The (Statistical) Confidence Game of Employee Engagement Surveys

Converge Consulting Group

Converge Consulting Group Inc.

Mailing: 2617-20 Street SW
Calgary, Alberta
Canada
T2T 4Z4

Phone: 403.266.0061
Web: www.converge-group.com
www.converge-group.net
www.voiceoftheemployee.com

Converge Consulting Group

engagement
statistical
employee
significance
data
companies
performance
research
survey
models
results
enterprise
factors
employees
important
one
people
company
correlation
economistic
growth
drivers
efforts
like
importance
much
engaged
provide
right
brivial
working
back
another
economic
better
score
measure
comparisons
feedback
benchmark
EEI8
HR
useful
capital
just
objective
organizations
anything
surveys
correlated
profitability
Statistical
Correlation
financial
organizational
benchmarking
analysis
point
practical
different
questions
business
evidence
work
human
Watson
making
judgment
regression
Company
HCI
good
likely
things
example
statistically
significant
relationship
confidence
information
scores
well
thinking
model
Deming
value
improving
single
story
create
also
less
yes
claims
meaningless
simple
true
problem
belief
beta
leadership
issues
years
case
time
Engagement
group
used
using
managers
action
improvements
management
critical
little
actionable
satisfaction
findings
product
difference
science
numbers
may
characteristics
flow
presentation
made
differences
levels
getting
high
rules
results
game
cherry
go
Tufte
techniques
efforts
drivers
years
belief
problem
answer
judgment
claims
meaningless
true
simple
critical
actionable
satisfaction
findings
product
difference

The (Statistical) Confidence Game of Employee Engagement Surveys

ABSTRACT

The aim of employee surveys is gathering feedback concerning strengths and weakness of the employee-employer relationship, enabling actions that strengthen this relationship and improve business performance. It's important the employee survey gets it right.

Most employee research gets it wrong. It's a statistical confidence game placing the public relations objective of telling a good story above the practical objective of telling the truth by mangling statistical technique and the overselling the results. This economisting uses:

- statistical significance to identify 'important' differences that amount to trivial artifacts,
- regression models of engagement that are cherry-picked statistical fairy tales, and
- performance benchmarking yielding absurd claims of being one of the best companies to work for.

Add to this, the belief that complex concepts can be reduced to a single number, and the result is evidence corruption, eliminating the possibility of rational decision making. Engagement surveys assign blame, promote tampering and destroy trust.

There is a cure. Organizations need to bring their employee surveys back to reality and that means getting back to the basics. It starts with HR departments wrestling back control, working with statisticians while bringing their knowledge of the unique conditions and circumstances of the enterprise into play.

The (Statistical) Confidence Game of Employee Engagement Surveys

In Northern California, a medium-sized company designs, manufactures electronic devices used for environmental monitoring. Business is good and the company enjoys steady growth and profitability.

Some of this success is attributable to the Human Resources (HR) group and the centerpiece of their employee feedback program, the annual employee engagement survey. The survey is constructed around 18 critical engagement drivers (the EE18 Model) developed by a large HR consultancy independently administering and analyzing the survey. The consultancy has statistically validated the 18 drivers or factors as leading to increased employee engagement, corporate growth, profitability and customer satisfaction using its results database of over 15,000 organizations.

Improvement efforts at the enterprise level focus on those EE18 drivers scoring low in both absolute terms and against benchmark comparisons with other companies in similar industries. When initiating this effort, the company's EE18 score was at the median of comparative companies. Now it stands firmly in the upper quartile. Recently selected as 'one of the 100 best companies to work for', executives are planning a trip to New York to accept the award, maximizing positive media coverage as an employer of choice.

Improvement activities go deeper into the enterprise. Each year, senior and mid-level managers receive departmental engagement reports highlighting year to year results and comparisons with competitive benchmarks and corporate targets. Rigorous statistical tests defining strength and weakness areas provide quality assurance and eliminate bias. Mangers review the analysis for their group and define improvement action plans. They are held accountable for these results and bonuses are determined, in part, by how well departmental engagement scores stack up against corporate targets and gains.

This all sounds very good, depicting the use of employee engagement research advocated by human resource, organizational development and effectiveness experts as well as the popular business media, but there's a fatal flaw. It's premised on a single assumption--that the employee engagement survey got it right.

