiff	λ =	1.00	1.37	
Ratings	λ =	1.11	1.81	
	Class			Size
	1	24.9%	75.1%	13.2%
	2	24.9%	75.1%	33.7%
	3	24.9%	75.1%	18.3%
	4	24.9%	75.1%	17.3%
	5	24.9%	75.1%	17.5%
	Total	24.9%	75.1%	

Again, for identification we ordered the sClasses such that the first sClass is associated with the least certain group, corresponding to a scale factor of 1 for the MaxDiff model. Compared to the earlier solution without use of the ratings (Table 8), the less certain scale class, corresponding to λ=1, is now the smaller of the 2 sClasses, consisting of only 24.9% of the respondents. sClass #2, with scale factor λ=1.37 is the more certain group. In addition, the standard errors for the estimated worths are much smaller than in the earlier model. These results are consistent with the utilization of the additional information provided by the ratings.

Regarding the ratings model, the scale factors are 1.81 for the more certain class and 1.11 for the less certain sClass. This shows that the ratings are consistent with those for the MaxDiff tasks in that respondents in the second sClass were again found to be more certain in their preferences via their assigned ratings as they are via their MaxDiff choices. In addition, respondents in *both* sClasses were somewhat more certain in providing ratings than providing MaxDiff selections (i.e., 1.11 > 1.00 and 1.81 > 1.37). However, there was less variation between the less and more certain groups with respect to the MaxDiff tasks than the Ratings (1.37/1.00 > 1.81/1.11).

As before, the worth estimates are shown (Table 12) for the less certain scale class, s = 1.

Table 12.
MaxDiff Worths for the 5 class Fused Model with 2 Scale Factor Classes – Effect Coding
(Worths shown are for s = 1)

	Class					sClass		
	1	2	3	4	5	1	2	
Item	13.2%	33.7%	18.3%	17.3%	17.5%	24.9%	75.1%	Average
2	2.05	1.42	2.23	1.79	1.72	1	1.367	1.77
1	1.90	1.32	1.78	2.37	1.58	1.110	1.810	1.71
3	2.02	1.44	1.69	1.67	1.28			1.57
7	1.73	.83	.55	.89	1.78			1.07
4	1.20	.55	2.38	1.13	.46			1.06
5	1.37	.82	.91	1.10	1.19			1.02
6	.25	.94	.12	1.03	1.19			.76
8	1.11	.09	1.35	.52	.95			.68
10	.60	.46	.53	1.24	.34			.60
12	.49	.76	-.36	.73	.40			.45
9	.44	.63	-.01	.73	.08			.41
13	.42	.61	.12	.41	.12			.37
11	.77	.11	.43	.19	-.06			.24
14	.27	-.22	-.47	.59	1.15			.18
15	-.17	.38	-.25	.09	-.07			.07
17	-.62	.36	-.48	.12	.38			.04
20	-.61	-.11	.65	-.33	-.18			-.09
16	.22	-.29	.33	.49	-1.16			-.12
23	.18	-.44	.28	-.60	.03			-.17
18	-.98	-.21	.77	-.24	-.52			-.19
19	.32	-.85	-.70	-.28	1.10			-.23
25	-.57	.00	-.75	-.25	-.18			-.29
21	-.32	-.09	-.77	-.24	-.43			-.33
24	-.71	-.93	.25	1.22	-1.10			-.34
22	-.64	-.41	-.39	-.61	.26			-.35
27	-1.17	-.14	-.29	-.51	-.46			-.42
26	-1.08	-.14	-.11	-.37	-.98			-.44
29	.29	-.40	-1.22	-1.76	-.14			-.65
30	-1.30	-.60	-.03	-1.49	-.16			-.66
32	-.15	-.27	-.95	-1.46	-.99			-.71
28	-1.45	-.43	-.82	-.98	-.66			-.77
31	.25	-.96	-.48	-1.20	-1.19			-.80
34	-.88	-1.04	-1.71	-.88	-1.57			-1.21
33	-1.61	-.82	-1.43	-1.47	-1.16			-1.21
35	-1.59	-1.02	-1.26	-1.76	-1.39			-1.33
36	-2.06	-1.35	-1.90	-1.89	-1.61			-1.68

The worths in Table 12 can be converted to the calibrated worths by applying Eq. (3). This has been done in Table 13. The estimates of the class effects, c_k, are shown in the 3rd row of Table 13. The largest class effect .27, is for class 2, the smallest, -.23, for class 3. Thus, the largest change due to the calibration is that the effect-coded MaxDiff worths associated with class 2 are increased relative to those for class 3. Note that for some attributes, the relative ordering between the segments change. For example, for attribute 10, the calibrated worth (Table 13) for segment 2 is higher than that for segment 1 (i.e., 1.24 > 1.13) while the reverse is true for the uncalibrated worths (i.e., .46 < .60), obtained from Table 12.

