

C

E



Center for  
Effective  
Organizations

---

**The Promise of Big Data for HR**

**CEO Publication  
G 13-08 (627)**

**Alec Levenson**  
*Senior Research Scientist  
Center for Effective Organizations  
Marshall School of Business  
University of Southern California*

**July 2013**

## The Promise of Big Data for HR

Alec Levenson  
Center for Effective Organizations  
Marshall School of Business  
University of Southern California  
[alevenson@marshall.usc.edu](mailto:alevenson@marshall.usc.edu)

July 5, 2013

“Big data” is all the rage these days. Companies have been making large advances in understanding their customers and markets as we gather more and more information on how people shop, work and live their lives. Now we are hearing many people advocate for the impact that “big data” is going to have on HR. But how much of this is hype versus reality, and where are the meaningful insights really going to come from? In this article I review recent big data developments and map out the likely paths for HR. Three types of innovation are addressed: new data, better access to existing data, and new types of analysis. Also discussed is what measures of employee engagement and linkage analysis tell us about the factors underlying business performance, and the role of causal models in big data and HR analytics.

### **What’s going on outside of HR**

Outside of HR there is a revolution taking place in our understanding of consumer behavior and business processes. On the consumer side, we are learning more and more every day about what drives people’s purchasing decisions from data collected from cell phones, social media and internet browsing. Huge databases are aggregating that information in ways never before possible, enabling an incredibly detailed description of how people live their lives and purchase goods and services. Grocery stores, department stores, and Amazon today all have very detailed information on consumer shopping carts – both virtual and physical – including what products are bought together, how much time people spend shopping, what they respond to from advertising, and what factors are likely to promote additional sales of specific products and services. In short order these pictures will become even more detailed, as the appliances and systems in our homes and our cars become even more linked into the internet and can report our minute-by-minute activities, even if we aren’t aware that any reporting is taking place.

For example, I recently took a trip to Chicago from Los Angeles. On the plane on the way to Chicago, I purchased in-flight wifi service at the beginning of the flight, with about 4 hours to go. I was presented an offer of three hours for the price of two, for about \$15, which was cheaper than buying

access for the entire flight; if I wanted to purchase only two hours the price would have been about \$10, and one hour about \$7. On my return trip home, I waited until there were about two and a half hours left in the flight, and was presented with a different offer: two hours for \$15; if I wanted to pay less than that I would have gotten only one hour of access time. This represents a type of dynamic, location-and-time specific pricing, which the wifi company was able to offer because they knew exactly which flight I was on, and how much time was left – all without me ever giving any personal information. Innovations in product offering and marketing like this are truly breakthrough – nothing like this was feasible before, and we are going to see more and more similar innovations as technology that gathers information on every aspect of our lives becomes only more pervasive.

On the business process side the evolution is also impressive, though more incremental. The proliferation of internet enabled computers, machines and vehicles using GPS devices, RFID chips and the like enables companies to monitor many aspects of business remotely and remove a lot of guesswork that used to be involved in managing the business. Now companies can know with incredible accuracy how their supply chain is performing, what is happening with the sales of their products and services, and if their equipment is operating optimally. Vehicles can summon emergency help after a crash, even if the driver is unconscious, and can report back on the how safe the driver is under normal driving conditions. Real-time congestion maps are being integrated with route optimization software to cut down on driving time and idling in traffic. Corporate headquarters can monitor the uptime of manufacturing lines 24x7 across the globe. All of these developments have led to much more efficient allocation of resources across the enterprise, saving tons of money that otherwise might be wasted on excess inventory, costly repairs, and inefficient usage of vehicles and equipment. In contrast to the breakthroughs in product and marketing described above, these innovations in business process management do not amount to a revolution, yet they are still very important for business success.

