How to Identify and Use Human Capital Analytics

Executive Summary

The purpose of this tool is to provide information about how to identify and use analytics to manage a major organizational resource—human capital—and its effect on organizational performance. This tool is targeted primarily to human capital executives, managers, specialists, and analysts. This includes the Chief Human Capital Officer, whose role it is to coordinate data-driven reviews that focus on key human resource management metrics that support mission accomplishment (U.S. Office of Management and Budget (OMB) Circular A-11, Section 200.14, July 2013). It may also be of interest to Performance Improvement Officers, Chief Operating Officers, and their staff, given their responsibilities to work with Chief Human Capital Officers and other leaders to conduct data-driven reviews to learn from experience, descriptive and predictive analyses, and evaluations and to make corrections to help their agency operate more effectively and efficiently (OMB Circular A-11, Sections 200.10-200.12, July 2013).

Agencies have access to a tremendous amount of data, but a major obstacle to agency top management is determining how to use the data to develop information to help them improve organizational performance. Organizational leaders need analytics that will help them identify the most powerful predictors of performance—predictive analytics. They need to understand the causes for performance improvements or declines to identify what actions they can take to improve future effectiveness. They also need leading indicators that will help them foresee changes, patterns, and trends that will affect the organization's future performance so they can address those changes in a timely way.

Creating an Analytics Group and Function In the Agency

1. Understand Leaders' Information Needs - Analytics group should:
    - Understand the agency's business objectives and processes.
    - Have familiarity with organization's strategic plans, operational plans, budget documents, and performance reports.
    - Be aware of the executives' mindset, concerns, and terminology.
    - Conduct a stakeholder meeting and ask them about the types of human capital information they would most like to see and the organizational outcomes they are most concerned about.
    - Gain support for the analytics effort from top leadership
    - Set up a feedback mechanism for internal customers so that the analytics group can improve the quality, utility, and presentation of analytics.
2. Understand and Work with Other Functions – Analytics group should:
    - Be familiar with the terminology, data, and analytical methods of other organizational functions, such as finance and operations.
    - Demonstrate to other functions how their data helps to create human capital analytics that connects human capital to critical organizational outcomes (e.g., productivity, performance, costs).

- Gain buy-in from analytics colleagues from the other functions.
- Coordinate its efforts with analysts in other functions to ensure that different functions are not reporting different results for the same type of metric, which can quickly damage the credibility of an analytics team and its results.
- Gain support to obtain the budget to supply the staff time, software, training, and other resources needed to implement a human capital analytics program.
- Start off with simple software (Excel and/or R) to demonstrate value and then make the case for more sophisticated software.

3. Establish a Human Capital Analytics Steering Committee – Analytics group should:
    - Form a steering committee made up of individuals from different departments to gain their buy-in, cooperation, insights, and expertise.
    - Provide highly knowledgeable and skilled individuals in statistical analysis and interpretation to advise the committee. Necessary knowledge and skills include:
        - Ability to clean data and evaluate data quality (e.g., reliability, validity)
        - Knowledge of different types of data (e.g., dichotomous, categorical, interval, continuous, ratio) and the appropriateness of various statistical methods for such data types
        - Familiarity with human capital and human resource data and other types of data (e.g., financial, accounting, productivity) likely to be involved
        - Skill in manipulating data and matching and merging data sets
        - Solid knowledge of and skill in applying and interpreting both descriptive analytics and more complex statistical methods (e.g., multivariate, longitudinal, multi-level)
        - Experience communicating results from complex statistical methods to non-statisticians in oral presentations and written documents
        - Interpersonal communication skills to work effectively with team members from a variety of different programmatic and functional backgrounds
    - Also include on the committee human capital managers and leaders with a broad knowledge of human capital management, organizational operations and other functions, and the organization's mission and strategy.
4. Conduct a Data Inventory to Determine What Data and Resources You Have to Work With – Analytics group should:
    - Consider establishing a few basic metrics using easily accessible data to build a foundation for more complex analytics.
    - Based on the business questions identified by the steering committee, determine what data you currently have and how you would access it.
    - Interview the steering committee members to learn about the types of data they have in their functional areas. What data definitions do they employ and data collection methods and sources they use?
    - Build the inventory using accurate, consistent definitions and measurement to help develop a foundation of credibility.

