EMERGING EVIDENCE ON THE USE OF BIG DATA AND ANALYTICS IN WORKPLACE LEARNING
A Systematic Literature Review

Lisa A. Giacumo
Boise State University

Jeroen Breman
Northwest Lineman College

This article provides a systematic literature review about nonprofit and for-profit organizations using "big data" to inform performance improvement initiatives. The review of literature resulted in 4 peer-reviewed articles and an additional 33 studies covering the topic for these contexts. The review found that big data and analytics could be used to improve selection of staff and training interventions, cut costs, improve retention, and performance as well as meet strategic business goals, but little evidence of big data analytic methodologies actually employed in workplace training and performance improvement projects was found. Additional robust research with methodological descriptions are needed to reveal the extent to which these methodologies yield improved human performance results that positively affect strategic business goals and the bottom line. Opportunities for further research are also suggested.

INTRODUCTION

Charities, nonprofits, and international nongovernmental organizations (iNGOs) recognized early the potential of distance learning to help develop staff competencies in business critical areas. This is especially true for international organizations with staff and offices located in various, geographically separated parts of the globe. For example, the World Wide Fund for Nature started its blended learning Conservation Leadership Program in 1998 (Stoel, 2004), far earlier than many corporate universities began seriously considering online learning as an educational option. Even for those organizations with fewer resources than one such as the World Wide Fund for Nature, there exist several interorganizational

• **Lisa A. Giacumo**, Assistant Professor of Department of Organizational Performance and Workplace Learning, 1910 University Dr., Boise, ID, 83725-2070. Telephone: (208) 426-1377. E-mail: lisagiacumo@boisestate.edu • **Jeroen Breman**, Learning and Development Consultant, Grid University, Northwest Lineman College, 7600 S. Meridian Road, Meridian, ID 83642. Telephone: (208) 888-4817. E-mail: jbreman@lineman.edu

The Quarterly Review of Distance Education, Volume 17(4), 2016, pp. 21–38 ISSN 1528-3518

initiatives to increase access to available online learning. One example is the Cornerstone OnDemand Foundation, which has a distance learning course library that is freely available on the disasterready.org website (Cornerstone OnDemand Foundation, 2016). Another example is LINGOs, a membership-driven organization, which offers a complimentary learning management system for members as well as access to Last Mile Learning, a distance learning library with courses on financial management, people management, and project management (Last Mile Learning," 2016; LINGOs, 2016).

Such nonprofit organizations are no strangers to the use of data to guide relief, development, fundraising, and marketing efforts. For example, in marketing and fundraising efforts, data analysts are essential resources to determine what different audiences are interested in (e.g., types of programs) or to measure the effectiveness of a specific campaign (Toal, 2014). In addition, large amounts of data are gathered in program delivery within such organizations. For example, humanitarian organizations use technology to register individuals affected by disasters and better manage certain services that are offered to them (Chibafa, 2014). Additionally they can be used to create end-to-end visibility in their supply chain data (Helios Foundation, 2014). That being the case, we hypothesized, *would it not make sense for such organizations to analyze if and how performance improvement, including staff development programs, resulted in operational gains?*

Having observed such examples in the authors' combined 17 years of experience working in performance improvement, training, distance education, and capacity building initiatives in nongovernmental organizations (NGOs) and iNGOs, we were interested to determine *to what extent learner analytics are driving management decisions about such initiatives*. We also were interested in determining the potential uses of "big data" (to be defined below) and analytics to improve the practice of instructional designers and learning and development project managers.

A systematic review of recognized academic publications yielded very few results (specifically, one) on our primary research question. While there were examples that big data are being employed in nonprofit organizations (e.g., the United Nations), the use of big data in projects focusing on training and development were almost nonexistent in the literature. Therefore, the review was broadened in scope to the use of big data and analytics in workplace learning, performance improvement, and learning and development in business and governmental settings. This broadening of scope yielded a greater number of documents covering our topic, and therefore serves as the basis for this literature review.

We describe this literature review as *a systematic literature review of how big data and analytics are currently being used in institutionally based formal workplace learning, talent development initiatives, and training departments*. Additionally, we examine why many organizations are not yet analyzing their own big data, and what resources are required to begin utilizing big data and analytics to improve outcomes, minimize costs, increase revenue, and achieve strategic business outcomes for the typical organization. Finally, implications for practitioners, policy makers, and decision makers are presented. We conclude with suggested directions for future research.

BACKGROUND

Economic Evidence that Data-Based Decisions Improve Business Results

As recently as 2011, Brynjolfsson, Hitt, and Kim developed research methods and authored a study that showed evidence of firms managed with a strong data driven decision making approach have output and productivity levels about 5%–6% greater than would normally be expected, given investments beyond traditional internal and external data collection,

input analysis, and IT resources (Brynjolfsson et al., 2011). Tambe and Saraf (2014) found that organizations that employ the use of big data technologies can show up to 1%–3% higher productivity than the average organization that does not use this technology, but only in those organizations that hold significant data assets and are situated in labor markets where a cadre of employees possessed complementary technical skills. It would seem that the analytical skills possessed by experienced performance improvement specialists and instructional designers, coupled with proficient software engineers, data analysts, and information technology (IT) professionals, could potentially serve as this "cadre" of employees possessing these technical skills.

Big Data, Analytics, and Workplace Learning

The use of big data and analytics in workplace learning appears to be one of the current big trends of organizational development, as a review of trade magazines, professional organization websites, and the blogs of leading learning and development authors and consultants will demonstrate. The Chartered Institute of Personnel and Development lists big data as one of ten key shifts learning and development professionals need to make: "There's more than enough information out there, but deciding what to use and how to make it valuable is a challenge" ("Get Ready", 2015, p. 9).