What if it didn't? What if the EE18 model and results are all wrong, a statistical fantasy as connected to reality as a fairy tale? That might paint a different picture, one in which;

- The EE18 score, used to evaluate engagement, is a meaningless number, leading executives into a delusion of improving engagement even as everyone else sees and feels the obvious deterioration.
- The eighteen drivers of engagement don't drive anything. The focused and disciplined improvement strategy yielding only employee bewilderment as to why the organization is fixing things that don't matter while ignoring things that do. Organizational issues grow and fester.
- The '100 best companies to work for' award is considered a bad joke and the trip for executives to accept it, as adding insult to injury. Increasingly, employees see a self-promoting leadership as out of touch with organizational realities.
- Managers develop engagement action plans knowing they are irrelevant to increasing workplace tensions and declining productivity. Increasingly, the employee engagement survey and program is seen as a waste of time by upper and mid-level managers, supervisors and employees alike.
- People are held accountable for year-to-year results that are statistical noise. Increasingly, people feel they are being abused by the numbers they resent it. Confusions is built, trust is demolished.
- With bonuses on the line, manipulating engagement scores sucks time away from running the business. Challenging assignments are tossed among departments like hot potatoes because of potential engagement repercussions. Endless project scope negotiations ensue. Witch hunting becomes an organizational pastime and threats of poor performance reviews are used to get departmental engagement numbers up. Managers spend money on shiny-penny projects in the hope of improving EE18 scores.

Within two years our California company is crumbling. Employee turnover starts climbing, productivity declines, critical projects fail to progress, growth and profitability suffers. The good news? Those EE18 scores are up again.

managing with coconut headsets

This story, and others like it, are becoming the norm. Real people and real organizations are suffering as a result. Well intentioned efforts at improving the employee-employer relationship are demolishing it, as well as the trust vital to successful enterprise.

The dirty secret of employee engagement survey research is that it's cargo cult (junk) science--a statistical confidence game corrupting employee feedback by placing the public relations objective of telling a good story, above the practical objective of getting it right.

The dirty secret of employee engagement survey research is that it's junk science--a statistical confidence game corrupting employee feedback by placing the public relations objective of telling a good story, above the practical objective of getting it right.

The forms and symbols of statistical science provide the confidence, but there is no substance behind it. Like the cargo-cults of the south Pacific trying to build airplane engines out of bamboo and radio headsets from coconuts (Rosenzweig 2007), the employee engagement cargo-cult dress employee surveys to look like science, claim the authority of science, but lack scientific logic or rationality. The result is crud. Three methods dominate this economisting (Appendix A).

- Using statistical significance testing to identify important findings or differences. What's identified are trivial statistical artifacts. Small irrelevant findings are classified as statistically significant and pursued, while real issues are ignored.
- Using statistical regression in building models of employee engagement that are statistical fairy tales. Engagement drivers are really cherry-picked factors ensuring action plan accountability falls exclusively on the shoulders of front-line and middle levels.
- Making benchmarking comparisons yielding absurd claims of being one of the 'best companies to work for'. Leaders can't understand why workplace problems persist even as the company rises in the rankings. That these comparisons are computational nonsense, is no more considered than the possibility that only 100 of worst companies to work for were competing for the honor.

Add to these the belief that complex concepts such as engagement can be reduced to a single number, and the result is evidence corruption, eliminating the possibility of understanding or improving the employer-employee relationship. Earnest efforts to use survey results as a basis for evidence-based action are perceived by employees as the hard sell of statistical fantasies by leaders that just don't get it. As this goes on, employee doubt and cynicism grows, while executive management and HR credibility are diminished.

The aim of employee survey research is simple enough, obtain feedback so management can gain some understanding as to what is going on in the organization, what people are thinking and feeling, what issues are affecting the employee-employer relationship and productivity to identify areas where improvement actions can be taken. Organizations have lost sight of this aim. Blinded by oversold statistical techniques lending shiny scientific veneers to corrupt analysis, organizations are inadvertently destroying trust, engagement and productivity.

the pun of statistical significance

The major challenge in research, including employee engagement, is identifying what is, and what is not, important to employees and to the business--separating the wheat from the chaff. A department may have an employee engagement score

of 4.6,on a five point scale, and another a score of 4.4, but is this difference important? Should we be expending organizational development effort on the second department and not the first, on both, or neither? What if 81% of your employees are engaged compared to a 77% of a benchmark comparison group? What if 84% of our employees were 'engaged' last year and only 79% this year? Are these differences important? Should we do something? How do you tell?

For employee survey cargo-cults, the simple answer is found in statistical significance. If it's statistically significant, then it's deemed important. No judgment, is required. Statistical significance provides an air of scientific rigor and legitimacy to the conclusions simultaneously stifling critical thinking and common sense. This is reinforced through complex formulas, computations and sciency sounding words of which logistic regression and factor analysis are perennial favorites.

