

Questions & Answers:

Selected Articles from *Research Design Review*

Margaret R. Roller

Research Design Review is an online blog that was first published in November 2009. RDR currently includes nearly 50 posts concerning quantitative and qualitative research design issues. This paper is a compilation of seven selected articles from RDR specific to question design and the peculiarities of responses. It is impossible to design an effective research question without a complete understanding of the quality of response it elicits. The importance of good question design and its impact on response cannot be underestimated. It is, after all, the only thing that ultimately enables the researcher to make sound conclusions from the research data.

Roller Marketing Research

www.rollerresearch.com

rmr@rollerresearch.com

January 2012

Table of Contents

Integrating Cognitive Interviewing into Research Design	1
The Net Promoter® Score: The Burden of Many Questions Within Its Single Question.....	3
The Vagueness of Our Terms: Are Positive Responses Really That Positive?.....	5
The Questions of Race & Ethnicity.....	7
The Question of Happiness.....	10
Ask Someone a Question, You'll Get an Answer.....	11
The Quality of Responses to Open-ended Questions.....	13

Integrating Cognitive Interviewing into Research Design

January 15, 2012

If maximizing our understanding of how people think is fundamental to research design – a common theme throughout *Research Design Review* – then why is so little attention paid to the idea that thinking is not stagnant but something that is continually changing from moment to moment. If I ask a survey respondent to name the primary reason she likes store A over store B, the response may be something entirely different than if I ask the same question the following day, or possibly even later the same day. And if I ask how many miles you drive to the office each day, you might say 10 miles today but 15-20 miles if I ask the same question tomorrow.



Why is that? In the February 2010 RDR post, [“Qualitative Research & Thinking About How People Think.”](#) I reference the four-step cognitive process (posited by others) considered necessary to respond “optimally” to research questions: interpretation, searching for relevant information, integration towards a judgment, and translation of a judgment into a response. These four stages alone suggest that isolating the key reason for choosing store A over store B may be more complex, requiring more thoughtful contemplation than the quick response researchers typically encourage in order to keep the interview to a manageable length. And that is why my judgment about how many miles I drive to work may change depending on whether I have included in my search for relevant information (the second stage of the cognitive process) the shortcut I take on certain days or whether I calculated the door-to-door distance versus some other parameter. The judgments derived by research respondents can change from moment to moment for any number of reasons.

George Bishop and Stephen Mockabee in their December 2011 piece [“Comparability of Measurement in Public Opinion Polls”](#) discuss the “peril” of “ignoring the incomparable responses given to most survey questions,” and particularly “how the meaning-and-interpretation of survey questions can vary across respondents and over time even when the wording and context of the question itself remains identical.” Meanings and interpretations (i.e., judgments) are continually changing within and across respondents depending on the social, economic, and political environment at any point in time. As one solution, Bishop and Mockabee encourage the use of random probes to reveal (and track) the particular meanings respondents give to question wording.

In addition to social-economic-political contexts, the ever-changing landscape that gives birth to respondents’ judgments is equally impacted by the human condition. Question interpretation, information retrieval, judgment formulation, and ultimate response may shift depending on a person’s general mood, physical condition, or surrounding stimulus and external cues. All of these factors impinge on how we think and why identical questions at two different moments in time can elicit unequal responses.

All of this suggests that there is no truth to be gained from our respondents, just a judgment at one point in time. Yet there *is* an important truth to be found in the meanings and interpretations that respondents give to our questions at the moment of asking. Similar to Bishop and Mockabee, I believe that by building [cognitive interviewing](#) into our research designs – not just during the pre-

field, testing phase but for the entirety of fielding – researchers will not only discover the basis from which responses are given but also gain the ability to segment respondents beyond the typical demographic and lifestyle characteristics to include how respondents group based on their interpretations of research questions. And, as Bishop and Mockabee state, this would enable researchers to track fluctuations in meanings over time.