Table 13.
Calibrated Worths for the 5 class fused model with 2 Scale Factor Classes
(Worths shown are for s = 1)

	Class					
	1	2	3	4	5	
c(k)=	-0.06	0.27	-0.23	0.15	-0.12	
Item	13.2%	33.7%	18.3%	17.3%	17.5%	Average
2	2.74	2.31	2.80	2.63	2.33	2.51
1	2.58	2.20	2.31	3.27	2.17	2.45
3	2.70	2.33	2.20	2.49	1.84	2.30
7	2.38	1.65	.93	1.63	2.39	1.74
4	1.80	1.34	2.97	1.89	.93	1.72
5	1.99	1.64	1.34	1.86	1.74	1.69
6	.75	1.77	.46	1.78	1.73	1.39
8	1.70	.84	1.83	1.21	1.48	1.31
10	1.13	1.24	.92	2.01	.80	1.22
12	1.01	1.58	-.08	1.45	.86	1.05
9	.96	1.43	.31	1.45	.51	1.00
13	.93	1.41	.46	1.09	.55	.97
11	1.32	.86	.80	.85	.36	.82
14	.77	.49	-.20	1.28	1.70	.75
15	.28	1.16	.05	.74	.34	.63
17	-.22	1.14	-.21	.77	.84	.60
20	-.20	.61	1.05	.27	.22	.46
16	.72	.41	.69	1.18	-.86	.41
23	.67	.25	.63	-.04	.46	.36
18	-.62	.51	1.18	.37	-.16	.34
19	.83	-.21	-.45	.32	1.63	.30
25	-.16	.73	-.51	.36	.22	.23
21	.11	.63	-.53	.37	-.05	.19
24	-.32	-.29	.60	1.99	-.81	.17
22	-.25	.28	-.10	-.05	.70	.16
27	-.84	.58	.00	.06	-.09	.08
26	-.73	.58	.20	.23	-.67	.06
29	.79	.29	-1.03	-1.32	.26	-.17
30	-.98	.07	.29	-1.02	.24	-.19
32	.31	.43	-.73	-.98	-.69	-.24
28	-1.14	.26	-.59	-.45	-.32	-.31
31	.75	-.34	-.21	-.70	-.90	-.33
34	-.51	-.43	-1.57	-.34	-1.32	-.79
33	-1.32	-.18	-1.26	-1.00	-.87	-.79
35	-1.29	-.40	-1.07	-1.32	-1.13	-.93
36	-1.82	-.77	-1.78	-1.46	-1.37	-1.32

COMPARISON OF CALIBRATED VS. UNCALIBRATED INDIVIDUAL WORTH COEFFICIENTS

In this section we compare the results obtained from uncalibrated (effect coded) individual worth coefficients (based on Table 12) with those based on the calibrated coefficients (Table 13). The segment level coefficients were converted to individual worth coefficients using equation (1). Figure E plots the uncalibrated and calibrated coefficients for attribute #10. (For simplicity, only respondents classified into the more certain sClass #2 are plotted).

Based on uncalibrated worths (horizontal axis of plot in Figure E), it appears that attribute #10 is more important to many Class #1 respondents than Class #2 respondents. That is, many Class #1 respondents are plotted to the right (i.e., higher value) of those respondents in Class #2. However, according to the calibrated results (vertical axis), the opposite is true. That is, many Class #1 respondents are positioned below (i.e., lower value) those in Class #2.

Figure E.
Relationship between Effect-coded and Calibrated Individual Parameters for Attribute #10.