Using statistical significance to determine practical importance is pure economisting, confirming H. L. Mencken's observation that, "*for every complex problem, there is a solution that is simple, neat, and wrong*". Statistical significance has little to do with practical importance. The statistical meaning refers to the level of detectability given the quality of the measurement system under certain assumptions that rarely apply. (Tufte 2009, Deming 1975)

Using statistical significance to determine practical importance is pure economisting, confirming H. L. Mencken's observation that, "for every complex problem, there is a solution that is simple, neat, and wrong"

A statistically significant difference between departments, companies, or across time periods, then, is one where the measurement system is capable of detecting a difference. It should go without saying that just because a difference is detectable, doesn't mean it's important. Statistical significance tests always produce far more 'significant' findings than findings of real and practical importance to employees or the business (Appendix B). Thus, most statistically significant findings are trivial, often of no practical or economic importance.

Organizations though, believe and behave otherwise. Rewards and punishments are distributed on trivial, but statistically significant, 2% differences in engagement scores. Actions taken to address 'significant' benchmark gaps, or correct 'significant' year to year engagement declines, amount to tampering, wasting resources and diminishing performance (Appendix C).

Cargo-cult employee engagement surveys sell the confusion between the everyday and statistical meanings of significance, enabling the repackaging of mundane observations into grand and mistaken conclusions. It's an economisting pun, and the joke is on those buying into it (Tufte 2009).

meaningless models

Misusing statistical significance pales against the levels of evidence corruption and economisting involved in defining employee engagement models. Anchored to the economic concept of human capital, engagement models treat people as input-

output machines, just another economic factor of production. Increasing economic output can be had simply by investing in the right set of competencies, much like the economic output of a milling machine can be improved by investing in an oil change.

This treatment works well and is useful in the realm of macroeconomic theory, but in the real world, on the shop floor, treating people like machines is an academic conceit. Just because it's useful for economists to analyze work in this way, doesn't mean it's useful to manage it in this way. People are more complicated than milling machines, laptops, or toaster ovens. That a foundation of HR practice is believing otherwise, treating people like machines, is one of the better ironies of modern HR management.

Nevertheless, this transition from economic theory to management practice in employee survey research largely took place after publication of "The Employee-Customer-Profit Chain at Sears" in the Harvard Business Review (Rucci, Kirn and Quinn 1998). Analyzing aggregated data from 800 stores, it concluded that improving employee attitudes led to increased customer satisfaction and revenue growth, providing not only empirical support for human capital management practices but also providing what HR managers had longed for, hard numbers connecting investments in people with improvements to the bottom-line. Any lingering doubts of HR practitioners were removed shortly thereafter by a Wall Street Journal article describing research at Nortel Networks that found working on employee satisfaction issues increased customer satisfaction rates and financial returns (Shellenbarger 1998).

The cargo cult confidence trick in this, and subsequent human capital/employee engagement research, is calculating the probability of unique or peculiar events after they occur.

Soon, cargo-cult research claiming to establish causal relationships between investments in human capital and output flooded a market convinced that a magic genie had come out of the lamp. Watson Wyatt Worldwide Human Capital Index (HCI) study claimed "*dramatic evidence*" of "*the 49 specific HR practices that play the greatest role in creating shareholder value*"(Watson Wyatt 2002). Gallop maintains that engaged workgroups, defined by their Q12 statements are; "*18% more productive, 12% more profitable, 12% better at engaging customers,. . . 62% less likely to have an accident, 27% less prone to absenteeism, 51% less likely to be a source of inventory shrinkage*" (Fleming and Asplund 2007). Aon/Hewett identified the ten national drivers of engagement stating, "*highly engaged employees provide higher value. They are more effective at producing high quality, innovative products/services, and they more positively impact customer satisfaction, cost, and revenue growth*" and go on to conclude that "*20% of the organization's employees create 80% of the value*" (Aon Hewitt 2010).

Does anyone really believe that 20% of employees create 80% of the value? If so, why not fire the 80%

Not to be outdone, the new Towers Watson, formed as a result of a merger between Watson Wyatt and Towers Perrin, dropped HCI in favor of '*Exponential Engagement*', a new and improved model, building on a; "*growing body of*

evidence . . . that validates the quantifiable relationship between levels of organizational engagement and financial performance." Specifically, highly engaged companies have, "*operating margins 1.4 times as great as companies with low levels of engagement while companies with high Exponential Engagement . . . have 3 times the operating margins of low engagement companies*" (Towers Watson 2011 p. 2-4).

The past decade has seemingly been less about validating '*the quantifiable relationship*' and more about the marketing of statistical analysis as a Keystone Cops remake, complete with consultancies frantically falling over themselves by making ever more exaggerated claims, inevitably attributable to 'secret' statistical methods recently discovered from mining extensive proprietary data bases. Real science of course, doesn't do secrets, because these protect sloppy, mangled and inept analysis from critique. Techniques that; "*crunch and grind vast data matrices . . . are perhaps useful for those with lots of data but no ideas*" (Tufte, 2006 p. 142). No clue either.