However, from experience and involvement with hundreds of colleagues connected through professional societies and conferences, the authors have not yet personally witnessed much direct evidence of training opportunities to develop skills in working with big data and data analytics, nor implement those skills consistently in general practice. Young (2015) reported that as of 2014, Deloitte found that only 14% of companies were conducting statistical analysis on employee data. However, the companies conducting these analyses report an increased investment in analytics and modeling and these findings reflect a growing industry trend. The question is: if there are opportunities for learning and development managers and instructional designers to harness the potential benefits of big data and analytics for purposes of increasing efficiency and effectiveness of training and workplace learning efforts, what evidence can be found through an objective, systematic review of the state of scientific research and professional practice to guide researchers, practitioners, and students who are interested in acquiring these skills?

What Is Big Data?

A preliminary search for content associated with big data conducted in most any scientific academic database will return thousands of results. Mayer-Schönberger and Cukier (2013) suggest that big data will result in "the ability of society to harness information in novel ways to produce useful insights, goods, and services of significant value" (p. 2). IT researchers are developing data analytics methods to use big data and to solve data-related problems in contemporary business organizations (Chen, Chiang, & Storey, 2012). Recently, governments, e-commerce, and health organizations have begun to systematically leverage the vast amount of web-based, mobile, and sensor-generated data for insights leading to new science and development of systems (Chen et al., 2012).

While academic database search results for big data will return a vast amount of research articles, there is nevertheless no widely agreed-upon, concrete definition of the term (Dumbill, 2013). One of the more popular early definitions of big data focuses on volume, velocity, and variety; the "three 'Vs'" of data management (Laney, 2001). In this schema, there must first be a large amount of data to be collected and analyzed, called *volume*. *Velocity* speaks to the timeliness of the data collection and analysis in order to maximize its value. The various types of data, both semistructured and unstructured data such as audio, video, web, and text, along with

traditional structured data constitutes *variety*. More recent discourse about big data indicate a fourth "V," or the need to realize the hidden *value* (Gantz & Reinsel, 2011). Chen, Mao, and Liu (2014) posit that this is the more important piece of handling big data, being able to uncover the value of large-scale data sets. However, researchers agree that the definition for big data may adapt to the different needs of researchers across diverse disciplines and will also continue to evolve as technology evolves (Chen et al., 2014; Dumbill, 2013; Gantz & Reinsel, 2011; Manyika et al., 2011).

While the four "Vs" of big data schema can be used to define current research endeavors in traditional education, training, or workplace learning setting, it should be noted that a gap exists within the literature reflecting the extent to which most researchers and practitioners in these areas have access to data points, technology resources, and the human resources required to process vast quantities of information. While some of this access is related to management culture, costs, and capacity, privacy is also more protected in an employer-employee or business-business relationship than in a market-consumer relationship and therefore inherently limits the amount of information actively collected for decision-making. Therefore, the sheer amount of information that is considered big data in a Fortune 500 company is not necessarily aligned with what is considered big data in a more traditional education, workplace learning, training, performance improvement context.

For example, to some researchers in information management, big data is often defined in tens to hundreds of terabytes of information pulled from a variety of sources, resulting in data sets that are so large and complex that traditional processing software is not robust enough to extract meaningful and valuable information. Meanwhile, the volume and variety of data that has been described by individuals conducting big data research and practitioners working in educational technology, learning sciences, human performance technology, and human resource development, have generally been employing smaller and less complicated datasets. We suggest, therefore, that a beneficial definition of big data is *the ethical collection of data points from a volume of learners in a given population over enough time to show patterns or changes in patterns, with the velocity required for timely policy, program, or organizational management decisions, that come from a variety of both primary and secondary sources, intended to result in added-value in the learners' or organizational performance*. We suggest that this definition would be of use to researchers and analysts serving individuals' and organizational needs in traditional education, training, informal workplace learning environments, or other noninstructional performance improvement interventions.

Why Are Analytics Important?

Analytics are the methods that specialists use to derive meaning from the raw big data and rich layers of information available to inform data-based decisions (Gandomi & Haider, 2015). These methods are a process of gathering, cleaning, organizing, and analyzing data available from multiple databases to answer questions about employees' performance, social network and community of practice participation, training interventions, use of online learning resources, e-learning, and other electronic learning systems. The term "learning analytics" has been used when researchers investigate workplace learning or instructional systems. Bersin (2013) calls this "datafication" process *talent analytics* or *people analytics* in order to make a connection with professionals supporting talent development. Others have called this process *HR analytics* or *intelligent analytics* (Pape, 2016; Young, 2015). The precise descriptive term placed in front of the word analytics can be customized to meet the expectations of one's target audience. These analytics can be used to answer desired questions with a descriptive, predictive, or even present prescriptive recommendations.

Workplace Learning

For the purposes of this article, we will use the term "workplace learning" to describe the primary context of interest. The term workplace learning encompasses all efforts made by professionals of various roles in organizations, who support the strategic performance improvement, training, learning and development, and talent development needs of adults. Today, these functions are most frequently supported by providing employees or volunteers with access to electronic performance support systems, e-learning, online learning, and other types of talent development support that are traditionally associated with distance learning in NGOs, not-for-profit organizations, governmental, and military organizations, as well as for-profit organizations.

RESEARCH QUESTIONS

In order to determine to what extent big data and analytics are being used by organizations and review the scientific research available to support decision makers, policymakers, adoption by practitioners, this work investigates the following questions:

- Question 1: Does the use of big data and analytics either reduce costs or improve iNGOs' training efficiency, workplace learning effectiveness, and performance outcomes?
- Question 2: What training and workplace learning outcomes, can be informed by the use of big data and analytics?
- Question 3: What factors are involved with an organization's ability to leverage big data and analytics for training and workplace learning?

METHODS

We implemented a formal, systematic literature review process of academic literature to find evidence of the use of big data in relation to learning and development within organizations to direct further research efforts (Petticrew & Roberts, 2008). A protocol and steps taken to conduct the literature review process was carefully designed to extract the scholarly information available on the topic. The method employed is explained in the sections below.

While conducting the literature review, we found only one scientific, academic study, which reported the use of big data and analytics in iNGOs' internal training, learning and development, or performance improvement initiatives. Because a single source cannot reasonably be used to answer a research question in a literature review, research question one is not answered in this study.

