G Model
INFMAN 2928 No. of Pages 16

ARTICLE IN PRESS

Information & Management xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im

Toward the development of a big data analytics capability

Manjul Gupta[a,*], Joey F. George[b]

[a] Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL 33199, United States
[b] Department of Supply Chain and Information Systems, College of Business, Iowa State University, 3311 Gerdin Business Building, Ames, IA 50011, United States

ARTICLE INFO

Article history:
Received 7 July 2015
Received in revised form 8 June 2016
Accepted 19 July 2016
Available online xxx

Keywords:
Big data analytics
Firm performance
Capability
Resource-based theory
Instrument development
Formative constructs

ABSTRACT

The era of big data has begun such that organizations in all industries have been heavily investing in big data initiatives. We know from prior studies that investments alone do not generate competitive advantage; instead, firms need to create capabilities that rival firms find hard to match. Drawing on the resource-based theory of the firm and recent work in big data, this study (1) identifies various resources that in combination build a big data analytics (BDA) capability, (2) creates an instrument to measure BDA capability of the firm, and (3) tests the relationship between BDA capability and firm performance. Results empirically validate the proposed theoretical framework of this study and provide evidence that BDA capability leads to superior firm performance.

1. Introduction

The information technology (IT) productivity paradox, which refers to the failure to establish a positive relationship between IT investments and firm productivity, has been the focus of several studies since the early 1990s. Eventually, the paradox was resolved, and more than two decades of research suggested several resources (e.g., managerial and technical IT skills, IT infrastructure, and a firm's intellectual capital) that were required to realize the true value of investments in IT. Although we have not yet witnessed the "big data productivity paradox," given the speed at which organizations in all industries and of all sizes are jumping on the bandwagon of big data (i.e., the new forms of data that need sophisticated technology to find meaningful patterns in them), it is likely that we, as information systems (IS) researchers, are waiting for it to happen. While in the 1980s, IT was touted as a competitive weapon, currently, it is big data that is heralded as the next big thing for organizations to gain the competitive edge. According to a recent global survey of 720 firms, 64% of organizations have already invested in or have plans to make investments in big data [1]. This is surprising, given that the research into the economic benefits of big data remains in an embryonic state. While the popular press, which is primarily written by technology consultants (or vendors), is rife with articles defining the characteristics of big data, there is little knowledge about how organizations build big data analytics (BDA) capabilities.

This is important since we know from previous research that firms achieve competitive advantage by building capabilities, which in turn are created by combining and deploying several firm-level resources [2,3]. Following this stream of research, the present study considers (big) data as one such resource, which is necessary but not sufficient to create a BDA capability. In other words, big data on its own is unlikely to be a source of competitive advantage, since all firms (of comparable sizes) will likely be collecting hordes of data from a variety of sources [4]. Similarly, investments alone are unlikely to create superior BDA capabilities [5]. A firm needs a unique blend of its financial, physical, human, and organizational resources to create a capability, which will be difficult to match by competitors [3,6,7]. Moreover, firms need to continuously reconfigure their resources according to changing market conditions [8,9]. However, to do so, it is imperative for firms to be aware of the various resources that are required to build a capability.

This study examines the resources that are needed to build a BDA capability, which is defined as a firm's ability to assemble, integrate, and deploy its big data-specific resources. Drawing upon the resource-based theory (RBT) of the firm, past IT capability literature, and recent work in big data, several resources are suggested. These resources are then categorized into tangible, human skills, and intangible types. Furthermore, this study develops a survey instrument to measure a firm's BDA capability.

* Corresponding author.
E-mail addresses: mgupta@fiu.edu (M. Gupta), jfgeorge@iastate.edu (J.F. George).

http://dx.doi.org/10.1016/j.im.2016.07.004

We followed rigorous scale development procedures described in the management information systems (MIS) literature [10]. Consequently, data were collected into two phases. The first phase was a pilot study in which psychometric properties of all measures were assessed using a sample of 232 BDA managers. The initial pool of items was then refined and another set of survey data was collected from 108 technology executives. The second set of data was used to revalidate the scale properties and to empirically test the relationship between BDA capability and firm performance.

This paper is organized as follows. We begin with a brief review of relevant literature pertaining to RBT and big data. We then discuss different resources that create a BDA capability. Next, we describe our research methods and data analysis for the BDA capability instrument. The paper then presents a discussion section, which includes theoretical and practical implications of this research. This paper ends with a conclusion section.

2. Literature

2.1. The RBT of the firm

What started primarily as a general view of the firm has become “widely acknowledged as one of the most prominent and powerful theories for describing, explaining, and predicting organizational relationships” across the business disciplines (p. 1300) [11,12]. This was no easy task, and it took the resource-based view more than three decades to evolve into the RBT. While several other theories of the firm (e.g., transaction cost economics, agency theory, network view, and institutional theory) have been proposed in the strategy and management literatures, only RBT considers an organization as a collection of resources and presents a powerful framework for uniting several and dissimilar resources, which in turn can be combined to generate competitive advantage [13].

This is important, given the context of our study, because knowing what big data-specific resources a firm has and finding out an appropriate way to utilize them is a key to attaining competitive advantage [14]. Moreover, scholars across business disciplines assert that RBT has the potential to subsume different research streams and other theories of the firm to create one overarching strategic theory of the firm [13,15,16]. Besides RBT, the knowledge-based view of the firm, dynamic capabilities, and contingency theory have gained attention from the IT strategy scholars. While the knowledge-based view of the firm and dynamic capabilities are RBT spin-offs [11], an empirical comparison of contingency theory and RBT by Oh and Pinsonneault [17] found that contingency theory has a lower explanatory power than RBT in describing firm performance with respect to revenue and profitability.

This empirical evidence is consistent with several recent (and past) studies that consider RBT as one of the most compelling theories in the IS and other business disciplines to explain the relationship between organizational resources and firm performance. For instance, Chae et al. [18] suggest that the value of IS resources in firm performance can be further expanded by using RBT. Melville et al. [19] recognize that RBT provides empirically testable propositions, assessment of which will advance our knowledge pertaining to the role of IS resources and organizational performance. Similarly, Gu et al. [20] call RBT a robust framework that enables identification and categorization of IS resources, in addition to measuring the effect of these resources on a firm's competitive advantage and performance. Besides the MIS field, RBT is a well-accepted theory in other business disciplines such as marketing [12], operations management and supply chain [21,22], management [11], and strategic management [23].

In sum, RBT is a principal paradigm for theoretically and empirically assessing the relationship between organizational resources and organizational performance. Given that the main objective of this study is to identify several resources that will allow organizations to create BDA capabilities, which in turn may lead to superior firm performance, the choice of RBT as a theoretical framework for this study seems appropriate. This is consistent with Wade and Hulland [24], who argue that RBT not only provides a useful means to assess the strategic value of organizational resources but also lays out a clear association between resources as an independent variable and firm performance as a dependent variable.

Specifically, RBT proposes that a firm has a collection of tangible and intangible resources, but only the ones that are valuable, rare, inimitable, and nonsubstitutable (or simply VRIN) are capable of generating competitive advantage [25]. Although RBT does not explicitly differentiate between resources and capabilities, Amit and Schoemaker [6] define resources as assets that are owned and controlled by a firm. By comparison, capabilities are defined as a special type of resource [26] that enables firms to aggregate and deploy their resources (in combination) to achieve a desired end [6]. There are several types of resources that have been suggested in the extant literature. As Barney [7,p. 50] puts it:

> Financial resources include debt, equity, retained earnings, and so forth. Physical resources include the machines, manufacturing facilities, and buildings firms use in their operations. Human resources include all the experience, knowledge, judgment, risk taking propensity, and wisdom of individuals associated with a firm. Organizational resources include the history, relationships, trust, and organizational culture that are attributes of groups of individuals associated with a firm, along with a firm's formal reporting structure, explicit management control systems, and compensation policies.

Grant [3] further classifies these resources into tangible (e.g., financial and physical resources), human skills (e.g., employees' knowledge and skills), and intangible (e.g., organizational culture and organizational learning) types. The classification of resources into tangibles, human skills, and intangibles has been actively used in the IS capability literature [2,18,27]. Consistent with this, we follow the same classification to categorize different resources that will be discussed in the following sections.

2.2. Big data

The term “big data” is often used to describe massive, complex, and real-time streaming data that require sophisticated management, analytical, and processing techniques to extract insights [28]. While there is no consensus on the definition and characteristics of big data, the term “big data” was initially coined to reflect the “bigness” or voluminous size of data generated as a result of using new forms of technology (e.g., social media, radio-frequency identification (RFID) tags, smart phones, and sensors). This definition was then extended to include variety (i.e., structured or unstructured data formats) and velocity (i.e., the speed at which data are created). Over the years, others have further dimensionalized big data into veracity (i.e., messiness of data) and value (i.e., the previously unknown insights) [29].

According to some accounts, there are as many as 10 such vs of big data in the practitioner literature [30]. While these severalvs have enhanced our understandings of the big data phenomenon, most of these vs emphasize the technical dimension of big data and understate the importance of several other resources (e.g., human skills, organizational culture) that are equally important to reap the benefits of big data [14,29,31]. We do not suggest that discussing (big) data characteristics is not important; however, the overemphasis on big data characteristics steers focus away from the critical question that the organization must be asking: How to create big data capabilities, which in turn may lead to superior firm performance?

Moreover, a recent report from the Executive Office of the President of the United States [31] suggests "the technical capabilities of big data have reached a level of sophistication" (p.5) and "what really matters about big data is what it does" (p.3). In a similar vein, Markus [30] opines that the real potential of big data lies not in its Vs, but in its affordance to a firm. Marr [14] further states that the major issue faced by today's business leader does not pertain to the characteristics of (big) data, but it relates to how to make best use of it. In line with these evolving views, while this study considers data as an important required resource for creating BDA capabilities, we further identify several other resources needed by a firm to create a BDA capability as discussed next.

2.3. Toward the development of a big data analytics capability

While the published research on big data is limited, there are some studies that have identified challenges associated with the success of big data projects. For instance, Kaisler and colleagues [32] identified data storage and data transport as long-term technology issues pertaining to big data. A survey by New Vantage Group (2012) found that companies were more worried about the unstructured nature of data rather than the volume of data. Zhao et al. [33] suggested that firms must deal with challenges pertaining to the integration of internal (e.g., transactional records) and external data (e.g., social network data). Clearly, new technology is needed to address new challenges caused by characteristics of big data; however, big data-specific technology has progressed immensely in the last few years [31,34].

While we are certain that big data-specific technology will continue to progress, it is time for organizations to focus on other resources, besides technology, which are needed to build firm-specific "hard to imitate" BDA capability [5,35]. For instance, Ross and colleagues [5] assert that the majority of the big data investments fail to pay off because most companies are either not ready or do not make decisions in response to the intelligence extracted from data. McAfee and Brynjolfsson [35] emphasize the importance of adopting data-driven decision-making culture where the senior-level executives make decisions based on data rather than on their instincts. Lack of managerial support is also cited as a critical factor affecting the success of big data initiatives [36]. Another challenge is to recruit fresh talent and train current employees in big data-specific skills, since working with big data requires new kinds of technical and managerial abilities, which are not commonly taught in universities [35,37].

The research discussed so far lists several resources that an organization may need to possess to reap benefits from big data; however, it does not yield insights into how firms can create a BDA capability. Relying on the original RBT, additional work on RBT by several other researchers [3,6,38][e.g.,3,6,38], and recent work highlighting big data challenges as discussed above [5,29,32,33,35,37], we propose seven resources (see Fig. 1) that will allow firms to create a BDA capability. Tangible resources include data, technology, and other basic resources (e.g., time and investments), while human resources consist of managerial and technical big data skills. Data-driven culture and the intensity of organizational learning are suggested as two critical intangible resources needed to build a BDA capability. We next discuss each of these proposed resources in detail.