While all these advances are impressive, there are some substantial gaps that we continue to have in understanding consumer behavior and business processes. On the consumer side, we can monitor what catches people's eyes when they walk through a store, and the parts of the brain that are stimulated by advertising and images. We can see how they respond to choices of different options for cars or beauty products. But understanding what drives a lot of purchasing decisions is still part art and part science. There are a million new product and service ideas innovators dream up every year, yet only a small fraction are ever successful commercially, whether it's new product development or ideas on how to drive incremental sales from existing products lines. Trial and error are still the best ways of

measuring what's going to be a success in the marketplace. That's the art side; predictions are abysmally inaccurate.

On the science side, our existing models of consumer behavior do an adequate job of predicting how people will respond in many situations even though they are not perfect. The growing mountains of data have not fundamentally changed our understanding of what drives many purchase decisions. We know that someone who buys peanut butter is highly likely to want jelly as well, so they are located next to each other in the grocery store, and right down the aisle from bread (at least in the U.S.); using sexually explicit imagery in advertising will attract people's attention; and warmer clothes sell better during the fall and winter. No amount of new "big data" collection is going to change those conclusions because we have solid models of consumer behavior that make those predictions highly accurate.

On the business process side, for every wonderful new technology that tells us what's happening with our products and services, it seems like there are twice as many barriers to improving organizational effectiveness. Most of the barriers come down to people: how do we get them to do the right things in the right places at the right times. For example, it is easy to identify that customer service drives a lot of purchasing decisions and revenue growth, and that innovation can greatly increase market share. Yet we can't just train employees and tell them to provide top notch customer service or be innovative: they have individual motivation that may or may not be aligned with our work systems, the job design, and the performance management and rewards. If we could really understand and improve what drives employee engagement and motivation, that would go a long way toward improving both sales and margins, and turn the big data promise into a reality.

### **The promise of big data for HR: new data vs. access and analysis**

To improve what is happening on the employee side, we need to expand and improve what we know. Three ways we can do it in this new big data world are through collecting new data, using existing data more effectively, and better strategic analysis.

**Collecting new data:** Some of the biggest gains on the consumer and business process side have been made by gathering new data on what people do and how things operate. Before the current digital era, these data were not available at low cost and certainly not in (close to) real time. When many people hear "big data" this is the probably the first thing that comes to mind: the massive collection and digitalization of information that previously was not available for easy analysis. On the employee side, there are some examples of new data collection that shed important insights into key business processes. Direct monitoring of call center workers' time spent per call coupled with immediate ratings

from customers about their experience can increase both productivity and quality at the same time. The conversion of medical records to electronic format offers the promise of better health care delivery through coordination of care with fewer errors. These both are examples of how technology can enable the direct observation of business processes in ways that reduce errors and increase efficiency. The information collected primarily is about when, where and how work gets done.

A different type of new data collection will come from social media platforms at work. The proliferation of social media platforms outside of work (Facebook, LinkedIn, etc.) has encouraged companies to experiment with adapting in-house versions, such as Yammer, that can be used for communication and peer-to-peer networking in ways that can enhance business performance. Like the early days of cell phones and the internet, we are at an inflection point where it is clear that businesses are almost certain to redefine key processes around the usage of social media platforms at work. The challenge at this point is we don't fully know what those innovations will be, though we can make educated guesses.

One potential contribution of work-based social media platforms is in knowledge management. Social media platforms can provide a user-driven exchange of information that spans functional, geographic and cultural boundaries. Personalized professional profiles, like those found on LinkedIn, may also provide an innovation in internal project assignment, promotion, and career planning. If companies build in-house platforms for employees to share information about their experiences and desired assignments and employees willingly participate, the resulting array of information can be used for better matching of employees to teams and projects.

**Better access to existing data:** Though the new data described above often provide insights into when, where and how employees do their jobs, they typically do not explain *why* employees do what they do – their motivation and engagement, which are the foundations for high performance and successful business processes. On this front, there is not much new data to be collected that hasn't been collected before: we know the questions to ask and have collected these types of data for two generations using interviews and surveys. What is different today is the ease of getting the data: data collection methods have improved greatly and offer the promise of deeper insights on larger groups of employees at more frequent intervals – information that is more “on demand” than in the past, which should improve managerial decision making and organizational effectiveness.