- Realize that integrating data from numerous sources (e.g., human resource, payroll, learning management, operational, customer, and financial systems) is daunting and requires considerable patience and persuasion.
- Ask those who manage and maintain the data about the:
    - Frequency and timing (e.g., time of year) of data collection
    - Original sources of data and methods to collect or derive the data
    - Level of data (e.g., individual, unit, component, agency)
    - Sampling procedures and response rates
    - Categories, response scales (e.g., 5-point scale from "strongly disagree" to "strongly agree") or units of analysis (e.g., years, days, educational credits)
    - Relationships with other data in the organization
    - Ability to identify data for individuals or units so it can be linked to other data for the same individuals or units
    - Codebooks or records that define data elements and give the possible values for each data element (for an example, see OPM's Data, Analysis & Documentation at https://ehr.nbc.gov/datastandards/list)
    - Length of time data are maintained
    - Data cleaning (a.k.a., scrubbing) and quality control procedures
    - Procedures for obtaining data extracts and the lead time necessary for requesting extracts
    - Privacy and security restrictions for data
- Record team members' and data holders' thoughts and hypotheses about how one variable might be related to or influence another variable.

5. Examine the Data – Analytics group should:

- Have skilled analysts examine the quality and integrity of the data that you have to work with:
    - Examine how many data points are missing.
        - What percentage of cases (e.g., employee records) have missing information for a variable?
    - Look at the distribution of values on variables, such as the minimum, maximum, and average values and frequencies for the data.
        - Does the distribution look reasonable?
        - Are there extreme outliers, values that are out of range (e.g., 0 value for age), codes that are not correct, large percentages of cases categorized in an "other" category, or distributions that do not look appropriate?
    - If there is more than one source for the same type of data (e.g., HR information system and payroll), compare the data from the different sources.
        - Do you get the same results from the different sources?

        - Is one source more complete and accurate than the other source?
        - Do the different sources use different data definitions (e.g., who is counted as an employee), collection methods, time periods, or calculation methods?
    - Catalog leading and lagging measures and the frequency with which they are updated
- Have the steering committee resolve differences in measuring/calculating metrics among the different functions.
- Encourage all functions to use common definitions to make it easier to make optimal use of data.

6. Explore the Data with the Tools You Have Available – Analytics groups should:
    - Learn the capabilities of their current analysis tools.
    - Train on how to create effective charts, graphs, diagrams, and infographics.
    - Utilize good presentation skills to tell a story with the analysis.

How to Decide What to Measure

1. In the Short-Term, Gain Support with More Easily Attainable Projects – Analytics groups should:
    - Consider the organization's unmet information and analysis needs and pain points.
    - Use notes from top leader interviews to identify their data needs and thoughts on how the organization's human capital affects its performance.
    - Identify small-scale, easily attainable analytics projects to get started and prove analytics' value.
2. In the Long-Term, Develop a Systematic Approach and Plan – Analytics groups should:
    - Determine what the ultimate questions are that need to be answered and what steps you will need to take to be able to get the data and resources needed to answer them.
    - Develop a longer-term plan that outlines how your human capital analytics function will evolve over time. Consider the following:
        - Review business strategy and goals with C-suite executives and determine how HR can best support these strategies.
        - Identify the HR functions to be measured that align with objectives. Show how these HR metrics would provide the most relevant information about how those functions impact business objectives and strategies.
        - Define each HR metric and its formula. Not all organizations define metrics the same way or use the same formulas. For example, some organizations measure the cost for each new hire without including payment for recruiters as a related expense.
        - Decide what data must be collected and what collection methods are available. For instance, will all data for cost-per-hire be captured in human resource information systems? Or will some data need to be obtained from another department?
        - Decide how often HR metrics information will be collected and reported. Will it be monthly, quarterly or annually?

3. Include Predictors, Not Just Outcomes – Analytics groups should:
    - Include metrics that show actions taken to produce results, not just end results.
    - After identifying predictor metrics, determine if new data collection methods and/or improvements to current data collection methods are needed.

How to Identify Analytics that Matter

Identifying human capital analytics that matter should be considered an iterative process, not a one-time project. Initially, an organization will likely need to focus on determining stakeholders' needs and the availability and quality of data. Once an analytics team has identified what it has to work with, it can begin to understand what the data mean and what types of analyses and other data may be needed to respond to the questions and issues that stakeholders really want and need to know about.

If an organization has just begun to collect certain types of data, it may be limited to cross-sectional analyses because the data were measured from one time period. This will limit the ability to examine what causes or predicts outcomes of interest, because one cannot draw inferences or conclusions about cause and effect relationships unless the data are appropriate for such analyses.