The cargo cult confidence trick in this, and subsequent human capital/employee engagement research, is calculating the probability of unique or peculiar events after they occur. Employee survey and corporate performance databases are 'mined' in the search of correlations. Once found, the probabilities against the correlation are calculated and found to be high, that is, statistically significant, prompting the analyst to declare something important has been discovered. Of course the probabilities against the event will always be high because the correlation itself, is a unique or peculiar event, which is why the data had to be mined to find it. Thus, employee engagement models trample on one of the most fundamental rules in science--the data that gives you the idea, can never be used to validate the idea. "*For example, I had the most remarkable experience this evening. While coming in here, I saw the license plate ANZ 912. Calculate for me, please, the odds that of all the license plates in the state of Washington I should happen to see ANZ 912. Well, it's a ridiculous thing*" (Feynman 1998). Scientifically ridiculous yes, but from a business perspective of selling employee engagement research, a gold mine..

Managers readily accept that next year's corporate profitability can't be predicted, even by finance and accounting. Yet this same manager believes that responses to 12 questions on a survey can not only predict profitability improvement, but can do so to two decimal points, and do the same for sales growth and inventory shrinkage? Does anyone really believe that 20% of employees create 80% of the value and if they do, why not fire the remaining 80%? What is amazing here, is not the sophistication of the statistical analysis used, or that would be required if the economisting claims were even remotely true, but the level of naiveté demonstrated in believing it. The power of cargo cult science allows otherwise

reasonable people to forget the time honored advice that, if it appears too good to be true, it probably is.

confusing correlation with causation

Statistical models mathematically link survey questions or factors to an overall effect variable, such as employee engagement. Statistical correlation determines the questions or factors ultimately comprising the model. These are typically referred to as drivers implying both importance and causality but constituting neither. This, because statistical tests can't determine practical importance and because correlation isn't causation.

Correlation is important. You can't have causation without it. But selling one for the other is survey snake-oil, a statistical confidence game. It may be true that engaged employees make for profitable companies, but it may also be true, and more than likely is, that profitable companies make for engaged employees. Trying to improve profitability by improving engagement, therefore, is just as likely to reduce profitability. The same for improving one or more of the EE18 variables in an effort to increase engagement in our California company. It's just as likely to have the opposite effect.

Some efforts have been made to overcome the lack of causality in engagement models, but these efforts degenerate into data torture and if you torture the data long enough, it will tell you anything. The following effort is both typical and instructive:

"*To see which way the relationship truly runs, we simply compared two different correlations:*

Correlation A represents the relationship between the 1999 HCI score and 2001 financial performance.

Correlation B represents the relationship between 1999 financial performance and 2001 HCI scores.

If better financial performance is what creates superior HR practices, Correlation B should be larger. If, in fact, the way companies manage their human capital is what drives financial success, Correlation A should be larger.

Our results were dramatic. Correlation A, .41, is statistically significantly larger than Correlation B, .19". (Watson Wyatt 2002).

This is the classic confidence trick of misdirection, where the measures of financial performance and HCI are switched for one another. For example, the 1999 HCI score was likely also correlated with 1999 financial performance. So Correlation A could be interpreted as the correlation between financial performance in 1999 and 2001. Likewise, 1999 financial performance was also likely correlated with 1999 HCI

scores. So Correlation B could be interpreted as the ability of HCI scores in 1999 to predict HCI scores in 2001.

The results are, therefore, more humorous than dramatic. The correlation between past and current financial performance, Correlation A, is running at about 0.41--analogous to making decisions by coin flip. The correlation between past and future HCI scores, Correlation B, is worse at 0.19, meaning HCI can't predict much of anything, even itself. The extent to which Correlation A is stronger than Correlation B is not so much an indicator of causality as it is a measure of non-existent HCI reliability.

Gallop makes the same error in claiming; "*Relative to their competition, high engagement organizations exhibited an upward trend, with EPS in 2004 to 2005 that was 18% higher than their competitors EPS. . . . Taken together the growth rate of high engagement companies was 2.6 times the growth rate of low engagement companies.*" (Fleming and Asplund 2007, p 170). In essence, Gallop is concluding that high growth companies tend to out-grow slow growth companies.

cherry picking

It's bad enough that factors in these models aren't causal factors or drivers of engagement (or anything else) but HR professionals are often surprised to learn that factors comprising the engagement model aren't even those most correlated with engagement. Variables in regression equations are those that collectively represent the most efficient solution to explaining variation in a response variable (engagement, profitability, etc.), not the factors most correlated with it. The difference is subtle, but it means some factors more strongly correlated with engagement, are not included in the regression model.