2.4. Tangible resources

According to RBT, tangible resources are the ones that can be sold or bought in a market. Examples include financial resources (e.g., debt, equity) and physical assets (e.g., equipment and facilities) of the firm. Moreover, the firm's financial statement clearly describes its stock of tangible resources [3,8]. Since tangible resources, to some extent, are readily available for all firms of comparable size [25], these resources are unlikely to provide any competitive advantage on their own. Yet tangible resources are required to create capabilities.

2.4.1. Data

According to a recent McKinsey report, in addition to labor, capital, and land, organizations across all industries now consider data an important factor of production [39]. While organizations in the past have primarily focused on enterprise-specific structured data (i.e., data that can be stored in relational databases) to make business decisions, today's organizations tend to capture every bit

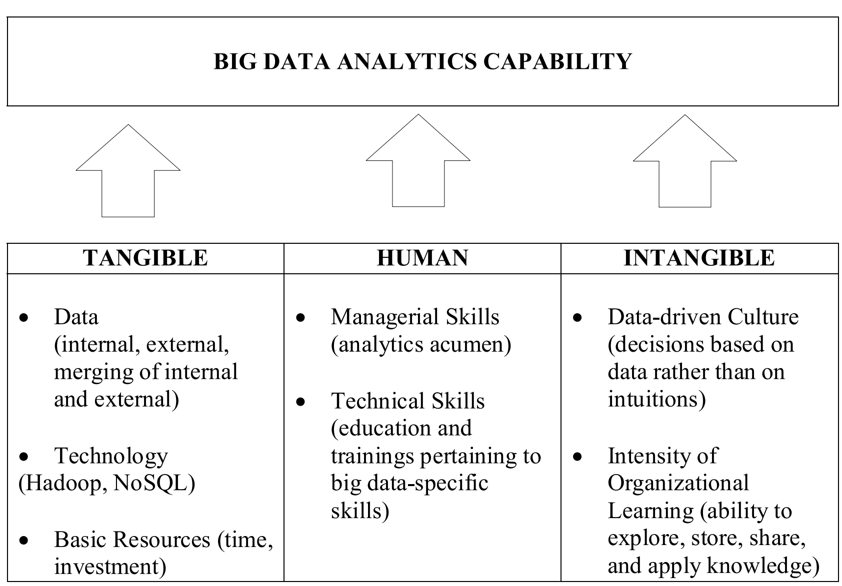


Fig. 1. Classification of Big Data Resources.

Please cite this article in press as: M. Gupta, J.F. George, Toward the development of a big data analytics capability, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.07.004

of information regardless of the size of data, structure of data, and speed at which data are created [39].

George et al. [40] identify five sources of (big) data: public data, private data, data exhaust, community data, and self-quantification data. Public data refer to government-owned datasets pertaining to healthcare, climate change, and consumer spending that are available to businesses (or individuals) for no cost. Private data are the firm-owned data that are actively collected by the firms. Examples include customer transactions, clickstreams, and data generated from the use of RFIDs. Data exhaust refers to the data that do not have a direct value attached to them. However, when combined with other sources of data, data exhaust can yield new insights. Examples include random Internet searches and location data generated from mobile phone usage. Data generated by users on online social communities such as Facebook and Twitter are considered community data. Finally, self-quantification data are personal data generated from wearable technologies such as fitness bands and smart watches.

More broadly, a firm's data can be categorized into internal data and external data [33]. Internal data refer to enterprise-specific data, which are created as a result of the firm's internal operations such as inventory updates, accounting transactions, sales, and human resource management. Consequently, the firms relying only on their internal data to make business decisions are less likely to gain an edge over the competition [33]. On the other hand, data collected from the sources external to an organization such as the web, e-commerce communities, mobile phones, and sensors can be termed as external data. External data, which are also referred to as population-level data, "may not be directly related with the firm's business operations but can provide novel and more flexible perspectives" in comparison to internal data and "can also provide additional information about customers' personal tastes on certain types of products such as movies, music and books" [33, p. 172]. Firms interested in creating BDA capabilities must integrate their internal and external data.

2.4.2. Technology

New forms of data call for novel technologies that are capable of handling the challenges posed by gigantic, diverse, and fast-moving data. Relational database management systems (RDBMS) have remained a popular choice for organizations to store structured data such as employees' records, customer orders, inventory management data, and financial transactions. Further, to gain insights from these disparate sources of organizational data, organizations have relied on extraction, transformation, and load methods to design data warehouses (or data marts). A data warehouse is a collection of enterprise-specific data, which are extracted from various organizational functions and are then made to conform to a standard structure. Key performance indicators are then extracted from the data using online analytical processing, database queries, and other reporting services. This approach is useful and efficient as long as the data that firms are dealing with are structured or in a format on which a structure can be easily imposed.

According to some estimates, as much as 80% of an organization's data exist in an unstructured format [41]. Consequently, organizations have started to move beyond traditional RDBMS methods of storing and analyzing data. New technologies such as Hadoop, a Java-based software framework, have emerged that allow distributed storage (via Hadoop Distributed File System or HDFS) and parallel processing of massive unstructured datasets. HDFS is the lower level layer for distributed databases, commonly known as Not Only SQL (or NoSQL) databases that can efficiently store and retrieve non-relational unstructured data. Some examples of NoSQL databases include Cassandra, HBase, and MongoDB. Apple's 2015 acquisition of FoundationDB, a company that produces NoSQL databases, further emphasizes how critical these new forms of technology have become for organizations interested in exploring insights from big data [42]. Besides Hadoop and NoSQL database technologies, organizations further need several other technologies to store, process, analyze, and visualize big data [32].

In the past, proprietary technology has been considered a source of competitive advantage. A firm that can keep its proprietary technology secret is likely to have an edge over its competition [4,43]. However, in most cases, it is difficult for firms to keep their proprietary technology hidden due to reasons such as labor force mobility and reverse-engineering [43]. Moreover, the emergence of social media-based communities such as LinkedIn groups and Meetups enables individuals from different and sometimes competing organizations to engage in informal interactions as never before. As a result, it is difficult for organizations to keep their (big data) technology completely secret from their rivals.

2.4.3. Basic resources

Besides data and technology, firms need to make adequate investments in their big data initiatives. Given the newness of big data and its related technology and tasks, most organizations are yet to explore a standard procedure to implement these initiatives. Therefore, it is likely that a firm's big data investments may not start yielding the desired results immediately. It is important that firms are persistent and devote enough time to their BDA initiatives to achieve their analytical objectives. Based on this and consistent with prior IS research [43,44], this study suggests investments and time as two tangible resources required by a firm to create a BDA capability. While Wixom and Watson [44], in their seminal work on a data warehousing success model, referred to investments and time as "resources," in order to differentiate these two resources from other resources described in this study, we have put them under the label of "basic resources."

2.4.4. Human resources

A firm's human resources consist of its employees' experience, knowledge, business acumen, problem-solving abilities, leadership qualities, and relationships with others [25,45]. Prior IT capability research has suggested technical and managerial skills as the critical dimensions of human resources with respect to IT [2,18,43]. Along the same lines, this study proposes big data-specific technical and managerial skills as two important aspects of a firm's human big data resources.

2.5. Technical skills

Technical "big data" skills refer to the know-how required to use new forms of technology to extract intelligence from big data. Some of these skills include competencies in machine learning, data extraction, data cleaning, statistical analysis, and understanding of programming paradigms such as MapReduce [29,46]. Although some universities have started to offer courses in these skills, there is still a significant shortage of individuals with big data-specific technical skills [47]. This is further vindicated by a recent McKinsey's report that claims the United States alone will need 140,000–190,000 individuals with big data skills by 2018 [39]. Technical IT skills such as programming, database skills, and system analysis and design in general are not considered rare, since these skills to a degree can be explicated (or codified) in procedures, documents, and manuals [43]. We believe the same will apply to technical big data skills; however, given the newness of big data technology and the skills associated with it, organizations with big data-skilled employees are likely to have some advantage over their rivals. However, this advantage may not

Please cite this article in press as: M. Gupta, J.F. George, Toward the development of a big data analytics capability, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.07.004

last long since, like technical IT skills, big data-specific technical skills may eventually get dispersed among individuals working in the same (or different) organizations, thereby making this resource ordinary across firms over time [48].

2.6. Managerial skills

While firms can develop technical skills by hiring new talent and/or by training their current employees, managerial skills are highly firm-specific and are developed over time by individuals working in the same organization [43]. These skills are developed as a result of strong interpersonal bonds between organizational members working in the same (or different) departments [2]. These skills are deep-rooted in an organization setting and can be described as taken-for-granted norms through which managers perform their everyday work and make decisions [2,43]. Stated simply, managerial skills are tacit and thus are heterogeneously dispersed across firms [43].

Within the context of a firm's big data function, the intelligence gleaned from data will be of little use to an organization if its managers fail to foresee the potential of newly extracted insights. Thus, it is imperative for managers to have a sharp understanding of how and where to apply the insights extracted by their technical teams. To do so, big data managers should have the ability to understand the current and predict the future needs of other business units, customers, and other partners [43]. Moreover, mutual trust and a good working relationship between big data managers and other functional managers will likely lead to the development of superior human big data skills, which will be difficult to match by other firms.

2.7. Intangible resources

Of the three principal types of organizational resources classified by Grant [3] and other strategic management scholars, intangible resources are considered central to a firm's performance, especially in dynamic markets [38]. Yet, unlike tangible resources, intangible resources are not documented on firms' financial statements [3]. This is because intangible resources do not have clear and visible boundaries, and their value is highly context-dependent [7,8]. While most intangible resources are not easily tradable in a market, there are, however, some exceptions such as trademarks, copyrights, and other intellectual capital (e.g., patents), which can be sold or bought legally by organizations [3]. In general, most intangible resources meet the VRIN status of the RBT, thereby making them highly heterogeneous across firms. [8]. This study suggests two such intangible resources that are likely to be a source of major heterogeneity across firms looking to reap benefits from big data. These resources are data-driven decision-making culture and intensity of organizational learning.

2.7.1. Data-driven culture

Organizational culture is a highly complex notion to understand and describe. Over the years, management scholars have suggested several definitions of organizational culture, yet there is no consensus on a single definition [49]. While some suggest that organizational culture encompasses nearly all areas of an organization, others call it a glue that keeps an organization together [50,51]. Prior studies in management strategy have identified organizational culture as a source of sustained firm performance [7,38,52]. On the same lines, recent work in big data suggests that organizational culture is critical for the success of the firm's big data initiatives. For instance, Lavalle et al. [36] indicate that the reasons why big data projects are often unproductive relate to organizational culture rather than to the characteristics of data and lack of technology. Ross et al. [5] opine that culture has the ability to inhibit (or enhance) an organization's ability to benefit from big data.

This emerging stream of research on big data further asserts that while organizations in all industries are collecting hordes of data, only a small percentage of organizations have actually benefitted from their BDA investments [5]. This is because most organizations rely on the past experience and/or intuition of their top executives to make important decisions, which is commonly referred to as the highest paid person's opinion [35]. To realize the full potential of data owned by firms, it is critical that firms develop a data-driven culture, which this study, following Ross et al. [5] and McAfee and Brynjolffson [35], defines as the extent to which organizational members (including top-level executives, middle managers, and lower-level employees) make decisions based on the insights extracted from data. A firm in which decisions are influenced by the title (or designation) of some individuals is unlikely to gain any return on its big data investments. Consequently, the efforts to collect massive amount of data, acquire technology, and build technical and managerial skills will be in vain. Moreover, given that employees at all levels in an organization are required to make some decisions, it is pertinent to diffuse the culture of data-driven decision-making to all levels such as that organizational members, regardless of their job titles, have the ability to make good decisions that are grounded on some tangible evidence as suggested from data [5].

