Rules of Engagement:

The war against poorly engaged respondents, guidelines for elimination

by Steven H. Gittelman and Elaine Trimarchi

Those that adhere to strict probabilistic models will grimace at the idea of removing respondents subjectively on the weight of an engagement (quality) metric. We propose a simple "better practice" whereby the number of faults derived from a quality metric is coupled with tests of data difference to identify which respondents are to be removed. Most questionnaires have sufficient measures of engagement to deploy an engagement tool and we also recommend using reference points to ground the respondent elimination process.

Introduction

Efforts to reduce the number of poorly engaged respondents have become an industry obsession. Various metrics have been deployed to catch the offenders. Straight lining, trap questions, inconsistencies, and other, more esoteric measures have come to the fore. Users of research are creating standards for eliminating the poorly engaged with haphazard results and often subjective criteria. It is our belief that engagement can be measured through simple metrics that are often innate to the studies we perform.

There is growing sentiment that the unengaged do damage, but who and how many should be eliminated? We have seen clients eliminating some 30% of their sample on the "look and feel" test, while others deem it reasonable and expected that without question the worst 5% should go. The sample sources have an understandable predilection to protect their respondent resources but are faced with the dim prospect of building in a price factor for clients who seemingly revel in their less-than-surgical strikes. Others question the validity of any sample frame which is merely the remainder that survives the clipping. Clearly, in an effort to remove respondents who generate biased data, others with better intent and valid responses can be eliminated by error. In the film "Rules of Engagement" (Paramount, 2000) actor Guy Pearce makes the less than profound statement, "Sometimes you just can't win no matter what you do." Here we take a stab at providing guidelines for when, how many and who should be left on the cutting room floor. No matter what the outcome, it is our intent to diffuse a growing atmosphere of tension between researchers and their sample suppliers.

Disengaged respondents

We in market research who conduct online research are confronted by disengaged respondents that represent a force that is still poorly defined. Our lives are complicated by the absence of clear guidelines as to which respondents should be kept in the data set and which must be removed. Without such rules, the data that we hold so sacred is at risk. Perfectly acceptable respondents could be removed based on insufficient criteria or meaningless responses could be included due to insufficient safeguards.

We advocate proper treatment of respondents in all respects. There is an abundance of literature referencing the myriad ways that we could do better by those so essential to completing online surveys (e.g. Du Jong, 2010, MarketTools, 2010). This is an old story that needs constant attention. Our surveys are convoluted, long, boring, and on esoteric subjects. Respondents are enticed with robust incentives and sent packing with lotteries. We screen them in a never ending sequence of questions until we can fit them to our purposes. Sadly, we get what we encourage.

When it is all said and done there are often respondents who "satisfice" their way through surveys. They provide us with little or no attention to the questions that we offer them and instead "complete" the task with less attention than is required. In the end we must establish rules for the elimination of respondents whose efforts, or lack thereof, are not fit for our purposes.

As researchers, we must understand the potential impact of poorly engaged respondents on our work and take appropriate action, but there exists no guideline for making such decisions although some elegant methods have been suggested (Garland, 2012). Previously, (Gittelman and Trimarchi, 2012) provided an exhaustive overview of global online panels and the degree to which they had demonstrated consistent results in a multi-wave tracking study. Those that did not prove consistent appeared to have high levels of poorly engaged respondents.

Metrics for calibrating engagement levels have been increasing in use during the past three years and the respondents condemned by them have at times been deemed worthless (Courtright and Brien, 2009). Our *QMetrics™* (Gittelman and Trimarchi, 2009 b) has evolved over the years into a fluid model that we adapt to the survey at hand. With some changes, we adopted QMetrics™ to fit the test instrument that we have deployed in 35 countries where we audited over 300 online panels as part of the *Grand Mean Project™* (Gittelman & Trimarchi, 2009 c). Respondents were subjected to two consistency questions, one trap question, and an analysis of speeding and straight lining (also known as non-differentiation). Our questionnaire is a diagnostic tool intended to measure ten largely behavioral segmentations relating to purchasing behavior, psychographics, and media, as well as a number of market-specific classifications. In addition, it diagnoses problems in questionnaire execution and the frequency of panel membership/survey taking. If a satisficing metric is deployed, those who are identified as the worst respondents may or may not provide data different from that collected in the balance of the sample, that is, from those who are not satisficing. Inherently, satisficers that do not influence the data are less damaging than those who are driving severe data changes. As a general rule, respondents who fail quality metrics but are not providing different data from the mainstream should not be removed with the same broad cut as those who are. Thus, it is not enough to identify satisificers; one must qualify them as game changers.

Many who were reared on strict adherence to probabilistic models will grimace at the idea of removing respondents subjectively on the weight of an engagement (quality) metric. Here we propose a simple "better practice" whereby the number of faults derived from a quality metric is coupled with tests of data difference (assumed to represent data bias) to identify which respondents are to be removed from a data set. We suggest that most questionnaires have sufficient measures of engagement to deploy an engagement tool. We also recommend the inclusion of various reference points in order to ground the respondent elimination process. (See Chiang & Krosnick, 2009, for a study demonstrating data accuracy through benchmarks.)

Our model is driven by the concept that if it is not broken, don't fix it. The chronic satisficers should be removed, but those who are less severe in their habit and have no apparent impact on data interpretation should have a chance at remaining in the sample frame. The less we manipulate our samples the better, but there is a threshold where too much is just too much and it is toward determining that threshold that we shall direct our attention.

We find that there is a conflict of interests inherent in the system. Not only are the sample sources pitted against the clients but the respondents could be considered to be pitted against all of us. The term satisficing takes its origins from the thought that the degree of cognitive attention applied to a task is variable and by satisficing respondents are applying the amount of attention to our inquisitive devices that they deem satisfactory for their purposes. They may seek to get through a burdensome and poorly written instrument or have no interest in the subject at hand. Accordingly, they decide on the appropriate or satisfactory amount of attention to expend on the task to maximize their reward, be it sharing their opinions or reaping an incentive. We, on the other hand, seek to glean information from those who apply effort sufficient to generate well-considered answers. We can either determine to use their responses or remove them from consideration. The difficulty in this endeavor is that the determination of whether to do so grows dangerous for us as it becomes more subjective. The less satisfactory our standards governing the decision are, the more likely we are to create a sample frame that is biased by our own rather subjective standards. It is thus important to present a uniform and justified approach, as we have attempted to do in the tests and methods presented here.