HR professionals are often surprised to learn that factors comprising the engagement model aren't even those most correlated with engagement

This is because regression is a statistical tool solving a statistical problem. The business problem is different. The business needs to know the factors having a potential impact on engagement, regardless of whether these factors are correlated with one another. Knowing all the possible solutions gives decision makers options, allowing the pursuit of different improvement strategies in different conditions, circumstances and constraints. Engagement models built on regression techniques are a result of statistical cherry picking, burying the very information the business needs to know, and typically, believes it is buying.

Not all the cherry picking is statistical though. In selecting its final 12 questions for the Q12, for example, Gallop focused on those questions that are; "*simple and easy to effect. They had to be 'actionable' not emotional outcome questions like, 'Overall how satisfied are you with your work environment?' or 'Are you proud to be working for your company?'*" (Buckingham and Coffman 1999, p. 282).

This search for actionable data is another conceit. Who says what is actionable and what isn't? How is this determined? For example, how employees respond to; *Are you proud to be working for your company?* is both important information and actionable. It's just not actionable by those in the front lines or middle levels of the enterprise. The question speaks to the overall image of the company, an executive responsibility. Removing questions like this, not only hides information of potential critical importance from organizational leadership, it also leaves a cherry picked collection of variables or factors for which employees can be held accountable, but for which executives cannot. Engagement research is becoming the means by which organizational leaders are seen to mandate accountability, while absolving themselves of any. No wonder employee cynicism grows.

Engagement research is becoming the means by which organizational leaders are seen to mandate accountability, while absolving themselves of any.

meaningless numbers

Reducing complex concepts like engagement to a single number has considerable appeal, especially in business. An employee engagement index is easily deployed on electronic dashboards and is highly digestible, requiring virtually no independent thinking or judgment beyond; 'big number good, small number bad'.

Douglas Adams (1979) tells the story of hyper-intelligent pan-dimensional beings that build the supercomputer Deep Thought to provide "*The Ultimate Answer to the Ultimate Question of Life, The Universe, and Everything*". It takes Deep Thought 7½ million years to compute the answer, giving considerable time for speculation and suspense to build. What will be the answer to the Ultimate Question? Deep Thought computes it to be 42. When the pan-dimensional beings threaten to riot over this computational non-sequitur, Deep Thought assures them that while 42 is *The Answer*, their disappointment is likely due to their not really understanding *The Question*.

The search for a single employee engagement number is much like this. Asking the question, guarantees a meaningless answer, because the concept is just more complicated than the question allows. Boiling down employee engagement to a single number only means you don't understand employee engagement.

This is evidenced, in part, in having so many different models all claiming to predict the same thing. The Conference Board reviewed studies of the leading employee engagement consulting firms including Gallup, Hewitt/Aon, Corporate Leadership Council and Towers Perrin. Across 26 drivers of engagement only eight were consistent, and then, for only four of the twelve studies examined (Deutsch 2001). This is a definite lack of operational definition. The bottom line here is that competing employee engagement models measure whatever it is they measure, and apparently, no one knows what that is exactly.

. . . competing employee engagement models measure whatever it is they measure, and apparently, no one knows what that is exactly

corrupt comparisons

That employee engagement models can't predict much of anything, let alone anything important, would be a problem if the claimed capabilities were put to the reality test--closing the empirical, scientific loop of Feynman's license plate example. This would include making accurate predictions of economic returns arising from real world investments in human capital. After all, if these are valid models, they should actually work. This doesn't happen.

Instead, companies are encouraged to compare themselves with groups of businesses in similar industries. This subtle benchmarking misdirection diverts attention away from what employee engagement models were supposed to do, measure the impact of investing in human capital and toward improving a meaningless engagement score to win a non-existent zero-sum game of those selling the engagement model. Thus, the enterprise is transformed into a lab rat pursuing non-existent cheese in a maze with no exit.

Adding insult to injury is going back to the poisoned well of using statistical significance to identify important differences or gaps between the enterprise and the benchmark comparison group. Here again, statistical significance is used as a test of importance, ensuring reams of statistically significant, but ultimately trivial, competitive gaps the enterprise can pursue--to negative economic consequences.

Typically, data is presented out of context, usually in table format, with only the most recent data visible, as presented in Table 1. Comparisons between years, as well as between the company and the benchmark comparison group, are made and conclusions drawn concerning the performance of the company.

