JOURNAL OF

Database Management

IGI PUBLISHING

Publisher of Peer-Reviewed, Timely, and Innovative Research Since 1988

www.igi-global.com

Journal of Database Management (JDM)

Volume 31, Issue 3
July - September 2020

Keng Siau - Missouri University of Science and Technology, USA

9781799804710

Journal of Database Management

Volume 31 • Issue 3 • July-September 2020 • ISSN: 1063-8016 • eISSN: 1533-8010

IGI PUBLISHING
AN IMPRINT OF IGI GLOBAL
WWW.IGI-GLOBAL.COM

EDITORIAL REVIEW BOARD

CONTINUED

EDITORIAL REVIEW BOARD

CONTINUED

EDITORIAL REVIEW BOARD

CONTINUED

Call for Articles

Journal of Database Management

Volume 31 • Issue 3 • July-September 2020 • ISSN: 1063-8016 • eISSN: 1533-8010

MISSION

The **Journal of Database Management (JDM)** publishes original research on all aspects of database management, design science, systems analysis and design, and software engineering. The primary mission of JDM is to be instrumental in the improvement and development of theory and practice related to information technology, information systems, and management of knowledge resources. The journal is targeted at both academic researchers and practicing IT professionals.

COVERAGE AND MAJOR TOPICS

The topics of interest in this journal include, but are not limited to:

Agile systems development • Artificial intelligence • Augmented and virtual reality • Bio-informatics • Business Analytics • Business Intelligence • Cloud computing • Cognitive modeling • Conceptual modeling • Cybersecurity • Data Science • Data quality • Data warehousing and data mining • Database management and administration • Database models and query languages • Database security and integrity • Design science • E-business, m-commerce, and social-commerce models and architectures • Empirical software engineering • Enterprise systems and supply chain integration • Financial technology • Geographical information systems • Human-computer interaction and user experience • Heterogeneous and distributed database • Information and knowledge modeling • Internet of Things • Intelligent agents and agent-based applications • Knowledge engineering and management • Machine learning • Method engineering and metamodeling • Object-oriented methods and methodologies • Requirements engineering • Service-oriented architecture/service-oriented computing • Semantic Web and ontology • Smart technology • Software engineering • System analysis and design • Unified modeling language and unified process • Web database and Web-based information systems • Web design methods and methodologies

ALL INQUIRIES REGARDING JDM SHOULD BE DIRECTED TO THE ATTENTION OF:

Keng Siau, Editor-in-Chief • JDM@igi-global.com

ALL MANUSCRIPT SUBMISSIONS TO JDM SHOULD BE SENT THROUGH THE ONLINE SUBMISSION SYSTEM:

http://www.igi-global.com/authorseditors/titlesubmission/newproject.aspx

IDEAS FOR SPECIAL THEME ISSUES MAY BE SUBMITTED TO THE EDITOR(S)-IN-CHIEF

PLEASE RECOMMEND THIS PUBLICATION TO YOUR LIBRARIAN

For a convenient easy-to-use library recommendation form, please visit:
http://www.igi-global.com/JDM

Table of Contents

Journal of Database Management

Volume 31 • Issue 3 • July-September-2020 • ISSN: 1063-8016 • eISSN: 1533-8010

Research Articles

The *Journal of Database Management* is indexed or listed in the following: ABI/Inform; ACM Digital Library; Aluminium Industry Abstracts; Australian Business Deans Council (ABDC); Bacon's Media Directory; Burrelle's Media Directory; Cabell's Directories; Ceramic Abstracts; Compendex (Elsevier Engineering Index); Computer & Information Systems Abstracts; Corrosion Abstracts; CSA Civil Engineering Abstracts; CSA Illumina; CSA Mechanical & Transportation Engineering Abstracts; Current Contents®/Engineering, Computing, & Technology; DBLP; DEST Register of Refereed Journals; EBSCOhost's Business Source; EBSCOhost's Computer & Applied Sciences Complete; EBSCOhost's Computer Science Index; EBSCOhost's Computer Source; EBSCOhost's Current Abstracts; EBSCOhost's STM Abstracts; Electronics & Communications Abstracts; Emerald Abstracts; Engineered Materials Abstracts; Gale Directory of Publications & Broadcast Media; GetCited; Google Scholar; INSPEC; Internet & Personal Computing Abstracts; ISBIB; JournalTOCs; Library & Information Science Abstracts (LISA); Library Literature & Information Sciences; Materials Business File - Steels Alerts; MediaFinder; Norwegian Social Science Data Services (NSD); PubList.com; SCOPUS; Solid State & Superconductivity Abstracts; The Index of Information Systems Journals; The Standard Periodical Directory; Ulrich's Periodicals Directory; Web of Science; Web of Science Science Citation Index Expanded (SCIE)