Search Strategy

As part of the systematic search approach used in this study, we identified major terms arising from the research questions, context, and associated variables. In addition, synonyms and related terms were tested for their ability to retrieve desired literature. Those terms that did not add relevant articles were then excluded. We used the Boolean OR to include synonyms and the Boolean AND to link the major terms with the desired contexts and fields of study. Primary sources were also checked for other supplementary references they might provide.

Search String

The terms used in the searches in this review included:

"big data" OR "analytics"

These were searched in conjunction with terms for workplace learning and for social sector organizations. The search string used for the search by topic in the Web of Science and by subject in the Academic Search Premier, for workplace learning was:

> training OR "workplace learning" OR "human performance improvement" OR "learning and development"

The query for social sector organizations included:

> "nonprofit" OR "not for profit" OR "non governmental organization" OR "non governmental organization" OR NGO

The combination of all three sets of terms yielded only one result in only the Academic Search Premier database search from a non-peer-reviewed journal. Therefore, as previously stated, we responded by deleting the context restriction for social sector organizations, and subsequently focused on reviewing the available literature on big data and workplace learning.

The Boolean NOT was used to help limit the search from retrieving results that were not directly related to the use of big data, analytics, and workplace learning. These excluded terms returned research literature that was related to computer science, training of data scientists, computer chip development, bioinformatics, and advanced medical specialists. Additionally, searches were initially limited to refereed peer reviewed academic journals.

Resources Searched

The search of resources was initially conducted using two digital libraries that each search different combinations of written resources, the ISI Web of Science and Academic Search Premier. In the ISI Web of Science, the core database was searched. In the Academic Search Premier, the electronic databases searched included those identified as relevant for social sciences, learning and training in the workplace, governance, business and not-for-profit organizations, including: Academic Search Premier, Applied Science & Technology Full Text (H.W. Wilson), Business Source Premier, Education Research Complete, Environment Complete, ERIC, Family & Society Studies Worldwide, GreenFILE, LGBT Life with Full Text, Library Literature & Information Science Full Text (H.W. Wilson), Military & Government Collection, Philosopher's Index, Professional Development Collection, PsycARTICLES, PsycINFO, Psychology and Behavioral Sciences Collection, Public Administration Abstracts, Race Relations Abstracts, Small Business Reference Center, Social Work Abstracts, Vocational and Career Collection, Web News. Because only four relevant peer-reviewed articles were returned from these databases, we next broadened the search scope to include magazines and trade publications. This alteration in scope returned an additional 33 results.

Search Process

The searching of the digital libraries was conducted by the two authors, who were also the primary reviewers of the collected information. The authors collaborated on developing the search strategy and search terms. The full list of articles identified by the searches was then further refined according to a set of inclusion and exclusion criteria. Table 1 depicts the total number of publications returned for each library search conducted. The relevant number and the percentage of the total relevant returned publications from each library is also shown.

TABLE 1
Library Search Results

Library	*Relevant*	*Not Relevant*	*Total*	*Percentage Relevant*
Web of Science	4	197	201	1.99%
Academic Search Premier	33	346	379	8.71%

Inclusion Criteria

Articles that described examples of the use of big data and analytics to inform workplace learning initiatives and outcomes associated with its use were included in this review. This literature base includes publications, technical reports, or other "gray" literature that describes empirical studies or other study designs in which big data or analytics were applied in the primary workplace learning context of interest. To be included in this review, workplace learning was required to be the primary context of the study.

Exclusion Criteria

Articles that described examples of big data and analytics used in educational organizations, such as schools for children, colleges, universities, and other higher education organizations, were excluded from the analysis. These studies were excluded because assessment of learning is often focuses on academic performance that may not find direct application beyond the academic setting. Also, the needs of the audience, as well as the information gathered from the analytic tools used by educational organizations, generally serves purposes quite different than the context of noneducational businesses, who are developing adult abilities (i.e., to produce measured improvements in efficient and effective work output resulting in cost savings or revenue increases associated with strategic business outcomes). In addition, articles were excluded from the search results when they were focused on machine learning, machine pattern recognition, data science or computer science, training data scientists, railway safety, genetics, biotechnology, medicine, bioinformatics, predictive marketing analytics, and conference announcements.

RESULTS

The results of this literature review confirmed that a surprisingly small amount of work has been published in scientific academic journals on this topic, including training and performance improvement journals. These results are summarized below according to each research question. As mentioned previously, we were not able to answer research question 1 ("Does the use of big data and analytics either reduce costs or improve iNGOs' training efficiency, workplace learning effectiveness, and performance outcomes?") due to a lack of available scholarly literature.

Question 2: What Training and Workplace Learning Initiatives Can Be Informed by the Use of Big Data and Analytics?

Staff Development Intervention Selection, Candidate Selection, and Retention

Many of the initiatives that can be better informed by the use of big data and analytics center on a more granular, efficient, and effective staff performance development and retention. Big data and analytics allows for better pattern recognition and identification of high performers and critical behaviors when they are recorded by employees, managers, and online user navigation information (Dutton, 2014; Gardner, McGranahan, & Wolf, 2011; Hall, 2013; Paine, 2015; Young, 2015). These individual patterns can also be aggregated to reveal highly localized staff retention distributions, which can be modeled to inform organizational management decisions (Bersin, 2013; Freifeld, 2014; Hyatt, 2009; Roberts, 2007).

For example, big data and analytics can be used to generate information valued by chief executive officers, chief financial officers, and learning and development managers about workplace learning interventions and resources that employees seek out, those that are used most often by employees and their managers, and those that each would recommend to others (Bersin, 2013; Everson, 2015; Freifeld, 2014; Gardner et al., 2011; Hall, 2013; Hartley, 2013; Kettleborough, 2014; Paine, 2015; Pease & Beresford, 2013; Tambe & Saraf, 2014). With innovative performance

development software, learning and development specialists and managers can track employee actions, contributions, and output results, post intervention (Baldassarre & Finken, 2015). Over time, a selection of interventions, by design, type, and content, can be balanced against associated costs and productivity through the use of big data and analytics. As more individuals are tracked, and as more data is collected on each individual within the organization, intervention selection and associated productivity could eventually be predicted by comparing similar profiles, backgrounds, and workplace learning opportunities, in combinations that prove to have significant influence on performance outputs (Bersin, 2013; Pease & Beresford, 2013).