2.7.2. Intensity of organizational learning

One of the shortcomings of RBT is that it does not address why some firms perform better than their rivals, especially in rapidly changing market conditions [53]. According to Teece et al. [9], firms that have the ability to reconfigure their resources according to the changes in their external environment will likely have a sustained competitive advantage. Grant [54] asserts that this ability of a firm will likely be affected by its intensity of organizational learning, which is a process through which firms explore, store, share, and apply knowledge [55,56]. This makes sense because organizational knowledge never wears out [57]. Grant [54] extended RBT by proposing the knowledge-based view of a firm that considers firms as institutions in which the specialized knowledge of individuals is integrated to form organizational-level knowledge that in turn leads to sustained business performance.

Though knowledge does not wear out, it may become outdated due to the emergence of new technologies (or inventions) [57]. Therefore, firms need to make concerted efforts to exploit their existing knowledge and explore new knowledge to cope with uncertain market conditions [38,55]. Based on this, it is safe to suggest that firms with high intensity of organizational learning are likely to have stocks of organizational knowledge that can be used toward creating a BDA capability. These stocks of (new and old) knowledge can be combined with the insights extracted from big data to make informed decisions. We know that any analysis of data does not tell the whole story; it is always the theory that explains. In the same manner, firms with high intensity of organizational learning will likely have an advantage of applying their stocks of knowledge to further validate the initial insights gleaned from big data.

2.8. Big data analytics capability

Drawing on RBT, we have proposed that firms need a combination of certain tangible, human, and intangible resources to build a BDA capability. Prior studies, citing that tangible resources can be acquired from a market, have emphasized the importance of human and intangible resources in creating organizational capabilities. While we agree that big data-specific tangible resources on their own cannot create a BDA capability, we believe this is true for human and intangible big data resources as well. To create a BDA capability, a firm needs not just one or two of these resources, but it is the unique combination of all three that generates a firm-specific BDA capability. For instance, a firm that has a corpus of data and powerful computational technology, but lacks managerial and technical big data skills, is unlikely to benefit from its data and big data technology. Similarly, the mere presence of tangible resources (e.g., data and technology) and human big data skills will not be rewarding if an organization lacks learning intensity and adopts a culture where decisions are made based on people's opinions.

Having defined the notion of BDA capability and identified the resources that collectively build this capability, we next develop an instrument to measure a firm's big data capability and then test its relationship with firm performance.

3. The big data analytics instrument

3.1. Conceptualization of constructs

As discussed previously, this study defines BDA capability as a firm's ability to assemble, integrate, and deploy its big data-based resources. Following this, the BDA capability construct is conceptualized as a multidimensional third-order aggregate (or formative construct) of big data-specific tangible, human skills, and intangible resources constructs, which in turn are conceptualized as second-order formative constructs comprising seven first-order constructs (see Table 1).

While the measures of all the first-order constructs were either adapted or created from existing literature on digital capabilities, the BDA capability is significantly different from other digital capabilities such as IT capability. Digital technologies broadly refer to IT-enabled resources (hardware and software) that support enterprise resource planning, supply chain management, knowledge sharing, virtual communication, etc. [58–60]. These digital technologies when combined with other organizational-level resources create digital capabilities [61]. Although (big) data can be considered a digital resource, it requires several other big data-specific organizational resources to collectively create a BDA capability. This is reflected in our proposed theoretical framework (see Fig. 1) and in the items of our first-order constructs, which are unique to BDA (see Appendix A).

For instance, items of the data construct to capture the extent to which an organization has access to large, unstructured, and fast-moving data and whether an organization integrates its internal data and external data. The technology construct identifies whether an organization possesses sophisticated data storage (e.g., Hadoop, No SQL), data visualization (e.g., Tableau, SAS Visual Analytics), and other cloud-based and open-source data analytics technologies. Technical and managerial skills measure the degree to which the technical and managerial staff has BDA-specific skills. The data-driven culture construct assesses whether an organization considers its data a tangible asset and the extent to which organizational decisions are made in response to the insights extracted from data.

3.2. BDA capability as a formative construct

This study conceptualizes the BDA capability construct as a higher-order formative construct. While the concept of formative constructs is not new, as it was first proposed by Blalock [62] more than four decades earlier, the use of formative constructs in business disciplines has been discouraged by some. The major criticisms pertain to the conceptualization of formative constructs and the presence of interpretational confounding (i.e., the difference between empirical and nominal meanings of a construct) and weakened external consistency (i.e., the extent to which indicators of a construct correlate with other variables) in formative models [63]. However, several recent commentaries and studies have demonstrated that these concerns can be successfully addressed [10,64–69]. As Petter et al. [65] put it, “When grounded theoretically and analyzed properly, formatively specified constructs can play a valuable role in IS research” (p. 154). Following this stream of research, we next establish the formative nature of the BDA capability construct.

Consistent with IT capability literature, which is primarily based on RBT, this study conceptualizes the BDA construct into three dimensions (tangibles, human skills, and intangibles). Since we borrow extensively from the IT capability literature, we first examined if the higher-order IT capability construct has been identified as a reflective or a formative construct in the IS literature. Surprisingly, we did not find a clear consensus. For instance, Lu and Ramamurthy [70] and Kim et al. [71] defined IT capability as a higher-order reflective construct, while Ravichandran and Lertwongsatien [72] and Wang and colleagues [73] identified IT capability as a higher-order formative construct. This may seem a little confusing, given that the IT capability construct has the same label in all of these studies. However, Diamantopoulos [74,75] brilliantly explained this conundrum:

> A construct with the same name, but varying in terms of being measured reflectively versus formatively, will not necessarily be the same construct [65,p. 151].

More importantly, despite having the same label, the measures (or indicators) of a construct will significantly differ if the construct is identified as formative versus reflective [65]. A careful review of the studies described above further validates this perspective. There is a considerable difference across these studies in terms of the measures of formative (vs. reflective) higher-order IT capability constructs. This is consistent with the well-accepted view in the methodological literature that a construct is not inherently reflective or formative; instead, it depends upon how a researcher has defined a concept based on theoretical or research objectives [10,64,65].

We next applied four decision rules identified in the literature on formative constructs to conceptually assess the formative

Table 1
Latent Constructs and Sub-dimensions.

Third-order	Type	Second-order (sub-dimensions)	Type	First-order (sub-dimensions)	Type
BDA Capability	Formative	Tangible Resources	Formative	Data	Formative
				Technology	Formative
				Basic Resources	Formative
		Human Resources	Formative	Managerial Skills	Reflective
				Technical Skills	Reflective
		Intangible Resources	Formative	Data-driven Culture	Reflective
				Intensity of Organizational Learning	Reflective

Please cite this article in press as: M. Gupta, J.F. George, Toward the development of a big data analytics capability, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.07.004

nature of the BDA capability construct [10,65,66,76]. *First*, of the three proposed indicators (tangible resources, human skills, and intangible resources), no single indicator can adequately explain the phenomenon of BDA capability, thereby yielding support that tangible, human, and intangible resources are defining characteristics rather than the manifestations of the BDA capability. In a similar vein, Chen [77] recently suggested that given the broadness of the IS capability constructs, it is preferable to model capability constructs as formative.

Second, each of the three indicators of the BDA capability construct captures a very distinct facet of an organization's BDA capability. In other words, these three dimensions (indicators) are not interchangeable since dropping one of them would significantly alter the meaning of the BDA capability. Simply put, the three measures jointly make up the BDA construct. Dropping items to achieve the desired reliability (or Cronbach α) is a common practice for a reflective construct because reflective indicators are defined as the manifestations rather than a distinct attribute of the construct. The main point here is that the indicators of a reflective construct, unlike a formative construct, are interchangeable since they have been similar content [65,66,76]. Therefore, dropping an indicator does not necessarily change the meaning of a reflective construct. However, in this study, if the tangible resource indicator of the BDA capability construct is dropped, the distinct aspect pertaining to this dimension is unlikely to be captured by human skills and intangible resources.

Third, the indicators of a formative construct are not required to covary, which is expected in the case of a reflective construct. Theoretically, the indicators – tangible resources, human skills, and intangible resources – need not covary [3,8,38,78]. For instance, having access to tangible resources does not necessarily imply that organizations also have the needed human skills and intangible resources [2,3]. In addition to the theoretical justification, a simple way to rule out covariation (or multicollinearity) among the indicators of a formative construct is to calculate the variance inflation factor (VIF) [10,76,79]. In this study, VIF values were calculated as a test of multicollinearity for each formative construct. These tests will be discussed later in more detail.

Fourth, the three proposed dimensions of the BDA capability construct have very different antecedents. For instance, the measures of tangible resources (i.e., data, technology, and basic resources), human skills (i.e., managerial and technical skills), and intangible resources (i.e., data-driven culture and organizational learning) are very distinct from each other. In sum, the higher-order BDA capability construct satisfies the four decision rules in the formative methodological literature [10,66,76]. The same guidelines were followed for the lower-order formative constructs.

Recently, Kim et al. [71] have suggested that formative constructs are likely to suffer from interpretational confounding and weakened external consistency. However, a number of studies [10,65,76,80,81][e.g.,10,65,76,80,81] assert that the presence of interpretational confounding is not limited to formative constructs, and that reflective constructs are equally susceptible to interpretational confounding. Moreover, two decades earlier, Chin and Marcolin [82] provided empirical evidence of interpretational confounding in reflective constructs. Thus, it is surprising that the use of formative constructs is still discouraged in business disciplines. One reason for this can be attributed to the lack of agreed-upon recommendations and empirical tests to assess interpretational confounding in formative constructs [81,83]. On the other hand, scholars such as Bollen [80,84,85] and Bagozzi [86,87] completely disagree with this perspective and suggest that the presence of interpretational confounding arises due to model misspecification rather than due to the formative or reflective nature of a construct. Bollen [80] further proved that interpretational confounding is unlikely to be a problem if a formative construct has at least two paths leading to other reflective constructs (or 2+ emitted paths rule). Similarly, Bagozzi [86] suggests that a MIMIC model (i.e., a formative construct predicting two or more observed variables) is "estimable, testable, and meaningful" (p. 276).

In addition to the 2+ emitted path rule, two empirical tests have been proposed by Kim and colleagues [63] and Cenfetelli and Bassellier [81] to assess the problem of interpretational confounding and external consistency in formative constructs. We conducted these tests to rule out potential interpretational confounding and weakened external consistency in the higher-order BDA capability construct. The details pertaining to these tests are discussed in detail toward the end of the paper.

3.3. Hierarchical model specification

Following the guidelines of Wetzels et al. [88], the hierarchical model was formally specified, representing the relationships between the indicators, sub-dimensions, and higher-order constructs (see Fig. 2). We first constructed the first-order latent variables and connected them to their corresponding indicators. Data, technology, and basic resource constructs were modeled as mode B "formative," while the remaining first-order constructs were connected to their indicators as mode A "reflective." The second-order latent variables were then constructed by repeating the indicators of their underlying first-order latent variables using mode B "formative" method. Thus, the tangible resource construct was made up of the indicators of basic resource, data, and technology constructs, while the human resource construct was connected to the indicators of managerial skills and technical skills constructs. The intangible construct was linked to the indicators of data-driven culture and the intensity of organizational learning constructs. Similarly, the third-order latent variable, BDA capability, was constructed by repeating the indicators of its second-order constructs.

