Sawtooth Software

RESEARCH PAPER SERIES

Sparse, Express, Bandit, Relevant Items, Tournament, Augmented, and Anchored MaxDiff—Making Sense of All Those MaxDiffs!

Bryan Orme
Sawtooth Software, Inc.

3210 N. Canyon Rd., Provo, Utah
+1 801 477 4700
www.sawtoothsoftware.com

Sparse, Express, Bandit, Relevant Items, Tournament, Augmented, and Anchored MaxDiff—Making Sense of All Those MaxDiffs!

May 2019
Bryan Orme, President
Sawtooth Software, Inc.

MaxDiff (best-worst scaling) has become a widely used measurement technique in the marketing research industry since Sawtooth Software released its MaxDiff tool 15 years ago. 73% of Sawtooth Software customers reported their firm conducted a MaxDiff study within the last 12 months. This article assumes the reader is already familiar with standard MaxDiff analysis.

Over the last 15 years, new flavors of MaxDiff have been devised to extend its capabilities to measure more items or to address drawbacks. While not an exhaustive list, here are seven different flavors of MaxDiff that have been discussed at Sawtooth Software conferences, followed by references so you can read more details about them.

Sparse MaxDiff: Used when the number of items is about 40 or more. The approach involves showing each item to each respondent about 1 or fewer times. For example, with 60 items in the study, 15 sets per respondent and 4 items per set, each item can show 1x per respondent. Or, with 120 items in the study, 15 sets per respondent and 4 items per set, it then takes 2 respondents to show each item once. Because the data are so sparse (or even missing) at the individual level, we estimate scores typically via pooled analysis such as aggregate logit or a latent class MNL. However, some authors have shown that HB-MNL can do a reasonable job even if each item is shown just 1x per respondent. (Though more accurate estimates certainly could be obtained by showing each item 2x or 3x per respondent.)

Express MaxDiff: Used when the number of items is about 40 or more. The approach involves randomly (or via a blocked design) drawing a subset of the items (such as 30 out of 60 items) for each respondent such that each item drawn per respondent can be shown 2x or 3x to each respondent. There are two potential advantages for Express MaxDiff over Sparse MaxDiff: 1) respondents don't need to orient themselves to so many items within the same questionnaire leading to some potential cognitive efficiencies, 2) if running HB-MNL, a fit statistic (RLH) can be computed for each respondent leading to the possible identification of inconsistent respondents. Despite these potential benefits, we've found under a variety of tests that Express MaxDiff performs a bit worse than Sparse MaxDiff. The typical approach for analysis is aggregate logit to summarize scores for the sample, though some researchers have used HB-MNL analysis and allowed HB to impute the preferences for items not seen by each respondent.

Bandit MaxDiff: Used when the number of items is about 50 or more, extending potentially into the 100s of items. Like Express MaxDiff, we draw a subset of the items (such as 30 out of 60 items) for each respondent such that each item drawn per respondent can be shown 2x or 3x to that respondent. Unlike Express MaxDiff, this is an across-respondents adaptive approach that draws items for respondents (using Thompson Sampling) that have tended to be preferred by earlier respondents. Thus, better items for the population are oversampled dramatically compared to worse items. When the goal is to identify the top few items for the population, Bandit MaxDiff can be 2x to 4x more efficient than Sparse or Express MaxDiff.

Relevant Items MaxDiff: This has been called "Constructed MaxDiff" by Bahna and Chapman (2018) but it seems easier to call it *Relevant Items MaxDiff*, since it mainly asks respondents to evaluate items that are more relevant (and more important) to them. A preliminary series of questions establishes whether each item has relevance and at least some minor importance to each respondent. Relevant items (plus typically just a few irrelevant/unimportant items) are carried forward and shown in each respondent's MaxDiff sets. The few irrelevant/unimportant items are included in each respondent's MaxDiff exercise to ensure stability of the solution and to reduce potential selection bias. HB-MNL analysis may be done, though a decision needs to be made regarding how to set the utility for items not seen by respondents (e.g. set to missing, or set to a very low utility such as -99).