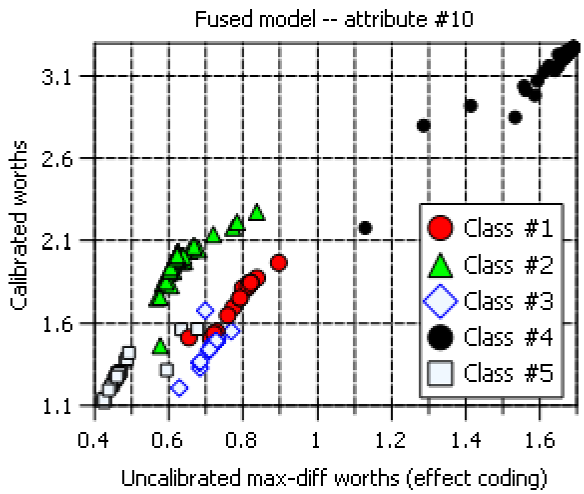
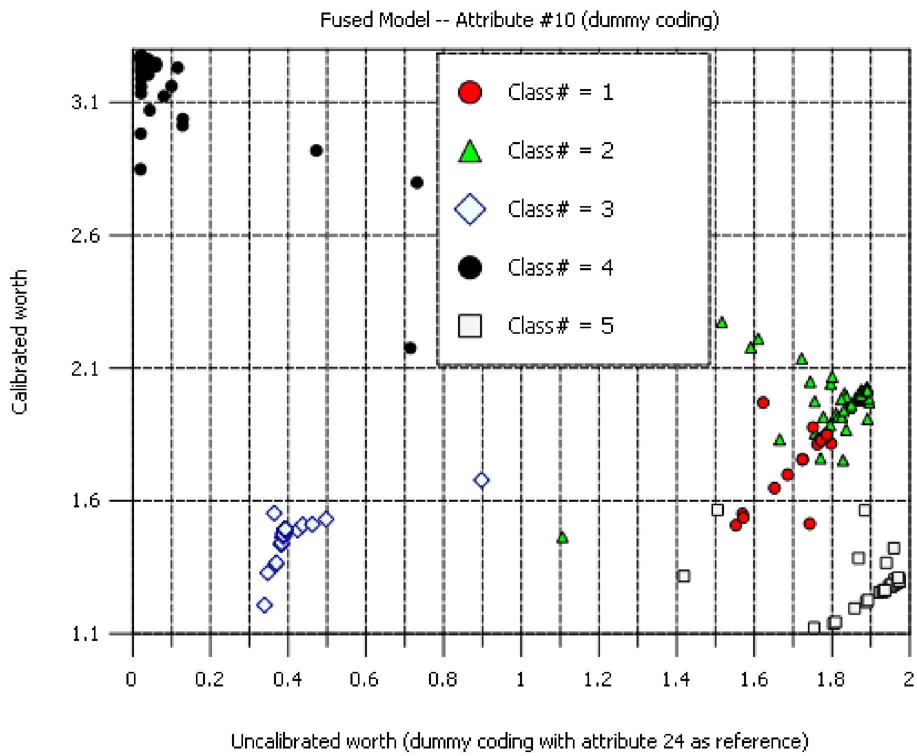


Figure F presents a similar scatterplot based on dummy-coding with attribute #24 as reference. Note from the worth estimates in Table 12 or Table 13, Class #4 considers attributes #10 and #24 to be about the same in importance. Thus, if we use attribute #24 as the reference, the (dummy coded) worth for attribute #10 will be about zero for Class #4 respondents. This is depicted in Figure E by the symbols in the upper left of the plot associated with Class #4. Now, according to this plot, attribute #10 appears to be less important to many Class #4 respondents than the other respondents, when in fact the calibrated results show that the opposite is true – it is *more* important. What is true is that the other respondents view attribute #10 to be *much* more important than attribute #24, which is one way that they differ in preference from Class #4 respondents.

Figure F.
Relationship between Dummy-Coded and Calibrated Individual Parameters for Attribute #10.
(Reference Attribute #24).



Next, we examine the correlation between calibrated and uncalibrated individual MaxDiff worths where the uncalibrated worths are based on different identifying restrictions. The correlation between individual coefficients obtained from uncalibrated MaxDiff worths and the *calibrated* worths in Figure F assesses the extent to which all respondents consider attribute #24 to be equally important. A correlation of 1 would occur if the calibrated worths for attribute #24 were equal for all respondents. As can be seen in Table 13, the worths for attribute #24 differ considerably across segments. Thus, it should be no surprise that the correlation turns out to be -.39. The negative correlation can also be seen in Figure E as it is clear that the best fitting line through the points would have a negative slope.