3.4. Data collection

Developing a psychometrically sound survey instrument is a rigorous process. Consequently, we followed the scale development procedure described by MacKenzie and colleagues [10], who suggested collecting two sets of data: one for evaluating the scale properties (i.e., construct validity, reliability, discriminant validity, multicollinearity, etc.) and item refinement and the other set for reevaluating the scale properties and establishing its nomological validity. Consequently, two studies were conducted – Study I and Study II. Study I was the pilot study such that data collected in this study were used to purify and refine the scale items, in addition to assessing the higher-order model. Based on the findings from (pilot) Study 1, data (Study II) were collected from a new sample (see Table 2 for sample characteristics) and the scale properties were re-examined, along with the test of nomological validity.

3.5. Study I: pilot study

The first set of data was collected from BDA managers, who were members of the *Big Data and Analytics* group on LinkedIn. In total, 232 responses were received. Respondents represented a variety of industries (e.g., computers, financial services, Internet, communications, and utilities), and their job titles included chief information officer, chief technology officer, vice president of technology, director of IT, and managers of analytics. Given that we had data collected on 34 indicators, we first conducted an exploratory factor analysis using principal component analysis and varimax rotation. Seven factors emerged (eigenvalues >1) from this analysis. All items loaded on their related factor as predicted,

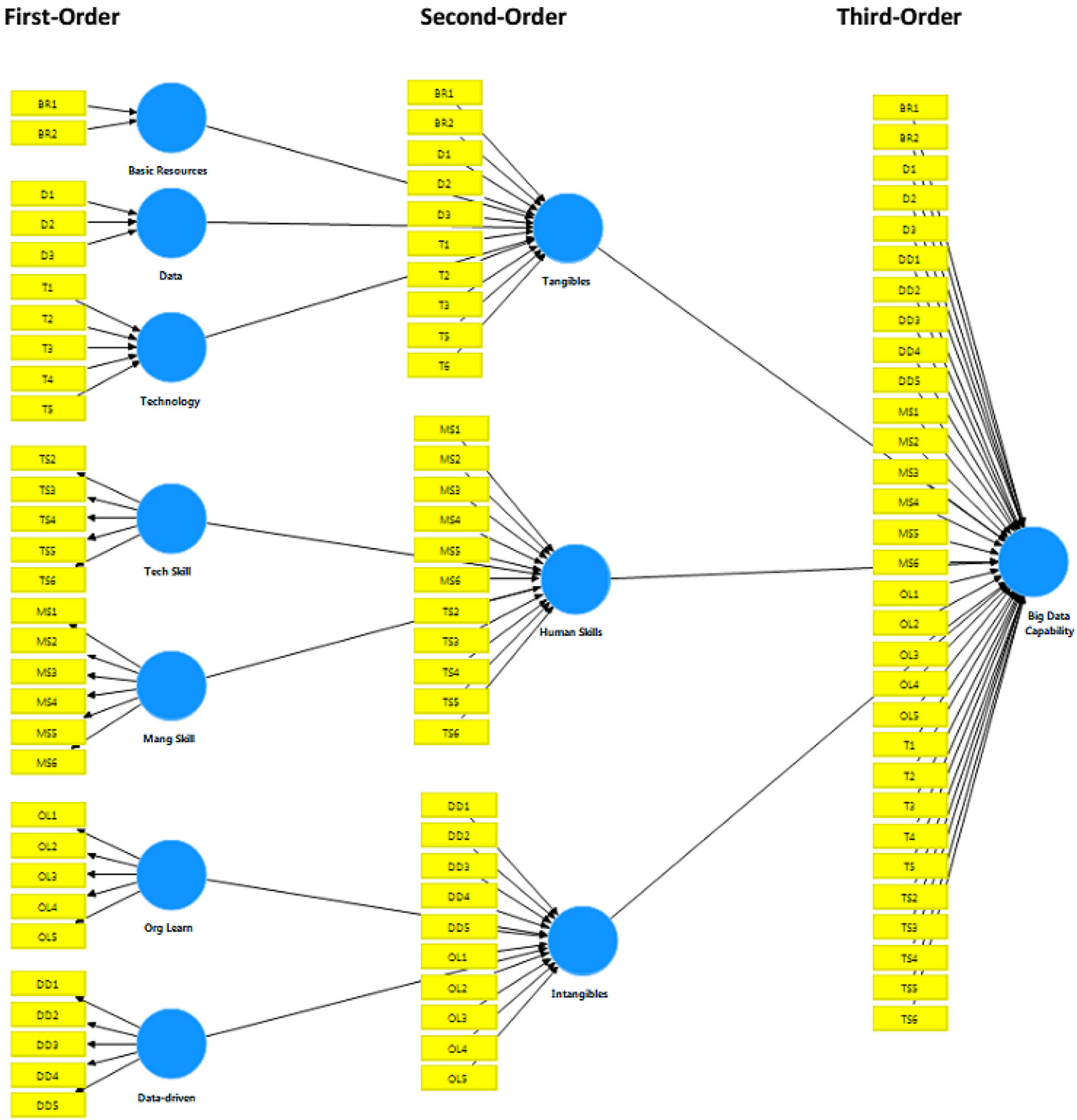


Fig. 2. Hierarchical Model Specification Using Repeated Indicators Approach.

except for TS1, which had significant loadings (>0.5) across multiple factors. Consequently, TS1 was dropped from further analysis [89]. The hierarchical model was then estimated.

3.5.1. Model assessment

The assessment criteria for formative and reflective constructs are different. We first assessed the validity of the indicators at the construct level. For reflective constructs, all of the items had outer loadings above 0.7 and the average variance extracted (AVE) of all the measures exceeded 0.50 [67]. While all of the indicators' weights of data and basic resources were statistically significant, two (i.e., T3 and T4) of the five indicators' weights of the technology construct were found to be nonsignificant. Cenfetelli and Bassellier [81] suggest that a formative construct with many indicators is likely to have few indicators with nonsignificant weights. They further suggest that the nonsignificant indicator of a formative construct, unlike a reflective construct, can be kept in a model as long as the researchers can justify the contribution of it. Given that the technology construct is proposed as an aggregate of five items where each item captures a different big data-related technology, we believed it was appropriate to keep the nonsignificant indicators in the model, as each item made an important and distinct contribution to the overall technology construct.

Following MacKenzie et al. [10] and Schmiedel et al. [90], we then evaluated the validity of the items of the formative constructs using Edwards' [91] adequacy coefficient ($R^2{}_a$) by summing the squared correlations between the formative construct and its indicators and then dividing the sum by the number of indicators.

Table 2
Sample Characteristics.

	Study 1	Study 2
Industries	N = 232	N = 108
Computer/Software	23%	16%
Manufacturing	12%	4%
Finance, Insurance, Real Estate	10%	21%
Retail, Wholesale	8%	5%
Services	8%	8%
Healthcare	5%	9%
Others (Transportation, Electronics, Services, Communication, etc.)	33%	38%
Total BDA experience		
Less than 3 years	26%	19%
3–6 years	47%	45%
More than 6 years	27%	36%
Number of employees in the organization		
Fewer than 1000	44%	43%
Between 1001 and 2500	14%	13%
Between 2501 and 5000	12%	9%
5000 to 10,000	8%	5%
More than 10,000	22%	30%

All R^2_a values were above 0.50 (see Table 3), suggesting that the majority of the variance in the indicators is shared with the formative construct, and thus, the indicators of the formative construct were valid [10]. Like the first-order formative constructs, we first evaluated the weights of the formative indicators on their respective higher-order constructs (three second-order and one third-order constructs). All weights were highly significant. The Edwards' adequacy coefficients (R^2_a) for the higher-order constructs were then calculated. All R^2_a values were greater than the recommended values of 0.50 [10].

We then examined the extent to which the indicators of the formative constructs were multicollinear with each other. While mulitcollinearity is desired among the indicators of a reflective construct, it is problematic for formative constructs. VIF values below 10 in general demonstrate low multicollinearity [10]; however, Petter et al. [66] suggest a more restrictive cutoff of 3.3 for formative constructs. VIF values for all the measures of the first-order, second-order, and third-order formative constructs in this study were less than 3.3, indicating that multicollinearity was not a major concern [66,81].

3.5.2. Reliability and discriminant validity

The concept of internal consistency reliability (ICR) does not apply to formative constructs; however, for reflective constructs, the reliability was assessed using ICR and Cronbach's α, both of which were above 0.8 for all constructs (see Table 4). Discriminant validity of the reflective constructs was established using Fornell and Larcker's [92] criteria. The square root of the AVEs of each latent variable was greater than its correlation with any other constructs. Examination of cross-loadings further yielded support for discriminate validity (see Table 5). Recently, Hensler et al. [93] have criticized Fornell and Larcker's [92] criterion of assessing the discriminant validity and have suggested a new criterion called the heterotrait–monotrait ratio (HTMT). The HTMT ratio is based on the average of the correlations of indicators across constructs measuring different phenomena relative to the average of the correlations of indicators within the same construct. According to Henseler and colleagues [93], the HTMT ratio below 0.85 demonstrates sufficient discriminant validity. We ran this test on the first-order reflective constructs, and the HTMT values for all the reflective constructs were found to be below 0.85. It should be noted that the HTMT method can only be used to assess the discriminant validity of reflective constructs [93].

To the best of our knowledge, there are no established tests in prior literature to assess the discriminant validity of formative constructs [93]. There are two recommendations, however. MacKenzie et al. [10] suggest that the formative construct should less than perfectly correlate (i.e., less than 0.71) with other constructs. Klein and Rai [94] suggest that like reflective items, indicators of the formative constructs should load highly on their corresponding constructs in comparison to other constructs. All the first-order formative (and reflective) constructs in our study satisfy both these conditions (see Tables 4 and 5).

Overall, except for TS1, all the formative and reflective items demonstrated good psychometric properties (construct validity, reliability, and discriminant validity) in the pilot study. We next gathered data from a new sample to further validate the BDA instrument and to establish its nomological validity by assessing the relationship between BDA and firm performance.

Table 3
Higher-order Construct Validation.

Construct	Measures	Weight	Significance	VIF	R^2_a [a]
Data	D1	0.53	$p < 0.001$	1.376	0.78
	D2	0.26	$p < 0.05$	1.466	
	D3	0.48	$p < 0.001$	1.281	
Technology	T1	0.18	$p < 0.05$	1.795	0.70
	T2	0.70	$p < 0.001$	1.529	
	T3	0.02	ns	1.541	
	T4	0.12	ns	1.61	
	T5	0.20	$p < 0.05$	1.887	
Basic Resources	BR1	0.54	$p < 0.001$	2.249	0.88
	BR2	0.28	$p < 0.01$	1.977	
Tangibles	Data	0.34	$p < 0.001$	1.72	0.84
	Technology	0.37	$p < 0.001$	1.82	
	Basic Resources	0.47	$p < 0.001$	1.65	
Human	Managerial Skills	0.40	$p < 0.001$	1.79	0.91
	Technical Skills	0.69	$p < 0.001$	1.79	
Intangibles	Data-driven Culture	0.49	$p < 0.001$	1.50	0.88
	Organization Learning	0.63	$p < 0.001$	1.50	
BDA	Tangibles	0.42	$p < 0.001$	2.34	0.91
	Human	0.31	$p < 0.001$	2.84	
	Intangibles	0.37	$p < 0.001$	2.88	

[a] Edwards [91] adequacy coefficient.