The engagement tests

No industry-accepted standard exists for the cleaning and editing of online data (Baker et. al. 2010). While the concept of "Real, Live and Engaged" (MarketTools, 2009) has been broadly promoted, the tests of engagement have been allowed to float without formal adoption. For the most part, these tests are easily understood.

The easiest way to identify these respondents is to include measures inside a survey that can indicate aberrant or illogical behavior. In most cases, these indicators are nothing more than suggestive evidence of an underlying issue and the determination of whether an individual is truly "behaving badly" cannot be made based on any one with great certainty. However, when a respondent accumulates many such faults across the entirety of a survey, we can identify him as disengaged with greater confidence. The number of such tests depends on the content of the survey, but the presence of more than three provides sufficient data to create a suitable engagement metric. The ARF utilized a “Bad Behavior Score” comprised of 16 variables (Walker, Petit, Rubenstein, 2009). We propose here that a single variable has use but that four or more is sufficient to create a usable metric. We have purposefully chosen to be nonspecific about the choice of variables that are used in constituting the metric to broaden the applicability of the method to include more studies without having to add time to the questionnaire. Thus, some three fourths of the questionnaires fielded can avail themselves of a four point metric and the rest can get some value of lesser instruments.

Speeding - It makes sense that some respondents take longer to complete a survey than others. In addition to differences in reading comprehension and other distractions in the respondent's environment, skip patterns and different concepts can drastically change the length of the survey he or she sees. However, when comparing sections of a survey where all respondents see the same (or nearly the same) questions on the same topics, those who complete portions in less than half the median time may not be giving sufficient thought to their answers, or may not be reading directions thoroughly.

Open-Ends - Open-ended questions require more effort than most question formats. As such, they represent a more stringent test of engagement. While strongly disengaged respondents may not even enter sentences composed of actual words, those who are only mildly disengaged may give responses that have less meaningful content than the norm (as measured by number of words). Respondents giving less than half the median number of words are marked for a quality fault. Those that provide gibberish should be omitted

Trap Questions - Trap questions are questions to which only a single response is acceptable. A ''directional trap'' directs respondents to provide a specific answer. Failure to do so indicates a failure to read the text of the question. A ''factual trap'' directs a respondent to rate or indicate ownership of a product that does not exist. If they indicate that they have ever used the item in question, it suggests they are not giving truthful answers, possibly to avoid being screened out of the survey.

Inconsistencies - Through the insertion of questions that ask a respondent to answer the same question in two opposite ways, we can tell whether they are sufficiently engaged to represent a coherent body of thought. On a scale question, providing responses in the bottom two or top two categories for both of a pair of logically contradictory questions causes us to mark a respondent as inconsistent.

In the consistency test below, respondents who answer a 6 or 7 to both questions or alternately, a 1 or a 2, would be marked as inconsistent.

Please indicate your level of agreement or disagreement with the following statements:

I am perfectly happy with my standard of living	7 6 5 4 3 2 1

I am not really happy with my standard of living	7 6 5 4 3 2 1

Straight-Lining (Non differentiation) - Long grids or matrix questions provide a tempting opportunity to easily complete a portion of a survey without proper consideration of the attributes within them. In most cases, straight-lining is marked for those respondents who provide identical responses for all attributes on a grid with at least a 5-point scale. However, in instances where respondents are asked to evaluate a combination of negative and positive attributes, we would not expect them to give the same answer throughout. In these cases, responses with less than half the average standard deviation (in other words, an unusually tightly grouped response pattern) are marked as straight-liners. This method ensures that the criteria for straight-lining are sensitive to the likelihood that a tightly-grouped but not identical pattern of responses represents a legitimate body of opinions.

Rare Items – While rare items questions are not generally present in the original draft of surveys, they represent a useful test of respondent attentiveness and honesty. These tests ask respondents to identify which of a series of behaviors they engage in, or which of a series of products they own, with a set of items that are common and a set that are extremely uncommon, either individually or in combination with one another. A general positivity bias among disengaged respondents means that many of the worst respondents can be caught by such a test, possibly because of a desire to qualify for the study, or possibly only because the question was not read thoroughly. We often recommend these in place of more obvious trap questions to avoid respondent disengagement as a result of our evident distrust. In a health study we might add:

Do you or anyone in your household currently take any of the following medications? (randomize attributes)

Vicodin	Yes	No
Prilosec	Yes	No
Claritan	Yes	No
Plavix	Yes	No
Jakafi	Yes	No
Kalydeco	Yes	No
Stablon	Yes	No
Dormalin	Yes	No

The first four items are fairly common, and the last four are experimental drugs or treatments for very rare diseases (the order was randomized for respondents). Respondents who indicate even a single one of the rare items tend to commit other quality faults.

Figure 1: Respondents failing the rare items test show shorter completion times (*t*-test, $p<0.01$).

While it does not serve as a particularly strong test on its own, speeding provides useful corroborative evidence of the effectiveness of the rare items test. Respondents who indicated using even one of the rare items in the example above demonstrated shorter completion times

Creating a metric

For our purposes here, the combination of any of the above engagement tests constitutes the formation of a "metric." We are not advocating any particular combination of tests to form a metric. We only prefer that they are as diverse and numerous as practical, preferably numbering at least four within a survey. Not all questionnaires contain trap questions, logical inconsistencies, rare items tests, or grids that lend themselves for straight lining, but the reliability of an engagement metric is likely to increase as the variety and number of individual measures are increased. As a matter of practice, not all questionnaire writers are willing to surrender valuable survey real estate to quality metrics. The degree to which a researcher will provide space for diagnostics is likely to be correlated with their perceived view of how well such diagnostics work.