Tournament MaxDiff: Originally called "Adaptive MaxDiff" by the author (Orme) in 2006, but to avoid confusion it's probably easier to think of it as *Tournament MaxDiff*, because it proceeds similar to a round-robin tournament in sports competitions. For each respondent, items that are selected "worst" are dropped from that same respondent's later MaxDiff sets. Later sets compare winners vs. winners, until an overall winning item is identified. The sets can decrement in complexity, from 6 items at a time, to 5 items, etc. until a final ranking is done among the remaining 2, 3, or 4 winning items. The utilities are typically estimated via HB-MNL, though aggregate logit and latent class MNL are also possible. The benefits include increased respondent engagement in the exercise and improved accuracy at the individual level for the best items per respondent.

Augmented MaxDiff: Additional ranking, rating, or sorting tasks are completed outside the MaxDiff exercise and then added (augmented) as new choice tasks to each respondent's MaxDiff data for utility estimation. For example, respondents might be asked to rank-order the 6 items chosen "best" across the previous 6 MaxDiff sets. The rank-order judgments may be exploded into paired comparisons (or other related coding approaches) for those 6 items and added to the choice data set. Augmented approaches can lead to even more accurate measurement of the top few items for each respondent, assuming the augment focuses on obtaining more information about best items. Other augmentations are possible and have been proposed, including augments that focus on obtaining more precision for the worst few items. Typical analysis is HB-MNL; aggregate logit or latent class MNL may also be used.

Anchored MaxDiff: Most MaxDiff approaches lead to relative (ipsative) scores such that we don't know whether any of the items is good or bad in any absolute sense. This often isn't a big issue, especially if a wide variety of items have been included in the experiment. With *Anchored MaxDiff*, we can establish whether each of the items is above or below some threshold anchor representing an important/not important, good/bad, or buy/not buy threshold. For anchoring, we ask additional questions wherein respondents indicate whether selected items are important/not important, good/bad, etc. Respondents do not need to evaluate all items with respect to the anchor. The additional anchoring questions are added to the choice data set as additional comparisons vs. the "anchor item" (threshold). Utility estimation may be done with HB-MNL, aggregate logit, or latent class MNL. The utility of the anchor threshold is typically set to zero, such that items with positive utilities indicate that they are important, good, or a "buy." Items with negative utilities indicate items that are not important, not good, or not a "buy." Anchored MaxDiff may be combined with any of the other six flavors of MaxDiff described in this article.

Summary of MaxDiff Methods

Method:	Description:	Strengths:	Analysis:	Available within Lighthouse Studio?
Sparse MaxDiff	Each item shown ~1x or fewer per respondent.	For measuring many items, such as 40+.	Aggregate Logit, *Latent Class, *HB	Yes
Express MaxDiff	Randomly select a subset of the items (e.g. 30) for each respondent to evaluate, typical 2 or 3 times.	For measuring many items, such as 40+.	Aggregate Logit, *Latent Class, *HB	Yes
Bandit MaxDiff	Draw a subset of items (e.g. 30) for each respondent, oversampling items tending to be the best per previous respondents.	For determining the best few items for the total sample, among 50+ total items. 2x to 4x more efficient than sparse MaxDiff.	Aggregate Logit	Yes
Relevant Items MaxDiff	Focus is on including the relevant and at least somewhat important items in each respondent's MaxDiff sets.	Greater respondent engagement and ensures relevancy of the MaxDiff questions.	HB, Latent Class, Aggregate Logit	Yes, but only with additional customization and data processing
Tournament MaxDiff	"Worst" items dropped in later sets, until respondents identify the "winning" item.	To obtain greater precision at the individual-level for best items than standard MaxDiff.	HB, Latent Class, Aggregate Logit	Yes, but only with additional customization and data processing
Augmented MaxDiff	In addition to MaxDiff questions, ask respondents to rank or sort items (typically the best items).	To obtain greater precision at the individual-level for best items than standard MaxDiff.	HB, Latent Class, Aggregate Logit	Yes, but only with additional customization and data processing
Anchored MaxDiff	In addition to MaxDiff questions, ask questions to determine if items are important or liked in an absolute sense.	For learning whether items are important or liked in an absolute sense.	HB, Latent Class, Aggregate Logit	Yes

*Apply with caution, depending on the data conditions and goals.