In contrast, the correlation between individual coefficients obtained from uncalibrated MaxDiff worths and the calibrated worths in Figure E assesses the extent to which all respondents consider the 36 attributes as a whole to be equally important. A correlation of 1 would occur if the average worth across all attributes were equal for all respondents. As can be seen in Table 13, the average worths across the 5 classes are about the same. Thus, it should be no surprise that the correlation turns out to be .89. This can also be seen in Figure E, where the slope of the best fitting line would be close to 1.

Table 14 provides the associated correlations between the calibrated and uncalibrated individual worths for all 36 attributes where the uncalibrated worths are obtained under effect coding -- for the traditional MaxDiff model estimated under HB ('HB') and LC ('MD1'), LC/scale adjusted ('MD2'), and LC fused ('Fused') -- as well as dummy coding with attribute

#24 as reference under the traditional MaxDiff model estimated under LC ('MD1'), LC/scale adjusted ('MD2'), and fused ('Fused'). For this application, correlations based on the fused model are about twice as high as obtained from the traditional MaxDiff models (HB and MD1) as well as the scale adjusted model (MD2) under effects coding. Under dummy coding, correlations are much more variable than under effects coding, and correlations based on the fused model again tend to be somewhat higher than those based on the other approaches.

Table 14.
Correlations between Calibrated and Uncalibrated Individual Worth Coefficients where the Uncalibrated Worths are Based on Different Approaches

		Correlations (effect coding)				Correlations (dummy coding)		
		Model				Model		
attribute	HB Effect	MD1	MD2	Fused	attribute	MD1	MD2	Fused
1	0.25	0.55	0.43	0.89	1	-0.43	-0.22	-0.20
2	0.30	0.38	0.37	0.82	2	-0.10	0.04	0.08
3	0.40	0.19	0.28	0.84	3	-0.27	-0.11	0.10
4	0.38	0.63	0.60	0.96	4	0.22	0.26	0.25
5	0.35	0.27	0.44	0.73	5	-0.16	0.04	0.24
6	0.46	0.51	0.51	0.96	6	0.11	0.15	0.56
7	0.49	0.54	0.50	0.92	7	0.32	0.38	0.74
8	0.52	0.31	0.33	0.93	8	0.08	0.19	0.32
9	0.41	0.39	0.35	0.97	9	-0.25	-0.24	0.29
10	0.64	0.50	0.50	0.89	10	-0.59	-0.51	-0.39
11	0.40	0.26	0.37	0.75	11	-0.16	-0.13	0.09
12	0.54	0.41	0.40	0.98	12	0.03	0.06	0.56
13	0.54	0.30	0.29	0.97	13	0.00	0.00	0.39
14	0.39	0.49	0.47	0.95	14	0.10	0.12	0.60
15	0.24	0.41	0.44	0.98	15	0.06	0.03	0.47
16	0.59	0.47	0.52	0.96	16	-0.26	-0.36	-0.13
17	0.36	0.53	0.52	0.97	17	0.24	0.22	0.62
18	0.48	0.63	0.61	0.95	18	0.15	0.10	0.22
19	0.24	0.57	0.51	0.98	19	0.37	0.40	0.77
20	0.39	0.43	0.41	0.90	20	0.22	0.21	0.27
21	0.27	0.47	0.45	0.97	21	0.03	0.04	0.46
22	0.54	0.35	0.42	0.85	22	0.39	0.43	0.76
23	0.42	0.37	0.50	0.92	23	0.44	0.54	0.67
24	0.26	0.65	0.66	0.98	24	N/A	N/A	N/A
25	0.24	0.30	0.38	0.96	25	0.16	0.15	0.55
26	0.26	0.55	0.56	0.95	26	0.11	0.09	0.25
27	0.31	0.34	0.19	0.93	27	0.14	0.01	0.44
28	0.43	0.53	0.61	0.94	28	0.35	0.35	0.60
29	0.34	0.63	0.66	0.97	29	0.58	0.61	0.96
30	0.47	0.49	0.45	0.96	30	0.50	0.48	0.81
31	0.47	0.44	0.43	0.95	31	0.30	0.32	0.56
32	0.36	0.47	0.51	0.96	32	0.42	0.47	0.83
33	0.22	0.53	0.55	0.96	33	0.36	0.34	0.67
34	0.26	0.36	0.33	0.97	34	-0.14	-0.10	0.34
35	0.34	0.52	0.43	0.92	35	0.41	0.38	0.68
36	0.52	0.19	0.36	0.96	36	0.26	0.32	0.62