As a matter of practice most questionnaires appear to contain multiple diagnostic tools. In a survey of all forty online studies conducted in our fielding facility during July, 2012, 78% were found to contain 4 or more faults that we could use in creating an engagement metric, a number sufficient for our purposes here. Only one of the forty did not lend itself to a speeding test, 85% provided useful material for straightlining, 22% had inconsistencies and about a third had open ended questions suitable for an analysis.

Speeding is always present, often grids for straight lining are evident, and at times representatives of the other categories can be found. The argument for adding an example of the other tests can often be mounted and in some instances won. The point here is that these measures are either innate to questionnaires or can be added easily.

If, in the last resort, the only test available is speeding, then we propose that sections of the questionnaire be timed independently and a metric created. This would clearly be the weakest of tests. A practitioner relying on a speeding only metric should avail himself of the growing literature on the subject (Gittelman & Trimarchi, 2009a, Beckers, Siegers & Kuntz, 2011, Gutierrez et al., 2011)

Determining the rules of engagement: which respondents stay and which go

Beyond creating the metric, the crucial question becomes one of determining which respondents shall be removed. Often, an arbitrary rule is deployed. When it comes to speeding we have seen the quickest 10%, below two standard deviations from the mean, and others in use. Straight lining is often considered to be a series of identical answers or some other measure of small variance. Instead we advocate here that an analysis of how the data changes, both in terms of statistical significance and in substance of meaning, as the number of engagement faults committed by the respondent increases. The intent is to find shifts in the data that corroborate the engagement metric and then remove respondents that are confirmed by both criteria. The following examples are standard consumer studies where we deployed the above method, with no changes needed to the original client script that was provided to us.

Study examples

1. Laundry detergent

In a study on laundry detergent there were six tests native to the questionnaire. There were two speeding sections, two straight-lining tests, an open end test and an inconsistency. Subjectively we determined that the peak at three faults corresponded to a sufficient change in the data to warrant removal of those with three or more faults. The change from 0 faults was greater than 20%. While the study was in progress the determination was made to remove all respondents with three faults and above. Accordingly 5.44% of respondents were replaced. (Figure 2).

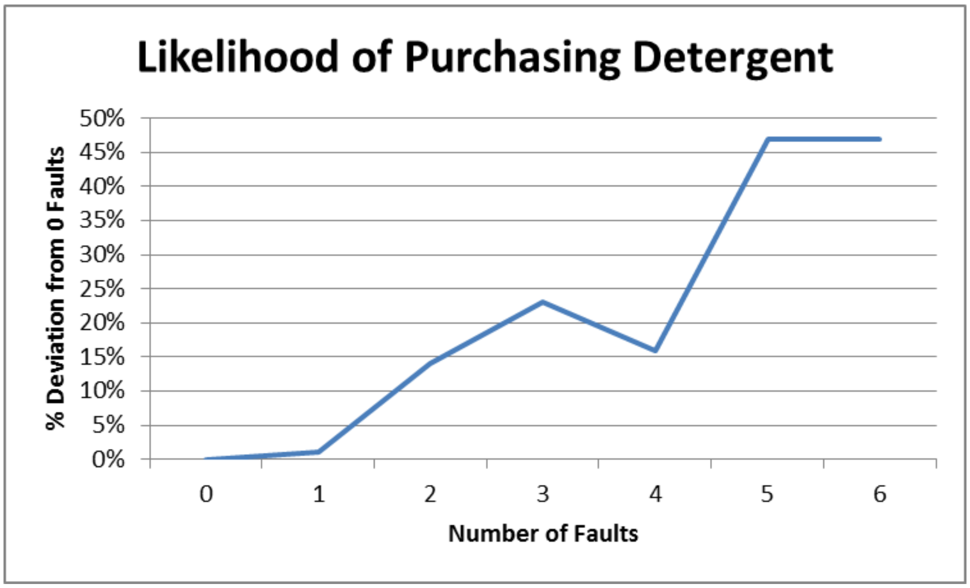


Figure 2. Relationship between laundry purchase intent and engagement faults. Respondents with three or more faults were statistically different (*t-test,* $p<0.01$) and were eliminated from the sample set.

Age, sex, income and race, all had less influence on the likelihood of laundry detergent purchase then did the quality (engagement) of a respondent.   It is not an outcome that is often part of the report given to a client (Figure 3).

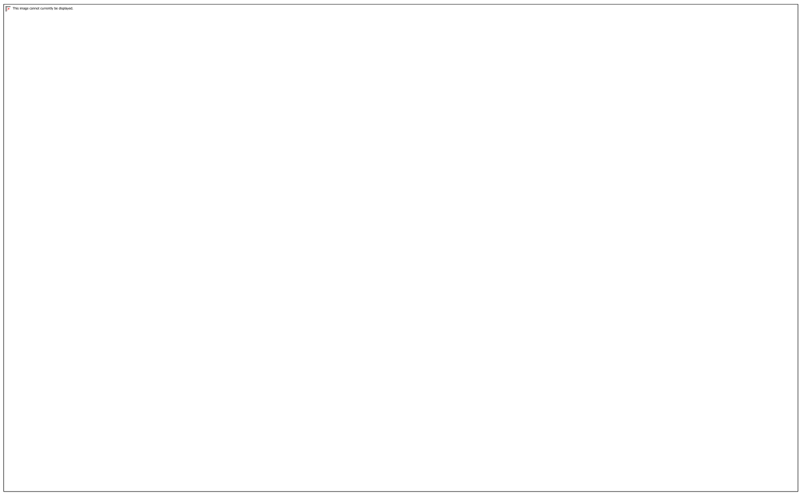


Figure 3. Table 1. Few analysts would disclose to their clients that the predicted likelihood of product purchase was influenced more by the engagement level of the respondent then by age, sex, income and race (Ordinary Least Squares Regression, P<0.01).

