Customer Carewords®
Top Task Management for Websites

Old survey methods are broken: here's a fix

Gerry McGovern
gerry@customercarewords.com
www.customercarewords.com

November 01, 2010

Contents

Those severe anchoring effects

Did you ever get a question on a website that asked you to rate, on a scale of 1 to 5, the "interactivity" of the website? And did that sort of question puzzle you? You wouldn't be alone. This sort of question has two fundamental flaws that make any results coming from it not just useless, but misleading.

Let's look at the least serious flaw first. To the vast majority of people a website's "interactivity" is irrelevant. People don't want to interact with websites; they want to complete tasks on them. Interactivity is classic organization-centric language. It is an utterly meaningless thing to try to measure and survey about. Only slightly less useful is asking people about the visual design of the website. Most people simply don't care that much, and they certainly nearly always care much less than the web team.

In study after study we have found a focus on the visual design by web teams, marketers and communicators that borders on obsession, while customers essentially couldn't care less. Much more important to customers are the quality of the search results and the simplicity and clarity of the menus and links.

But the original question about "interactivity" has an even deeper flaw. Asking people to choose from a scale of 1 to 5 or 1 to 10 leads to faulty data because, as Stuart Sutherland puts it in his book Irrationality, "almost everyone is influenced by the two end points of a scale, tending to pick a number that is near the middle". He wrote this in 1992, so this problem has long been known. "Presented with two numbers at either end of a scale, people tend to opt for a number that is near the middle, regardless of whether it is correct."

If you ask a person a question they don't understand or really care about, and tell them to give a score based on a scale, it is even more likely that they will choose a number near the middle. This is because they want to answer the question as quickly as possible and by choosing a score near the middle they are essentially giving no opinion.

"These phenomena are known as 'anchoring effects'," Sutherland wrote. "In picking a number, people tend to pick one close to, or anchored on, any number with which they are initially presented or in the case of a scale one close to the midpoint. The cause of the anchoring effect is probably people's reluctance to depart from a hypothesis. If they start with a number, even one determined by the random spin of a wheel, they adopt that number as a working hypothesis and although they do move away from it, usually in the right direction, they are reluctant to move too far. Similarly, when picking a point on a scale or selecting a number from a series of consecutive numbers, they are reluctant to depart too far from either point and hence plump for a point near the middle. They unconsciously assume that the end points are likely to be approximately equidistant from the true value. Allowing one's judgment to be influenced by the initial anchoring point causes inconsistency: different judgments are given with different anchoring points although the anchoring point has no bearing on the correct judgment."

The apple and the chocolate

Years ago, British Airways (BA) decided to introduce a new service for its first class passengers on long haul flights. It was basically a mini fridge full of goodies so that if you woke up in the middle of the night feeling a little hungry, you could get something nice for yourself. The question was: What do we put into this little fridge?

BA dutifully went about doing its market research and assembled several focus groups of first and business class passengers. What would you like, they asked? The response was universal. People wanted fruit or perhaps some light salads. All very good. All very healthy.

On the first flight with the new service an air hostess noticed as the fridge was being filled up. "What are you doing," she asked? And the person dutifully explained what was happening. The hostess laughed. "They're lying!" she said. "They don't want salads. Listen, I've being doing the London to LA route for years, and when they wake up in the middle of the night the last thing on their minds is salads."
"But the focus groups all said …"
She shook her head and walked away. A couple of minutes later she came back with some chocolates and cakes. "Please put these in as well," she said. "Trust me. I know my customer."

And they did put some chocolate and cakes in and when they checked at the end of the flight, they were all gone and nobody had touched the salads.

The worst way to design a website is to have five smart people in a room drinking lattes. The longer you leave them in the room the worse the design becomes. The next worst way to design a website is to have 15 customers in a room drinking lattes. What people say they do and what they actually do are rarely the same thing.

It's not that people deliberately set out to lie. They're thinking: 'I should eat better. I am putting on a bit of weight. Salads are really good for you. It's ages since I ate an apple.' But when they wake up in the middle of the night, wipe the sleep from their eyes and open that fridge door their hand reaches out as their mouth mumbles c-h-o-c-a-l-a-t-e.

A cardinal rule in web management is:

Do as I do not as I say.

Of course the best way to know what people do is to observe them, but are there survey approaches that might get closer to the chocolates? That was a problem I tried for years to solve. And I stumbled on the answer accidentally.

