



Impact of Number of Brands Rated

Overview

In this study, we looked at how many brands respondents rated at a time and if this impacted the research results. The null hypothesis would be that there is no difference, and we were able to reject this hypothesis.

Prior Work

Exhaustive secondary research uncovered only one somewhat related study. Working off the classic study by Miller (1956), Hulbert (1975) applied the information processing rule of thumb of “seven units (plus or minus two)” to scale usage. This rule of thumb briefly summarized: researchers have found that humans have the ability to hold or evaluate seven individual items in mind at once, whether that be remembering a sequence of numbers, counting dots on a screen, perception of speech variations, et cetera. After seven (+/- two) items, we must organize the material into chunks in order to continue to retain or evaluate it.

Hulbert wanted to assess whether this principle would be true in scale ratings. And so rather than presenting respondents with a preset scale, they were allowed to assign any positive number they wished in rating the stimuli, on three different scales [“scale” used here in the test construction sense], each over 50 items each. He hypothesized the number of distinct assignments respondents would use would be less than or equal to 10.

Indeed, respondents (97 salesman) used between 6 and 10 ratings, on average, to express their opinions of dissatisfaction, motivation or satisfaction with their job. Hulbert writes,

One of the goals of scale design is generally to avoid preventing the respondent from expressing his true feelings because of some property of the scale itself. Thus, a necessary though not sufficient condition to attain measurement at some level equal to or higher than ordinal is that the scale used should enable preservation of strict monotonicity between obtained measures and the underlying (latent) continuum. This condition is met simply by ensuring that the number of categories in the scale is greater than the number of stimuli to be rated.

In applying his results to market research, he suggests that the small number of items (often brands) usually rated should avoid measurement error, but due to information capacity limits, rating more could lead to more measurement error.

Our Study

In our study, we asked respondents to rate either up to three brands or up to six brands on a five-point scale (the one very top firm, world class, stronger than most, average, weak), for sixteen attributes regarding brands in the financial industry, as in Figure 3.

In your experience, how would you rate each firm's performance on...?

Demonstrating exceptional integrity and honesty in all their dealings

Control

	The One Very Top Firm	World Class	Stronger than Most	Average	Weak
<Brand 1>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<Brand 2>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<Brand 3>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

In your experience, how would you rate each firm's performance on...?

Demonstrating exceptional integrity and honesty in all their dealings

Test

	The One Very Top Firm	World Class	Stronger than Most	Average	Weak
<Brand 1>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<Brand 2>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<Brand 3>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<Brand 4>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<Brand 5>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<Brand 6>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 3:
Control and Test questions for Number of Brands Rated Study

To qualify for this study, respondents needed to have voted, attained a certain minimum level of investments and income, and be actively involved in expressing their opinions publicly on financial issues. The screener averaged three minutes to complete, followed by a six-minute questionnaire, half of which was the brand rating series. The e-Rewards panel provided the online sample.

Differences in the control and test groups were as follows:

	Control	Test
Number of brands rated	Up to 3 familiar with	Up to 6 familiar with
Interviews (n)	272 of which 222 qualified for more than three brands	155 of which 127 qualified <i>and were asked</i> more than three brands
Field dates	November 5 – December 2, 2008	December 5 – 12, 2008

Results

If respondents are indeed constricted by information processing capacity, or simply overwhelmed or fatigued by an increased requirement to process and provide information, what may happen? Measurement error. While we can't necessarily label it error, we can definitely label it "different" – respondents gave lower responses when presented with more brands:

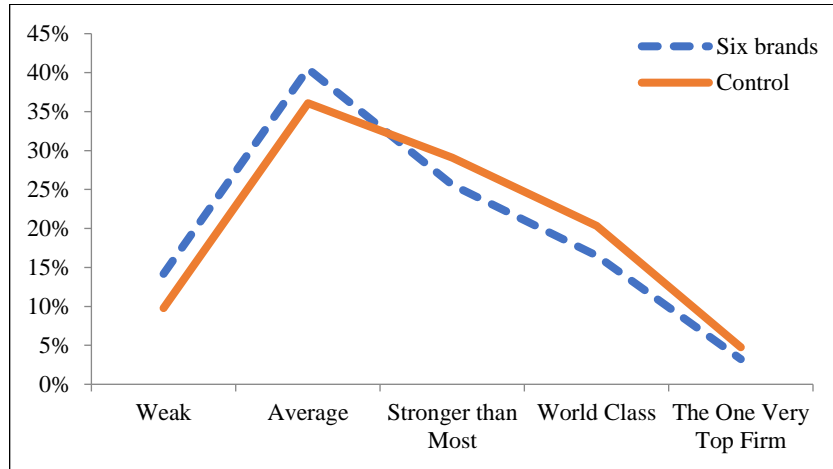


Chart 4

However, this difference could have been driven by familiarity. Brands the respondents rated were those with which they were “very familiar” or “somewhat familiar”, with those with which they were “very familiar” receiving priority. Chart 5 shows that brands with which a respondent is very familiar received consistently higher scores than those brands with lower familiarity.

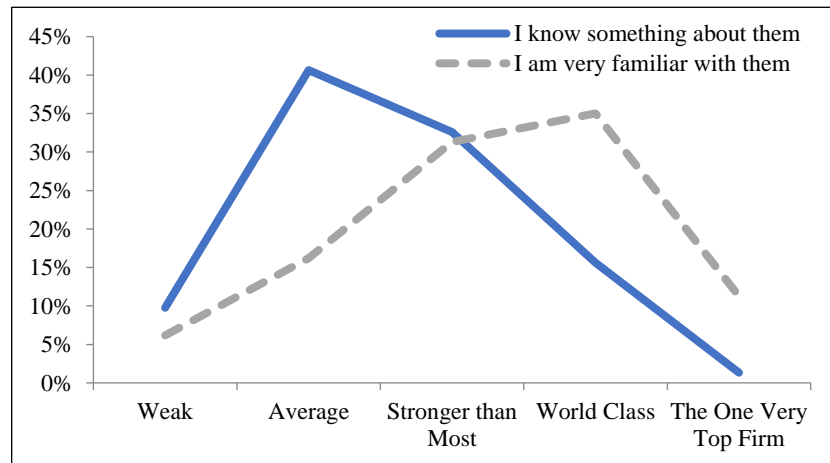


Chart 5

As it turned out, familiarity did not impact the results. For the key client brand respondents were more likely to give lower ratings for that brand when asked in the context of more brands, even when controlling for familiarity. Chart 6 shows the differences in ratings for the main brand, according to levels of familiarity. Observe the six brand very familiar line (long dashes). It is noticeably offset (and significantly different) from the three-brand very familiar distribution.