Parameter Estimates					
		Parameter	Standard		
Variable	DF	Estimate	Error	t Value	Pr > |t|
Intercept	1	3.76359	0.37875	9.94	<.0001
Age**	1	-0.25651	0.08108	-3.16	0.0016
gender	1	0.04357	0.10356	0.42	0.6741
Engagement**	1	0.17059	0.04205	4.06	<.0001
Income	1	0.02884	0.02236	1.29	0.1976
Hispanic	1	-0.09824	0.15021	-0.65	0.5133

Table 1.

2. Restaurants

In this study, there were a total of eight tests, all innate, with two speeding sections, two straight-lining tests, an open end test, and 3 inconsistencies. Respondents with four faults were determined to have begun satisficing as the survey progressed, while respondents with five or more faults were thought to be disengaged throughout. The client chose only to replace the latter set, which consisted 8.2% of the total completes in the data (Figure 4).

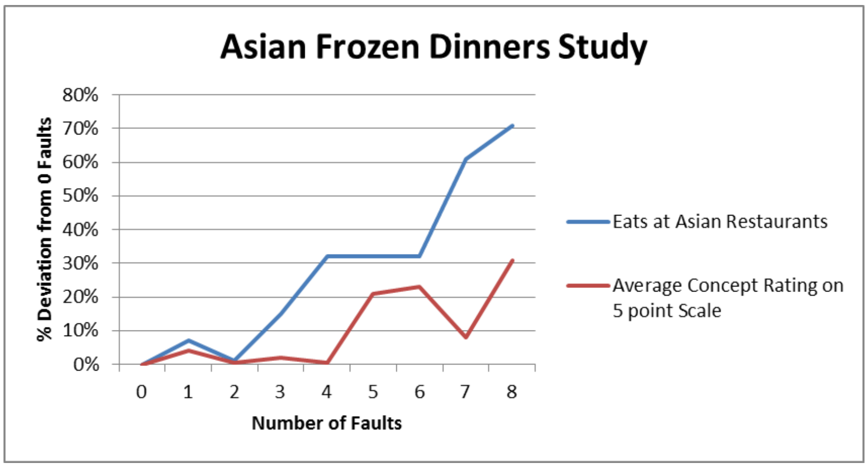


Figure 4. Respondents with 5 quality faults or more rated a food concept differently than those with four or less (*t-test,* $p<0.01$) and were removed upon concurrence with the client.

3. Banking

In this case, respondents were determined on the basis of two different questions to have provided significantly different answers at 3 faults or more on a five point scale. Average ATM withdrawal and mobile banking usage were clearly different at three faults and above. There was no discriminating trend at both online banking and teller usage. The scale consisted of three speeding sections and two straight-lining tests. Accordingly, 6.7% of respondents were found to be problematic and the client chose to have them replaced (Figures 5 and 6).

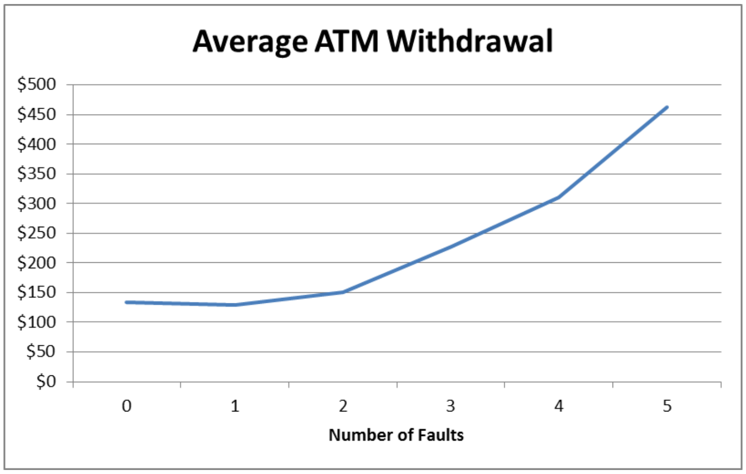


Figure 5. The average ATM withdrawal grew as the number of quality faults increased (*t-test,* p<0.01).

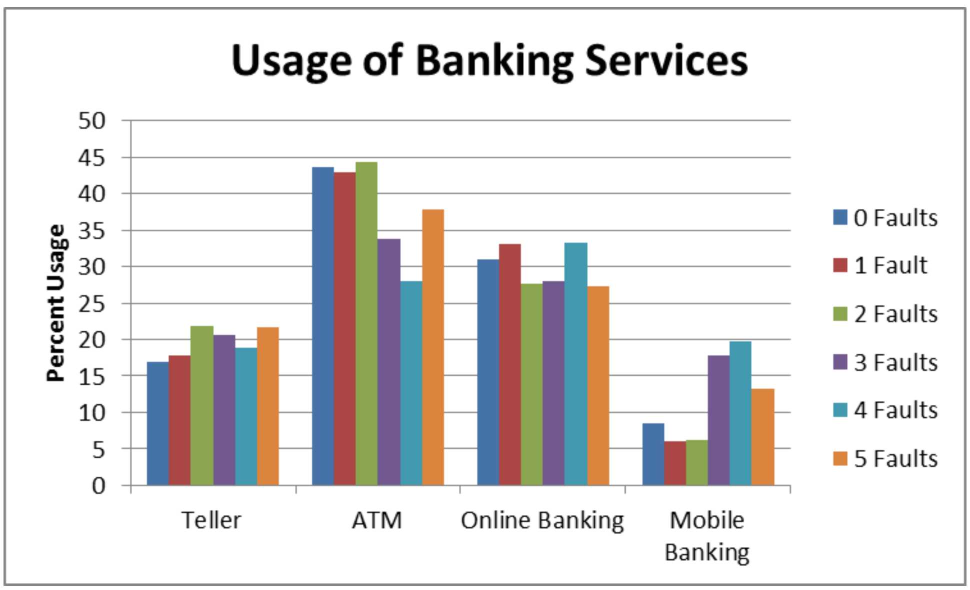


Figure 6. Respondents with three or more faults claimed to use ATM services less and Mobile banking services more(both *t-test,* p<0.01) but teller and online banking the same.