An accidental discovery

For years I did workshops on card sorting. I had a case study of an Irish tourism website and 146 cards with relevant phrases on them. Some examples can be seen below:

- About Ireland
- Accommodation
- Best of Ireland
- Dublin
- Irish Vacation Packages
- Midlands East
- Planning a Trip
- Sights & Activities
- Special Offers
- What to See & Do

I would go through a time-consuming process of getting people to put these cards into groups with the objective of helping this website design a more intuitive navigation. But the exercise was cumbersome and it was often hard to see clear results coming out of it.

I was looking for clear, repeatable trends and I wasn't getting them. Were there universal top tasks for tourism websites? I started experimenting. Instead of getting people to sort the cards into bunches, I started asking them to pick out the cards they thought contained the most important words that should appear on the homepage of this imaginary Irish tourism website.

Some trends began to emerge. Then I started limiting the number of cards they could choose to 10. Stronger trends emerged. Then I started giving people less time to choose. Even stronger trends emerged. Then I started getting people to vote. They had to choose the most important one, the next most important one, and so on. Very strong, clear trends emerged. People in different countries were choosing the same cards for their top 10 and when they had to vote, certain choices really began to stand out. The same choices rose to the top in New Zealand, Australia, Sweden, Switzerland, the United States, Canada. And the funny thing was that the less time I gave people the more consistent the voting patterns became.

Again and again, people told me that they couldn't possibly choose 10 out of 146, and certainly not in the time that I was giving them. That's common sense, isn't it? Classic marketing research theory would say that it's impossible. Classic marketing research theory is wrong.

It was amazing. After going through this process with over 1,000 people in 11 different countries, there was an extraordinary consistency of choice. Two tasks—'accommodation' and 'special offers'—got 21% of the vote. This needs to be placed in perspective. There were 146 cards to choose from. Two got as much of the vote as the bottom 108.

When I compared these choices to what the leading websites in the travel industry were doing, they mapped very well. On the homepages of these websites were booking processes for accommodation along with deals and special offers.

In order to allow people to record their votes for their top 10 choices I created a simple one page printout with all the phrases listed in three columns in alphabetical order. After having chosen their 10 cards, I would ask people to place a 10 beside their most important choice, a 9 beside their next most important choice and so on. Then something a bit strange began to happen. Certain participants started cheating.

A cheaters discovery

Instead of going through all the hassle of spreading the cards out and diligently sorting through them, certain attendees went straight to the alphabetical list and started putting their scores in directly. I was annoyed. I told them that they needed to go through the card sorting process because there was something special in this process that would get much better results.

And then someone challenged me. "Why do I have to sort these cards? Why can't I just vote based on this list?"
"Well, you can't", I said to him. "You can't. You can't because card sorting is a proven method that works.

But the truth was that I didn't know why card sorting worked. To be more exact, I didn't know why this new method of just voting from the list didn't work. So one day I decided to leave the cards in my bag and simply hand out the list. To my great surprise, when I added up the results, I found that the voting patterns were almost identical to the ones from workshops where I had used the card sorting approach. I tested again. I found similar patterns. The same top tasks of accommodation and special offers kept coming out in front. The process was much faster than card sorting and the trends were much stronger.

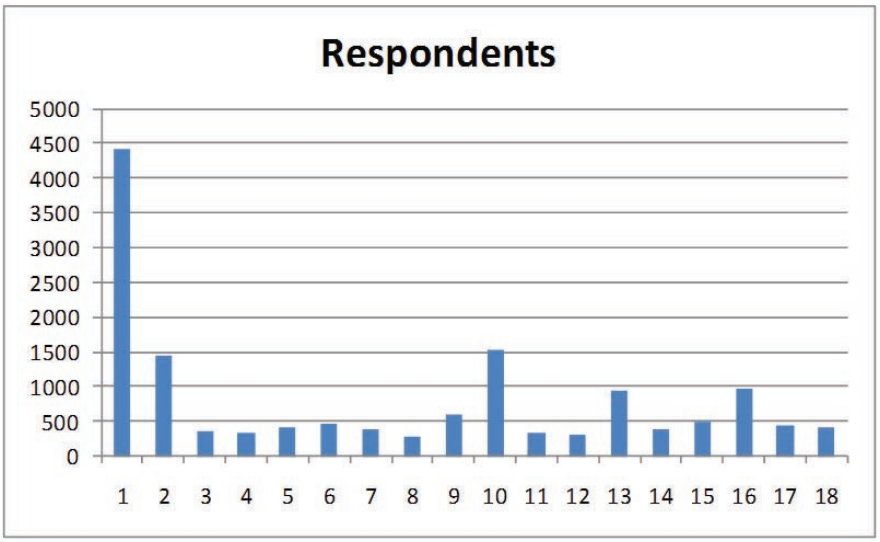
Why this method works

Over the last eight years we have done well over 100 task list surveys in six languages with more than 70,000 people participating. Repeatable trends have emerged, For example, by surveying 400 customers you will have identified your top 10 tasks with reasonable certainty and your top three tasks with high certainty. Your overall top task will have emerged by about 100 voters and sometimes by as early as 50.