Please cite this article in press as: M. Gupta, J.F. George, Toward the development of a big data analytics capability, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.07.004

Table 4
Inter-correlations of the Latent Variables for First-Order Constructs[a].

	Construct	ICR	Alpha	AVE	1	2	3	4	5	6	7
1	Data	*NA*	*NA*	*NA*	*NA*						
2	Basic Resources	*NA*	*NA*	*NA*	0.54	*NA*					
3	Technology	*NA*	*NA*	*NA*	0.60	0.58	*NA*				
4	Managerial Skills	0.92	0.87	0.79	0.45	0.57	0.47	*0.89*			
5	Technical Skills	0.93	0.89	0.76	0.53	0.61	0.55	0.67	*0.87*		
6	Data-driven Culture	0.90	0.86	0.70	0.54	0.48	0.52	0.64	0.71	*0.84*	
7	Organization Learning	0.94	0.92	0.75	0.53	0.53	0.57	0.56	0.58	0.58	*0.87*

[a] Square root of the AVEs on the diagonal.

Table 5
Cross-loadings.

Items	Data	Technology	Basic Resources	Managerial Skills	Technical Skills	Data-driven	Org Learning
D1	**0.82**	0.51	0.46	0.36	0.42	0.44	0.46
D2	**0.73**	0.44	0.37	0.42	0.42	0.42	0.37
D3	**0.79**	0.45	0.43	0.46	0.42	0.42	0.40
T1	0.45	**0.66**	0.37	0.30	0.38	0.28	0.36
T2	0.55	**0.93**	0.53	0.50	0.51	0.50	0.53
T3	0.31	**0.59**	0.38	0.35	0.36	0.23	0.30
T4	0.34	**0.65**	0.42	0.32	0.39	0.29	0.40
T5	0.42	**0.68**	0.41	0.32	0.40	0.29	0.39
BR1	0.52	0.55	**0.93**	0.54	0.57	0.44	0.48
BR2	0.44	0.47	**0.82**	0.49	0.49	0.40	0.46
MS1	0.47	0.43	0.48	**0.85**	0.53	0.61	0.53
MS2	0.50	0.49	0.48	**0.87**	0.62	0.61	0.58
MS3	0.46	0.44	0.52	**0.90**	0.62	0.62	0.56
MS4	0.42	0.39	0.50	**0.85**	0.60	0.50	0.50
MS5	0.41	0.46	0.51	**0.84**	0.60	0.59	0.47
MS6	0.39	0.41	0.51	**0.83**	0.60	0.61	0.52
TS2	0.46	0.44	0.50	0.48	**0.73**	0.40	0.53
TS3	0.48	0.53	0.56	0.64	**0.91**	0.55	0.64
TS4	0.46	0.49	0.53	0.66	**0.91**	0.55	0.66
TS5	0.46	0.45	0.53	0.60	**0.89**	0.52	0.63
TS6	0.46	0.52	0.62	0.64	**0.91**	0.53	0.63
DD1	0.35	0.25	0.28	0.40	0.39	**0.69**	0.47
DD2	0.48	0.45	0.37	0.52	0.45	**0.84**	0.42
DD3	0.39	0.45	0.42	0.51	0.48	**0.83**	0.49
DD4	0.52	0.45	0.44	0.69	0.55	**0.84**	0.56
DD5	0.41	0.39	0.38	0.60	0.46	**0.79**	0.46
OL1	0.48	0.43	0.42	0.52	0.61	0.50	**0.86**
OL2	0.41	0.45	0.42	0.51	0.60	0.50	**0.90**
OL3	0.45	0.49	0.43	0.55	0.61	0.52	**0.90**
OL4	0.51	0.52	0.49	0.56	0.61	0.57	**0.88**
OL5	0.45	0.58	0.56	0.54	0.64	0.52	**0.81**

3.6. Study II: scale revalidation and nomological validity

Data for Study II were collected from the *Chief Data Officer (CDO)* group on LinkedIn. A new survey was created in Qualtrics that included questions about market performance (MP) and operational performance (OP), in addition to the refined list of BDA capability questions from Study I. Industry and firm size questions were also included as controls. The survey link was then sent to the group members by the owner/moderator of the CDO group. The members were also informed that by participating in this survey, they had a chance to win an Apple watch. A total of 108 responses were received. For sample characteristics, please refer to Table 2. We then examined the responses to check if there was anyone from the initial survey who also responded to this survey. No common respondents were found in the two studies.

Similar to Study I, the psychometric properties of the measures of the first-order constructs were assessed (See Table 6). We first evaluated the loadings (and weights) of the first-order constructs. All the first-order reflective constructs had significant loadings (>0.7). Similarly, the indicators' weights of data and basic resources were significant. For the technology construct, while T2 and T5 had significant weights, T1, T3, and T4 had nonsignificant weights. Moreover, T3 had a negative weight. Unlike reflective indicators, it is not recommended to cull any formative indicators based solely on nonsignificant weights [66,95]. As discussed previously, the number of indicators of a formative construct is likely to affect the magnitude of the indicators weights. Consequently, we decided to keep both T1 and T4 for further analysis. In the past, while the co-occurrence of negative and positive indicators' weights was considered surprising (e.g., [96]), according to Cenfetelli and Bassellier [81], negative weights are caused by the pattern of correlations among the indicators of a formative construct. They further assert that if a negative weighted item is not multicollinear, it is appropriate to include the indicator in the analysis; however, the indicator should be culled in the future if it repeatedly behaves similarly [81]. Given that T3 had a significant positive weight in the first survey and was not multicollinear (VIF = 1.34) in the second survey, we decided to keep T3 in the analysis as well.

For reflective constructs, the AVE and ICR of all measures were above 0.6 and 0.9, respectively [67]. The square root of the AVEs of each latent variable was greater than its correlation with any other constructs [92]. The HTMT values for all reflective constructs were

Table 6
Inter-correlations of the Latent Variables for First-Order Constructs[a].

		ICR	α	AVE	1	2	3	4	5	6	7	8	9
1	Data	*NA*	*NA*	*NA*	*NA*								
2	BR	*NA*	*NA*	*NA*	0.57	*NA*							
3	Tech	*NA*	*NA*	*NA*	0.63	0.42	*NA*						
4	MS	0.92	0.89	0.65	0.60	0.54	0.57	*0.81*					
5	TS	0.94	0.92	0.77	0.53	0.58	0.61	0.62	*0.87*				
6	DD	0.90	0.87	0.65	0.51	0.46	0.49	0.61	0.56	*0.81*			
7	OL	0.95	0.93	0.79	0.66	0.51	0.67	0.62	0.66	0.67	*0.89*		
8	OP	0.96	0.94	0.85	0.57	0.54	0.53	0.77	0.65	0.53	0.67	*0.92*	
9	MP	0.96	0.94	0.85	0.45	0.44	0.38	0.72	0.39	0.44	0.49	0.49	*0.92*

[a] Square root of the AVEs on the diagonal.

found below the threshold value of 0.85, thereby yielding additional support for the discriminant validity of the reflective constructs [93]. The discriminant validity of the formative constructs was assessed by examining the correlation between first-order formative constructs (data, basic resources, and technology) and other constructs. All correlations were below 0.71 [10], indicating sufficient discriminant validity per MacKenzie et al. [10]. Additionally, the examination of cross-loadings of all the first-order formative (and reflective) constructs was satisfactory such that indicators of the constructs loaded highly on their corresponding constructs in comparison to other constructs [67,94].

We next evaluated the hierarchical model. The indicators' weights of the second-order constructs – (1) data ($\beta = 0.39$, $p < 0.001$), basic resources ($\beta = 0.40$, $p < 0.001$), and technology ($\beta = 0.41$, $p < 0.001$) on the tangible construct, (2) managerial skills ($\beta = 0.61$, $p < 0.001$) and technical skills ($\beta = 0.50$, $p < 0.001$) on the human skills construct, and (3) data-driven culture ($\beta = 0.49$, $p < 0.001$) and the intensity of organizational learning ($\beta = 0.60$, $p < 0.001$) on the intangible construct – were statistically significant. Similarly, the indicators' weights of the third-order construct – tangibles ($\beta = 0.25$, $p < 0.001$), human skills ($\beta = 0.54$, $p < 0.001$), and intangibles ($\beta = 0.27$, $p < 0.001$) – were significant.

Having established the psychometric properties of the BDA capability scale twice, we then assessed the nomological validity of the BDA construct by examining the relationship between BDA capability and the two separate dimensions of firm performance (i.e., MP and OP). Consistent with the literature on firm performance, we define firm performance as the extent to which a firm generates superior performance with respect to its competitors [79,95,97]. Operating performance and market-based performance have been proposed as two important and distinct dimensions of firm performance in the IS literature [72,73]. The performance measures were asked with respect to the competitors (see Appendix B). This is consistent with the tenets of RBT, which argues that the firm needs to collectively employ its VRIN internal resources to outperform their rivals [11]. Bharadwaj [2], in her seminal paper on the relationship between IT capabilities and firm performance, contended that to assess the firm's organizational capabilities, it is imperative to make inter-firm comparisons. This view has been widely accepted across a plethora of studies, particularly the ones focused on IS capabilities and related constructs, which have been conducted in the field of IS and other business disciplines over the years [63,71,77,79,95,98–101]. Moreover, a recent study by Wu, Straub, and Liang [97] suggested that firm performance is "best measured relative to competition" (p. 507).

As shown in Fig. 3, we found a significant, positive effect of BDA capability on both MP (β=0.86, $p < 0.001$) and OP (β=0.67, $p < 0.001$). Consistent with past IS studies, we had also included firm size and industry as control variables; however, their relationships with MP and OP were nonsignificant. The model accounted for 46.2% of the variance in MP and 74.4% of the variance in OP.

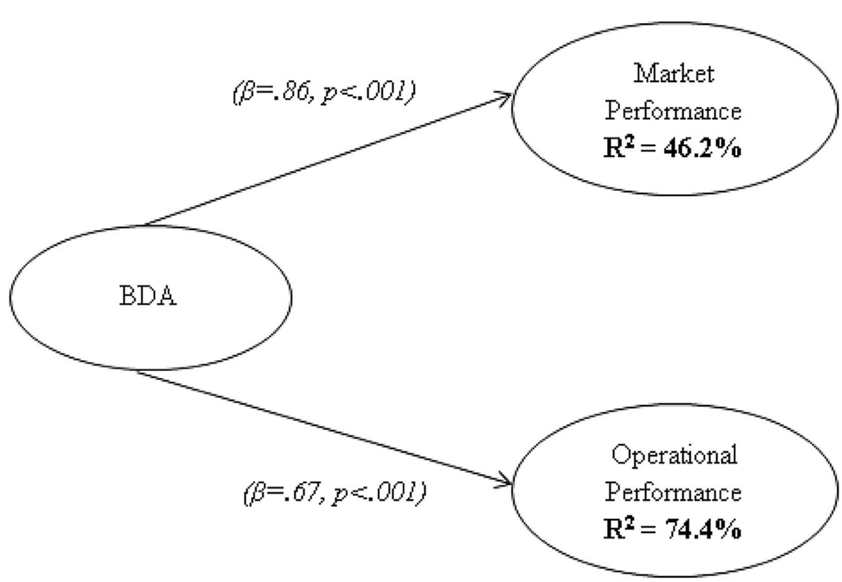


Fig. 3. Results.