The tools that are available for conducting electronic employee surveys and providing real-time feedback to improve managerial decision making continue to proliferate and improve. Today anyone with access to the internet can design a low or no cost survey. Of course, designing and collecting

accurate and actionable data on employee motivation and engagement requires more than just throwing a bunch of questions up on Survey Monkey. But whether the survey is designed by in house experts or farmed out to an external consultant or survey company, the technological barriers to conducting extensive surveys have been removed for good. Today the only real barriers are created by survey fatigue – too many employee surveys – and the challenges of rallying the organization to take action on the data collected by the survey. The out-of-pocket costs of fielding a survey from a technological perspective are simply too small to be a barrier to collecting new data.

More recent innovations center around shorter “pulse” surveys that are used for more frequent monitoring of employee attitudes. One use of these surveys is more real-time measurement of key questions from an annual employee survey so that managers and the organization more broadly can see how they are doing on a quarterly or more frequent basis, and not be operating off of information that is up to a year old. A second use for these surveys is to measure the progress and improve the success of organizational change initiatives: this can include open ended questions that enable employees to raise issues and contribute ideas to improve the change process. The pulse surveys can complement actions that are taken in response to the annual employee survey, including monitoring key parts of the organization to see if any change has occurred.

A final innovation in the reporting area relates to manager feedback of employee survey data. Survey consulting companies have developed automated routines for providing summaries to each manager of how their direct reports responded on the survey. The individualized feedback reports can be used for benchmarking against other managers’ performance in the organization, and for comparison with previous year’s performance and trend analysis.

**Analysis:** There is very little that is new in the way of how to analyze data, with two possible exceptions. The first exception is social network analysis, which provides a mapping of how information flows in organizations, and the relationships that people rely on to do their work. You can think of it as a way to map the informal organizational structure, identifying people who are disproportionately responsible for transmitting key information and/or who are turned to for advice more often than others. While such mappings can provide some interesting ways of looking at the organization, the time needed to collect the information to do a social network analysis is relatively laborious, and the upside benefit needs to exist to justify the effort. One potential benefit of this type of work lies in the organization’s ability to understand who, from a formal and/or informal perspective influences organization attitudes, behaviors and change. Organizations could use this information when working on a change process. Knowing who influences others would allow the organization opportunity to bring

them into the loop on a change process and would potentially facilitate faster change as they are networked into the social influence system.

A second relatively new practice links employee survey responses to other data sources such as employee turnover, sales, customer satisfaction, etc. and analyzing the relationships to identify drivers of talent retention and business performance. Proponents of this approach include those arguing that engaged employees are critical for business success, particularly in customer-facing roles, and include the authors of the *The Service Profit Chain* (Heskett, Sasser and Schlesinger) and many consulting companies that provide survey hosting services.

Other potential gains on the analysis side are likely to come from making it easier for anyone with minimal training to use traditional statistical software routines. Examples of routines include regression analysis, anova, factor analysis, and even calculating simple correlations and means. HRIS and enterprise software providers are building more and more functions into their offerings that can be operated in turnkey fashion. This offers the promise of dramatically reducing the need to rely on statistical experts to do highly technical analysis. Yet, this creates a corresponding risk at the same time. Knowing how to specify hypotheses, clean data, set up and test models, and examine data for anomalies are all critical competencies that anyone should possess before attempting statistical analysis, even if a turnkey software program promises to automate some of the steps, such as cleaning the data of outliers. Anyone can walk into a surgical room, put on a gown and pick up a scalpel, but I would only trust a surgeon to operate on me. Setting up and running a complicated statistical analysis is not brain surgery. Yet doing it so that real, actionable insights are found requires a minimum level of statistical competence and knowledge about modeling individual and group behavior.