Proposals for Human Capital Investments Can Be Translated in Business Numbers

Talent development and performance improvement professionals can leverage big data and analytics to help them develop standard human capital metrics, indicators, and scorecards. Human capital metrics, indicators, and scorecards can then be used to describe the cost and added value of talent investments in the language of business numbers such as money, cost, and returns on investment (cf. Berk, 2004, 2005; David, 2006; Dell, 2012; Dutton, 2014; Hartley, 2013; Higgins, 2014; Kristick, 2012; Moore, 2005, 2011; O'Leonard, 2012; Pease & Beresford, 2013; B. Roberts, 2007; P. Roberts, 2015; Saunderson, 2014; Tozman, 2012; Young, 2015). Higgins (2014), for example, describes an organization's investment in promotions to develop internal employees' competencies cost the organization 30%–35% less than the hiring of external candidate replacements.

Social Networks and Big Data Analytics Support Real Time Performance and Productivity Changes

Social network technology can be used to connect individuals with those who have specialized expertise to quickly solve problems (de Laat & Schreurs, 2011; Dutton, 2014). Such technology can also track performance behaviors online and, more recently, offline, as well. De Laat and Schreurs (2011) developed an online tool focused on real and urgent problems in the workplace that require learning relationships with colleagues or known experts, and combines three streams of information: (a) an overview of current problems on which professionals are currently working; (b) a network visualization based on existing professional relationships; and (c) the organizations or subdivisions within the greater organization in which these network members are located. Through investigation, another social media type of learning software (Siadatya, Gašević, & Hatala, 2015) found that recommender system technology aids discovery of useful performance development opportunities, and underscores the critical need for seamless social media integration in modern workplace learning environments (Everson, 2015). The "design of work and learning environments needs to be intertwined" so that learners can "easily share their experience with their colleagues" (Siadatya et al., 2015, p. 1017).

When social network information is coupled with workplace performance, actionable, immediate feedback about progress, misconceptions, and knowledge and/or skills gaps can promptly be communicated between managers and employees. Training and performance improvement specialists, too, as well as executives, can use such information to make decisions about intervention selections and organizational management or development strategies. For example, Dutton (2014) reported that leveraging social networks and big data analytics in a change management project resulted in a $25 million savings in 6 months for NWH Global. For Halliburton, there was a 10% productivity increase, 22% revenue increase, and a 66% cost reduction within a 9-month period.

Question 3: What Factors Are Involved With an Organization's Ability to Leverage Big Data and Analytics for Training and Workplace Learning?

The *resources*, *culture*, and *talent* or skills that an organization possesses make up the three greatest factors that can impede any efforts to build an analytical organization (Dutton, 2014). As most organizational leaders will state, employees are an organization's most valuable assets (Higgins, 2014). However, the resources and culture can make or break the move to an organization that embraces data-driven decision-making.

Resource Assets Including Data Analysis Expertise, Analytics Software, and Data Storage

An interdisciplinary team that can work together to leverage data assets is required to derive value from big data. This team can get started by assessing the analytical capabilities of key learning and talent development leaders or human resources managers to determine their ability to acquire and manage the flow of big data as well as to derive value from it (Dutton, 2014; Hagel, 2012). O'Leonard (2012) advises that if the learning and development or human resource leaders in an organization do not possess such analytical capability, then these can be either developed or additional talent can be recruited. Paine (2015) describes three case studies in which at least three key individuals, each with a specialty in one of the following backgrounds would support this effort: workplace learning analytics, computer data scientists, and an IT specialist with experience in big data management.

Talent with workplace learning and analytics expertise can lead the effort to ensure that the right data are measured and they are measured accurately. While Hagel (2012) and Hartley (2004) note that analytics can be powerful business tools, enterprisewide human resource management systems are all too often dated legacies with different pieces layered together, and ultimately, unable to handle big data. Computer data scientists provide support with accessing and merging data streams to develop useful reports and data visualization. They accomplish this work with specialized analytics software tools and dashboards, sometimes acquired off the shelf or in the "cloud," other times developed in-house. These analytics tools are needed to process the large amount of data that are merged from a number of data streams, each organized and structured differently, which results in a highly complex network of information (Berk & Magee, 2005; Dolezalek, 2003; Dutton, 2014; Hartley, 2004; Moore, 2005). The massive amount of data assets that track employees' performances and resource consumption require considerable storage space as organizational systems are built and as the organization continues to grow (Berk & Magee, 2005; Hall, 2013; Moore, 2005). A strong IT department can help maintain these types of systems, as well as help ensure privacy of both the data and the analysis results.

Culture of Data-Based Decision-Making Aligned with Business Goals, Use of Metrics, and Performance Management

Dutton (2014) notes that if data is not employed to make better decisions, then an increase in an organization's analytical abilities will not significantly improve organizational performance. In such cases, an important cultural change may be required. Directors, who historically made decisions based primarily on intuition, conversations, or other existing knowledge, may be required to make changes in their practice by tapping into the massive data stream provided by big data and its analysis in order to deliver evidence-based decision-making, as well as to measure progress towards meeting strategic business objectives and performance development goals (Paine, 2015). Organization leaders are also an integral component to cultural change, and can support the cultural change by communicating value for an analytical initiative

and fact-based decisions that are aligned with business goals and made with confidence (Hagel, 2012; Young, 2015).

Along with this focus on decision-making aligned with business goals, Toterhi (2014) describes the importance of developing and using human resource metrics and scorecards with directional indicators and corresponding actions based on careful detailed analysis and conscientious interpretation. According to Dolezalek (2003) and Toterhi (2014), this type of practice and analytic information can be drawn from different sources to help demonstrate the value of training opportunities. For example, the utilization of analytics applications with data inputs including workforce metrics and training usage provides better insight when trying to evaluate whether or not a training is correlated with improved performance and it is more cost effective than traditional needs assessment and evaluation practices. Additionally, Baldassarre and Finken (2015) suggest managers at all levels can make better use of the real-time data to garner results by engaging in developmental check-ins with their direct reports more often than the formerly prescribed once or twice per year routine performance management approaches. Brief but routine check-ins with employees, spanning from the front-line employees through upper management, should be conducted to review successes, barriers, and to acknowledge changes in behaviors as well as add value and provide additional motivation for continuous performance improvement.