Table1: Comparing company engagement to benchmark group

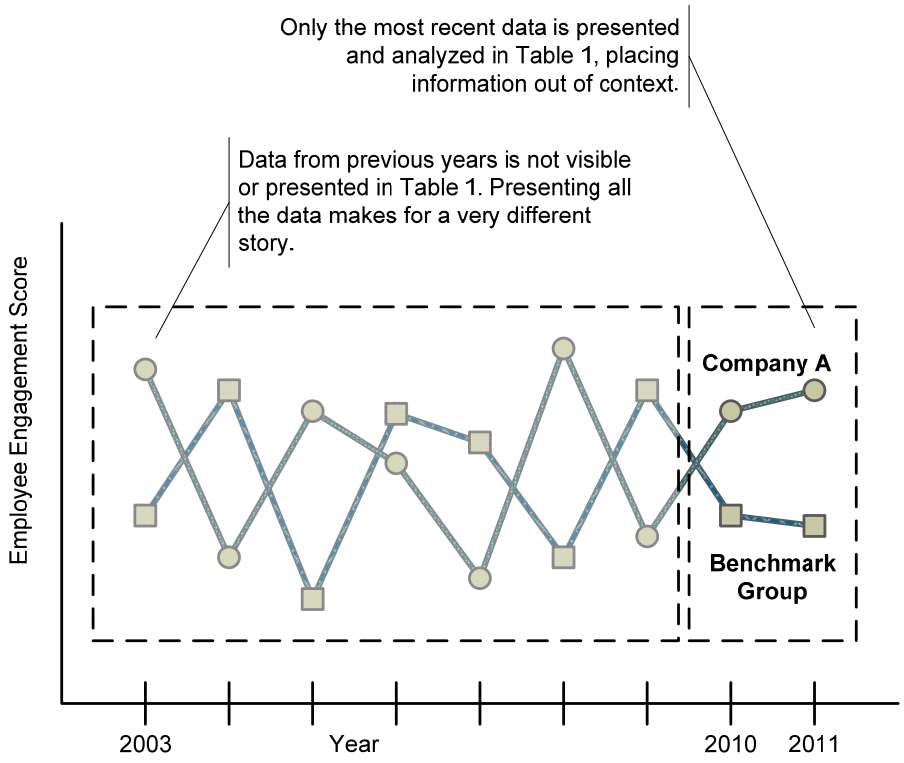
Data Table	2010	2011	Statistically Significant?	Conclusion
Company A	85	87	yes	trending up
Benchmark Group	81	79	yes	trending down
Statistically Significant?	yes	yes		
Conclusion	we're better	we're better still		

In this case, Company A, representing our nice little California company, was significantly better than the benchmarking comparison group in 2010 and again in 2011. The obvious conclusion is that Company A is more capable of generating

engagement among its employees than competitors. This inference is reinforced by the statistically significant increase in Company A performance between 2010 and 2011, during which time the industry benchmark recorded a statistically significant decline. Again, the seemingly obvious conclusion is that Company A is getting better at engaging its people even as competitors are getting worse at it.

Both conclusions are nonsense, a function of misusing statistical significance as a test for importance, or worse, as a test for truth, while using data presentation techniques that hide relevant information from decision makers. Simple graphical techniques, such as the run chart in Figure 1, provides a very different and more accurate story by presenting all the data in context.

Figure 1: Run chart of company and benchmark group engagement



Employee engagement results for Company A have risen and fallen randomly over the past 9 years. This is common cause variation, some years are better than others because of all the little things that come together at different times and in different ways to create different results. The same is true for the Benchmark Group. Engagement rises and falls with no obvious pattern, but generally has experienced employee engagement levels more or less equal to those of Company A. It's simply a matter of random chance that Company A's results improved over the past two years and were better than those of the Benchmark

Group during that period. The supposed statistically significant findings were all trivial, of no material significance to the business or its employees.

Where statistical significance fails to be sufficiently misleading, ranking can be used instead (i.e.; *the 100 best companies to work for*). This forced ranking is an old statistical trick performing the same task as significance testing, making "*much ado about practically nothing*" (Huff 1954, p. 53). Differences between two organizations may be scientifically imperceptible, but with ranking, only one can be the winner -- *We're Number 1!*. Add in problems concerning the definition and changing composition of the benchmarking group and employee engagement comparisons move from fairy tale, to illusion or delusion, depending upon who believes it.

Benchmarking employee engagement is rooted in the belief that the best way to exhibit leadership is to following everyone else, without being too picky as to who everyone else is. This seems less like a path to leadership, than evidence of a lack of it.

Benchmarking employee engagement is rooted in the belief that the best way to exhibit leadership is to follow everyone else and without being too picky as to who 'everyone else' is. This seems less like a path to leadership, than evidence of a lack of it.

getting back to the basics

The way data is treated in employee engagement research; framing historical data out of the analysis, using summary statistics rather than distributional graphics, misusing statistical significance, confusing correlation with causation, cherry picking, and model building by post hoc calculation of probabilities, doesn't stand up against Shewart's rules for data presentation (Appendix D) nor Feynman's related notion of scientific integrity--presenting "*all of the information . . . not just information that leads to judgment in one particular direction or another*" (Feynman 1999, p210).