Journal of Database Management
Volume 31 • Issue 3 • July-September 2020

Systems Analysis and Design Toolkit Based on Work System Theory and Its Extensions

Steven Alter, University of San Francisco, USA
Dominik Bork, Faculty of Computer Science, University of Vienna, Vienna, Austria
https://orcid.org/0000-0001-8259-2297

ABSTRACT

This article describes proposed content of an online toolkit for users of the work system modeling method (WSMM), an extension of the work system method (WSM), which was developed to help business professionals understand IT-reliant systems and collaborate with IT professionals. A summary of work system theory (WST) and WSM provides a background. A two-dimensional design space for modeling methods illustrates WSMM's context. Two limitations of WSMM imply the need for a toolkit that overcomes those limitations. An auto rental example is used to illustrate a series of modeling, analysis, and design modules related to different stakeholder purposes. Most of the modules reflect components of Word documents used as outlines to produce over 700 management briefings, mostly by MBA and Executive MBA students, between 2003 and 2017. A concluding section summarizes the main ideas, explains how the use of different modules based on the work system metaphor flexibly supports systems analysis and design, and identifies challenges for future research.

KEYWORD

Enterprise Modeling, Modeling Methods, Systems Analysis and Design, Systems Analysis Toolkit, Work System, Work System Method, Work System Theory

INTRODUCTION

This research note builds on Alter and Bork (2019), which proposed a work system modeling method (WSMM) that advances a long-term stream of research (e.g., Alter 1995, 2006, 2013) that developed work system theory (WST), the work system method (WSM), and various extensions of WST. Most of that research occurred in the general spirit of design science research (e.g., Rossi et al., 2013) and involved the design and testing of artifacts to be used in analyzing and designing systems. Alter and Bork (2019) was motivated by widely discussed difficulties in the practical application of formal modeling methods that stem from a mismatch between the details and rigor of formal software development techniques and the knowledge, interests, and inclinations of most business professionals. Experienced modeling experts sometimes overcome those difficulties, but even they often find it difficult to elicit full collaboration from business colleagues who lack sophisticated appreciation of modeling.

WSMM attempts to bridge methods designed for different purposes pursued by different stakeholders when moving from rough requirements to programming details. The evolution of WSM

DOI: 10.4018/JDM.2020070101

over two decades tried to help business professionals think about systems in terms of business interests and concerns, which does not require high specificity related to technical details. WSMM's attempt to bridge user and developer interests is based on relaxing assumptions about modeling and modeling methods, thereby permitting use of different metamodels to address different modeling purposes with different degrees of formality and specificity. Consistent use of a work system perspective as an invariant modeling metaphor (Ferstl and Sinz, 2013) supports coherence across different purposes.

This research note extends the initial conceptualization of WSMM by proposing an SA&D toolkit that treats WSMM as part of a broader SA&D process that includes much more than modeling. The toolkit builds on design artifacts used in over 700 management briefings related to problematic IT-reliant work systems produced mostly by MBA and Executive MBA students working individually or in teams between 2003 and 2017.

The next section provides more background by positioning WSMM's central work system metaphor in the context of conceptual modeling, SA&D, WST, and WSM. A two-dimensional design space for modeling methods illustrates the main idea of WSMM. An SA&D toolkit proposed here overcomes two important limitations of WSMM. An auto rental example illustrates that toolkit, which includes a series of modeling, analysis, and design modules related to different stakeholder purposes. A concluding section summarizes the main ideas, explains the main implications, notes a practical path to implementation, and identifies challenges for future research.