Lastly, Wroe (2012) points out the importance of questioning internal data and benchmarking against that data. This becomes possible when external benchmarking data is included for comparison with best-in-class peer organizations in analysis methods. Big data can support better executive decision-making by facilitating the combination of often isolated datasets, which include information on performance data, assessment data, recruitment data, recruiter ratings, manager ratings, and manager performance, as well as peer organizations, to get a complete picture of how an organization is performing, and measure the results of management decisions.

Responsive and Responsible Leadership; From the Top Executives to Line Managers

Most of the articles reviewed in this study suggested that learning and development, and human resources professionals would need to use value metrics and learn to speak the language of finance and become proficient in business acumen (Hall, 2013; Higgins, 2014; Hyatt, 2009). While this is sound advice and can be recommended to all those who have not yet embraced the communication across organizational silos in a language that incites results from the finance and executive leadership, these individuals also need support in developing such skills. In addition, finance professionals will also need to alter reporting practices to acknowledge the value of investment into human capital assets.

When data is integrated from multiple systems, including human resources, budgeting, accounting, learning databases, and employee information, as noted by Dutton (2014), executives should be able to benchmark that data against surveys, courseware, and facilities, and then determine how individual courses affect the organization's performance. However, based on our experience, line managers' responsibility for noting training participation and outcomes in human resource systems is not yet routine. Without a responsible role for monitoring participation in training and workplace learning programs, large gaps of data form and accurate findings may prove to be elusive. Therefore executive leadership may need to communicate these expectations for monitoring and embrace an analytical culture, balancing data with interpretation and qualitative stories, to build organizational enthusiasm for data-based decisions (Toterhi, 2014; Paine, 2015).

DISCUSSION

This section includes a discussion of what we learned, including the nature of the limited literature base focused on big data, analytics, workplace learning, and performance improvement, as well as what we can learn from existing, related big data and analytics research in computer science, business, and economics. We now examine each of these elements and discuss why we believe these findings to be the case, and how these findings might impact organizational learning. We will also make suggestions on how to overcome implementation barriers, based on the systematic literature search findings and our professional experience. In addition, current literature related to equipping organizations with the resources required to leverage the benefits that can be derived from the use of big data and analytics is also covered in this discussion, as this information was deemed imperative for policy makers, decision makers, managers, thought leaders, and advanced practitioners in workplace learning and distance education.

Systematic Literature Review

As previously mentioned, the results of this literature review reveals very little scholarly, peer-reviewed evidence to illustrate the benefits of using big data and analytics to achieve strategic business objectives, reduce costs, or otherwise improve human performance. Indeed, there are but a handful of articles available in popular training and talent development industry magazines and trade journals to suggest potential benefits that human resources, training department managers, and performance improvement consultants might enjoy if an organizational shift in culture was undertaken. We agree with several of the authors we cited asserting that changes in organizational culture surrounding decision-making, away from being based primarily on experience and intuition with more emphasis on massive information inputs, is part of the path forward for organizations (Dutton, 2014; Hagel, 2012; Paine, 2015; Young, 2015). We believe this to be true due to the dearth of descriptions of methods applied and evidence-based results in the literature. Without robust evidence of many success stories stemming from validated and replicable results, it is difficult to promote change as the "what's in it for me" becomes a greater risk. Additionally, we believe this culture shift requires more than overcoming senior organizational executives' (chief executive officer, chief financial officer, chief operating officer, chief learning officer, etc.) perceptions of the potential investment risk and rewards.

Such efforts also require upper management to champion changes to organizational systems, processes, and actual work tasks. Systems and process changes require front-line professionals to be entering the workplace anticipating projects in this area and equipped to address them. As Manyika et al. (2011) echo, we believe that there is an urgent need to prepare the next generation of instructional designers, human performance improvement specialists, and human resources development specialists to collaborate on multidisciplinary teams to identify, collect, and analyze multiple complex streams of data for valuable information in order to inform data-based decision making that either reduces operational costs or increases productivity. This would likely involve research grant funding and industry partnerships, where faculty researchers partner with organizations to acquire anonymized data streams to analyze, cocreate new systems, technology, and mentor teams of interdisciplinary participants who are given an opportunity to apply new methods, investigate models, and share the results in academic publications.

Current Recommendations and Comparisons With Related Work in Computer Science and Business

Researchers working for McKinsey Global Institute predicted that by 2018 there would be a shortage of 140,000 to 190,000 people

possessing deep analytical skills, and a shortage of 1.5 million data-savvy managers with the ability to analyze big data for the purpose of making expedient decisions (Manyika et al., 2011). Learning and development analysts with the skills to move beyond operational reports and be capable of answering questions about how spending is being allocated in different areas of an organization and its relationship with career progression, performance, retention, and product innovation, would better benefit decision making in organizations (Bersin, 2013). In the past, the ability to understand how investments in an organization drives results has been difficult to demonstrate simply from satisfaction ratings collected in workshops. However, such information can be now be discovered with the rich data layers currently available to organizations through big data. *This means that it is now possible to make a shift from linking specific, one-off or even a series of training interventions directly to individuals' performance,* historically a difficult connection to establish. The layers of data from groups of employees' involvement in social network development and communities of practice in an organization may be used to build personas of successful individuals, and thus recommendations for continued development, all of which can be tested against longevity and career progression in an organization. But this ability is only possible if an organization chooses to access or develop individuals with the knowledge and skills to collaborate on the storage, retrieval, and analysis of the data for valuable meaning.