You will need to test variations on models to identify those that perform the best, statistically speaking. Over time, you will want to re-test models and add new data to improve them, or modify models to consider new conditions or changes in organizational issues and objectives.

How Can You Effectively Communicate Results?

1. Decide what reports will be provided, in what formats, and who will receive them. Some results may be part of regular reports or scorecards, while others may be in stand-alone documents or special reports
2. Use the spearhead approach to communication (Pease et al., 2013): prepare an elevator pitch, executive summary, and a detailed report.
    - The elevator pitch is a concise statement of what was learned—short enough to be delivered during an elevator ride.
    - The executive summary should be a 2-6 page document that tells the story of the study by focusing on the impact of the results and recommendations for improvement. It should be written at a memo level—what an executive needs to know about the project. It should be accompanied with a slide deck that highlights major sections from the executive summary. The elevator pitch, executive summary, and slide deck should be in the language of the business. It should not include HR-speak or statistical "gobbledygook." Charts and graphics are more effective than big tables. One should put information about the benefits to their business up front, not where the data came from and what analytic methods were used (Pease et al., 2013).
    - The detailed report documents what happened during the project. This includes the more technical information about data sources, assumptions, methods, statistical models built, and various breakdowns of the data and results (Pease et al., 2013). The detailed report can be written with other analysts in mind as an audience. It should enable someone in the future to understand and even replicate what was done.

3. Overman (2008) stresses avoiding “death by metrics.”
    - Short reports with easy-to-interpret charts and a cover sheet highlighting good and bad “standout” metrics are ideal.
    - Figure out the metrics that organizational leaders care most about and customize reports for those areas. For example, for a call center manager, the concerns may be voluntary turnover of high performers or unscheduled time off. For information technology leaders, it might be progress against staff development plans. For managers of growing departments, it might be time to fill jobs.
4. Avoid HR jargon and speak plainly to what people outside of HR care about.
5. Present executives with a compelling, credible story that the data tell. Data presented to line executives needs to be seen as actionable—they need to understand it and know what to do with it (Robb, 2011).
6. “Use benchmarks as a tool, not as a rule” (SHRM, 2013).
7. Be prepared to defend the validity of benchmark data to internal stakeholders when questions arise about where a benchmark data came from and how comparable the benchmarks are to the organization. Use benchmarks based on quality data-gathering and validation methods—read the methodology and analyses sections of source reports (Shapiro, 2010). Ensure your calculation method and the period your data represents are the same as those for the benchmarks you use.

Appendix

A Brief Guide to Statistical Concepts for Human Capital Analytics

Descriptive Analytics

Descriptive Statistics and Cross-Tabulations

Once you have conducted a data inventory or audit, you need to begin to understand the data. Descriptive analytics are usually the first step. Descriptive statistics, including the number of respondents or records with non-missing data, averages (i.e., means), standard deviations, minimum and maximum values, and frequency distributions can help you ensure that the data are of good quality. At this point, having those who are knowledgeable about the area the data represent (e.g., a finance expert for financial data) can help you determine if the data are of good quality and how to make sense of the data.

Knowing the number of respondents or records that they have to work with and how the data are distributed can help analysts decide what statistical methods they can use to further analyze the data. For example, if there are few data cases (i.e., respondents or records), data with only two values ("yes" or "no," "true" or "false"), or highly skewed data (i.e., most of the data are at one end or the other of the distribution), these data characteristics may limit the types of analyses possible or indicate that they need to use particular methods appropriate for the type of data they have.

Some questions can be answered with descriptive analytics. For example, a compensation analyst for a group of 14 hospitals tracks the following metrics, all of which are based on descriptive statistics and simple analyses:

- Percentage of performance goals met or exceeded, which show whether the company is meeting the performance goals aligned with its mission.
- Percentage of employees rated at the top performance appraisal level who are paid above the average salary.
- Percentage of top-performing employees who resigned for compensation-related reasons.
- Turnover percentage of low-performing managers and employees within one year of receiving a low performance rating.
- Percentage of employees in any performance management program who improve at least one level within the year.
- The rate of involuntary turnover in key jobs. Tracking this metric provides the trend and shows whether actions are being effective in reducing the rate (Robb, 2011).

For some of these, the data can be segmented by group and analyzed through cross-tabulations. For example, the second metric above combines data for performance appraisal rating level and salary level for each individual in the same analysis. The third metric combines the performance appraisal level and type of resignation.