### **It's not just the data, it's what you do with it**

Which brings us to the real “bottom line” when we’re talking about the pros and cons of the current big data era: the critical need to ask the right questions. The old adage about the quality of information from computers – “garbage in, garbage out” – applies equally well to statistical analysis: fancier analysis does not necessarily mean deeper insights, unless the fancy analysis is directly tied to insightful questions that require advanced statistics. Asking the right questions means being clear about what causes what so you don’t confuse correlation and causation. And that in turn means setting up and testing a causal model.

Causal models address “why” things happen. Figure 1 provides an example of a causal model. Generally speaking, there are three main contributors to performance in a job or team, and they are represented in Figure 1:

- Job design: the design of the roles and responsibilities of the jobs or team, and how they are supposed to work together with other roles, teams, departments/functions and business units.
- Capability: the individuals’ or team’s capability to perform the tasks efficiently and effectively.
- Attitudes: how motivated and engaged the individuals or team are to perform the tasks.

In order to do a complete diagnosis of what drives employee motivation and productivity, all three types of contributors need to be considered, even if not all three are ultimately included in the measures used to conduct the analysis.

The challenges of setting up a “true” causal model can be demonstrated using two examples that have received a lot of attention and are closely related: employee engagement and linkage analysis. These both are good examples of the promise and pitfalls of big data for HR.

**Employee engagement.** Employee engagement is a concept that is both old and new at the same time. There is a growing research literature on the topic, which I will not describe in detail here, but will highlight some of the key points that are relevant to this discussion. “Engagement” broadly speaking has been used to refer to everything from traditional measures like job satisfaction and commitment, to new measurements that include thriving, energy, and discretionary effort.

The implied causal model says that higher employee engagement leads to better business performance. Arguing in favor of that causal model, there is evidence to suggest that having more engaged employees in certain customer-facing roles can directly contribute to improved business performance; this is the argument made by the authors of *The Service Profit Chain*, and appeals to the common sense idea that customers will buy more if they are treated well. Yet that common sense insight that is derived from an analysis of what drives bottom line performance in one part of the organization (customer-facing roles) is also the source of sometimes questionable conclusions regarding the data that is collected from employees more broadly and the implications of that data for what HR and line leaders should focus on. It is true that, all else equal, having more engaged employees should be better for organizational performance. Yet that is NOT the same thing as saying that if organizations improve measured employee engagement, that business results will improve; that is the causal conclusion that cannot be universally supported by either data or logic. The answer can be found by paying proper attention to what is being measured and the true causes of business performance.



As George Rose, EVP People and Organization, Sony Pictures notes, the first problem is that: “Focusing on employee engagement alone is not enough. I have known some groups that register as highly engaged on annual employee surveys yet are clearly not high performing. You have to get to the root of the issues that are the true causes of performance.” A group of employees can seem very “engaged” in survey measures because they are happy with the work they are given and the direction provided by their supervisors. But if their supervisors do a poor job focusing the team on tasks that support the business and/or don’t hold them accountable for business results, there can be a disconnect between measured employee engagement and actual business results. From a modeling perspective, what’s missing in this case is the information on goal setting and accountability: these turn what should be a positive relationship between engagement and productivity into a negative relationship.

A second problem comes from the challenge of measuring and comparing engagement across employees. Some jobs are more interesting to do, provide greater autonomy, and are involved in making key decisions (for example, many technical and professional roles); people who work in these jobs are much more likely to register high levels of engagement in a survey because the work is intrinsically motivating. Other jobs are more routine, less interesting, and even unpleasant to do (for example, past due bill or garbage collection); registered engagement in these jobs quite likely should be lower than other jobs in the organization. Thus it is very hard to come up with an objective standard of what is an appropriately high level of measured engagement, especially when comparing different groups of dissimilar employees: what’s missing is the baseline information on how intrinsically motivating each job is. One way around this is to look at trends over time within a group, but that still doesn’t answer the question of when engagement is high enough: there are more things than just engagement that contribute to performance. We need a way of knowing when attention should be focused elsewhere on other areas that could be improved, and which might contribute even more to improved business performance than more employee engagement.