3.7. Post Hoc Analysis to Assess the Formative Nature of the BDA Capability Construct

By having two paths leading from the BDA capability construct to MP and OP constructs, the model shown in Fig. 3 represents a MIMIC model and satisfies the 2+ emitted paths rule. Therefore, the formative model shown above is unlikely to have interpretational confounding [80,87]. However, to empirically assess interpretational confounding in our study, we created two models as suggested by Kim et al. [63]: Model1–MP as the sole dependent variable, and Model 2 – OP as the sole dependent variable. The weights of the three formative measures of the BDA capability construct were consistent (and statistically significant) across the two models (see Fig. 4), thereby suggesting that interpretational confounding was not a concern in this study. The same method has been recently applied by Wu, Straub, and Liang [97] to empirically validate the formative constructs in their study.

We next assessed external consistency for the BDA capability construct. Following Kim et al. [63], we developed a test model (TModel) consisting of the three formative indicators of the BDA capability construct and two endogenous constructs – MP and OP. The two models – Model 1 and Model 2–that were used to test for interpretational confounding became the baseline models. External consistency is attained when the formative measures of a construct have consistent correlation with the measures of the dependent variable in proportion to their correlation with the other construct [63,102]. Consequently, (1) the correlations between the three measures of the BDA capability construct and the four measures of MP were compared across Model 1 and TModel, and (2) the correlations between the three measures of BDA capability construct and the four measures of OP were compared across Model 2 and TModel. The difference in correlations between the BDA capability's measures and the measures of MP and OP across the baseline models and TModel were near zero (see Table 7). Hence, any problem with weakened external consistency can be ruled out in this study [63].

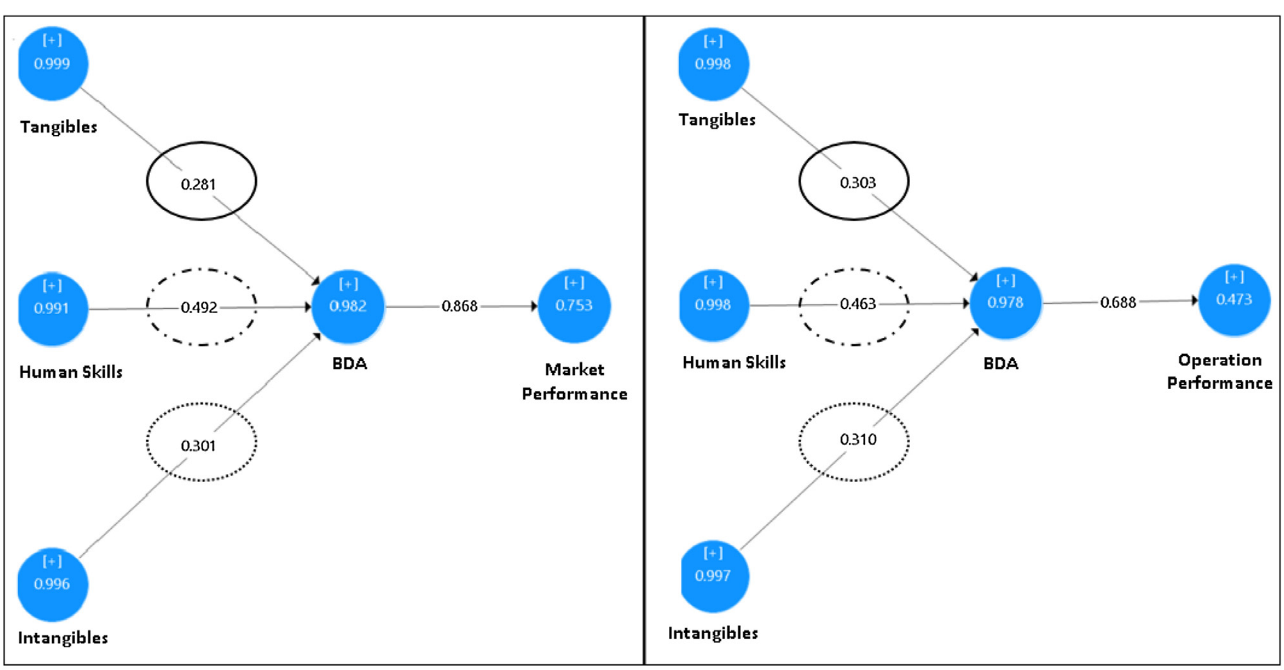
4. Discussion

The hype around big data is unprecedented; however, only a small percentage of companies have been able to realize the true potential of their data and big data investments [5]. This seems counterintuitive, given a number of articles that have appeared in almost all business publications (*Harvard Business Review*, *The Economist*, *Fortune*, etc.), discussing the transformative potential of big data. Part of the reason for this paradoxical situation is that the majority of the existing literature on big data has been contributed by technology consultants and thus lacks theoretical insights.

4.1. Research implications

This study is an early attempt to understand the big data phenomenon using the theoretical lens of RBT, a well-established strategic theory of the firm. While the practitioners of big data have made remarkable contribution to the existing big data literature, the majority of them have termed BDA exclusively as a technical capability [31]. We addressed this shortcoming in the existing literature by highlighting the importance of several nontechnical resources, in addition to other resources such as data and technology, which are needed to create a BDA capability.

This study makes an important contribution to big data literature by not only presenting a theoretical framework of BDA capability consisting of several technical and nontechnical resources classified across three categories, but also yielding empirical support to the proposed theoretical framework. Furthermore, the support for the relationship between a firm's big data capability and its performance has largely been anecdotal in the extant big data literature. Using survey data from 108 executive-level technology leaders, this study has empirically validated the relationship between BDA capability and firm performance. However, it should be noted that as discussed in this study, creating a BDA capability is a complex process as it



Model 1: MP as the sole dependent variable **Model 2**: OP as the sole dependent variable

Fig. 4. Test for Interpretational Confounding (Kim et al. [63]).

Please cite this article in press as: M. Gupta, J.F. George, Toward the development of a big data analytics capability, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.07.004

Table 7
Test for External Consistency (Change in Correlations).

Formative Indicators	Model 1–Tmodel				Model 2 – Tmodel			
	MP1	MP2	MP3	MP4	MR1	MR2	MR3	MR4
Tangibles	0.002	0.001	0.000	0.002	0.001	0.001	0.000	0.004
Human Skills	0.007	0.011	0.007	0.010	0.007	0.005	0.004	0.007
Intangibles	0.004	0.004	0.003	0.004	0.002	0.003	0.001	0.001

requires several firm-level tangible, human skills, and intangible resources.

Another important contribution of this study is the development of a theoretically grounded construct of the BDA capability and a psychometrically sound survey instrument to measure a firm's BDA capability. The BDA construct and its measures will enable researchers interested in studying the big data phenomenon to further the emerging research stream of big data.

This study also asserts that, theoretically, the BDA capability construct is different from digital capabilities such as IT capability. Simply put, IT capabilities, besides other organizational-level resources, include information, communication, and connectivity technologies [58]. Moreover, IT capability facilitates "day-to-day running of the firm" (p. 175) [2]. On the other hand, big data-specific technologies enable organizations to extract insights from data originating from a number of sources and make decisions based on newly gleaned intelligence. In addition to this, BDA professionals are likely to have significantly different skills, roles, and responsibilities from the ones possessed by regular IT staff. Although testing the relationship between IT capability and the BDA capability of firms was beyond the scope of this study, we did look at the correlation between organization size and BDA capability from the data collected in Study II. Organization size has been described as an important correlate of an organization's IT capability in the IS literature [2,18]. The correlation between organization size and BDA capability was insignificant ($r = 0.032$) in our study. This further suggests that IT capability and an organization's BDA capability may not necessarily correlate.

A decade earlier, Wade and Hulland [24] recognized three properties of RBT that could provide rare and valuable benefits to the IS community: (1) RBT provides the foundation to specify the firm-level technology resources, (2) RBT facilitates cross-functional research by clearly differentiating between mutually exclusive technical and nontechnical firm-level sources, and (3) RBT allows researchers to systematically test a relationship between resources and firm performance. By systematically applying these properties of RBT to big data, we have furthered the explanatory power and generalizability of RBT to the novel field of big data. This is consistent with Kozlenkova and colleagues [12], who predicted that, given the robustness of RBT, future business research should continue "to increase its applicability and breadth" (p.19).

Finally, this study contributes to the methodological literature on formative constructs. By applying several theoretical guidelines and empirical tests to correctly assess the formative nature of the constructs used in this study, we have provided a step-by-step approach to avoid misspecification of formative constructs, a topic that has gained considerable attention recently in the IS field [10,63,69].

4.2. Practical implications

The present study yields some interesting insights for practice. By highlighting the importance of human skills and intangible resources, this study has attempted to enlighten big data managers that gaining competitive advantage from big data is not only about making investments, collecting hordes of data, and having access to sophisticated technology but also about having availability of big data-specific technical and managerial skills, an intensity of organizational learning, and an organizational culture where insights extracted from data are valued and acted upon. It is an aggregate of all these resources that will create a firm-specific big data capability. More recently, some have started to address these issues by suggesting that BDA capability is not about data or technological advances [5,14].

The foremost step in building an organizational capability is self-assessment of the organization's strengths and weaknesses [2]. The survey instrument presented in this paper can be used by organizations to determine the resources they have in abundance and the resources they lack. For instance, scales pertaining to human skills will enable organizations to assess the extent to which they possess big data-specific managerial and technical skills. Indeed, employees are the greatest asset for any organization; however, hiring a wrong person in a job can be disastrous for organizations. According to some estimates, a mis-hire may cost an organization up to six times of his or her base salary during the mis-hire period [103]. Given that BDA courses are still in their infancy [104], it is imperative for organizations to assess their BDA human skills since in the short term big data-specific human skills are likely to be heterogeneous across firms.

As discussed previously, data-driven culture is a required intangible resource for organizations willing to make the best use of their big data. The data-driven culture scale can be used separately to capture the extent to which data drive decision-making in organizations. Managers of the organizations obtaining a low score on this scale would need to empower as well as coach their employees at all levels to make data-driven decisions [5]. Also, the organizations obtaining a high score on it should continue their existing efforts to remain data driven. In addition to this, the BDA instrument can be used by firms in its entirety to evaluate their current and future BDA capabilities and respond accordingly.

4.3. Limitations and future research

Like any other research, this study is not without limitations. First, we would like to emphasize that the BDA capability framework presented in this study should not be considered a universal model. We are in the early stages of understanding the big data phenomenon, and thus constructing an exhaustive list of organizational-level resources that, in turn, will lead to the creation of BDA capability is not easy. The current big data situation is similar to the 1990s when IS researchers were struggling to establish a model of the IT capability. Like IT capability, the proposed BDA capability in this study will continue to evolve (and possibly change), as it cannot anticipate every aspect of big data at this time. Moreover, since a majority of organizations are in the process of adopting (and developing) a BDA capability, an interesting avenue for future research is to enhance the work presented in this study by including other big data-specific tangible, human, or intangible resources, which this study might have missed.

Second, though the control variable “industry” in this study did not yield a significant effect on the BDA capability construct, we would like to highlight that a new wave of inherently data-driven companies, such as Uber, a multinational ride-sharing company, and Airbnb, a US-based accommodation rental company, have been completely reshaping the ways of doing traditional business in automotive, transportation, retail, and consumer goods industries. For instance, Uber’s surge pricing model uses real-time data from its vehicles and passengers to make real-time adjustments to the fare rates. Airbnb uses data from its users’ demographics, preferences, reviews, past stays, and social connections and other geographical data from the cities to display personalized search results to its users. Given that companies such as both Uber and Airbnb are built on the foundation of leveraging data and making data-driven decisions, the proposed BDA model in this study will offer limited insights to this new generation of data-driven companies. Thus, future IS research can compare traditional and new generation data-driven companies in terms of their adoption and development of BDA.