The Net Promoter® Score: The Burden of Many Questions Within Its Single Question

August 14, 2011

The summer 2011 issue of AMA's [Marketing Research](#) magazine includes two articles that discuss the ever-popular whipping boy of marketing research, the [Net Promoter® Score](#) (NPS). Randy Hanson ("[Life After NPS](#)") as well as Patrick Barwise and Seán Meehan ("[Exploiting Customer Dissatisfaction](#)") evaluate both the positive and not-so-positive attributes of NPS, along with their ideas for enhancing or actually circumventing the NPS model.

As most researchers know, NPS delivers a metric derived from responses to a single survey question –

How likely is it that you would recommend this company to a friend or colleague?

The NPS score is derived by subtracting the percentage of “detractors” (i.e., respondents who answer this question anywhere from 0-6 on a 10-point scale) from the percentage of “promoters” (i.e., respondents who answer either ‘9’ or ‘10’). Fred Reichheld, the developer of NPS, writes about the virtues of NPS in his book [The Ultimate Question](#). Reichheld asserts that the value of his ‘recommend question’ is that it focuses on behavior (“what customers would actually do”) and separates out drivers of “good profit” from “bad profit” thereby leading companies to future growth. As Reichheld puts it, the NPS metric produced from this one question is the “one number you need to grow.”

Hanson and Barwise/Meehan discuss many of the usual benefits associated with NPS – e.g., it is “intuitive” and easy to understand, and the built-in simplicity of the model—the single question, the simple calculation, the output of a single number—serves to gain the attention of top management who might otherwise ignore survey data – along with the oft-mentioned drawbacks – e.g., it is overly simplistic, reducing complex behavior and attitudes to a single question/number, and it is not widely correlated with its chief raison d’être, predicting growth.

These discussions leave out another all-important downside to the NPS. Namely, the NPS recommend question is frequently *not* a single question. While it may appear as a single, simple request, the recommend question is in reality embedded with multiple questions, each of which tugs at the respondent who weighs its appropriateness for the response. Entrenched in the recommend question are the questions of:

Who – Would I recommend this company to my best friend or to people [such as those at the office] who are friends but not close friends? Should I include my mother who I often think of as my best friend?



What – Under what circumstances would I recommend this company? If my “friend” needed one type of service or product from this company, I would give a high recommend rating; but if my “friend” needed something else, I would respond with a lower recommend rating. For instance, my bank offers great in-branch service as well as above-par rates on certificates of deposit but its online banking system is cumbersome and the standard checking account is laden with fees.

When – At what point in time should I base this response? Am I basing this recommendation on just one specific instance and not the other times I have purchased from this company? How can I honestly answer if I’m asked to base my answer on my most recent purchase which is not indicative of my overall experience with this company?

Like the donkey in *Shrek*, each of these sub-questions is shouting “pick me, pick me,” tormenting the respondent into either: a) opting for one scenario while ignoring all other possible situations (e.g., highly recommending my bank because my “friend” only cares about getting a good rate on a CD), or b) giving up and abandoning the survey.

My choice is typically to give up. Rather than muddy the researcher’s results with what amounts to a half-answer, I opt to drop out when confronted with this question as a survey taker. Because if you ask me if I would recommend Starbucks to a friend or colleague, I am thinking about: who to consider as a “friend or colleague”; whether this person actually drinks coffee or tea; how I really like Starbucks’ Caffè Mocha but not a fan of their cappuccino; whether this person likes Caffè Mocha, cappuccino, or neither; and how I received great service along with a great Caffè Mocha the last time I was in Starbucks but two earlier visits were disappointing with slow, unfriendly service and a Caffè Mocha that was mediocre.

So, should I answer ‘9 or ‘10’ and be categorized as a promoter, or give a rating somewhere between ‘0’ and ‘6’ and be labeled a detractor, or respond with a ‘7’ or ‘8’ and be branded “passive” – a “satisfied but unenthusiastic” customer? Or should I just not answer.