BACKGROUND

This section starts with issues related to conceptual models, mentions several modeling perspectives, and then summarizes the work system perspective that is the basis of WSMM.

Conceptual Modeling Issues

A special issue of the *Journal of Database Management* on ontological analysis in conceptual modeling (Burton-Jones et al., 2017b) starts with a research commentary in which Wand and Weber (2017, p. 2) expresses disappointment that "representation theory (RT), conceptual modeling, and ontology have attracted little attention" from IS researchers despite a belief (shared by the current authors) that conceptual modeling is at the core of the IS discipline. Burton-Jones et al. (2017a) summarizes the main ideas of RT and provides an extensive literature review covering articles that build on RT. Their critique notes, "a tacit assumption that appears to underlie RT is that the usefulness of information systems increases monotonically with increases in their representational faithfulness" (p. 1323).

Past research leads to questioning similar assumptions in the general realm of conceptual modeling, especially due to mismatches between user capabilities and the nature of many modeling tools. Domain experts often perceive business situations in imprecise ways and may or may not be able to capture their knowledge in conceptual models (cf. Bjeković et al., 2014; Figl, 2017; Zur Muehlen and Recker, 2013). Research on modeling method usage (e.g., Fettke, 2009; Mendling et al., 2010) and model comprehension (e.g., Haisjackl et al., 2018; Johannsen et al., 2014; Mendling et al., 2018) illuminates factors that exacerbate those issues. Modeling tools sometimes force users to express themselves using formalisms that are unfamiliar or difficult to use (Wüest et al., 2017)). Modeling methods often do not fit modelers' aptitudes, knowledge, and purposes (Hinkel et al., 2016; Zur Muehlen & Recker, 2013). Simões et al. (2018) notes that the "lack of intuitiveness of diagrammatic representations and the complementary role of text-based representations has been underlined in recent research." Cognitive load (Sweller, 1994) for stakeholders becomes increasingly important as unfamiliar symbols and icons proliferate. In combination, those issues support the desirability of recognizing that different purposes of different stakeholders imply different needs for specificity and formality – a reality that is reflected in WSMM.

Alternative Perspectives on Systems

Wand and Weber (1990, p. 62) say "information systems are primarily intended to model the states and behavior of some existing or conceived real world system," which can be done using concepts in what is now called the Bunge-Weber-Wand (BWW) ontology. This includes thing, properties, state, state law, event, event space, history, coupling, system, and so on (p. 64). General systems theory (GST) provides a different view of the most basic ideas for modeling. Skyttner (1996, pp. 20-21) describes "near total agreement on which properties, together, comprise a general systems theory." Those properties include interrelationship and interdependence of objects and their attributes, holism, goal seeking, transformation process, inputs and outputs, entropy, regulation, hierarchy, differentiation, and equifinality. It is very difficult to attain system-related insights from BWW, GST, and other foundational ontologies because general concepts at such a high level of abstraction seem far distant from business concerns. Activity system perspectives such as activity theory (Engeström & Miettinen, 1999), soft system methodology (Checkland, 1999), and the work system perspective address business concerns more directly by focusing on activities that are performed rather than objects and events.

Work System Perspective

The WS perspective assumes that the goal is to generate beneficial changes in work systems, which typically involves much more than developing or improving technical artifacts. A work system is a system in which human participants and/or machines perform processes and activities using information, technology, and other resources to produce product/services for internal and/or external customers. The first and/or implies that work systems can be sociotechnical (with human participants) or totally automated. A work system operates within an environment that matters (e.g., national and organizational culture, policies, history, competitive situation, demographics, technological change, other stakeholders, etc.). Work systems rely on human, informational, and technical infrastructure that is shared with other work systems. Work systems should support enterprise and departmental strategies. The definition of work system implies that work system is a very general case that includes many special cases such as information systems, supply chains, service systems, projects, and totally automated work systems. For example, an IS is a work system most of whose activities are devoted to processing information. Supply chains are work systems that extend across multiple organizations to provide resources for supply chain customers. Projects are work systems that are created to produce specific product/ services and then go out of existence. An enterprise can be viewed as a set of interacting work systems that serve overarching goals.