Typically, in surveying 100 tasks, the top 5 tasks will get an average of 25% of the vote, with the bottom 50 tasks also getting 25%. In other words, the top 5 tasks get as much of the vote as the bottom 50. After 400 customers have voted, the chances of a task that is in the bottom 50

becoming a top task are infinitesimal. Also, the chances of a top task dropping into the bottom 50 tasks are also infinitesimal

This is according to our analysis of 18 of our Customer Carewords task identification surveys, in several languages and countries, with 14,524 customers voting. There was an average of 807 customers per survey (or 595 excluding the largest one which had 4,407). Small surveys with less than 100 customers were excluded from this study.

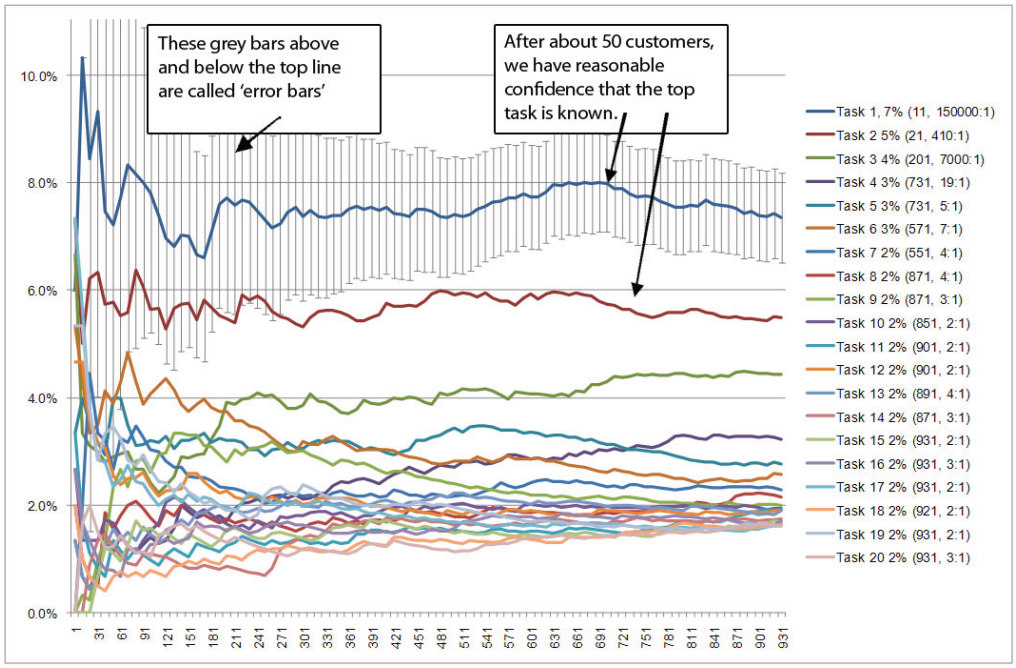


The following pages give the detail behind the summary statements below:

- After about 50 customers, we have reasonable confidence that the task that has emerged as first in the vote will remain first.
- After about 80 customers, the top two tasks are known within two rank positions.
- After about 200 customers, the top two tasks are known and ranked in order.

95% confidence

The following chart, from an anonymous representative survey, illustrates the percentage of the vote for each task to date. The voting trend shows initial random variations and then settles down to a reasonably stable pattern.



