TrustedAdvisor ASSOCIATES LLC




STATISTICAL SPECS ON THE TRUST QUOTIENT

TO: Whom it may concern

FROM: Sandy Styer, Charles H. Green

RE: Queries About the Statistical Validity of the TQ survey

DATE: June 24, 2010

As of this date, the TQ has had over 12,000 takers. It was assessed at three different points in different ways by different people. We are highly confident of its accuracy for the purposes we have set in using it.

Following are descriptions of analyses performed on the database.

DR. MIKE LINACRE, PHD, RESEARCH DIRECTOR, WINSTEPS.COM 6/2/2008:

Dr. Linacre is the creator of WINSTEPS; the Rasch measurement software used world-wide. He also developed FACETS software, for many faceted Rasch analysis widely used in constructing objective measures from judge-mediated observations. Linacre is dedicated to creating the tools for measurement that the rest of the psychometric world relies upon when doing their work.

Dr. Linacre has consulted extensively in educational and psychological measurement, assessment and testing.

He has an M.A. in Mathematics from Cambridge University, and a Ph.D. in Psychometrics from the University of Chicago. He worked closely with Benjamin D. Wright, the leading advocate of Rasch measurement, for over 15 years as a Research Associate at the University of Chicago.

Dr. Linacre saw the study online and was intrigued enough to volunteer to do analysis of it. On June 1, 2008 he examined the dataset when it had 1169 data records. In his words, that was "a generous amount of data for psychometric analysis."

Dr. Linacre's entire 3-page commentary is reproduced as an appendix to this document. Note, this was not a full report, but rather a voluntary offering on his part. He says, in his conclusion:

> *The 20 questions are successfully probing one main theme with some sub-themes. The respondents are responding in a reliable (reproducible way). The 5 category rating scale is functioning effectively. But there are indications that 20 questions may be too many. A reduction to 12 questions will maintain the psychometric properties of the instrument.*
>
> ***Congratulations – you have produced a gem, Charlie!*** [his emphasis]

More information about Dr. Linacre can be found at
http://www.meaningfulmeasurement.com/node/31 .

When we later wanted a more in-depth analysis, we went back to Dr. Linacre. He recommended a former student of his, William Fisher, whose report is listed below.

DR. WILLIAM FISHER, PHD, APRIL 2, 2009

At Dr. Linacre's suggestion, we engaged Dr. William Fisher for a more full-blown analysis when the dataset had grown to about 6,000. Dr. Fisher's executive summary statement is attached as Appendix B to this document. The opening paragraph states:

> *The instrument is a reliable and valid measure of trust, defined as a composite of credibility, reliability, intimacy, and self-orientation. It has a reliability coefficient (.89) that approaches the reliability coefficients required for high stakes educational examinations, and surpasses the reliability of a great many other instruments in use in comparable applications.*

ROBERT BOWERS, CEO SOLIANT CONSULTING, JUNE 2009

Bob Bowers is CEO of a firm, Soliant Consulting, that specializes in database management. He has helped us analyze the trust quotient material since 2008. We asked him to do a complete database analysis when the dataset hit 10,000, which then became the source of much of the Trust Temperaments' development.

In addition, in June of 2010, we asked him to conduct a brief statistical (not psychometric) overview of some of the key findings from the data. Bowers is an analyst, not a statistician, however he is familiar enough with statistics to apply some basic tools on the database. In his words:

> *"In addition to confirming the statistical validity of the TQ test as a whole with a respected psychometrician, we have used standard measures such as the t-test and F-test to determine the statistical significance of variations between groups. Because we have such a large sample size to work with (over 12,000 respondents), even relatively small differences in the means have proven to be statistically significant. For most analyses, in fact, the level of significance has been $p < .001$."*

FIELD TESTING WITH CLIENTS

A key part of the refinement of the Trust Temperaments was field-testing with clients. That there were statistical differences between the six temperaments was clear, but we used empirical testing with clients to refine just what those differences were and how to express them. We did this by grouping like-temperament groups and asking them to self-describe, and to describe other temperament-groupings.

THE COMMON SENSE TEST

Finally, we have tried to limit our analytical speculations to those with clear and simple hypotheses, for which the data appear obvious when graphed in simple ways. Our rule has been that if a point cannot be seen easily when graphed, and can only be supported by statistical analtyics, then it is not convincing enough for us to include.

