

Is Max/Diff Really All That ?

As long as people make mistakes, the answer is yes.

Max/Diff has been touted as the savior of ratings data, a miracle cure to the common monadic scale. It's appealing to our religious belief that all problems will eventually be solved by technology. It's appealing to our intellectual sense because it's complicated.

Author: Paul Richard "Dick" McCullough

Published by *the AMA* in
MARKETING RESEARCH MAGAZINE
Column "*BEG TO DIFFER*"
Fall 2010

Is Max/Diff Really All That?

As long as people make mistakes, the answer is yes.

Max/Diff has been touted as the savior of ratings data, a miracle cure to the common monadic scale. It's appealing to our religious belief that all problems will eventually be solved by technology. It's appealing to our intellectual sense because it's complicated.

And it's appealing to our sense of justice because we all know that ratings scales simply do a horrible job of measuring anything except adjacent ratings scales. We hunger for an avenger to right a wrong longstanding!

But is Max/Diff the valiant knight of whom we dream? Well, yes, it turns out it is. But I was worried for a while. Let me explain why.

By ratings scales, I mean the historically ubiquitous survey question that takes shape as one of a million variations on this theme: *Please rate each statement below on a scale of 1 to 10 where 10 means the statement completely describes the brand and 1 means the statement does not at all describe the brand. You may use any number between 1 and 10. Amen.*

The failings of ratings scales have been reported with such enthusiasm you'd think scales were an unfaithful celebrity bound for rehab. Their sins are many:

- Lack of discrimination between items and/or between objects being rated (e.g., brands)
- Scale usage bias—different people use different parts of the scale. This bias has been shown to be, at least in part, culturally driven. So, for example, Japanese people might use only the top end of a scale (because they are polite) where Norwegians, presumably made grumpy by their bitter winters, may tend to use the bottom end of the scale. Tahitians, given their weather and scenery, use only the number 10.
- Brand halo: Ah, yes, the elephant in the room no one wants to notice. I'm not going too far out on a limb to say that every—let me repeat, every—brand imagery study that uses ratings scales has overwhelming brand halo effects. Effects that are almost always ignored but effects that just as often would have profound impact on the results of the analysis, had they been somehow taken into account.

Onto this stage confidently strides our hero, Max/Diff. Max/Diff looks like CBC, which everybody knows is really cool. Max/Diff gives metric results that are largely uninterpretable unless you run them through a logit transformation (which is also really cool). Max/Diff solves all the problems that ratings data suffer from. This is totally cool and reaffirms our faith in technology. Amen.

Brief pause for caveat: Max/Diff will remove brand halo if and only if the Max/Diff exercise is run for each brand separately. But if you do that, brand halo disappears faster than beers at a hockey game.

Now let's take a closer look at our dragon-slaying prince, dear Max. Contemplate the following: three respondents, named True Blue, Scaley Q. Bias and Too True Two. Let's say that we asked these three brave souls to rate five items for importance. Here are their data:

Item	True Blue	Scaley Q. Bias	Too True Two
Item 1	9	5	10
Item 2	7	4	9
Item 3	6	3	3
Item 4	2	2	2
Item 5	1	1	1

Let's assume that True Blue and Too True Two both accurately rate these five items for importance. That is, they accurately report their true beliefs. True Blue spreads his ratings nicely, using the entire scale, as does Too True Two. Scaley Q. Bias, being of Scandinavian descent, crowds his answers down near the bottom of the scale.

Common wisdom would tell us that Max/Diff, like a good laundry, will clean the stains out of Scaley Q's hockey jersey yet leave the Sharks logo sparkling clean. And, in a way, that would happen. But what about T. Blue and T. T. Two? Are there stains on their jerseys, too? What will Max/Diff do to them? Notice that T. Blue puts twice the importance on Item 3 than does T. T. Two. Very different opinions from our two "true" responders for the importance of Item 3. How would Max/Diff launder these differing opinions? Let's see.

Assuming these three respondents have not lied, we could infer what their respective responses would be had we designed a Max/Diff exercise for these five items. For example, one Max/Diff task might look like this:

Most Important	Least Important
	Item 1
	Item 3
	Item 5

The choice data for each of our respondents would look like this:

	Most Important	Least Important
True Blue	Item 1	Item 5
Scaley Q. Bias	Item 1	Item 5
Too True Two	Item 1	Item 5

Even the not-so-careful reader will notice that all three of our respondents answered the imaginary Max/Diff question exactly the same way. And given a Max/Diff exercise of 10 Max/Diff tasks (the complete set of all possible combinations of five items, three at a time) they would have answered every one of the Max/Diff questions exactly the same way. But this is OK because Max/Diff is a relative (not absolute) measure, right? Without adding a Dual Response question, we can't tell which items are important or unimportant; we can just tell which items are more important and how much more important they are. We get order and magnitude but no zero point. Interval but not ratio. But do we really get interval data? How? How do we learn that T. Blue places twice the importance on Item 3 than T. T. Two does when they answered every Max/Diff question exactly the same?

Here's the problem in a nutshell: If we had collected just rank order importance data on these same respondents, we could still inform the Max/Diff exercise completely for each respondent. They would still all answer the Max/Diff questions in exactly the same way as above. And they would all get the same utilities. And while those utilities would appear metric, they would have been based solely on rank order data. T. Blue would appear to put the same importance on Item 3 as T. T. Two does. But we know that ain't so. T. Blue actually puts twice the importance on Item 3 than T. T. Two. Choice exercises only use rank order information.

Turning disaggregate rank order data into disaggregate metric smells like statistical alchemy. What's going on here?

I've used Max/Diff methods numerous times with seemingly excellent results. I've read the papers by Chrzan, Cohen, Orme and combinations thereof demonstrating the superiority of Max/Diff over other measurement techniques.

Am I missing something? Yeah, I am. What I'm missing is Error Theory. Respondents are not perfectly rational. That is, they make mistakes. T. Blue will more often correctly pick Item 3 over Item 4 than will T. T. Two because the relative importance of Item 3 over 4 is four times greater for T. Blue than T. T. Two. So T. Blue's utilities will reflect a larger gap between Items 3 and 4. The inclusion of error, ironically, allows for ratio-like measurement.

Back in the 90s, it was popular in some circles, including mine, to "explode" rank order preference data to inform a hypothetical choice exercise and then estimate a logit model. But using rank order data eliminates the benefits we get from respondent error in choice tasks and the problem described above comes back into play. "Exploding" data wasn't such a good idea then and it isn't a good idea now.

Is Max/Diff really all that? As long as people make mistakes, yeah, it is.

(My thanks to John Howell, formerly of Sawtooth Software and now at Ohio State University, for his invaluable comments on an earlier version of this column)

© 2010 / MACRO Consulting, Inc.

Published in *MARKETING RESEARCH MAGAZINE* - Column "BEG TO DIFFER", Fall 2010

We are an independent **marketing research consulting firm** dedicated to helping you make the most informed, insightful marketing decisions possible. We specialize in technology, consumer, and new product research, and are well recognized for our **State-of-the-Art Research** techniques.

Ultimately, we provide more than just technical expertise. We focus on developing **pragmatic solutions** that will have a positive impact on the profitability of our clients.

CONTACT US:

Telephone: 650-823-3042

General Inquiries:
info@macroinc.com

Advanced Analysis Inquiries:
analysis@macroinc.com

richard@macroinc.com

www.macroinc.com