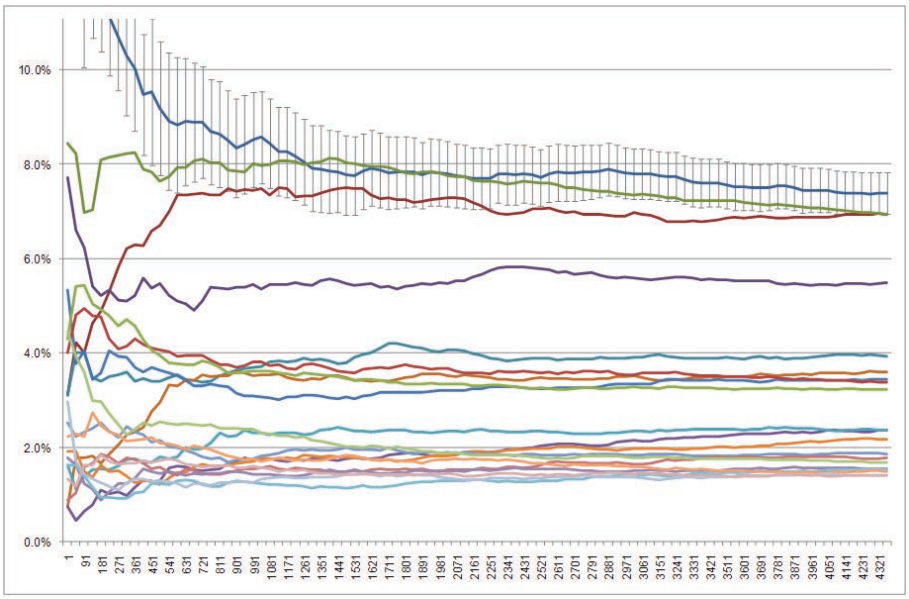
At first, we have a large margin of error around the score, but that narrows as the number of voters increases. That band of error is called our '95% confidence band'. We pick 95% because it is the most common measure used for statistical significance. It means that we expect our data to be within the bands shown in 19 out of 20 surveys of a similar nature. The bands narrow in proportion to the square of the number of voters; in other words, it takes 400 customers to obtain twice the precision of 100 customers (20 squared vs. 10 squared).

The error bars of the 95% confidence band initially cover the top three task lines but gradually converge, leading to the third preference dropping out after about 60 votes and the second preference after about 260. It is also clear that the first three tasks are clearly preferred, and as you go down in importance the remaining tasks bunch together into an undifferentiated tangle.

95% confidence means that the odds are 20:1 that the top score could be outside its error bars. That means 40:1 in either direction. So in this example the odds are 40:1 of the top score of 7.4% being as low as the band of 6.5%; and they are 150,000:1 of the top score being as low as the second preference; and the chances of it being any lower are infinitesimal. There is no chance of any of the items from fourth place down ever achieving the same score as the top task. Of course, all of this assumes that the nature of all participants in the survey is the same throughout.

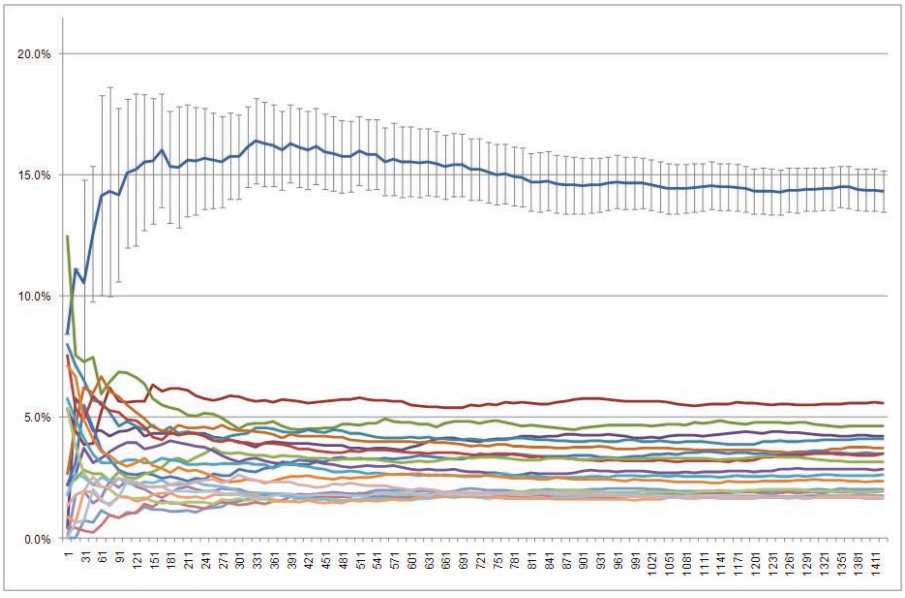
Top tasks emerge

Here is the chart of vote trends for the largest survey. The 95% confidence error bars around the top task show that the scores for the top three tasks are not statistically significant. And the fact that it took 2,161 customers for the top task to stabilize reflects this.



On the other hand, if you ask the question, "When were the top three tasks settled?", or the top five or 10, you can see that #1 and #3 were always in the top three, and #2 separated after 226 customers.

In some surveys, a clear leader is established right from the start. Here is the second survey:



And all of this is achieved just how?

You can identify your top tasks by giving your customers a randomized list of tasks that is anywhere from 20 to 150 items long and asking them to quickly vote for their top tasks. Doesn't make sense does it? But it works. And we have the evidence.