4. Makeup concept

In this case, there were five tests: two speeding sections, two straight-line tests, and one open end. Founded was 6.3% of respondents to be responsible for the obvious jump at 3 faults. The client chose to replace them (Figure 7).

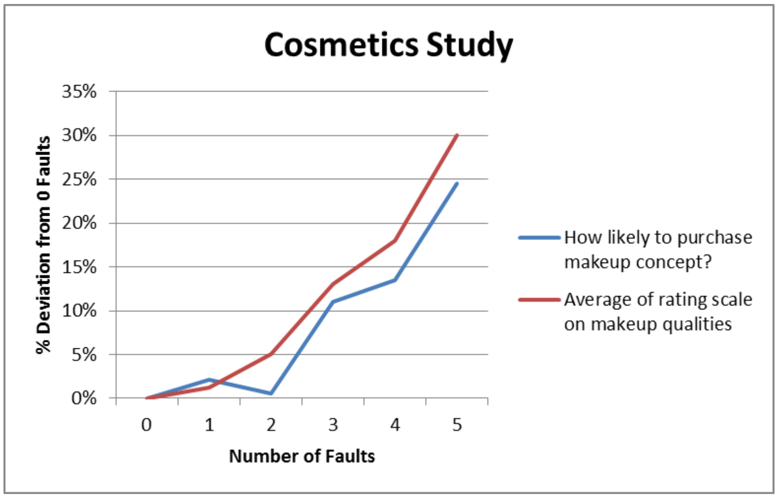


Figure 7. The number of faults corresponded to changes in purchase intent and the quality rating of a makeup concept (both, *t-test,* p<0.01).

5. Liquor

The amount of liquor consumed and the usage category preferred both changed dramatically at four faults. The questionnaire appeared to be popular with respondents resulting in higher engagement. There were six possible faults: two speeding sections and four straight-line tests, though no respondent failed all six. Those falling above 4 faults consisted only 1.6% of the sample, and the client chose to replace them (Figure 8).

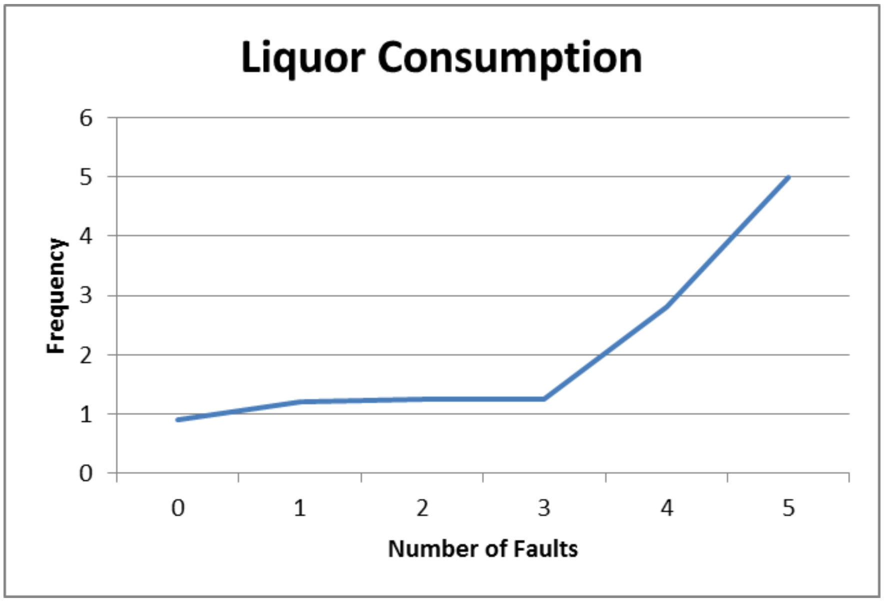


Figure 8. Liquor consumption was influenced by the number of faults. Respondents with four faults or more were eliminated (*t-test,* $p<0.01$).

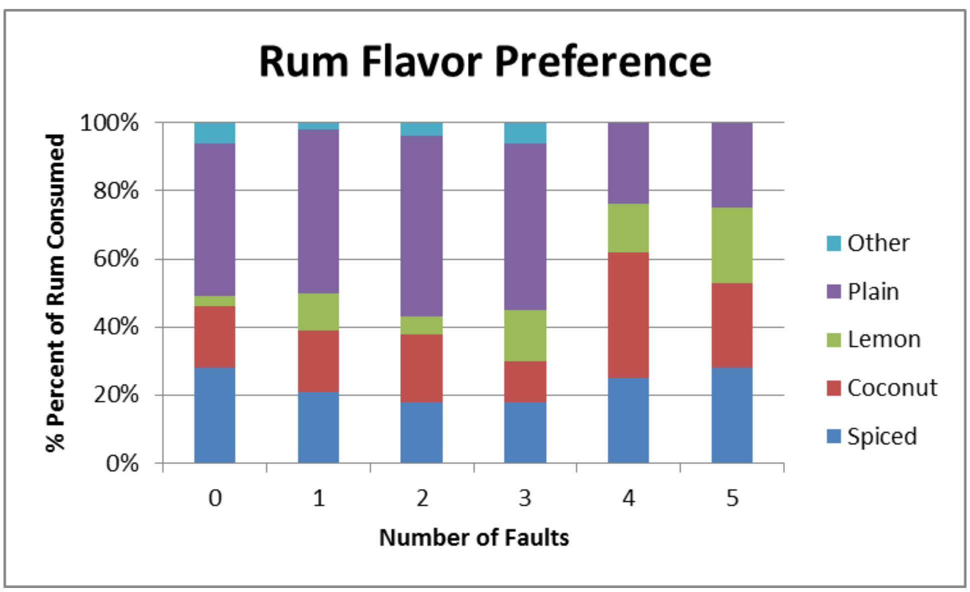


Figure 9. Those with four and five faults preferred different flavors of rum (χ^2-*test,* $p<0.01$).