Summary and Future Research Directions

Overall, the fused model provided much smaller standard errors than the MaxDiff worth coefficients and fewer respondents in the less certain sClass. The fused model can also yield *calibrated* worths which provide absolute as well as relative information. The more traditional

MaxDiff models provide worth coefficients that are very tricky to interpret and proper interpretation depends upon the identifying criterion that is employed.

The current fused model is estimable using only a subset of the ratings, and in fact can be applied when ratings are available for as few as 1 attribute (see e.g., Bochenholt, 2004).

Future research will examine how the quality of the solution is affected by

a. the number of ratings used, and

b. the particular attributes that are used in the model.

It may be that the highest rated attributes are best to use or it may be the highest and/or lowest rated attributes are best. Or, it may be that use of the middle rated attributes provides the best bang for the buck.

MODEL SPECIFICATION

Compared to the Bacon/Lenk approach:

we model MaxDiff somewhat differently (we treat it as first and second choice)

we use latent classes instead of individual effects (thus, it yields a segmentation)

we account for scale usage heterogeneity somewhat differently (we use a CFactor)

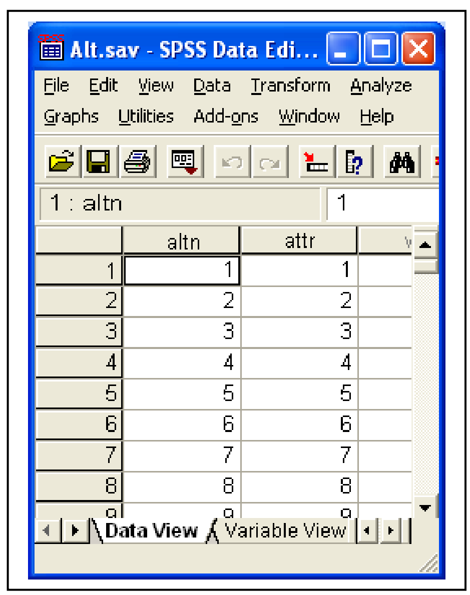
we also allow the scale factors in choice and rating to differ across persons (sClasses)

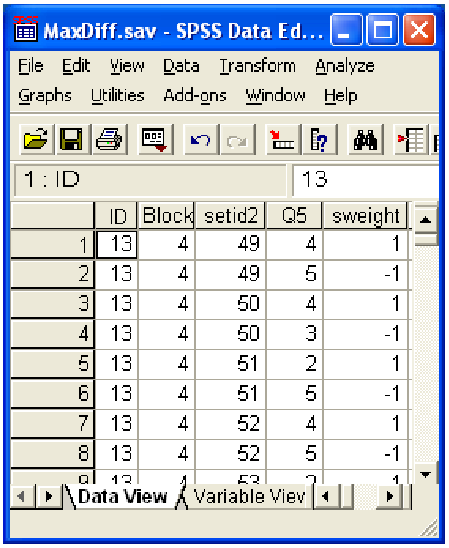
The specific model specifications are provided below.

LG 4.5 SYNTAX SPECIFICATIONS FOR MAXDIFF MODELS

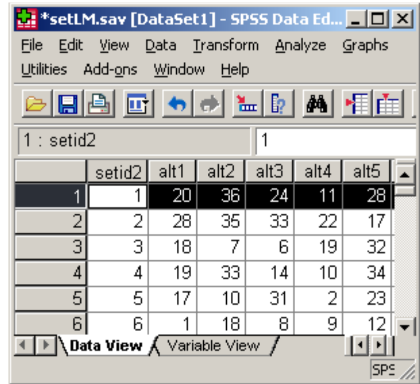
//Traditional LC MaxDiff
//**Class** denotes the nominal latent variable
//**sweight** distinguishes the choice of the
//MOST (+1) from the LEAST (-1) important,
//**Q5** represents the alternative (1-5)
//selected from the given choice set
//and **attr** is the attribute id

```
variables
  caseid ID;
  repscale sweight;
  choicesetid setid2 ;
  dependent Q5 ranking;
  attribute attr nominal;
  latent
    Class nominal 6;
equations
  Class <- 1;
```