We adapted this approach for a special survey we call the Customer Centric Index. Over the years we have sought to identify the critical factors that influence task success on a website. We identified 13 and broke them down into three groups:

1. Content
2. Social
3. Information architecture

CONTENT FACTORS		
Factor	Positive	Negative
Up-to-date	Up-to-date information	Out-of-date information
Accurate	Accurate information	Inaccurate information
Complete	Complete information	Incomplete information
Language	Plain language	Full of jargon, corporate speak

SOCIAL FACTORS		
Factor	Positive	Negative
Contact	Easy to contact a person	Hard to contact a person
Participation	Easy to participate / give feedback	Hard to participate / give feedback
Open	Gives me the facts / transparent	Misleading / not transparent
Recommendations	Has ratings, reviews, recommendations	Has no ratings, reviews, recommendations

INFORMATION ARCHITECTURE FACTORS		
Factor	**Positive**	**Negative**
Search	Helpful search results	Poor search results
Menus & Links	Clear menus and links	Confusing menus and links
Layout	Simple layout / easy to read	Cluttered layout / hard to read
Visual appeal	Looks attractive / appealing	Looks unattractive / unappealing
Speed	Fast to do things	Slow to do things

Of course, we didn't design the survey using the classical approach of asking people to vote based on a scale of 1 to 5 for each factor. Instead, we presented the survey as follows:

Please choose the THREE factors from the list below that best describe your actual experience with the PMA website.
Give a score of 3 to the factor which best describes your experience, 2 to the next best description, and then 1.
Please give only one score of 3, 2 and 1.
Leave the rest blank.

Fast to do things	☐
Full of jargon, corporate speak	☐
Complete information	☐
Confusing menus and links	☐
Looks attractive / appealing	☐
Out-of-date information	☐
Has no ratings, reviews, recommendations	☐
Accurate information	☐
Has ratings, reviews, recommendations	☐
Inaccurate information	☐
Poor search results	☐
Hard to participate / give feedback	☐
Looks unattractive / unappealing	☐

So, just why does this approach work?

This method allows people to vote for what's really important to them. If it's a task list for an intranet, for example, then many employees will vote for the task "Find people," because on a day to day basis, finding people is a really important task for them.

Equally important is what doesn't matter to them i.e. what they don't vote for. So, with the Customer Centric Index, if a customer is really happy with your website they might vote for "helpful search results," accurate information" and "easy to contact a person." If another customer is really unhappy they might vote for "poor search results," "confusing menus and links," "slow to do things." We're asking them to choose their carewords—the things they really care about. We're asking them to give their top vote to what matters most to them—and most people respond very positively to that.

It is essential that you restrict the number of choices. We tested situations where people could choose as many as they liked from the list and the results were not useful at all. It's important to limit the choices to five or less. (We use 3 choices for the Customer Centric Index.). We tested with 10 choices and found that a significant number of people struggled to select 10 things that really mattered to them. We found that after five choices, people began selecting things that might be of interest but that they didn't feel very strongly about, or else they got frustrated with the process. We need to find out what people really care about.

We found that it's important that people vote. Voting does something to people. If they just have to select five tasks, that's one thing, but if they have to select the most important task to them and vote for it as the most important, their choice takes on a new level of seriousness for them.

The less time you give people to vote the more accurate the choices become. The less they 'think', the better. If they scan the list, what really matters to them will jump out, and what doesn't will fade into the background. In a list of 100 we recommend that no more than 10 minutes is given to voting. With the Customer Centric Index, which has 26 choices, 1 minute is given.

Doesn't work in theory, works in practice

Giving people a list of 100 tasks and asking them to choose the five most important to them sounds mad. I've had lots of market researchers tell me that it is utterly impossible. I wouldn't have believed it myself but the results speak for themselves. So, why does it work? My colleague Gord Hopkins suggests one reason: the cocktail party effect. You're at a party. There are a lot of people in the room, a lot of buzz and noise. Across the crowded room someone mentions your name in conversation and out of the noise there's clarity; you hear your name being spoken. The words jump out at you.

What I have noticed is that people don't read the whole list. They just scan it, and out jump the things that really matter to them. It's not that they discover things on the list that interest them. It's that inside their brains there are things that really matter to them (their carewords). Seeing these words on the list is a reinforcement of something they already care a lot about.

About Gerry McGovern

Gerry has been consulting on the Web since 1994 and has worked for organizations such as Microsoft, Cisco, HP, Tetra Pak, BBC, Rolls-Royce, etc. Read the first chapter of Gerry's latest book, The Stranger's Long Neck:
http://www.gerrymcgovern.com/sln-ch1.htm

Gerry is founder of Customer Carewords, a global network that focuses on helping organizations implement top task management for websites:
http://www.customercarewords.com/

gerry@customercarewords.com