A cross-tabulation table for the second metric from the list above could show the percent of employees at each performance appraisal level who are above the average salary and who are at or below the average salary, as in Table 2. The compensation analyst is interested in the single cell for Outstanding Performance crossed with Above Average Salary. This cell is highlighted in the table. A similar type of cross-tabulation could be used to create a table with a cell related to the third metric for each performance rating level and each reason for resigning.

Table 2. Cross-tabulation of Performance Appraisal Rating Level by Salary Level

	Performance Needs Improvement	**Acceptable Performance**	**Above Average Performance**	**Outstanding Performance**	**Salary Group as a Percent of Total**
Above Average Salary	2 (10% at this rating level)	48 (30% at this rating level)	80 (50% at this rating level)	**50 (83% at this rating level)**	180 (45% of total)
At or Below Average Salary	18 (90% at this rating level)	112 (70% at this rating level)	80 (50% at this rating level)	10 (17% at this rating level)	220 (55% of total)
Performance Rating Level as a Percent of Total	20 employees (5% of total)	160 employees (40% of total)	160 employees (40% of total)	60 employees (15% of total)	400 employees (100% of total)

For these types of analytics, the challenge is more likely to be in getting the right data for the analysis, not the need to figure out how to use complex statistical methods or explain the results to organizational leaders. Often, it is a challenge to get multiple types of data for the same individuals. "Canned" reports from database systems may give you the percent of employees who were rated at each performance appraisal level and the percent who have above and below average salary, but they may not give you the percent at each performance rating level by salary level or by reason for resignation.

To obtain the data you really want, you may have to develop or request a special query and custom report. Or, you may need to request the data as individual cases, so that each employee's data are in a single row and each piece of data is in a column in that row. The same type of data for the employees in the data set is arrayed in the same column (e.g., performance appraisal rating in column 1, base salary in column 2, separation personnel action in column 3, reason for separation in column 4).

You may not need any information that identifies the employees, as long as all of the pieces of data that you need are provided together in the same data set. However, if you need to combine data from one source with data from another source for the same individuals or the same organizational units, you will need some kind of *key variable* or common connecting variable so that you can match and merge the data into a single data set.

Most of the metrics or analytics in dashboards and scorecards are descriptive statistics. Most benchmarking is based on comparing descriptive statistics for one organization to those for a set of organizations.

Correlational Analyses

Correlations are a type of statistical analysis typically used to examine the relationship between two types of data. Correlations range from -1.00 to +1.00, although the range can be narrower if data involved are binary. A negative correlation (e.g., -.30) signifies a negative relationship, such that when one variable is higher, the other tends to be lower (e.g., when job satisfaction is higher, intent to quit is lower). A positive correlation (+.40) signifies a positive relationship, such that when one variable is higher the other is higher (e.g., when employee engagement is higher, customer satisfaction is higher).

Statisticians and analysts often consider correlational analyses to be a type of descriptive analysis because “correlation does not imply causation” (Pease et al., 2010; Trochim, 2006). A correlation describes a relationship. The correlation may indicate that two types of data are related; however, if the correlation is statistically significant, that does not allow one to infer that a variable causes the other. To make conclusions about causality, the data need to be collected under certain conditions and other potential causes need to be considered. We discuss this more in the next section on inferential analyses.

A common example for illustrating the problem of inferring cause-and-effect relationships from a correlation analysis is that of a statistically significant, positive correlation between ice cream consumption and murder rates. A positive, statistically significant correlation does not mean that murder is the cause of ice cream consumption or that ice cream consumption is the cause for murder. There is a missing, third variable—hot weather—that is related to both higher ice cream consumption and higher murder rates that is considered to be the cause of both.

Before using more complex analytical methods, an analyst usually examines univariate statistics that involve one variable or type of data, such as the mean and standard deviation for base salary. The analyst then progresses to bivariate correlations, which show the relationship between pairs of variables. Knowing that two variables are very highly related can provide ideas for more complex analyses, as well as help explain unusual results from those analyses.

For example, if two types of data, such as job level and base pay, have an extremely high correlation, such as +.97 (with correlations, by definition, ranging from a possible -1.00 to +1.00), then they are so highly related to one another that including both of them in the same regression equation would be redundant and cause statistical instability. A regression model should not use both because they demonstrate such a high degree of correlation, or multicollinearity (Tabachnick & Fidell, 2012). A well-trained analyst would be aware of these common problems and examine the data using descriptive analytical methods before conducting more complex analytical methods (Berger, 2004).