The challenge of drawing a direct link between measured employee engagement and business performance can be seen in Figure 1. Measures of engagement at the individual level fall into the “attitudes” bucket, and thus have the potential to impact business performance. Yet the roles of capability and job design have to be considered as well:

- **Capability:** A group of lower-capability employees could have high measured levels of engagement if they are not pushed hard to perform beyond their comfort zones; yet once they are held accountable for performing at a high level of productivity, measured engagement could plummet as they resent having to work hard at tasks that are difficult for them to accomplish. In this case, high

measured engagement predicts lower business performance, and vice versa; thus engagement in this case does not “cause” performance.

- Job design: Even a highly competent group of employees could be held back from contributing to business success if they are not supported by the rest of the organization because of lack of alignment or resource support. When there is lack of organizational alignment and resource support, any link between employee engagement and business performance can be totally severed.

**Linkage analysis.** Linkage analysis suffers from similar problems. In fact, the statement that employee engagement improves business performance comes from a type of linkage analysis. In its simplest form, linkage analysis attempts to link people measures (employee attitudes and capabilities) with business performance metrics. The implicit assumption is that if the people measures and business metrics are correlated, then making improvements on the people side will improve business performance: people metrics “cause” business metrics. Examples of people measures can include HR interventions and metrics such as training, coaching, competencies and more.

The reason why linkage analysis is so appealing is because it is almost always easy to construct an argument for why the people measures should matter for business performance. Yet, just because something should contribute to improved organizational effectiveness and performance, there typically is no guarantee that it will – and often many scenarios under which it won’t. The basic problem is that while many people measures are important contributors to organizational effectiveness, they are part of an interdependent system that relies on multiple HR contributions to be successful. For example:

- Training plays an important role in closing competency gaps, so measures of training incidence should be correlated with improved performance; yet training is not the only contribution, and often not the most important one.
- Measures of competency demonstration should also be correlated with improved performance; yet some of the most important competencies that drive performance cannot be measured well and thus are missing from the set of measurements (for example: decision making ability).
- Coaching is an important part of effective feedback and performance management, yet it is rarely cited as the most important or critical barrier to performance; instead, it is a contributor to while rarely, if ever, being “the” cause of improved performance.

The problem with linkage analysis that focuses on only one of the measures above is that the analysis is highly likely to show a statistical correlation between the people measure and organizational effectiveness and performance. Yet as the discussion above demonstrates, such correlation cannot be attributed to true causation.

Some organizations have recognized that engagement alone is not enough to help explain or understand the interrelated perspective that results in observed relationships between people data and organization performance. Jack in the Box has adopted a model similar to the one in Figure 1. As Mark Blankenship, SVP and Chief Administrative Officer, explains, “employees can be engaged, but if they aren’t aligned with the organization’s goals (through communication, compensation and accountability) and they don’t have the right team or tools in place to perform (through equipment, training, processes, hiring of the right talent), it’s difficult to determine what organization decisions to make to further improve performance.”

These two examples – employee engagement and linkage analysis – highlight the general challenges of big data facing HR and organizations today: a lot of analysis is being driven by ideas that “make sense” yet which fall short of providing truly actionable insights. The availability of new data that is easier to access more frequently opens the door for lots of interesting analyses. Yet in order for the analyses to move beyond interesting to providing actionable insights that can improve performance, causal models have to be constructed and tested. And whenever a correlation is run that shows a statistically significant relationship between people measures and business outcomes, alternative explanations for why that correlation exists – potential alternative causal models – need to be considered and to the best extent possible ruled out, before any conclusion can be reached about true causality. Only at the point at which we subject the data to scientific testing against alternative explanations and causes of performance will we have the confidence of knowing that big data is moving us forward and not just providing an alluring distraction.