	ID	Block	setid2	Q5	sweight
1	13	4	49	4	1
2	13	4	49	5	-1
3	13	4	50	4	1
4	13	4	50	3	-1
5	13	4	51	2	1
6	13	4	51	5	-1
7	13	4	52	4	1
8	13	4	52	5	-1
9	13	4	53	2	1



	setid2	alt1	alt2	alt3	alt4	alt5
1	1	20	36	24	11	28
2	2	28	35	33	22	17
3	3	18	7	6	19	32
4	4	19	33	14	10	34
5	5	17	10	31	2	23
6	6	1	18	8	9	12

//LC MaxDiff with 2 scale factors

```
variables
   caseid ID;
   repscale sweight;
   choicesetid setid2 ;
   dependent Q5 ranking;
   independent set nominal inactive;
   attribute attr nominal;
   latent
      Class nominal 5, sClass nominal 2 coding=first, scale continuous;
equations
   Class <- 1;
   sClass <- 1;
   Scale <- (1)1 + (+) sClass;
   (0) Scale;
   Class <-> sClass;
   Q5 <- attr scale | Class;
```

Note: Since the nominal latent factor sClass has 2 categories (2 classes), ‘coding = first’ causes the first category to have a coefficient of ‘0’ in the equation for ‘Scale’, and the coefficient for the second category is estimated. The ‘(+)’ causes these 2 coefficients to be non-decreasing, and since the first equals 0, the second must be ≥ 0. The first term in the equation for Scale ‘(1) 1’ is an intercept restricted to equal ‘1’. Thus, we have ‘1 + 0’ for the scale factor associated with the first sClass, and ‘1 + a non-negative coefficient’ for the scale factor associated with the second sClass.

Latent GOLD Choice Syntax Specifications for the Fused Model

```
Maxdiff fused with 2 scale classes
//index is the attribute id for attributes rated
variables
```

```
    caseid ID;
    repscale sweight;
    choicesetid setid2;
    dependent RATING cumlogit, Q5 ranking;
    independent Index nominal;
    attribute attr nominal;
    latent
        Class nominal 5, sClass nominal 2 coding = first, Scale continuous,
                Scale2 continuous, CFactor1 continuous;
equations
    Class <- 1;
    sClass <- 1;
    Scale <- (1)1 + (+) sClass;
    Scale2 <- 1 | sClass;
    (0) Scale;
    (0) Scale2;
    (1) CFactor1;
    Q5 <- (b1)attr scale | Class;
    RATING <- 1 scale2 + Class scale2 + CFactor1 scale2 + (b2)Index scale2 | Class;
    b2=b1;
```

REFERENCES:

Bacon, L., Lenk, P., Seryakova, K., and Veccia, E. (2007). Making MaxDiff more informative: statistical data fusion by way of latent variable modeling. October 2007 Sawtooth Software Conference Proceedings,

Bacon, L., Lenk, P., Seryakova, K., and Veccia, E. (2008). Comparing Apples to Oranges. Marketing Research Magazine. Spring 2008 vol.20 no.1.

Bochenholt, U. (2004). "Comparative judgements as an alternative to ratings: Identifying the scale origen," Psychological Methods, 9 (4), 453-465.

Louviere, Marley, A.A.J., Flynn, T., Pihlens, D. (2009) *Best-Worst Scaling: Theory, Methods and Applications*, CenSoc: forthcoming,

Magidson, J., and Vermunt, J.K. (2007a). Use of a random intercept in latent class regression models to remove response level effects in ratings data. Bulletin of the International Statistical Institute, 56th Session, paper #1604, 1-4. ISI 2007: Lisboa, Portugal.

Magidson, J., and Vermunt, J.K. (2007b). Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. October 2007 Sawtooth Software Conference Proceedings

Vermunt and Magidson (2008). LG Syntax User's Guide: Manual for Latent GOLD Choice 4.5 Syntax Module.