Reichheld states that “this single [recommendation] question allows companies to track promoters and detractors, producing a clear measure of an organization’s performance through its customers’ eyes.” It would be one thing if the NPS was actually asking just a single question. But the invisible questions that lie beneath Reichheld’s “ultimate question” present a real design issue. Requiring survey participants to consider various scenarios (undetected by the researcher) is both confusing and frustrating for the respondent and, of course, impossible to analyze.

The Vagueness of Our Terms: Are Positive Responses Really That Positive?

May 19, 2011

[John Tarnai](#), [Danna Moore](#), and [Marion Schultz](#) from Washington State University presented a poster at the recent [AAPOR](#) conference in Phoenix titled, “Evaluating the Meaning of Vague Quantifier Terms in Questionnaires.” Their research began with the premise that “many questionnaires use vague response terms, such as ‘most’, ‘some’, ‘a few’ and survey results are analyzed as if these terms have the same meaning for most people.” John and his team have it

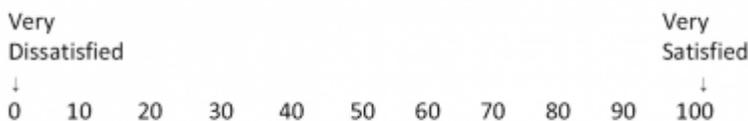


absolutely right. Quantitative researchers routinely design their scales while casting only a casual eye on the obvious subjectivity – varying among respondents, analytical researchers, and users of the research – built into their structured measurements.

One piece of the Tarnai, et al research asked residents of Washington State about the likelihood that they will face “financial difficulties in the year ahead.” The question was asked using a four-point

scale – very likely, somewhat likely, somewhat unlikely, and very unlikely – followed by a companion question that asked for a “percent from 0% to 100% that estimates the likelihood that you will have financial difficulties in the year ahead.” While the results show medians that “make sense” – e.g., the median percent associated with “very likely” is 80%, the median for “very unlikely” is 0% – it is the spread of percent associations that is interesting. For instance, some people who answered “very likely” also said that there was a 100% chance of financial difficulties but other people who answered “very likely” stated that the likelihood of being in financial difficulty was *less than 50%*. And the respondents who responded “somewhat likely” indicated that their chance of facing financial difficulty was anywhere from 100% to as low as 0%! So what on earth does “somewhat likely” actually mean?

I returned from the AAPOR conference and found in my mailbox a customer satisfaction survey from [Amtrak](#). This questionnaire is comprised of 36 rating questions *all but three of which* involve 11-point scales that look something like this:



So here is a question design that actually defines the vague term “very” by telling the respondent that “very satisfied” is equal to 100 and “very dissatisfied” is equal to 0. The problem of

vague terms looks like it’s solved. Or is it? As a respondent I found the design disconcerting. It almost felt like a double-barreled question asking, on the one hand, if I was satisfied or dissatisfied with the train service and, on the other hand, if I was 100% satisfied, 90%, 20%? Frankly, I was stopped in my tracks (pun intended).

Maybe there is some value in the vagueness of our terms, at least from the standpoint of the respondent. Maybe it is a good thing allowing respondents to answer from their own understanding of terms, saving them the anguish of fixing their satisfaction to a metric. But if the researcher is to have any hope of providing usable data to the client, attention has to be paid to clarifying survey responses. After all, wouldn't you want to know if the finding that 90% of your customers are "very likely" to buy from you again *really* means there is a 50-50 chance of a repeat purchase?