To gain value from using big data, one needs "big information technology," as well as computer science resources. In the early 2000s, researchers started to show the value of productivity and output growth for firms that invest substantial resources into IT (Brynjolfsson & Hitt, 2003; Daveri & Mascotto, 2006; Kim, Lin, & Simpson, 2015; Matteucci, O'Mahony, Robinson, & Zwick, 2005). Therefore, we now understand that firms and market sectors with vast IT resources and significant budgets realize increased productivity and growth outputs. However, there appears to be a "missing link" in the research literature that specifically investigates the changes in productivity and growth that firms realize from investing in IT. This is similarly true for other performance support initiatives related to measuring upgrades to performance support systems, talent development initiatives, training interventions, and other performance improvement interventions.

Off-the-shelf big data analysis methods and applications for learning and development or performance improvement initiatives are not currently widely available. Increasingly there are cloud-based services where very large data sets are uploaded to the cloud for analysis, such as SkillSoft SumTotal, SAP SuccessFactors, Saba Analytics, and Oracle Taleo. However, when combining information sources from legacy systems, as well as other cloud-based systems, bespoke (i.e., custom made-to-order) software is often required before data from all corners of an organization can be merged, analyzed, and reported. Creating bespoke analyses and applications are costly, and require advanced computer science and IT skill sets. For example, key challenges in the development of big data applications include data representation, redundancy reduction and data compression, data life cycle management, analytical mechanism, data confidentiality, energy management, expandability and scalability, cooperation of interdisciplinary researchers to complete analytical objectives, storage capacity, database design, and security (Chen et al., 2014).

The rich data layers from structured business data will provide better insight into the contextual clues about the results of performance improvement opportunities offered to employees through distance learning and educational technologies. These include:

- surveys;
- demographic information;
- semistructured data from project management reports;
- employee profiles;

job performance data; and
- unstructured data, such as internal social media, computer usage, access of internal documentation, participation in professional development, calendar entries, et cetera.

In addition to learning more about the context and how to better design and select performance support systems and allocate training resources to individuals who will endeavor to grow their skillsets, the business value and return on investment will be easier to assess when big data methods are employed. After the performance support system is implemented, big data can provide better insight into the increased productivity and output realized as a result of interventions. The realization of patterns of optimal workplace performance for individuals will allow managers to provide individualized work structure and assignments that are best suited to abilities and interests.

Doan, Ramakrishnan, and Halevy (2011) noted that Web 2.0 content, which was being authored by users on various forums including newsgroups, social media platforms, and crowdsourcing systems, offers insight on different stakeholder groups' attention and focus through information integration for interested researchers and data analysts. For example, as Bo and Lillian (2008) note, text analysis and sentiment analysis techniques are frequently employed to garner information about opinions for social media analytics. Analytic techniques have been developed to mine association rules, discover database segmentation and clustering, detect anomalies, and mine graphs, all of which can produce information about target product use and users' opinions (Adomavicius & Tuzhilin, 2005). Similar data analysis techniques could be applied to organizations' internal electronic dialogue tools and when compared with employee roles and backgrounds, performance records, and professional development activities, which yield important information on morale, departmental focuses, skillsets at different levels of the organization, reactions to training, patterns of productivity, performance, retention, promotions, and change in costs and billable outputs.

While these methods have been used for e-commerce, market intelligence, e-government, and politics, it is a small leap to consider how these very techniques could be further leveraged for training and performance support systems design as well as personalized training recommender systems. For example, potential outcomes of these analytic methods for all stakeholders include improved transparency, participation, and satisfaction, resulting from a better fit between employees' innate interests, abilities, and performance with improved selection of performance improvement supports that are valued and validated. Managers could then feel empowered and able to match employees' interests and optimal performance areas with performance support systems optimized for efficiency and effectiveness, along with training programs, that have been recommended according to increased productivity and more transparent human resource patterns.

Limitations to the Findings of This Literature Review

General limitations

While research exists to document the barriers to adoption, and while these barriers need to be considered, additional scholarly evidence is needed to validate effective methodology and expectations for financial value. The use of big data and analytics is not a direct path to performance improvement, however, and thus does not guarantee benefit to an organization. Big data's integration into an organization's practices is advocated as a way to improve cost-effective resources strategic decision making to recruit, retain, and maintain employee productivity and engagement. Additionally, to date, its suggested use, as well as a very small number of case studies, may be a surrogate for measures of perceived potential value.

Threats to Validity

It may be possible that not all relevant publications were identified. The extent of the findings are based on the search criteria used, as well as the limitations of the research databases harnessed for this project. No preexisting known set of references were used to validate the search terms, and, as described previously, the search was broadened as additional terms were suggested in articles. The combination of searching through several databases relevant for the fields of study, in addition to searching references of included studies along with the use of Google Scholar, would suggest that all relevant work was identified. Additionally, future work published from 2016 and forward is not included in this work. Lastly, publication bias is also a potential threat to validity. It is possible that scholars have found negative results and then not publishing their findings, leaving those findings invisible to the current study. (This could include practitioners who are not publishing their work in scholarly venues, and have developed case studies with various results that are not as widely shared.)

IMPLICATIONS

Those who are formally trained to engage in an evidence-based professional practice in roles associated with the fields of education, learning, human resource development, training, human performance technology have been utilizing various methods of data analysis to gauge learners' newly acquired knowledge, skills, and attitudes that were signaled by changes in behavior long before computers were widely available. What is a new and emerging area of practice for researchers and practitioners are the analytic approaches afforded by digital big data collection and improved processing power. In concert, these two affordances support the potential for finer grained analyses and a richer understanding of human performance for improved policy, program, and decision-making. Based on the literature uncovered in this systematic review, it is clear that a consensus is building: such data can be used at all levels of capacity development, performance improvement, and performance support, from societal policymaking to organizational decision-making. This data can also be used towards the development of individuals.

The methods used to gather and analyze big data can be used to improve learning environments, training outcomes, and the experiences of individual users. A tight economy, which has been very slow to recover from a severe economic crash in 2008 and decreased budgets for training efforts, has perhaps stalled a more widespread development and integration of these methods. Yet, we know that at this time the opportunity to capture and analyze data from educational technology and electronic performance support systems is growing rapidly, and organizations that have mastered a culture of continuous improvement can use such tools to help survive periods of economic downturns.