Appendix A: Voluntary report by Dr. Mike Linacre, June 2008

Report on Trust Quotient Instrument

By Mike Linacre, Ph.D., mike@winsteps.com – Research Director, Winsteps.com. 6/2/2008
website: www.winsteps.com

Data file received: 6/1/2008
Data file: Excel workbook dated 6/1/2008

Data records: 1169.
Comment: a generous amount of data for psychometric analysis

Data scan of 20 questions:

Code	Count	Meaning
0	12	Unknown – processed as "missing – not administered"
1	105	never
2	1255	rarely
3	5476	often
4	10291	almost always
5	6241	at all times

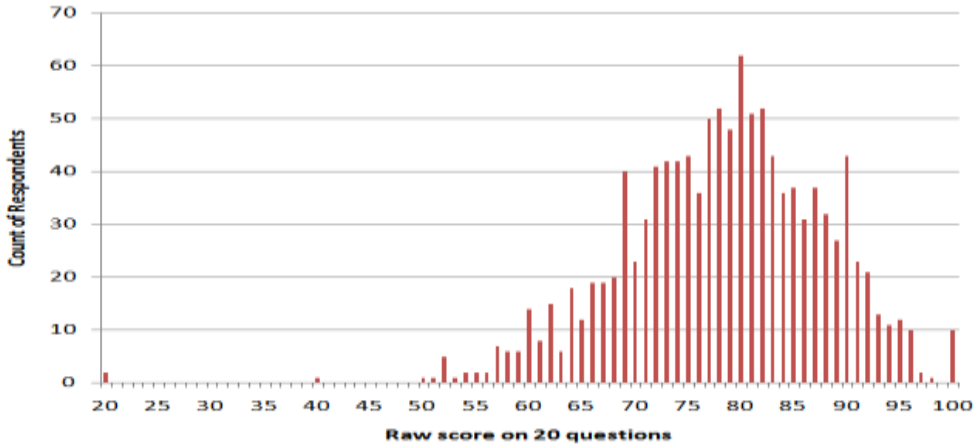
Response frequencies: a very nice distribution. There is no evidence that the sample are being asked to discriminate too many categories (e.g., "on a scale from 1 to 10, …"), nor is there evidence of misunderstood or overlapping categories (e.g., categories labeled "2=rarely" "3=occasionally")

Comment: Invalid data: much less than usually encountered in these situations.

Summary of the respondent sample:

Count (N) = 1169. Raw scores on the 20 items: Mean score of the sample = 78.2, standard deviation = 9.6

Sample distribution:



The Cronbach Alpha "Test" reliability (statistical reproducibility) of this instrument for this sample is 0.89 (with a possible range 0 – 1.0 and the higher the better). The target value for well-controlled educational tests is 0.9. Survey instruments aim at 0.8. Your instrument is very close to 0.9, so is unusually reliable for an instrument of its type. *Comment:* reducing the instrument to 12 questions would produce a reliability above 0.8 value.

Behavior of the sample:

Spikes in the sample distribution may indicate people who started "response setting" (no longer reading the questions, but responding the same, or close to, every time). *This may indicate that 20 questions is too many for casual respondents.*

2 respondents rated every item "1", including two items Q12 and Q19 which no one else rated a "1". This suggests that these people were checking to see what happens with all 1's. 10 respondents rated every item "5". Only 12 extreme scores in a sample of 1169 suggests that respondents were reading the questions (to some extent).

47 people responded in a way that was clearly self-contradictory if a common theory underlies all the questions. The most extreme self-contradiction was person 995 who responded "5" to all questions except "1" to qu. 6 "consistent".

Another 114 respondents were erratic or idiosyncratic. For example person 1024 responded 3, 4 or 5 to all questions except for a 2 on q15 "losing short", an unexpectedly low rating in this context.

In total, 173 respondents or the 1169 (15%) produced noticeably response strings with noticeable unpredictability. A low percentage for an uncontrolled data collection.

Summary of the items:

Average rating (1-5)	20 questions	Item hierarchy	Strongest factor cluster	Strong factor cluster	Grab-bag
4.32	q10 bond	Easiest to say "always"	1		
4.15	q18 discreet				3
4.15	q19 promises		1		
4.10	q12 expert		1		
4.09	q20 byproduct	"Professional items" ?	1		
4.07	q16 credentials		1		
3.98	q15 losing short				3
3.97	q17 no surprises		1		
3.97	q4 honest				3
-	-	-			
3.89	q5 emotional risks			2	
3.88	q6 consistent		1		
3.88	q11 wedded				3
3.80	q14 empathize			2	
3.80	q13 personal risks	"Inter-personal items" ?		2	
3.78	q3 not blaming				3
3.73	q1 at ease			2	
3.70	q2 communicator				3
3.68	q8 confide			2	
3.68	q7 curiosity			2	
3.61	q9 relate	Hardest to say "always"		2	

Comment: Does the item hierarchy from "easiest to say always" to "hardest to say always" make sense according to your substantive theory? This indicates the "construct validity" of the instrument.

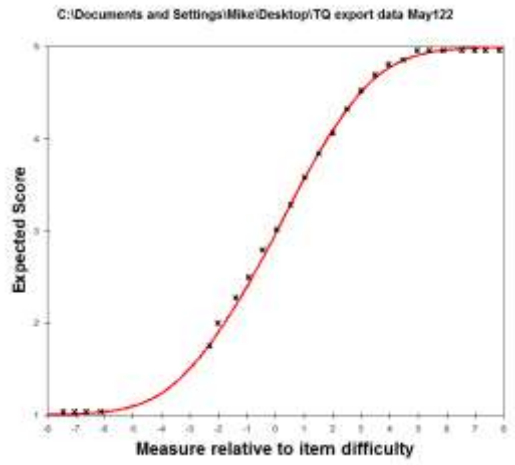
The one question which best summarizes the respondent's entire set of 20 responses is: q17 no surprises.