Third, both sets of data in this study were collected from members of LinkedIn communities. Given the proliferation of social media, IS scholars have been exploring online sources such as Xing and LinkedIn to collect data to test their hypotheses (e.g., [90,105]). However, another avenue for future research is to further validate this instrument by collecting data from non-LinkedIn sources.

Finally, this study only focused on companies from the United States. Since big data is a global phenomenon, this study can be expanded by including a broader sample of firms outside of the United States. It will be interesting to see if the country-level differences affect the relationship between BDA capability and firm performance.

5. Conclusion

This work was motivated by the phenomenal influx of interest in BDA by both practitioners and academics. While practitioners have long been contributing to the literature on BDA, academics have only recently begun to understand the big data sensation [30,33,37,40]. Consequently, a majority of the extant big data literature talks about the transformational potential of BDA without clearly defining the notion of BDA capability and how firms can create one. This study took insights from RBT, past IT capability literature, and recent published work in big data and suggested seven resources that are likely to create a BDA capability. Specifically, data, technology, and basic resources (e.g., sufficient investments, adequate time) are suggested as three necessary tangible resources, and managerial and technical big data skills are identified as two important human skills. In addition to tangible and human resources, firms need to construct intangible resources such as data-driven culture and the intensity of organizational learning to create a BDA capability. Finally, this study developed a survey instrument to measure a firm’s BDA capability, which was then used to empirically validate the relationship between BDA capability and firm performance.

Appendix A

First-Order Constructs of the BDA Capability Construct and their Items

Construct		Item	Source
Data	D1	We have access to very large, unstructured, or fast-moving data for analysis	[29]
	D2	We integrate data from multiple internal sources into a data warehouse or mart for easy access	[29]
	D3	We integrate external data with internal to facilitate high-value analysis of our business environment	[29]
Technology	T1	We have explored or adopted parallel computing approaches (e.g., Hadoop) to big data processing	[29]
	T2	We have explored or adopted different data visualization tools	[29]
	T3	We have explored or adopted cloud-based services for processing data and performing analytics	[29]
	T4	We have explored or adopted open-source software for big data analytics	[29]
	T5	We have explored or adopted new forms of databases such as Not Only SQL (NoSQL) for storing data.	[106]
Basic Resources	BR1	Our big data analytics projects are adequately funded	[44]
	BR2	Our big data analytics projects are given enough time to achieve their objectives	[29,44]
Technical Skills	TS1	We provide big data analytics training to our own employees	[43]
	TS2	We hire new employees that already have the big data analytics skills	[43]
	TS3	Our big data analytics staff has the right skills to accomplish their jobs successfully	[44,107]
	TS4	Our big data analytics staff has suitable education to fulfill their jobs	[107]
	TS5	Our big data analytics staff holds suitable work experience to accomplish their jobs successfully	[107]
	TS6	Our big data analytics staff is well trained	[107]
Managerial Skills	MS1	Our big data analytics managers understand and appreciate the business needs of other functional managers, suppliers, and customers.	[43]
	MS2	Our big data analytics managers are able to work with functional managers, suppliers, and customers to determine opportunities that big data might bring to our business	[29,43]
	MS3	Our big data analytics managers are able to coordinate big data-related activities in ways that support other functional managers, suppliers, and customers	[43]
	MS4	Our big data analytics managers are able to anticipate the future business needs of functional managers, suppliers, and customers	[43]
	MS5	Our big data analytics managers have a good sense of where to apply big data	[29]
	MS6	Our big data analytics managers are able to understand and evaluate the output extracted from big data	[108]
Data-driven Culture	DD1	We consider data a tangible asset	[109]
	DD2	We base our decisions on data rather than on instinct	[5]
	DD3	We are willing to override our own intuition when data contradict our viewpoints	[35]
	DD4	We continuously assess and improve the business rules in response to insights extracted from data	[5]
	DD5	We continuously coach our employees to make decisions based on data	[5]
Intensity of Organizational Learning	OL1	We are able to search for new and relevant knowledge	[55]
	OL2	We are able to acquire new and relevant knowledge	[55]
	OL3	We are able to assimilate relevant knowledge	[55]
	OL4	We are able to apply relevant knowledge	[55]
	OL5		[55]

Please cite this article in press as: M. Gupta, J.F. George, Toward the development of a big data analytics capability, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.07.004

(*Continued*)

First-Order Constructs of the BDA Capability Construct and their Items	
	We have made concerted efforts for the exploitation of existing competencies and exploration of new knowledge.

Appendix B

Measures of Firm Performance

Construct		Item	Source
Market Performance	MP1	We have entered new markets more quickly than our competitors.	[72,73]
	MP2	We have introduced new products or services into the market faster than our competitors.	
	MP3	Our success rate of new products or services has been higher than our competitors.	
	MP4	Our market share has exceeded that of our competitors.	
Operational Performance	OP1	Our productivity has exceeded that of our competitors	[72,73]
	OP2	Our profit rate has exceeded that of our competitors.	
	OP3	Our return on investment (ROI) has exceeded that of our competitors.	
	OP4	Our sales revenue has exceeded that of our competitors.	

*All questions were asked with respect to the past 3 years.

References

[1] Gartner, Gartner Survey Reveals That 64 Percent of Organizations Have Invested or Plan to Invest in Big Data in 2013, Gartner, 2013.
[2] A.S. Bharadwaj, A resource-based perspective on information technology capability and firm performance: an empirical investigation, MIS Q. (2000) 169–196.
[3] R.M. Grant, Contemporary Strategy Analysis and Cases: Text and Cases, John Wiley & Sons, 2010.
[4] N.G. Carr, IT doesn't matter, Educause Rev. 38 (2003) 24–38.
[5] J.W. Ross, C.M. Beath, A. Quaadgras, You may not need big data after all, Harv. Bus. Rev. 91 (2013) 90–98.
[6] R. Amit, P.J. Schoemaker, Strategic assets and organizational rent, Strateg. Manage. J. 14 (1993) 33–46.
[7] J.B. Barney, Looking inside for competitive advantage, Acad. Manage. Exec. 9 (1995) 49–61.
[8] D.J. Teece, The foundations of enterprise performance: dynamic and ordinary capabilities in an (economic) theory of firms, Acad. Manage. Perspect. 28 (2014) 328–352.
[9] D.J. Teece, G. Pisano, A. Shuen, Dynamic Capabilities and Strategic Management, John Wiley & Sons, 1997.
[10] S.B. MacKenzie, P.M. Podsakoff, N.P. Podsakoff, Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques, MIS Q. 35 (2011) 293–334.
[11] J.B. Barney, D.J. Ketchen, M. Wright, The future of resource-based theory revitalization or decline? J. Manage. 37 (2011) 1299–1315.
[12] I.V. Kozlenkova, S.A. Samaha, R.W. Palmatier, Resource-based theory in marketing, J. Acad. Market. Sci. 42 (2014) 1–21.
[13] R.W. Palmatier, R.P. Dant, D. Grewal, A comparative longitudinal analysis of theoretical perspectives of interorganizational relationship performance, J. Market. 71 (2007) 172–194.
[14] B. Marr, Big Data: Using SMART Big Data. Analytics and Metrics to Make Better Decisions and Improve Performance, Wiley Hoboken, 2015.
[15] M.A. Peteraf, The cornerstones of competitive advantage: a resource-based view, Strateg. Manage. J. 14 (1993) 179–191.
[16] J.T. Mahoney, J.R. Pandian, The resource-based view within the conversation of strategic management, Strateg. Manage. J. 13 (1992) 363–380.
[17] W. Oh, A. Pinsonneault, On the assessment of the strategic value of information technologies: conceptual and analytical approaches, MIS Q. (2007) 239–265.
[18] H.-C. Chae, C.E. Koh, V.R. Prybutok, Information technology capability and firm performance: contradictory findings and their possible causes, MIS Q. 38 (2014) 305–326.
[19] N. Melville, K. Kraemer, V. Gurbaxani, Review: information technology and organizational performance: an integrative model of IT business value, MIS Q. 28 (2004) 283–322.
[20] J.-W. Gu, H.-W. Jung, The effects of IS resources, capabilities, and qualities on organizational performance: an integrated approach, Inform. Manage. 50 (2013) 87–97.
[21] M.S. Sodhi, Conceptualizing social responsibility in operations via stakeholder resource-based view, Prod. Oper. Manage. (2015).
[22] M.A. Hitt, K. Xu, C.M. Carnes, Resource based theory in operations management research, J. Oper. Manage. 41 (1) (2015) 77–94.
[23] M.J. Mannor, J. Shamsie, D.E. Conlon, Does experience help or hinder top managers? Working with different types of resources in hollywood, Strateg. Manage. J. (2015).
[24] M. Wade, J. Hulland, Review: the resource-based view and information systems research: review, extension, and suggestions for future research, MIS Q. 28 (2004) 107–142.
[25] J. Barney, Firm resources and sustained competitive advantage, J. Manage. 17 (1991) 99–120.
[26] R. Makadok, Toward a synthesis of the resource-based and dynamic-capability views of rent creation, Strateg. Manage. J. 22 (2001) 387–401.
[27] R. Santhanam, E. Hartono, Issues in linking information technology capability to firm performance, MIS Q. (2003) 125–153.
[28] M.A. Beyer, D. Laney, The Importance of 'Big Data': A Definition, Gartner, Stamford, CT, 2012.
[29] T. Davenport, Big data at work: dispelling the myths, Uncovering the Opportunities, Harvard Business Review Press, 2014.
[30] M.L. Markus, New games, new rules, new scoreboards: the potential consequences of big data, J. Inform. Technol. (2015) 1–2.
[31] E.O.o.t. President, Big Data: Seizing Opportunities, Preserving Values, Executive Officer of the President, Washington, DC, 2014.
[32] S. Kaisler, F. Armour, J.A. Espinosa, W. Money, Big data: issues and challenges moving forward, System Sciences (HICSS), 2013 46th Hawaii International Conference On, IEEE (2013) 995–1004.
[33] J.L. Zhao, S. Fan, D. Hu, Business challenges and research directions of management analytics in the big data era, J. Manag. Anal. 1 (2014) 169–174.
[34] R.L. Mitchell, 8 Big Trends in Big Data Analytics, Computerworld, 2014.
[35] A. McAfee, E. Brynjolfsson, Big data: the management revolution, Harv. Bus. Rev. 90 (2012) 60–66.
[36] S. LaValle, E. Lesser, R. Shockley, M.S. Hopkins, N. Kruschwitz, Big data, analytics and the path from insights to value, MIT Sloan Manage. Rev. 21 (2014).
[37] J. Chen, Y. Chen, X. Du, C. Li, J. Lu, S. Zhao, X. Zhou, Big data challenge: a data management perspective, Front. Comput. Sci. 7 (2013) 157–164.
[38] D.J. Teece, Intangible assets and a theory of heterogeneous firms, in: intangibles market failure and innovation performance, Springer (2015) 217–239.
[39] M. Manyika, B. Chui, J. Brown, R. Bughin, C. Dobbs, A.H. Roxburgh, Big Data: The Next Frontier for Innovation, Competition, and Productivity, M.G. Institute, 2011.
[40] G. George, M.R. Haas, A. Pentland, Big data and management, Acad. Manage. J. 57 (2014) 321–326.
[41] Accenture, Eighty Percent of Consumers Believe Total Data Privacy No Longer Exists Accenture Survey Finds, Accenture, 2014.
[42] TechCrunch, Apple Acquires Durable Database Company Foundation DB, TechCrunch, 2015.
[43] F.J. Mata, W.L. Fuerst, J.B. Barney, Information technology and sustained competitive advantage: a resource-based analysis, MIS Q. (1995) 487–505.
[44] B.H. Wixom, H.J. Watson, An empirical investigation of the factors affecting data warehousing success, MIS Q. (2001) 17–41.
[45] J.W. Ross, C.M. Beath, D.L. Goodhue, Develop long-term competitiveness through IT assets, Sloan Manage. Rev. 38 (1996) 31–42.
[46] P. Russom, Big data analytics,TDWI best practices report, Fourth Q. (2011).
[47] H. Chen, R.H. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact, MIS Q. 36 (2012).
[48] I. Nonaka, R. Toyama, N. Konno, SECI, *Ba* and leadership: a unified model of dynamic knowledge creation, Long Range Plann. 33 (2000) 5–34.
[49] R. House, M. Javidan, P. Hanges, P. Dorfman, Understanding cultures and implicit leadership theories across the globe: an introduction to project GLOBE, J. World Bus. 37 (2002) 3–10.
[50] J. Iivari, M. Huisman, The relationship between organizational culture and the deployment of systems development methodologies, MIS Q. 31 (2007) 35–58.
[51] G.R. Dowling, Developing your company image into a corporate asset, Long Range Plann. 26 (1993) 101–109.
[52] J.B. Barney, Organizational culture: can it be a source of sustained competitive advantage? Acad. Manage. Rev. 11 (1986) 656–665.
[53] K.M. Eisenhardt, J.A. Martin, Dynamic capabilities: what are they? Strateg. Manage. J. 21 (2000) 1105–1121.
[54] R.M. Grant, Toward a knowledge-based theory of the firm, Strateg. Manage. J. 17 (1996) 109–122.
[55] G.D. Bhatt, V. Grover, Types of information technology capabilities and their role in competitive advantage: an empirical study, J. Manage. Inform. Syst. 22 (2005) 253–277.
[56] W.M. Cohen, D.A. Levinthal, Absorptive capacity: a new perspective on learning and innovation, Adm. Sci. Q. (1990) 128–152.
[57] I. Nonaka, D.J. Teece, Managing Industrial Knowledge: Creation, Transfer and Utilization, Sage, 2001.
[58] A. Bharadwaj, O.A. El Sawy, P.A. Pavlou, N. Venkatraman, Digital business strategy: toward a next generation of insights, Mis Q. 37 (2013) 471–482.