"*Making a presentation is a moral act as well as an intellectual activity.*" (Tufte 2009, p141). Given the standards offered by Shewart, Deming and Feynman, what happens in employee engagement research is a dishonest, corrupt representation of data (and people), a confidence game where statistical method is used to impress rather than inform. Normally, few would care, but in this case, the impact on people and organizations is materially significant and destructive.

Organizations need to change their approach. Computing off-the-shelf engagement indexes, using statistical significance as a measure of importance, building statistical models and making benchmarking comparisons only creates a fantasy complete with pretty multi-colored three dimensional charts and graphs. Leaders operating this way have simply forgot the point of it all.

The aim of employee research is gathering feedback from our people on how they feel about things, what's working and what isn't, and identifying areas of strength and weakness in the employee/employer relationship to create a motivated and

engaged workforce. In short, gathering information to support rational, productive action. Doing this adds value to the enterprise. Here's some things to consider.

(i) If engaged in employee research employing statistical significance testing, regression models or benchmarking, *stop*! Try using your judgment instead. It's not perfect but it's better than letting a computer do your thinking for you. Your people will thank you for it, and if conditions are right, you just may hear a gigantic, collective sigh of relief from the front lines.

(ii) When ready to return to employee survey research, start by thinking about what the organization needs to know. Engage in a little qualitative research, examine different measurement frameworks, do some reading, create a list and organize your thoughts. By all means, get the help of a consultant to support you in these efforts, but don't let the tail wag the dog. Subject matter experts and statisticians must work together. Ultimately, you must build your own employee engagement framework comprised of survey questions that make sense to you and the leadership team. Some of these questions will be of a general nature common to all organizations, perhaps borrowed from other surveys. Others will be contingent on existing organizational conditions and circumstances, like gaining feedback on recent changes to the benefit plan or travel policy.

(iii) Never assume that organizational characteristics are all of the same type (another faulty assumption made by regression models). They're not. Some characteristics motivate, others de-motivate. Some operate as an on-off function, others are linear, varying by the degree to which they're provided. Frederick Herzberg made this point brilliantly in his famous HBR article;"*One More Time, How Do You Motivate Employees?*' differentiating between hygiene and motivator factors" (Herzberg 1987). Every HR professional should read this article once a year. Discussing it with the CEO wouldn't hurt either.

If Herzberg's model feels a little too simple, Dr. Kano's Model of value characteristics is more recent and distinguishes among five types of characteristics relevant to engagement research. (Kano, Nobuhiku, Fumio and Shinichi, 1984). A formal Kano study probably isn't necessary, just thinking about the various types of characteristics is useful. If you can't understand why 'having a decent job with decent pay' isn't fundamentally different from 'having your opinions count at work', and further, why they can't possibly be compared to one another on a five point scale, go back and think about it some more.

(iv) Speaking of scales, use seven or nine point scales in your surveys. Scales with fewer than seven points do not provide the granularity to indicate much of anything. The problem is compounded when data is summarized using the top two box scores. Scales beyond nine points stretch the individual's ability to differentiate. Seven or nine point scales balance these competing trade-offs. Provide opportunities for open ended questions and work with the data that

results. Text analysis routines can help, but never replace, the act of reading and reworking qualitative data. To hear the *voice of the employee*, you must listen.

(v) Once the data are obtained, areas of importance need to be identified. Statistical significance testing cannot do this although it can help by identifying detectable results. Material significance requires clear decision rules specified in advance by experts in the field. Some statistical tools can support your efforts. Analysis of Means (ANOM) and Control Charts, especially *X, mR* charts for the analysis of data over time, are particularly useful in identifying material significance. Analytic Tables provide a convenient presentation of summary statistics and distributions in a manner consistent with Shewart's rules. This facilitates the Inter-ocular Trauma Test (ITT)--if it's there, it should hit you right between the eyes.

Trying to make sense of it all is certainly more difficult than plugging survey results into a model and mindlessly marching in lockstep to computer generated priorities. It requires HR professionals and managers to gather reliable feedback, sit back and actually think about things.

Fortunately, that's what these folks are paid to do and the key to how HR can add real value to the enterprise.

HR won't add value to the enterprise by plugging survey results into a model and mindlessly marching in lockstep to computer generated priorities.