Inferential Analyses: Drawing Conclusions about Cause-and-Effect Relationships

Terms like *big data*, *data mining*, *predictive analytics*, and *leading and lagging indicators* have become popular in the business press. It is hard to read a business periodical or open an Internet browser or email box without seeing some mention of big data. The relevance of these terms and trends for human capital analytics depends partly on what one means by them and how much data you need to analyze. In this section, we discuss definitions for these and related terms and discuss their relevance for human capital analytics and HR. We also discuss a critical underlying factor: the analytical conditions necessary to draw conclusions about cause-and-effect relationships.

Big Data and Data Mining

Some data-centric industries and organizations have huge datasets, labeled *big data*. Organizations like Twitter have big data—billions of tweets. As of Twitter's fifth anniversary in 2011, there were 140 million tweets per day. Each tweet contains metadata on date, time, number of followers, account creation date, geodata, and more (Watters, 2011, Kindle location 1779-1782).

Big data is sometimes described in the number of petabytes (a quadrillion or 1,000,000,000,000,000 bytes of data). For example, Google processed about 24 petabytes of data in 2009 (portal.acm.org, Aug. 16, 2009). Microsoft migrated 150 petabytes of data in 7 weeks when it migrated Hotmail accounts to the Outlook.com email system (Wikipedia, "Petabyte," June 15, 2013).

One definition for big data is "when the size of the data itself becomes part of the problem" (Loukides, 2011, Kindle location 274-275). Traditional database and data warehouse systems stop being effective when data are stored across a "horde" of database servers (Watters, 2011, Kindle location 1779-1782).

Companies like Facebook are developing methods to manage big data—as of August 2012, their Hadoop clusters included a 100 petabyte physical disk space in a single file system. In January 2012, Cray began construction of a supercomputer that will have a capacity of 500 petabytes of storage, making it the largest storage array yet realized (Cray, November 14, 2011;Wikipedia, "Petabyte," June 15, 2013).

Data mining has developed from the need to analyze big data: from the "crushing practical needs of business to extract knowledge… in very large data sets with an unknown distribution" (Nisbet, Elder, & Miner, 2009, p. 11). Data mining is, "The use of machine learning algorithms to find faint patterns of relationship between data elements in large, noisy, and messy data sets, which can lead to actions to increase benefit in some form (diagnosis, profit, detection, etc.)" (Nisbet et al., 2009, p. 17).

Thus, data mining is the process of using algorithms and computers to identify valid, novel, and potentially useful patterns in data. "Data mining doesn't start with a model; it builds a model with the data…data mining methods discover patterns in data inductively" (Nisbet et al., 2009, p. xxv). The objective is to find actionable patterns that have utility for decision making aimed at getting something done. It finds patterns and classifications that look forward and predict the future. Data mining can discover patterns in data not previously seen and make models that predict, enabling decisions and action.

Where to draw the line between big data and ordinary data is not entirely clear. Whether your HR function has big data to manage and needs data mining to analyze it depends on the size of your organization, the nature of your business, and the data and questions to be examined. If you are using multiple years of human resource data from hundreds of thousands of employees to predict outcomes represented in hundreds of thousands or millions of cases, the amount of data you have to deal with will be part of the problem. In many organizations, big data is not yet the problem for human capital analytics, and data mining is not the analytical method to start with.

Leading Indicators and Predictive Analytics

Leading indicators are “typically nonfinancial values that tell you if you are on the right track,” such as numbers that let you know if you are on the way to having a successful year. For example, in sales, the number of customer contacts, appointments, product presentations and customer satisfaction scores may be leading indicators for sales volume (Pease et al., 2013, p. 42). For HR, this period’s employee engagement scores, new hire selection test and training scores, employee retention rates, and position vacancy days may be leading indicators for next period’s productivity.

Predictive analytics are related to leading indicators, because both are concerned with prediction. The objective of predictive analytics is to help managers be more proactive and less reactive. Some definitions of predictive analytics are fairly narrow, and authors focus on technology and computing, such as data mining and machine learning. One such definition for predictive analytics is, “technology that learns from experience (data) to predict the future behavior of individuals to drive better decisions” (Siegel & Davenport, 2013, p. 11). In this type of definition, the machine (i.e., technology) is doing the learning and predicting. In contrast, another author suggests that “multimillion-dollar software” is not required and shows how to use Excel to solve problems with predictive analytics (Carlberg, 2012).