[59] V. Sambamurthy, A. Bharadwaj, V. Grover, Shaping agility through digital options: reconceptualizing the role of information technology in contemporary firms, MIS Q. (2003) 237–263.
[60] K. Du, Research note—parenting new acquisitions: acquirers' digital resource redeployment and targets' performance improvement in the US hospital industry, Inform. Syst. Res. 26 (2015) 829–844.
[61] R. Kohli, V. Grover, Business value of IT: An essay on expanding research directions to keep up with the times, J. Assoc. Inform. Syst. 9 (2008) 1.
[62] H.M. Blalock, Causal Inferences in Nonexperimental Research, (1964) .
[63] G. Kim, B. Shin, V. Grover, Investigating two contradictory views of formative measurement in information systems research, MIS Q. 34 (2010) 345–365.
[64] C.B. Jarvis, S.B. MacKenzie, P.M. Podsakoff, The negative consequences of measurement model misspecification: a response to aguirre-Urreta and marakas, Mis Q. 36 (2012) 139–146.
[65] S. Petter, A. Rai, D. Straub, The critical importance of construct measurement specification: a response to Aguirre-Urreta and Marakas, MIS Q. 36 (2012) 147–155.
[66] S. Petter, D. Straub, A. Rai, Specifying formative constructs in information systems research, MIS Q. (2007) 623–656.
[67] J.F. Hair Jr., G.T.M. Hult, C. Ringle, M. Sarstedt, A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM), SAGE Publications, Incorporated, 2013.
[68] E.E. Rigdon, J.-M. Becker, A. Rai, C.M. Ringle, A. Diamantopoulos, E. Karahanna, D.W. Straub, T.K. Dijkstra, Conflating antecedents and formative indicators: a comment on Aguirre-Urreta and Marakas, Inform. Syst. Res. 25 (2014) 780–784.
[69] C.M. Ringle, M. Sarstedt, D. Straub, A critical look at the use of PLS-SEM in MIS Quarterly, MIS Q. (MISQ) 36 (2012).
[70] Y. Lu, K. Ramamurthy, Understanding the link between information technology capability and organizational agility: an empirical examination, Mis Q. 35 (2011) 931–954.
[71] G. Kim, B. Shin, O. Kwon, Investigating the value of sociomaterialism in conceptualizing IT capability of a firm, J. Manage. Inform. Syst. 29 (2012) 327–362.
[72] T. Ravichandran, C. Lertwongsatien, Effect of information systems resources and capabilities on firm performance: a resource-based perspective, J. Manage. Inform. Syst. 21 (2005) 237–276.
[73] N. Wang, H. Liang, W. Zhong, Y. Xue, J. Xiao, Resource structuring or capability building? An empirical study of the business value of information technology, J. Manage. Inform. Syst. 29 (2012) 325–367.
[74] A. Diamantopoulos, The error term in formative measurement models: interpretation and modeling implications, J. Modell. Manage. 1 (2006) 7–17.
[75] A. Diamantopoulos, Reflective and formative metrics of relationship value: response to Baxter's commentary essay, J. Bus. Res. 63 (2010) 91–93.
[76] C.B. Jarvis, S.B. MacKenzie, P.M. Podsakoff, A critical review of construct indicators and measurement model misspecification in marketing and consumer research, J. Consum. Res. 30 (2003) 199–218.
[77] Y. Chen, Y. Wang, S. Nevo, J. Benitez-Amado, G. Kou, IT capabilities and product innovation performance: the roles of corporate entrepreneurship and competitive intensity, Inform. Manage. 52 (2015) 643–657.
[78] A.S. Bharadwaj, V. Sambamurthy, R.W. Zmud, T capabilities: theoretical perspectives and empirical operationalization, Proceedings of the 20th International Conference on Information Systems, Association for Information Systems (1999) 378–385.
[79] A. Rai, X. Tang, Leveraging IT capabilities and competitive process capabilities for the management of interorganizational relationship portfolios, Inform. Syst. Res. 21 (2010) 516–542.
[80] K.A. Bollen, Interpretational Confounding Is Due to Misspecification, Not to Type of Indicator: Comment on Howell, 2007, Breivik, and Wilcox, 2007.
[81] R.T. Cenfetelli, G. Bassellier, Interpretation of formative measurement in information systems research, MIS Q. (2009) 689–707.
[82] W. Chin, B. Marcolin, The holistic approach to construct validation in IS research: examples of the interplay between theory and measurement, Administrative Sciences Association of Canada-annual Conference, Administrative Sciences Association of Canada, 1995, pp. 34–43.
[83] R.D. Howell, E., Breivik, J.B. Wilcox, Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007), (2007).
[84] K.A. Bollen, Evaluating effect, composite, and causal indicators in structural equation models, MIS Q. 35 (2011) 359–372.
[85] K.A. Bollen, W.R. Davis, Two rules of identification for structural equation models, Struct. Equ. Modeling 16 (2009) 523–536.
[86] R.P. Bagozzi, Measurement and meaning in information systems and organizational research: methodological and philosophical foundations, MIS Q. 35 (2011) 261–292.
[87] R.P. Bagozzi, On the Meaning of Formative Measurement and How It Differs from Reflective Measurement: Comment on Howell, 2007, Breivik, and Wilcox, 2007.
[88] M. Wetzels, G. Odekerken-Schröder, C. Van Oppen, Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration, MIS Q. (2009) 177–195.
[89] J.F. Hair, W.C. Black, B.J. Babin, R.E. Anderson, R.L. Tatham, Multivariate Data Analysis, Pearson Prentice Hall Upper Saddle River, NJ, 2006.
[90] T. Schmiedel, J. Vom Brocke, J. Recker, Development and validation of an instrument to measure organizational cultures' support of Business Process Management, Inform. Manage. 51 (2014) 43–56.
[91] J.R. Edwards, Multidimensional constructs in organizational behavior research: an integrative analytical framework, Organ. Res. Methods 4 (2001) 144–192.
[92] C. Fornell, D.F. Larcker, Evaluating structural equation models with unobservable variables and measurement error, J. Market. Res. (1981) 39–50.
[93] J. Henseler, C.M. Ringle, M. Sarstedt, A new criterion for assessing discriminant validity in variance-based structural equation modeling, J. Acad. Market. Sci. (2014) 1–21.
[94] R. Klein, A. Rai, Interfirm strategic information flows in logistics supply chain relationships, MIS Q. (2009) 735–762.
[95] A. Rai, R. Patnayakuni, N. Seth, Firm performance impacts of digitally enabled supply chain integration capabilities, MIS Q. (2006) 225–246.
[96] S. Gosain, A. Malhotra, O.A. El Sawy, Coordinating for flexibility in e-business supply chains, J. Manage. Inform. Syst. 21 (2004) 7–45.
[97] S.P.-J. Wu, D.W. Straub, T.-P. Liang, How information technology governance mechanisms and strategic alignment influence organizational performance: insights from a matched survey of business and IT managers, MIS Q. 39 (25) (2014) 497–518.
[98] M.J. Tippins, R.S. Sohi, IT competency and firm performance: is organizational learning a missing link? Strateg. Manage. J. 24 (2003) 745–761.
[99] X. Tang, A. Rai, The moderating effects of supplier portfolio characteristics on the competitive performance impacts of supplier-facing process capabilities, J. Oper. Manage. 30 (2012) 85–98.
[100] H. Lee, B. Choi, Knowledge management enablers, processes, and organizational performance: an integrative view and empirical examination, J. Manage. Inform. Syst. 20 (2003) 179–228.
[101] W.Q. Judge, T.J. Douglas, Performance implications of incorporating natural environmental issues into the strategic planning process: an empirical assessment, J. Manage. Stud. 35 (1998) 241–262.
[102] J.C. Anderson, D.W. Gerbing, Some methods for respecifying measurement models to obtain unidimensional construct measurement, J. Market. Res. (1982) 453–460.
[103] B.D. Smart, Topgrading: How Leading Companies Win by Hiring, Coaching, and Keeping the Best People, Penguin, coaching, and keeping the best people, 2005.
[104] H. Chen, R.H.L. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact, MIS Q. 36 (2012).
[105] P. Tambe, Big data investment, skills, and firm value, Manage. Sci. 60 (2014) 1452–1469.
[106] K. Gordon, Principles of Data Management: Facilitating Information Sharing, BCS, The Chartered Institute, 2007.
[107] A. Carmeli, A. Tishler, The relationships between intangible organizational elements and organizational performance, Strateg. Manage. J. 25 (2004) 1257–1278.
[108] S. Athey, How Big Data Changes Business Management, Stanford Business, 2013.
[109] D. Laney, 3D data management: controlling data volume, velocity and variety, META Group Res. Note 6 (2001).

Manjul Gupta is an assistant professor of management information systems in the College of Business at Florida International University. He earned his doctorate in information systems from Iowa State University. His research focuses on big data, cross-cultural differences, and agile systems development.

Joey F. George is a professor of information systems and the John D. DeVries Endowed Chair in Business in the College of Business at Iowa State University. He received his doctorate in management from the University of California Irvine in 1986. His research interests focus on the use of information systems in the workplace, including deceptive computer-mediated communication, computer-based monitoring, and group support systems.

Please cite this article in press as: M. Gupta, J.F. George, Toward the development of a big data analytics capability, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.07.004