Note: we emphasize the utility estimations supported by Sawtooth Software tools; but many other utility estimation algorithms can provide similar results, such as either Mixed Logit or Latent Class Ensembles instead of HB.

References

Sparse and Express MaxDiff:

Chrzan, Keith and Megan Peitz (2019), "Best-Worst Scaling with Many Items," Journal of Choice Modeling, Vol. 30, March 2019, pp 61-72. Accessed at: https://www.sciencedirect.com/science/article/pii/S1755534517301355?via%3Dihub

Wirth, Ralph, Anette Wolfrath (2012), "Using MaxDiff to Evaluate Very Large Sets of Items," 2012 Sawtooth Software Conference Proceedings, Provo, UT. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2012Proceedings.pdf

Bandit MaxDiff:

Fairchild, Kenneth, Bryan Orme, and Eric Schwartz (2015), "Bandit Adaptive MaxDiff for Huge Number of Items," 2015 Sawtooth Software Conference, Provo, UT. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2015Proceedings.pdf

Orme, Bryan (2018), "Bandit MaxDiff: When to Use It and Why It Can Be Better than Standard MaxDiff," Sawtooth Software Research Paper. Accessed at: https://www.sawtoothsoftware.com/1943

Relevant Items MaxDiff:

Bahna, Eric and Christopher Chapman (2018), "Constructed, Augmented MaxDiff," 2018 Sawtooth Software Conference, Provo, UT. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2018Proceedings.pdf

Tournament MaxDiff:

Firestone, Howard (2016), "Can Adaptive MaxDiff Provide Better Results than Standard MaxDiff?" 2016 Sawtooth Software Conference, Provo, UT. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2016Proceedings.pdf

Orme, Bryan (2006), "Adaptive Maximum Difference Scaling," Sawtooth Software Research Paper. Accessed at: https://www.sawtoothsoftware.com/support/technical-papers/maxdiff-best-worst-scaling/adaptive-maximum-difference-scaling-2006

Augmented MaxDiff:

Hendrix, Phil, and Stuart Drucker (2007), "Alternative Approaches to MaxDiff with Large Sets of Disparate Items—Augmented and Tailored MaxDiff," 2007 Sawtooth Software Conference, Sequim, WA. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2007Proceedings.pdf

Jones, Urzula, and Jing Yeh (2013), "MaxDiff Augmentation: Effort vs. Impact," 2013 Sawtooth Software Conference, Orem, UT. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2013Proceedings.pdf

Anchored MaxDiff:

Horne, Jack, Bob Raynor, Reg Baker, and Silvo Lenart (2012), "Continued Investigation into the Role of the Anchor in MaxDiff and Related Tradeoff Exercises," 2012 Sawtooth Software Conference, Orem, UT. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2012Proceedings.pdf

Lattery, Kevin (2010), "Anchoring Maximum Difference Scaling against a Threshold—Dual Response and Direct Binary Responses," 2010 Sawtooth Software Conference, Orem, UT. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2010Proceedings.pdf

Lee, Jake, Jeffrey Dotson (2013), "A Simulation Based Evaluation of the Properties of Anchored MaxDiff: Strengths, Limitations, and Recommendation for Practice," 2013 Sawtooth Software Conference, Orem, UT. Accessed at: https://www.sawtoothsoftware.com/download/techpap/2013Proceedings.pdf

Orme, Bryan (2009), "Using Calibration Questions to Obtain Absolute Scaling in MaxDiff," Presentation at the 2009 SKIM/Sawtooth Software Conference.