HR professionals and managers must use reliable feedback and actually think about things.

appendices

appendix A: economisting

"economisting: (e kon' o mist' ing) *1. The act or process of converting limited evidence into grand claims by means of rhetorical ploys, especially punning. 2. The belief or practice that empirical evidence can only confirm and never disconfirm a favored theory. 3. Conclusions that are theory-driven, not evidence based. See also confirmation bias, painting with a broad brush, Iraqi weapons of mass destruction, marketing, post-modern critical theory, German meaning of 'mist'*" (Tufte 2009 p. 149).

appendix B: material significance & oomph

The distinction between practical importance and statistical significance is well understood by statisticians and although pounded into the heads of students in every first year college statistics class, it seems to be the lesson soon forgotten. I have convinced myself the reason is the lack of a suitable label. Authors Ziliak and McCloskey (2011) use the term oomph which brilliantly communicates the idea, but suffers from not sounding sciency enough, especially when communicating to industry. Borrowing from the accounting profession, I propose the phrase 'material significance'. It means oomph.

appendix C: tampering

Tampering is defined as treating common cause variation as if it were special cause variation. It invariability makes the performance of the system worse. The Funnel Experiment provides a good demonstration of tampering in action. (Deming 1994)

appendix D: Shewart's rules for data presentation

"*Rule 1. Original data should be presented in a way that will preserve the evidence in the original data for all the predictions assumed to be useful.*

Rule 2. Any summary of a distribution of numbers should not give an objective degree of belief in any one of the inferences or predictions to be made there, for that would cause human action significantly different from what this action would be if the original distribution had been taken as a basis for evidence." (Deming 1967)

Reinforcing this point Deming would write; "*Analysis of variance, t-test, confidence intervals, and other statistical techniques taught in books, however interesting, are inappropriate because they provide no basis for prediction and because they bury the information contained in the order of production.*"(Deming 1992, p. 132)

references

Adams, Douglas (1979), *The Hitchhikers Guide to the Galaxy*, Random House, Inc., New York,

Aon/Hewett (2010), *Engagement 2.0* p. 4 Available at: http://www.aon.com/attachments/engagement_2%200_oct2010.pdf

Buckingham, M. and Coffman, C. (1999), *First Break All the Rules--What the World's Greatest Managers do Differently.* Simon & Schuster, New York, N.Y.

Deming, E.W. (1967), "Walter A. Shewart 1891-1967," *The American Statistician* 21: 39-40

-- (1975), "On Probability As a Basis for Action," *The American Statistician*, Vol. 29, No.4

-- (1992), *Out of the Crisis*, Massachusetts Institute of Technology, Centre for Advanced Engineering Study, Cambridge Mass.

-- (1994), *The New Economics*, Massachusetts Institute of Technology Center for Advanced Engineering Study

Deutsch, M. M. (2001), *Employee Engagement: Understanding the Business Case and Providing Tools for Increasing Engagement in the Workplace,* ASQ Human Development & Leadership Division

Feynman, R. P. (1998), *The Meaning of it All; Thoughts of a Citizen Scientist.* Addison-Wesley, Reading Massachusetts

-- (1999), *The Pleasure of Finding Things Out*, Perseus Books, Cambridge Massachusetts.

Fleming, J. H., and Asplund, J. (2007), *Human Sigma* Gallop Press, New York.

Herzberg, F.I. (1987), "One more time: How do you motivate employees?", *Harvard Business Review*, Vol. 65 Issue 5

Huff, D. (1954), *How to Lie with Statistics*, W.W. Norton & Company, Inc. New York, N.Y.

Kano, N. Nobuhiku S., Fumio T. and Shinichi T. (1984), "Attractive quality and must-be quality," *Journal of the Japanese Society for Quality Control* 14 (2): 39–48. ISSN 0386-8230.

Rosenzweig, P. (2007), *The Halo Effect*, Free Press, New York.

Rucci, Anthony J., Kirn, Steven K., and Quinn Richard. T. (1998), "The Employee-Customer-Profit Chain at Sears," *Harvard Business Review* (January-February 1998).

Shellenbarger, S. (1998) "Companies Are Finding It Really Pays To Be Nice To Employees," *Wall Street Journal*, Page 1B, Wednesday July 22.

Towers Watson (2011), *The Power of Three: Taking Engagement to New Heights*. Available at: http://www.towerswatson.com/assets/pdf/3848/Towers-Watson-EmployeeSurvey_power-of-three(1).pdf

Tufte, Edward. *Beautiful Evidence*, Graphic Press LLC, Cheshire Connecticut, 2006.

Watson Wyatt (2002) *Human Capital Index®: Human Capital As a Lead Indicator of Shareholder Value*. Available at: http://www.watsonwyatt.com/research/printable.asp?id=W-488

Zilak, S. T. and McCloskey, D. N. (2011), *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice and Lives*, University of Michigan Press