Grounding our engagement metrics-the use of reference points

The frequency of those smoking a cigarette everyday is a commonly used reference point. Data on the subject can be obtained readily from the CDC. A simple benchmark such as this in a questionnaire could help determine which engagement metric segments should be retained in a data set. We couple the engagement metrics score to data bias and consider for elimination those segments that are most distant from the outside reference point being deployed. American respondents drawn from our consistency research who had three or more quality faults claimed to smoke far more than would have been expected. Respondents with three faults and above would be eliminated from a study sample set. (Figure 10)

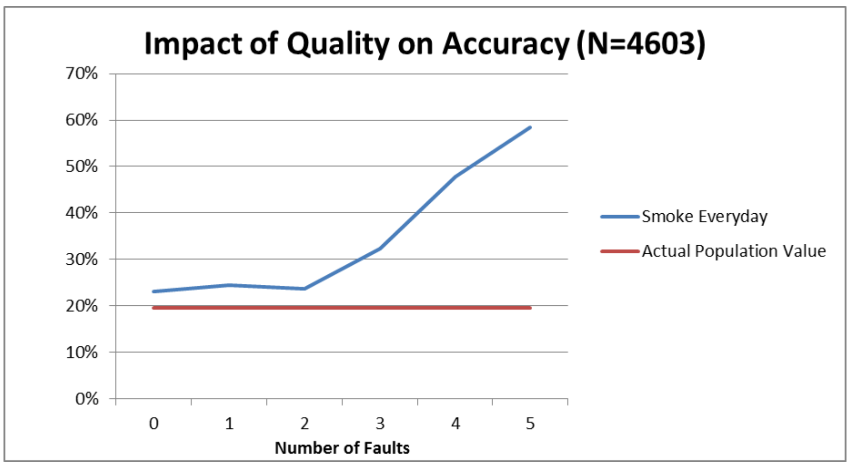


Figure 10. Those with three quality faults and above reported abnormally high levels of smoking. We view this to be an anomaly driven by their predilection to satisfice and recommended their removal from the study. Respondents with two faults differed significantly (t-test significant at $p<.05$) from those with three.

It is not always possible to include a question on smoking. Often the subject matter of the targeted study does not lend itself to the question. In figure 11, we provide an example of where we used ownership of a high definition television set as a discriminating question. The data is not quite as demonstrative as the smoking data. We decided once again to eliminate respondents with three faults or more (Figure 11).

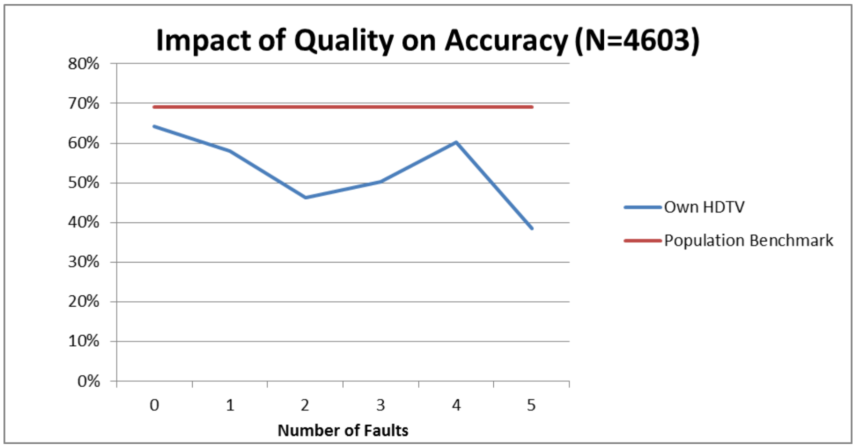


Figure 11. Ownership of an HDTV changes dramatically, although not with pure linearity, in an inverse relationship with the number of faults in our engagement metric. Those with three faults and above were recommended for elimination. (*t*-test significant at p<.01 for 0-1 faults versus five faults)

Those bearing a passport are well documented by the Federal Government thus making passport ownership a reasonable reference point. We find that respondents who are poorly engaged tend to report higher than normal passport ownership. Here, respondents with three faults and above would be considered for elimination (Figure 12).

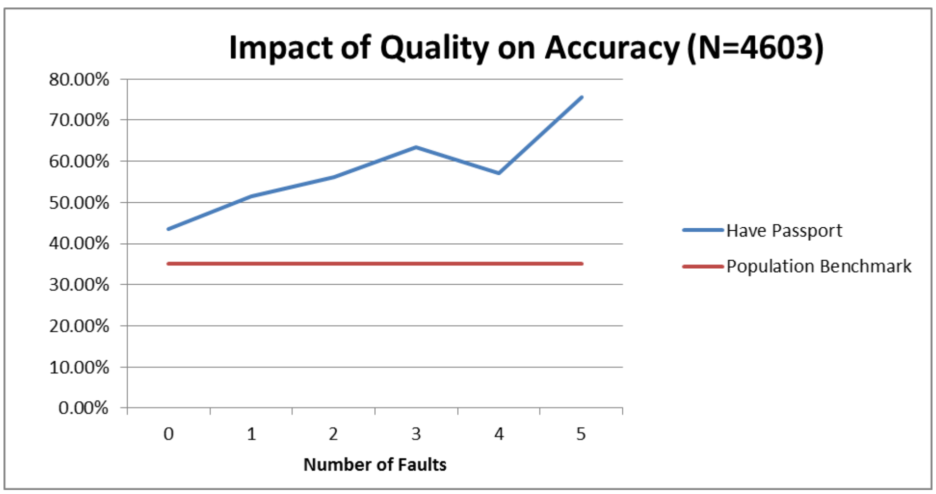


Figure 12. Respondents with even a single fault provide higher than expected passport ownership. Those with two faults differed from those with three significantly. (t-test significant at p<.05) Use of this metric can be used to discriminate between categories of engagement.

At times, it can be argued that we should expect some types of respondents to be poorly engaged. Travelers might be thought to fall into that category. Time restraints of traveling might make them less patient with long surveys. We find that those with three or more faults indicate such high levels of travel that it is sufficient for us to discount the data they provide.