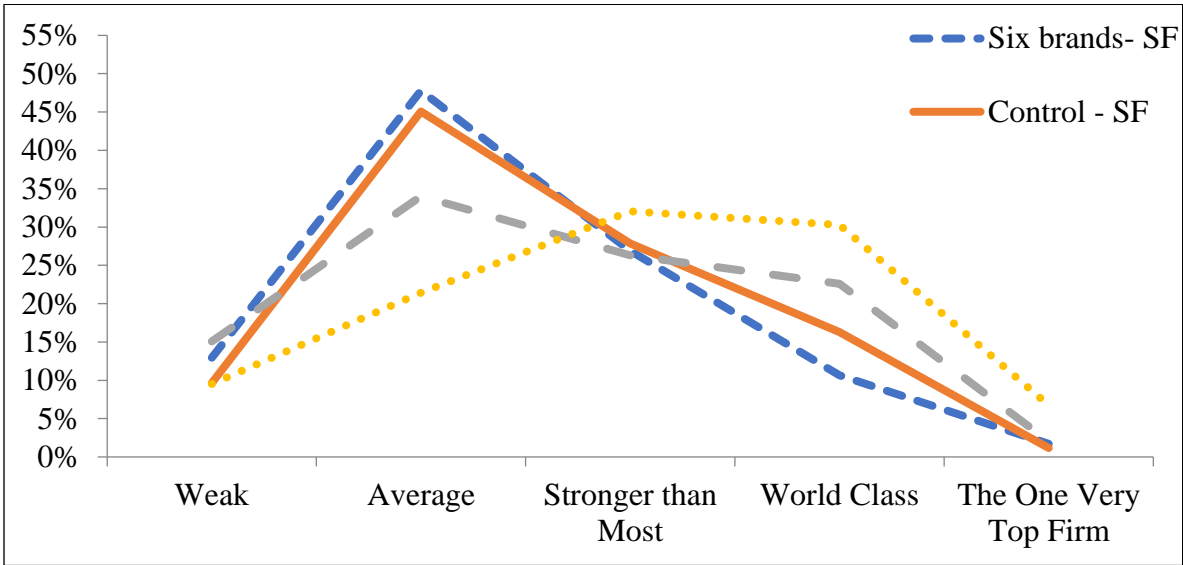


Chart 6

Finally, comparing apples to apples, the client brand's mean ratings (see Hays note above) were statistically higher on 11 out of 16 brands when compared to a smaller group of competitors. This was evaluating specifically those who would have rated more brands had they had the opportunity in the control group vs. those who did rate more than three in the test group.

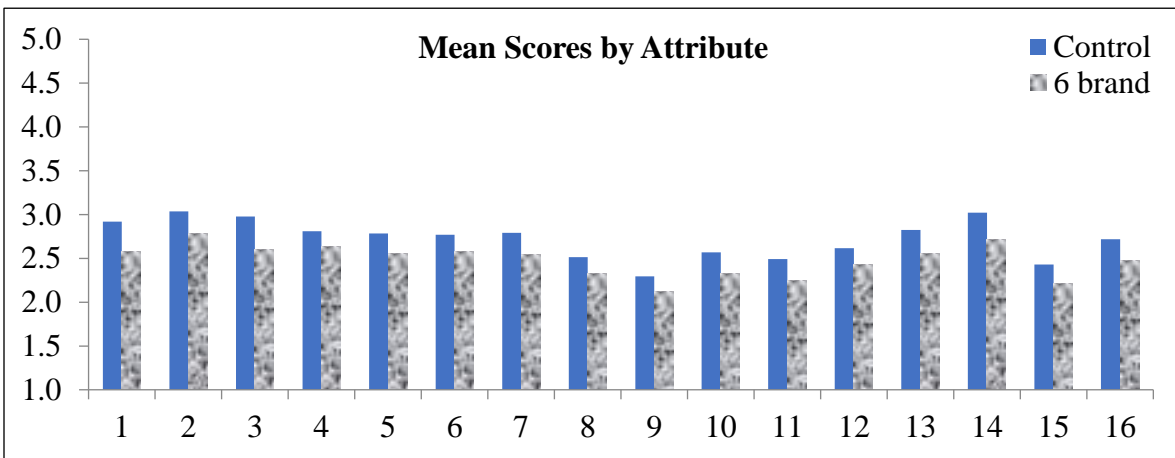


Chart 7

Conclusion

We cannot willy-nilly change question structure even if it appears on the surface to be collecting the same information. If we want to track results over time or compare against other studies, we need to ensure that the question is formatted in the same way consistently from time period to time period or from study to study. This also has strong implications for the client brand. The results could be unintentionally manipulated just by adding or subtracting a brand to the brand list. A shorter list leads to higher ratings ...

are respondents more thoughtful, less overwhelmed, feeling more favorable towards the brand or the questionnaire when not asked to rate as many? Or in view of a larger list does the entire industry look the same? The 'why' is missing without further investigation.