Business and organizational leadership needs to invest in the technology and the associated human resources that will enable tapping into the wealth of data that is available today. Leadership needs the technology to collect, store, process, and report on data. Ideally, this process should be automated as much as possible. The quality of the data, and accuracy of it, too, are essential, given the life-changing decisions that will potentially be based on it. With that requirement come the parallel issues of privacy and anonymity: it should not be possible, for example, to reconnect data with individual employees. Such data need to be stored securely as well, so that it is not accessible by parties who are not authorized to see and examine it.

Systematic checks to verify the quality of the data gathered, a healthy understanding of the difference between correlation with and causation, objectivity in reporting the data, along with keeping the data anonymized and secured are all key aspects of the responsible use of big data for decision-making (Kettlebor-

ough, 2014; Pease & Beresford, 2013). The analyses and conclusions that are possible with large data sets can produce reliable information only when the data is correct. While this information can be useful to make organizational decisions based on statistics derived from a large group of individuals (e.g., what type of training to fund, or skillsets required for specific roles), we do not recommended using correlational statistics to make predictive decisions about specific individuals without personal conversations between managers and their direct reports. Statistics are aggregated observations of large groups of data points and may not be applicable to all individuals' specific needs, as they are not derived with consideration of all potentially available information. For example when Google decides you are a prime target for specific marketing information, it has been shown that you are *likely to be a consumer but does not mean you will purchase, use, and benefit from each product* determined to align with your consumer profile. Therefore, analytic recommendations derived from available, comprehensive digital records can be useful starting points for systematic management conversations about hiring practices, individualized recommendations for professional development opportunities, and other systemic noninstructional performance improvement decisions. Further, to facilitate a culture of learning and continuous improvement, it is important to report what did *not* work as much as what *did* work, for often more is learned from failure than from success.

While leadership will need to develop the skills to interpret the data that are presented to them, at the same time, learning and development practitioners, too, need to learn the language spoken by business leadership. Executives who lead the learning and development function in organizations and strategists, and those who lead departments, need to document organizational policy (as well as allocate resources for human performance and learning data analysts practice) and be capable of presenting responses to data-driven resource requests in a manner that is understood by both chief executive officers and chief financial officers. Those who are hired to carry out the learning and development function in organizations, such as line managers, need to learn to ask the right questions, analyze the relevant data to help answer those questions, and manage others who are learning to do it as well.

Finally, learning and development analysts need to be brought into the organizational structure, either through training, or by hiring (O'Leonard, 2012). Ideally, these analysts have been trained to analyze social networks, career progression, use of electronic performance support systems, and participation in communities of practice, in addition to tracking investment in learning and development training opportunities. These learning and performance data analysts should be able to speak two languages: that of the learning and development field, and that of executive decision-makers and financial experts (Higgins, 2014).

FUTURE RESEARCH

Although the term big data is a common business term, there is very little published management scholarship that tackles the challenges of analyzing big data and the subsequent decision-making practices that such analyses could afford (George, Haas, & Pentland, 2014). The lack of reliable, rigorous research information suggests that workplace learning researchers have not yet even reached the early adoption phase of Rogers' (2003) diffusion of innovation model. Additional rigorous research and interdisciplinary collaboration amongst the fields of education and learning sciences, organizational development and performance improvement, computer science, business management, finance, and economics is needed to support further development of data collection and analytic methods, reporting instrumentation, and instructional practice. The outcomes of this call for research should aid in data-based decision-making to improve organizational policy,

individualized training program recommendations, and organizational performance. En route to these outcomes, we see vast opportunities for experts in each area to help educate organizational executives, business strategists, and workplace learning practitioners seeking to develop their skills and careers.

While additional research is needed in this area, the potential for a rapid increase in future research is limited at this time. Researchers often develop studies in partnership with advanced practitioners and consultants to further develop methods and tools required to guide development of novice practitioners. The current lack of knowledge and skills found in most practitioners working in the areas of training management, talent development, and performance improvement areas increases the challenge of finding suitable partnership projects. Additionally, the opportunity to further develop methods and tools is limited because very few researchers are working in this relatively new area.

One area in which researchers might start could be found in strategic business partnerships, or in the case of large research institutions with thousands of employees, in human resource departments on campus. Additionally, researchers might look to collaborate with hospitals and other large allied health organizations. Other industries with highly regulated, mandatory continuing education requirements, a technological skilled workforce, and a large number of employees, may also find value in leveraging big data and analytics to support improved training and workplace learning initiatives. Organizations with significant operations in the financial services and various engineering manufacturing industries may also be ideal partners in this endeavor.

CONCLUSION

This study may represent the first time a review of the existing literature base on the use of big data and analytics in training and workplace learning contexts was conducted. This study has directly documented almost 40 references located through a systematic literature review, along with an additional 16 references originating from research conducted on big data and analytics in related fields of study. Perhaps the most notable point ascertained from our review was the discovery that so little has been shared about recent practice improvements and research in this area. While it may seem counterintuitive for businesses to share their practices, we know from the open source movement in software development that everyone benefits from improving upon shared source code. We also know that an economy built on scarcity will eventually self-implode. Hence, there is much work yet to be done to determine the full potential value of such efforts to improve training and workplace learning initiatives. A great service would be done by researchers and practitioners who publish evidence in this area, documenting and sharing information derived from their scholarship.

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17*(6), 734–749.

Baldassarre, L., & Finken, B. (2015). GE's real-time performance development. *Harvard Business Review, 93*(7/8). Retrieved March 3, 2016, from https://hbr.org/2015/08/ges-real-time-performance-development

Berk, J. (2004). The state of learning analytics. *T+D, 58*(6), 34–39.

Berk, J. A. (2005). Identifying tangible business results. *Chief Learning Officer, 4*(3), 30–35.

Berk, J., & Magee, S. (2005). Technological considerations in learning analytics. *Chief Learning Officer, 4*(7), 42–45.

Bersin, J. (2013). The "datafication" of learning. *Chief Learning Officer, 12*(12), 14.