The Questions of Race & Ethnicity

March 31, 2011



The race and ethnicity questions in research design have been discussed and debated for many years. Two important issues in the debate are a) should the (esp., race) question even be asked? and b) if so, what is the best way to word these questions to gain some true (objective or subjective) measure. [Rubén Rumbaut](#) addressed the first issue in his [recent interview](#) on NPR's *Morning Edition* (as well as his 2009 paper, [“Pigments of our imagination: On the racialization and racial identities of ‘Hispanics’ and ‘Latinos’”](#)) where he maintains that asking the race question is more about “social status” and putting people “in their place” than about understanding “a natural,

fixed marker...inherent in human bodies.” In a similar vein, fellow research blogger Jeffrey Henning concludes his [February 2009](#) post on race and ethnicity by saying, “I look forward to the day when race and ethnicity are no longer standard demographic questions, and when skin color is no more important than hair color or eye color.”

If race and ethnicity questions *are* to be asked (and there are reasons for asking these questions, particularly when civil rights and money are involved), how should we ask them? It is generally accepted practice in the federal government – Census Bureau, FDA, Department of Education, Department of Defense – to distinguish between race and ethnicity, acknowledging that race refers to a population's physical characteristics while ethnicity speaks to [“social groups with a shared history, sense of identity, geography, and cultural roots.”](#) This is especially important as Hispanics become a commanding segment in the U.S. population. As anyone who completed last year's Census form can tell you, the convention is to first ask the ethnicity question – “Is this person of Hispanic, Latino, or Spanish origin?” – followed by the race question – “What is this person's race?”

Outside the federal government, the thinking behind race-ethnicity question design seems murky. In some instances no distinction is made between race and ethnicity*:

To which racial or ethnic group(s) do you *most* identify?

African-American (non-Hispanic)

Asian/Pacific Islanders

Caucasian (non-Hispanic)

Latino or Hispanic

Native American or Aleut

Other

Would you describe yourself as:

American Indian/Native American

Asian

Black/African American

Hispanic/Latino

White/Caucasian

Pacific Islander

Other

How would you classify yourself?

Arab

Asian/Pacific Islander

Black

Caucasian/White

Hispanic

Latino

Multiracial

Other

While others are confused by what is “race” and what is “ethnicity”*:

What is your race?

White

White, non-Hispanic

African American

Hispanic

Asian/Pacific Islander

Native American

What is your ethnicity?

African American

Asian

Hispanic

Pacific Islander

White

And [eHow](#) explains that race and ethnicity can impact survey responses citing the example that, “Someone raised in a Native American culture may respond to questions about topics such as the environment differently from someone raised in a city culture.” So, “city culture” is an ethnicity category?

Every researcher on every study needs to decide whether the race-ethnicity questions provide meaningful, actionable information. If these questions are deemed appropriate, careful design considerations are imperative to giving the respondent a clear path to response.

** Examples taken from actual survey questionnaires*

The Question of Happiness

December 21, 2010

What a great time of year to think about happiness. We can think about our own happiness, we can think about others' happiness, but happiness is really nonexistent if not for research design. A short look around the Web will prove the point that there are as many ways to think about the happiness question as there are positive psychologists, health providers, and a variety of researchers interested in our well-being. Here are just a few examples:

- [Todd Kashdan](#) tells us that the key to happiness is “harnessing and intensifying our curiosity” and that “curious people live longer and live better.” The happiness question in this case might be: ‘How open are you to new experiences?’
- A [study](#) conducted by researchers at the University of Chicago and Shanghai Jiaotong University links happiness to “busyness,” i.e., people who stay busy are happier than people who are idle (“people dread idleness”). So the happiness question could be, ‘What has been your degree of involvement with activities or tasks in the past week?’
- Psychologist [Harry Reis](#) and his colleagues have concluded from [their research](#) that sharing good news with other people can increase our pleasure (happiness) with the events as well as our personal relationships. So, if we want to assess happiness, maybe we should ask, ‘When and how often do you share news about something good in your life with another person?’
- [Stephanie Rosenbloom](#), from the New York Times, tells us in her article, “[But Will It Make You Happy?](#)”, that the road to happiness is lined with no more than 100 personal items, experiences (travel, hiking, entertainment) not things, *anticipation* of purchases not the purchases themselves, strong personal relationships, and a disregard for one-upmanship (keeping up with the Joneses). In this case, the appropriate happiness questions might be: ‘How many personal items do you own?’ How big is your house?’ How often do you go on vacation and for how long?’
- Geography researchers at the [London School of Economics](#) are looking at the relationship between how people feel and their environment. [Mappiness.org.uk](#) enables participants to download a free app to their iPhone from which they periodically report on their surroundings as well as how they are feeling at that moment. How happy am I because of where I am?
- [Tom Anderson](#) recently asked a happiness question on his [LinkedIn discussion group](#). The question was, “How happy/satisfied are you with your current job?” So, is happiness the same as satisfaction? If I am satisfied, am I happy?
- The [Gallup World Poll](#) finds a relationship between “life satisfaction,” happiness, and cold, hard cash (i.e., income). But Gallup makes the greatest admission of all when it states that “happiness is elusive to define,” requiring more than a measure of wealth but [six distinct indices of well-being](#) – ranging from emotional health to the necessities of life – as well.