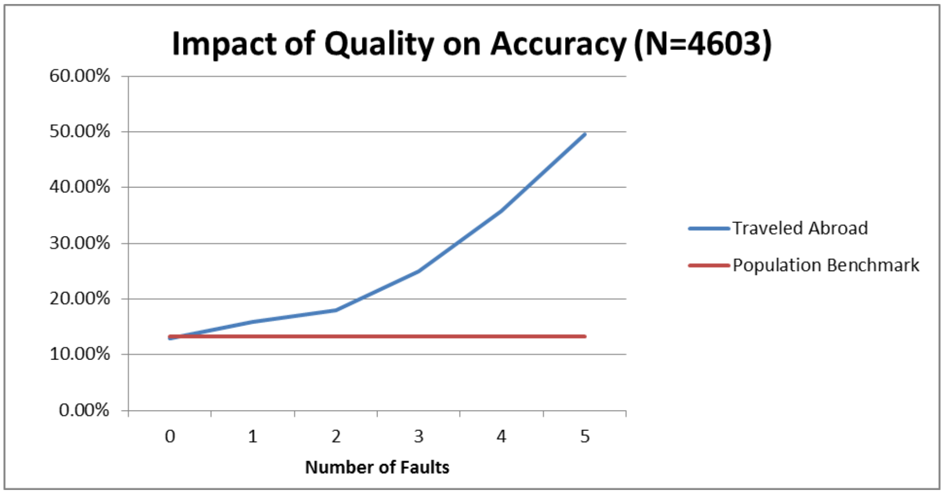


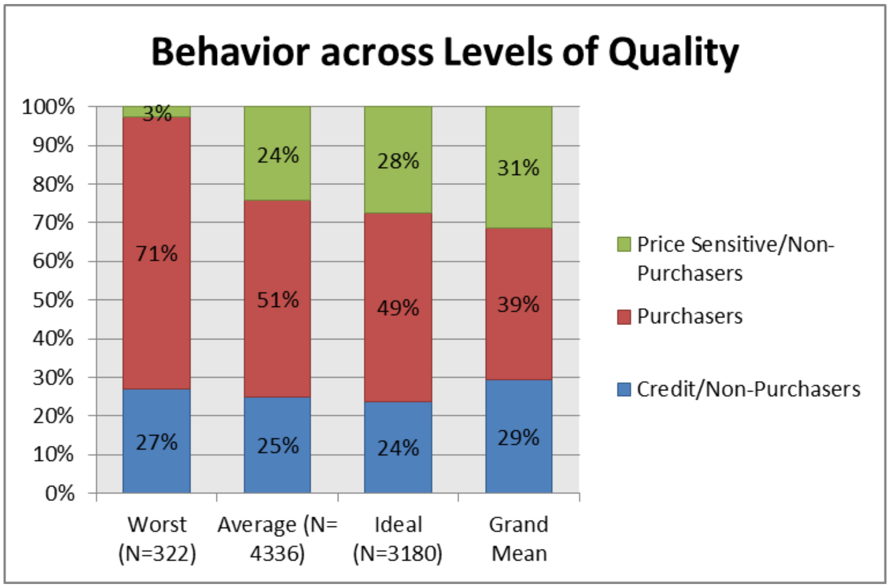
Figure 13. It could be argued that those who travel abroad may have less patience for the interviewing process and thus satisfice. While that may be true, it should give one pause to include respondents with three or more faults, who were significantly different from those with fewer (t-test significant at $p<.01$)

There is a crisis within the crisis. We are most often forced to utilize data internal to the questionnaire and are rarely afforded the luxury of comparisons to outside reference points. In that case, the data collected tends to float without connection to the real world. We tend to have little sense for the accuracy of the data we collect and must then rely on data consistency as a logical fallback. The reference points themselves are often collected through other modes (face to face, mail, telephone, etc.), which may cause differences in itself, and may lead to references being a bit old in a fast changing world.

To meet our own needs for reference points, we conduct the Grand Mean Standard™, a yearly multi-mode study which provides us with a source of reference material to compare current data from any study we might perform. By imbedding questions into a survey, we can conduct the kinds of tests that we believe are needed when calibrating the work of others or testing our own research.

Strategic removal of poorly engaged respondents provides a partial solution for data bias. If the most poorly engaged respondents are removed in favor of average and then best engaged, some data "correction" occurs (see Figure 14). In this case, restoration of the purchasing segment and re-appearance of the price sensitive segment begins to become evident. Poorly engaged respondents appear to have minimal price sensitivity, a situation which is altered on the removal of those respondents. Better engaged respondents provide a closer fit in this data to the Grand Mean Standard which is a combination of phone and online data optimized to a battery of third party reference points. Here we use the distribution of quality segments composed of some 37 questions to calibrate our data.

Figure 14. Poorly engaged respondents with high levels of engagement faults, a large fraction of panelists in this case, have different segment distributions from those who have fewer or no faults. Sample based on 4336 American panel respondents, census balanced.



Conclusion

The first step in dealing with unengaged respondents is to identify them. We argue that most studies have sufficient tools innate to their design that allow us to use a convenience model for a convenience sample. We recognize that there are occasions where questionnaire real estate is at such a premium that adding even a single external item appears out of the question.

Our first rule of engagement is to move forward with a metric custom crafted for each study whose origins are from the existing questionnaire. In our test sample of forty studies, 78% afforded us four usable data points to create a serviceable metric. In those cases, little is to be lost and much to be gained by examining the poorly engaged respondents and the data differences inherent in their numbers. We believe it should be a routine practice. Some practitioners will resist, but in the end, the benefits of removing unengaged respondents who are providing biased data should be self-evident. We cannot deny that there will be situations where the respondents removed create a bias of a different kind, but we are convinced that the way in which they taint our data now must be dealt with.

It is our view that this is an early step in a process toward the adoption of multivariate and uniform engagement models that have passed the rigor of analysis. Time is of the essence; until better practices are at hand we see the need for more easily deployed means. Further, simplicity of design is a high order requirement, as data cleaning should be done in real time to allow the replacement of respondents that are deleted.