Bo, P., & Lillian, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1-2), 1–135.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). *Strength in numbers: How does data-driven decision making affect firm performance? (*Technical report). Retrieved July 25, 2016, from the SSRN website: http://ssrn.com/abstract=1819486

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171–209.

Conertsone OnDemand Foundation. (2016). Explore all the free learning resources. Retrieved July 25, 2016, from http://disasterready.org/courses

Chibafa, K. (2014). Why not digital? Technology as an interagency tool in the Central African Republic. *Humanitarian Practice Network, 62*, 19–21.

Daveri, F., & Mascotto, A. (2006). The IT revolution across the United States. *Review of Income and Wealth, 52*, 569–602.

David, C. C. (2006). Revving up elearning to drive sales. *Econtent, 29*(2), 28–32.

de Laat, M., & Schreurs, B. (2013). Visualizing informal professional development networks building a case for learning analytics in the workplace. *American Behavioral Scientist, 57(*10), 1421–1438.

Dell, D. (2012). Learning analytics: How data provides direction. *Claims, 60*(10), 29–32.

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM, 54*(4), 86–96.

Dolezalek, H. (2003). Measure for measure. *Training, 40*(11), 72.

Dumbill, E. (2013). Making sense of big data. *Big Data, 1*(1), 1–2.

Dutton, G. (2014). What's the big deal about big data? *Training Magazine, 51*(2), 16–19.

Everson, K. (2015). Leave learning to employees. *Chief Learning Officer, 14(*11), 30–33.

Freifeld, L. (2014). Go figure. *Training Magazine, 51*(2), 4.

Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iview, 1142*, 1–12.

Gardner, N., McGranahan, D., & Wolf, W. (2011). Question for your HR chief: Are we using our "people data" to create value? *McKinsey Quarterly, 2*, 117–121.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal, 57*(2), 321–326.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144.

Get ready for the future of learning. (2015, June). *People Management, 9,* 9.

Hagel, J. (2012). Executives turn data into internal insight. *Journal of Accountancy, 213*(5), 26–27.

Hall, B. (2013). Will big data equal big learning? *Chief Learning Officer, 12*(3), 16.

Hartley, D. (2004). A love-hate thing. *T+D, 58*(6), 20.

Hartley, D. (2013). BM: Driving innovation. *Chief Learning Officer, 12*(6), 46–47.

Higgins, J. (2014). Bringing HR and finance together with analytics. *Workforce Solutions Review, 5*(2), 11–13.

Helios Foundation. (2014). *Visibility from end to end.* Retrieved July 25, 2016 from http://www.helios-foundation.org/humanitarian-supply-chain/oxfam-gb-case-study/visibility-from-end-to-end

Hyatt, J. (2009). The metric system. *CFO, 25*(7), 31-35.

Kettleborough, J. (2014). Big data. *Training Journal, 51*(3), 14.

Kim, G., Lin, W. T., & Simpson, N.C. (2015). Evaluating the performance of US manufacturing and service operations in the presence of IT: A Bayesian stochastic production frontier approach. *International Journal of Production Research, 53*, 5500–5523.

Kristick, J. (2012). Are you learning from your learning programs? *Chief Learning Officer, 11*(8), 30–31.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note, 6*, 70.

Last Mile Learning. (n.d.) *Last mile learning: Learning where it really matters.* Retrieved July 25, 2016, from http://pilot.lastmilelearning.org

LINGOs. (n.d.). Retrieved July 25, 2016, from http://lingos.org/for-anyone

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. New York, NY: McKinsey Global Institute.

Matteucci, N., O'Mahony, M., Robinson, C. & Zwick, T. (2005), Productivity, workplace performance and ICT: Industry and firm-level evidence for Europe and the US. *Scottish Journal of Political Economy, 52*, 359–386.

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work and think. London, England: John Murray.

Moore, C. (2005). Measuring effectiveness with learning analytics. *Chief Learning Officer, 4(*5), 34–39.

Moore, C. (2011). Diamonds in the analytics rough. *Chief Learning Officer, 10*(5), 38.

O'Leonard, K. (2012). Mind the global skills gap. *Chief Learning Officer, 11*(8), 50–52.

Paine, N. (2015). Game changers for learning. *Training Journal, 52*(3), 17.

Pape, T. (2016). Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research, 249*(1), 687–698.

Pease, G., & Beresford, B. (2013, June). Big data analysis drives not-for-profit performance. *T+D*, *67*(6), 22–24.

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide.* Oxford, England: Blackwell-Wiley.

Roberts, B. (2007). Data-driven human capital decisions. *HR Magazine, 52(*3), 105–108.

Roberts, P. (2015). The CFO and CHRO guide to employee attrition. *Workforce Solutions Review, 6*(1), 8–10.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.*)*. New York, NY: Free Press.

Saunderson, R. (2014). What happens when you add big data to L&D? *Training, 51*(2), 54–55.

Siadaty, M., Gašević, D., & Hatala, M. (2016). Associations between technological scaffolding and micro-level processes of self-regulated learning: A workplace stud*y. Computers in Human Behavior, 55*, 1007–1019.

Stoel, D. (2004, July). WWF's corporate university: A highly rewarding investment in human capital. *TD Links Newsletter.* Retrieved from https://www.td.org/Publications/Newsletters/Links/2004/07/WWFs-Corporate-University-a-Highly-Rewarding-Investment-in-Human-Capital.

Tambe, H., & Saraf, O. (2014, October). How analytics helps companies improve talent ROI. *Baseline,* 1.

Toal, R. (2014, August). *Fundraising evaluation and analysis*. Retrieved from https://www.fundsforngos.org/featured-articles/fundraising-evaluation-analysis/

Toterhi, T. (2014). Make sure big data adds up. *Training Magazine, 51*(4), 14.

Tozman, R. (2012). New learning analytics for a new workplace. *T+D, 66*(2), 44–47.

Young, K. (2015, October). Intelligent analytics. *Training Journal*, 56–59.

Wroe, N. (2012). Innovations in talent analytics. *T+D, 66*(8), 30–31.