So, are you happy?

Ask Someone a Question, You'll Get an Answer

April 30, 2010

A recent online survey questionnaire asked me to compare my salary in 2009 with that in 2010. Because the question was impossible to answer (I won't know my 2010 salary for another eight months), I didn't respond and attempted to move forward to the next question. Unfortunately the salary question was mandatory and so I was left with the choice of either abandoning the study or falsifying a response in order to qualify for the next question. I opted for the latter (I was keen on learning where the battery of questions would take me), then immediately upon submitting my completion sent a message to the survey's sponsor alerting him to "a real problem" with the salary question and its potential impact on the final data. Online research is replete with examples of questionnaire designs that force respondents into a response when a response is either not possible (given the inadequacy of the answer options) or extremely difficult (given the ambiguity or confusion with the question being asked). Whatever the reason, these types of must-answer questions contribute unnecessary error to the design and ultimate study findings.

However, as much as I am annoyed by online programmers' insistence on a response to unanswerable questions – along with the irritating error messages that attempt to coerce a response – I admit that forcing respondents to re-examine a question and think it through for a second time is not a bad thing. While I would argue that a second consideration still leaves the respondent with a dilemma (having to choose among inappropriate answer options or having to respond to a question that is not understood), at least the survey-taker is given the opportunity to reflect more closely and *possibly* make a response that approximates reality. Having said that, I would contend that it is more likely that the respondent will simply drop out of the survey or do exactly what I did – that is, pick an answer option based on no truth (in my case, I chose a neutral response) and move on.

The incentives and other motivations that we build into our designs make it likely that many respondents will choose to falsify their responses rather than dropping out of the study altogether. But regardless of what is driving people to respond one way or another the fact remains that – if you ask someone a question you will get an answer, any answer. This is a tenet that I and other researchers live by. It is not good enough to just ask a question, because you will for sure get an answer and it may not be what you bargained for. Getting responses to our questions can be pretty exciting; but once you realize that people will answer questions no matter how they are structured – confusing, misleading, nonsensical – it should give the researcher serious pause and spur deep consideration of question construction.

Even if the answer options are adequate and the question itself is clear, an added burden on the question designer is to keep the question neutral in tone or at least non-biasing. I recently received via USPS a "2010 Congressional District Census" from the Republican Party (Note: I do not want to talk politics here, but let it be said that I don't recall ever getting anything from the Republican Party in the past). There are lots I could complain about this direct-mail piece – not the least of which is frugging – but the "survey" questions are particularly interesting. One question reads,

“Do you think the record trillion dollar federal deficit the Democrats are creating with their out-of-control spending is going to have disastrous consequences for our nation?” How neutral is that? And yet I am fairly sure that the Republican Party will get the answer they are looking for, not only from Republicans but by affiliates from other parties as well who, with good intentions, simply answered the question they were asked.

So, be careful out there – because if you ask someone a question, you may very well get an answer.