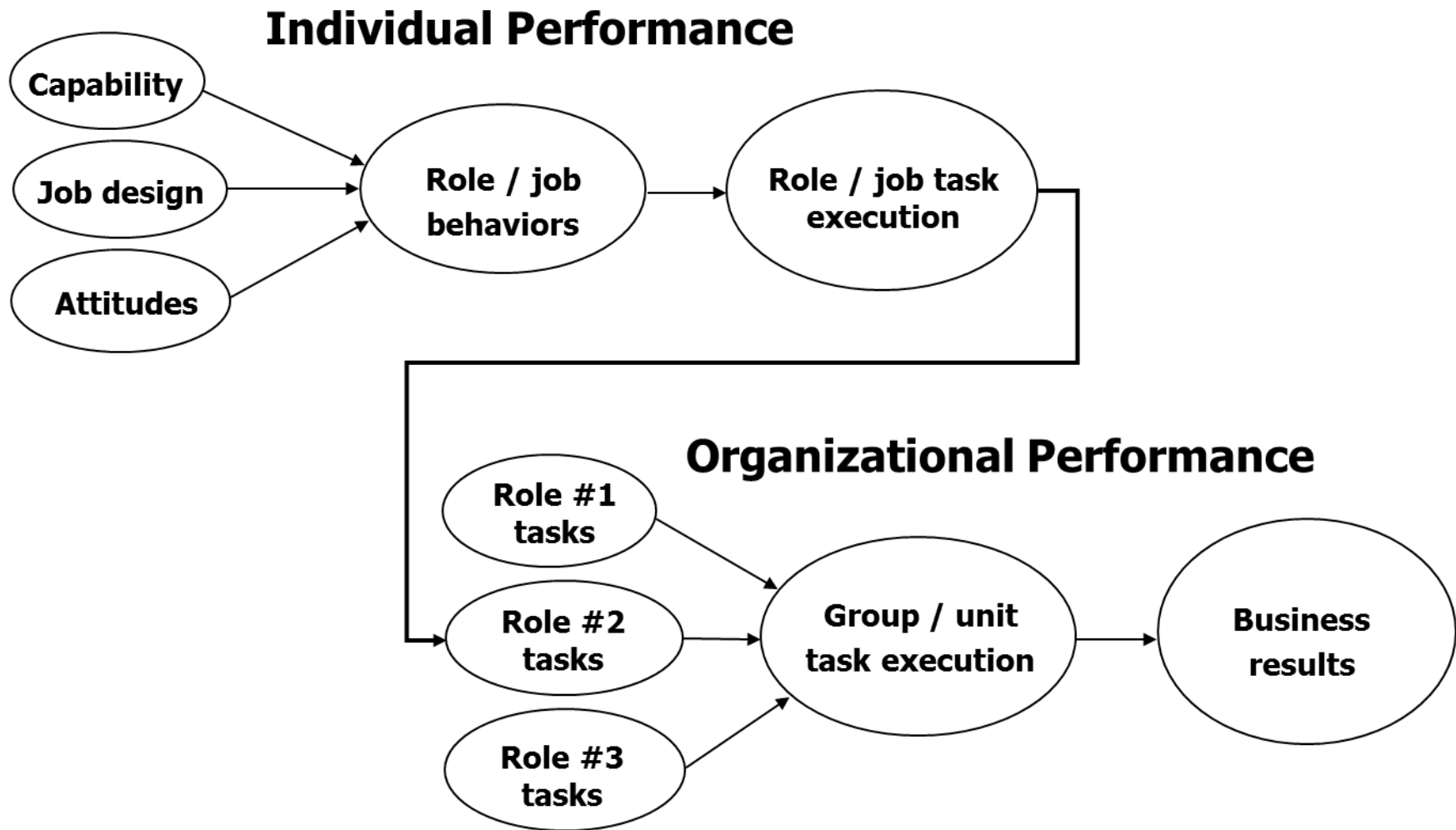


Figure 1