As our quality culture prospers we hope to see reference points incorporated that provide a new grounding of our data. We are confronted by the shortage of such reference points, however and call for the industry to create its own. In our case it is the Grand Mean Standard, in yours it could be anything of your choosing. Whatever it is, we find the need for better focus of proper rules of engagement.

The deployment of engagement models in convenience sample cleaning should become routine practice. The tools required are simple to understand, flexible and require little effort. In most cases surveys require no modification, in some slight additions would go a long way. There is a particularly bright future for ad hoc applications of this type, especially if a few strategic tests are routinely planned into questionnaire design. The addition of a rare items test or an inconsistency would improve precision in most cases. In time, we are confident that practitioners will find they agree with us that no study should go without real time removal of the poorly engaged.

References

Baker, Reg, Simon Chadwick, and Joel Rubinson.(2009). "Navigating the Research Quality Maze: A Complete Review of What Is Available Today." Presented at Peanut Labs Webinar.

Beckers, T., P. Siegers, and A. Kuntz. (2011). "Speeders in Online Value Research: Cross-Checking Results of Fast and Slow Respondents in Two Separate Samples Answering the 40 item "Portrait Value Questionnaire." Presented at General Online Research Conference, Düsseldorf.

Courtright, M. and D. Brien. (2009). *The Devil is in the Data* published by DMS Quirk's.

Du Jong, Kees. (2010). "Rules of Engagement" *Research World* June: 42-44.

Chang, Linchiat and Jon Krosnick. (2009). "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." Public Opinion Quarterly 73(4):641-678.

Garland, Philip. (2012). "Speed Doesn't Kill: A New Way to Detect Survey Satisficing". CASRO Online.

Gittelman, Steven and Elaine Trimarchi. (2009a). "On the Road to Clarity: Differences between sample sources." Casro Panel.

Gittelman, Steven and Elaine Trimarchi. (2009b). QMetrics – A New Online Quality Standard. White Paper. Mktg, Inc.

Gittelman, Steven and Elaine Trimarchi. (2009c). "Finding the Strength in Numbers: The value of consistency auditing of online panels." Quirk's: November.

Gittelman,Steven and Trimarchi. (2012). Consistency in Global Non-Probabilistic Online Samples. CASRO Online.

Gloeckner, Don. (2011). "How Procter & Gamble worked to develop online data quality guidelines." Quirk's July: pp 42-46.

Gutierrez, C., Wells, T., Rao, K., & Kurzynski, D. (2011). Catch Them When You Can: Speeders and Their Role in Online Data Quality. Paper presented at the Midwest Association of Public Opinion Research, Chicago, IL.

MarketTools. (2010). "When Good Respondents Go Bad: How Unengaging Surveys Lower Data Quality." http://www.markettools.com/KnowledgeCenter

MarketTools. (2009). "MarketTools TrueSample." http://www.markettools.com/KnowledgeCenter

Rogers, Felicia & Michael Richarme. (2009). "The Honesty of Online Survey Respondents: Lessons Learned and Prescriptive Remedies." Decision Analyst.

Walker, Robert, Raymond Pettit, and Joel Rubinson. (2009). "The Foundations of Quality: Response Data Quality." *Unpublished*.

Steven H. Gittelman, Ph.D. - President CEO, Mktg. Inc & SampleSolutions™

With a doctorate in Ecology and Evolutionary Science (University of Connecticut, 1976) Steve has a unique approach to statistics and sampling. The problems of sampling natural populations, where there is no census and no RDD sample from which to draw gave Dr. Gittelman a unique introduction to the challenge of creating a workable sample frame. His over thirty years of industry experience, has provided Steve with a career rich in traditional research methodology combined with a powerful desire to recapture ground lost as our more traditional modes of data collection have eroded. While he bemoans the erosion of probabilistic modes he confesses the inability to return to those halcyon days. Instead he sees a renewed challenge to create usable sample frames in a non-probabilistic setting making him the ideal ''standard bearer'' for the company's many years as the industry leader in online quality. Steve has a passionate approach to life; he writes historical biographies, having published two with two more in press, has been the president of numerous museums and societies, chases dinosaurs (he has two named after him), and restores antique John Deere tractors. Years ago he was named Socially Conscious Entrepreneur of the Year for Long Island, by Inc. Magazine for his work as President of the internationally known Dinosaur Society. There he created a major traveling exhibit, breaking records almost everywhere on the globe, in cooperation with Universal Studios and Amblin Entertainment, based on the movie blockbuster Jurassic Park. His efforts translated the power of Hollywood into a wave of funding for dinosaur science. He remains as a trustee of the Suffolk County Vanderbilt Museum on Long Island where he often can be seen driving a model A Ford, dressed in dirty overalls, in his role as ''John the Gardener'' a cantankerous character that is best known for telling stories about how it was when Mr. Vanderbilt was in residence. Along with Elaine Trimarchi, Steve is an award winning and frequent industry speaker and the author of numerous articles on internet sampling.

Elaine Trimarchi - Executive Vice President, Mktg. Inc & SampleSolutions™

Elaine has over thirty years of sampling expertise and serves as partner and Executive Vice President of Mktg, Inc. She is the co-founder of its Sample Source Auditors Division and the author of numerous papers on online sampling methodologies. Her papers, both published and unpublished, have provided cutting-edge insights into the impact of professional respondents, consistency of online panels around the world, the differences between online panels, just to name just a few. She is also the originator of the Grand Mean Project®, the largest quality control tracking study around the world, now reaching into thirty five countries and setting the standard for transparency in the panel industry. Her work on blending models, representative sampling through Real ID®, and pioneering work on the blending of social network respondents has distinguished her in all quarters. She is the joint recipient of "Best Presentation Award" at ESOMAR in Berlin, for her inspirational presentation: Online Research… and all that Jazz! The practical adaptation of old tunes to make new music". Her most recent presentation was at the ESRA in Lausanne, Switzerland. She enjoys all sports and participates in woman's teams for softball, soccer and volleyball, and enjoys skiing and gardening. She is the mother of three children.