The Quality of Responses to Open-ended Questions

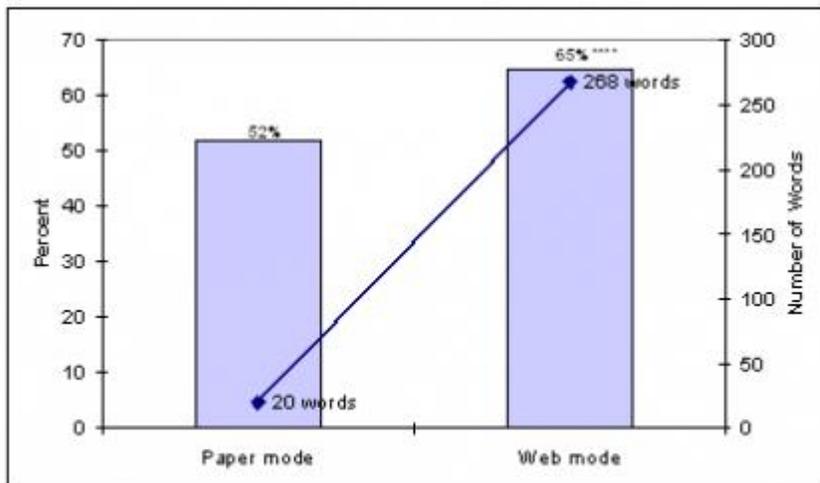
February 28, 2010

A few years ago I teamed up with Linelle Blais (my client at the American Cancer Society) to conduct a test of online versus traditional paper survey mode effects. At issue was whether the Volunteer Satisfaction Study, which we had been conducting for many years nationwide, could be converted from a self-administered paper questionnaire delivered via USPS to an online format. By shifting to online we hoped to save the Society money as well as provide a faster turnaround of the findings. This experiment set out to determine: the rate at which volunteers would respond to a Web survey; the rate at which sub-groups would respond; the completeness of volunteers' responses; and, the degree to which item responses vary as a result of Web vs. paper completion.

A fairly complete account of this research was published in *Quirk's* last year (["A Volunteered Response"](#)), and much of the following comments are taken from this article.

Among other things this research confirmed what researchers elsewhere had found, including a lower response rate on the Web but also a lower rate of item nonresponse. Interestingly, respondents to the Web survey not only answered more questions but also were significantly more likely to respond to the open-ended question asking for their suggestions to improve volunteer satisfaction. Sixty-five percent (65%) of the volunteers in the Web mode answered this question compared to 52% of the volunteers in the paper mode. The sentiment of these comments (i.e.,

positive vs. negative vs. neutral remarks) did not differ greatly across modes; however, the



Percentage of volunteers providing open-ended comments and the average number of words per comment by mode.
Note. n=590 for paper mode, n=576 for Web mode.
**** $p < .001$.

length of these comments varied hugely by mode. The average word count of comments made by volunteers responding to the online survey was 13 times higher than the word count among volunteers responding on paper – 268 words per comment vs. 20 words per comment, respectively.

There is also some indication that the quality of the open-end comments in the Web mode may be superior to that in the paper mode. A cursory analysis of comments in both modes suggests that comments from Web respondents are more detailed (e.g., references to specific examples or names) and tend to be more constructive (i.e., offer suggestions for improvement) than comments from the paper questionnaire.

The readability of these comments, however, appears to be a different issue. Looking at the readability scores, based on the Flesch-Kincaid Grade Level analysis, the comments from the paper survey read at an 8th grade level while comments from the online respondents read at the 7th grade level. Whether this is a function of the younger age of Web respondents or the informal (even, sloppy) writing style many email users have adopted or something else is left for further research.

Open-ended questions – both in qualitative and quantitative research – are typically added to interview and questionnaire designs to enrich the data, bringing meaningful insight that might otherwise be lost in a completely close-ended approach. Given that the quality of these responses can vary across modes should serve as an important consideration in constructing our research designs as well as in our analyses and interpretation.