The least predictable question of these 20 good questions is: q14 empathize.

Response-level Variance decomposition indicates that:
36.6% of the variance in the response-level data is explained by the average responses of the persons
5.6% of the variance in the data is explained by the average responses to the items
6.9% of the variance is explained by characteristics of the strongest cluster of items (1)
5.7% of the variance is explained by characteristics of the strong cluster of items (2)
45.2% of the variance in the data is explained by individual characteristics of the "grab-bag" items (3) and personal factors within each respondent.

Comment: This variance decompositions accords with analyses of other social-science data sets.

Plot of item characteristic curves:



This shows the functioning of the rating scale (y-axis) plotted against the location of the respondent on a hypothetical "latent variable" (i.e., whatever the instrument is measuring) on the x-axis. The red line shows the ideal predictable relationship between the rating on an item and the respondent's location on the latent variable. The x's show the average rating (y-axis) of respodnents near that location (x-axis). The x's are close to the red line. Excellent!

At the bottom left corner are the respondents who were all 1's. At the top right all 5's

Conclusion:

The 20 questions are successfully probing one main theme with some sub-themes. The respondents are responding in a reliable (reproducible way). The 5 category rating scale is functioning effectively. But there are indications that 20 questions may be too many. A reduction to 12 questions will maintain the psychometric properties of the instrument.

Congratulations – you have produced a gem, Charlie!

APPENDIX B: ANALYSIS BY DR. WILLIAM FISHER

Analysis of the
Trust Quotient Self-Diagnostic Assessment
William P. Fisher, Jr., Ph.D.
2 April 2009

Executive Summary and Conclusions

The instrument is a reliable and valid measure of trust, defined as a composite of credibility, reliability, intimacy, and self-orientation. It has a reliability coefficient (.89) that approaches the reliability coefficients required for high stakes educational examinations, and surpasses the reliability of a great many other instruments in use in comparable applications.

The data incorporate some local dependencies, stemming from perhaps as many as three sources. None of these are serious problems in need of immediate attention, but they indicate ways in which the instrument could be improved.

A basic question here concerns the value and meaning of the Trust Equation. The issue boils down to whether or not the four component scales measure the same thing, or if they measure different things that can be meaningfully combined or put into a ratio. Analyzing each C, R, I, and S scale separately, I found that they produced measures correlating about .99 after disattenuation, suggesting that they all measure the same thing and ought to be expressed in a single number. The same result appears to be produced from the existing equation as from others I tried. This is explored in detail after establishing the basic instrument characteristics.

1. Summary statistics:
The Trust Quotient measurement data file was composed of 5,934 rows by 20 columns, with additional columns of various demographics. The scaled data are 99.9% complete. The mean score is 77.9, with a standard deviation of 9.8.

2. Substantively annotated construct maps useful as a basis for self-scoring forms:
Table 2.2 in the primary Winsteps output (below and in TQ.out.txt) has been annotated to show the meaning of the variation in the measures and calibrations.

3. Rating scale analyses and optimizations
In the analyses conducted, the codes of 0 were processed as missing, and the 1-5 codes were labeled following the categories used on the assessment. The categories functioned well in their task of consistently discriminating among the items, from the respondents' perspective, and among the respondents, from the items' perspective. The relative proportions of the category counts remained roughly constant with those observed by Linacre last year.

4. Data quality evaluations such as model fit analyses or differential item/person functioning analyses
The overall fit of the items to the measurement model was quite good, with all individual statistical values falling within an acceptable range. On average, respondents' data exhibits good consistency, but as individuals, they don't always seem to take the assessment as seriously as they might. Women and men differ in some striking respects on some items.

5. Principal Components Analysis of the standardized residuals

Tables 23.x in the primary Winsteps output show the results of these analyses. They again largely reproduce the results seen by Linacre last June.

6. Statistical analyses and graphs of the measures by any available demographic indicators
Open the enclosed HTM file in a web browser. The JPG files will show in the HTML as graphs. The most striking finding is the steady upward change in the measures across age groups.

7. Excel, SPSS, DBF, or text data files integrating your original data with the measurement results
In lieu of specific requests, Excel, SPSS, and CSV versions of the data are included, along with SPSS and HTML versions of the statistical analyses, and text files containing the Winsteps analyses.

8. Optionally, further guidance in taking the work towards publication in a peer-reviewed journal.
Publication in a scientific, peer-reviewed journal can lend a degree of credibility not easily obtained by any other means. If you're interested, we can discuss focus, target journal, time frame, etc.