A relatively broad definition for predictive analytics that seems relevant for human capital analysis is, analysis that “uses past behavior to predict future outcomes, telling what is likely to happen given a stated approach” (Pease et al., 2013, p. 155). Predictive analytics “not only measures impact but also helps optimize and prescribe future investments” (Pease et al., 2013, p. 13). In this definition, technology is not mentioned. Instead, the emphasis is on analyses that allow prediction.

HR metrics have typically been *lagging metrics* that measure past workforce-related expenditures or efficiencies, such as cost per hire or position fill rates. Lagging metrics describe the past. Predictive analytics are about identifying forward-looking metrics that help predict organizational performance (Pease et al., 2013).

One example of a potential predictive human capital analytic is employee engagement, which has been identified as an indicator of organizational performance (Coco et al., 2011). Declines in employee engagement scores may help predict declines in organizational performance and signal a need to make changes to influence engagement before it has had time to cause a decline in performance.

However, cross-sectional analysis of employee engagement and organizational performance data from the same period will not allow an organization to determine whether engagement is actually a predictor of organizational performance or how much time the organization has before declining engagement may cause declining performance. Without appropriate data and analytical methods, one may draw the wrong conclusion. It may instead be the case that declining organizational performance causes declining engagement, rather than the opposite. In that case, engagement is not a useful predictor of performance, but a symptom or result of performance.

References

Berger, D.E. (2004). Using regression analysis. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), *Handbook of Practical Program Evaluation (2nd ed.)* (pp. 479-505). San Francisco, CA: Jossey-Bass.

Coco, C.T., Jamison, F., & Black, H. (2011). Connecting people investments and business outcomes at Lowe's: Using value linkage analysis to link employee engagement to business performance. *People & Strategy, 34 (2)*. Retrieved from https://c.ymcdn.com/sites/hrps.site- ym.com/resource/ resmgr/p_s_article_preview/ps_34.2_peopleinvestments.pdf.

Cray (2011, November 14). Cray partner on sustained-petascale Blue Waters Supercomputer. Retrieved from http://investors.cray.com/phoenix.zhtml?c=98390&p=irol-newsArticle_pf&ID=1629676.

Loukides, M. (2011). Data science and data tools. In A. Noren (Ed.), *Big Data Now* (locations 270-1485). O'Reilly Media, Inc.

Nisbet, R., Elder, J., & Miner, G. (2009). Handbook of statistical analysis and data mining applications. Boston, MA: Elsevier, Inc.

Overman, S. (2008, October 1). Staffing management: Measure what matters. Staffing Management Magazine, 4 (4). Retrieved from http://www.shrm.org/Publications/StaffingManagementMagazine/EditorialContent/Pages/1008overman.aspx.

Pease, G., Byerly, B., Fitz-enz, J. (2013). *Human capital analytics: How to harness the potential of your organization's greatest asset*. Hoboken, NJ: John Wiley & Sons, Inc.

Robb, D. (2011, December 1). Creating metrics for senior management. *HR Magazine, 56 (12).* Retrieved from http://www.shrm.org/Publications/hrmagazine/EditorialContent/2011/1211/Pages/1211tech.aspx.

Shapiro, J. (2011, April 1). Benchmarking the benchmarks. *HR Magazine, 55 (4).* Retrieved from http://www.shrm.org/Publications/hrmagazine/EditorialContent/2010/0410/Pages/0410shapiro.aspx.

Siegel, E., & Davenport, T.H. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Hoboken, NJ: John Wiley & Sons, Inc.

Society for Human Resources Management (2013, March 26). *Benchmarking human capital metrics*. Alexandria, VA: Society for Human Resource Management. Retrieved from http://www.shrm.org/templatestools/toolkits/pages/benchmarkinghumancapitalmetrics.aspx.

Tabachnick, B.G., & Fidell, L.S. (2012). *Using multivariate statistics (6th ed.).* Pearson.

Trochim, W.M.K. (2006). *The Research Methods Knowledge Base (2nd ed.)*. Internet page retrieved from http://www.socialresearchmethods.net/kb/index.htm.
Also available from Cincinnati, OH: Atomic Dog Publishing.

Watters, A. (2011). How the Library of Congress is building the Twitter archive. In In A. Noren (Ed.), *Big Data Now* (locations 1750-1801). O'Reilly Media, Inc.