WST, the theoretical basis of WSM, consists of three parts: 1) the definition of work system, 2) the work system framework, and 3) the work system life cycle model. The work system framework (Alter, 2013) outline nine elements of even a rudimentary understanding of a work system's form, function, and environment as it exists during a time interval when its structure is basically static even though it might not have a well-defined business process and might not be IT-intensive. Processes and activities, participants, information, and technologies are internal to the work system. Customers and product/services may be partially inside and partially outside because customers often participate in work systems. Environment, infrastructure, and strategies are the other three elements of the framework. In practice, the realism of high-precision modeling is limited by the way processes fall along a dimension from unstructured to structured (Alter & Recker, 2017) that starts with largely unstructured creative processes (such as many design and management processes) that have no pre-specified sequence, may involve extensive iteration, and therefore are not amenable to detailed, high precision modeling. Furthermore, work system participants may make errors and may pursue adaptations and workarounds instead of following prescribed procedures.

Work System Method

WSM is a semi-formal systems analysis and design approach that was developed over several decades in the spirit of a long-term design science research (DSR) project. Its goal was to produce SA&D

artifacts that would help business professionals visualize work systems in their own organizations and collaborate more effectively with IT professionals. The research involved iterations of applying and improving work system analysis outlines (the design artifacts) that were used mostly by employed MBA and Executive MBA students working individually or in teams. While details of the IS or "systems in organizations" courses varied, the relevant assignments involved identifying a problematic work system in an organization (usually a student's employer) and producing an analysis and explanation guided by a version of a WSM Word outline that was tailored to a specific course. Those outlines contained many broadly applicable questions with responses in a variety of formats. Each question could be viewed as a separate modeling, analysis, or design module that applies to a broad range of work systems. A majority of the students were at the University of San Francisco or Georgia State University (e.g. reports in Truex et al. (2010, 2011) reporting results from 75 and 301 briefings by Georgia State students), although some of the briefings were produced in courses in China, Germany, Hong Kong, India, and Vietnam.

While details differ, every version of WSM is organized as follows: 1) identify the smallest work system that has the problem or opportunity; 2) summarize the "as-is" work system using a work system snapshot, a stylized one page summary (example shown later); 3) evaluate work system operation using measures of performance, key incidents, social relations, and other factors; 4) drill down further as necessary; 5) identify proposed changes using a work system snapshot of a "to be" work system that seems likely to perform better; 6) describe likely performance improvements.

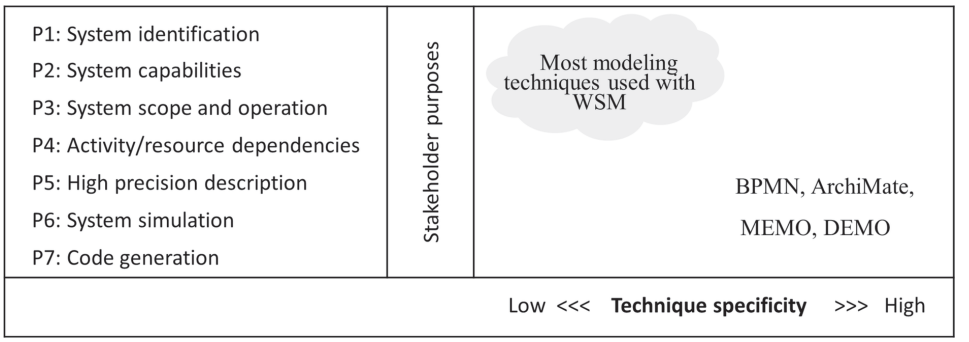
Work System Modeling Method

WSMM development started with discussions related to Sandkuhl (2018), in which enterprise modeling (EM) community leaders call for moving EM from an expert discipline towards "grass roots modeling" and "modeling for the masses." That research agenda proposes "softened requirements to completeness, coherence and rigor." Alter and Bork (2019) explains how WSMM supports the spirit of that agenda and addresses six of seven process modeling problems discussed in van der Aalst (2012). It follows a call in Karagiannis (2015) for "overcoming tendencies to view diagrammatic modeling methods and languages" as "stable, even standardized, artefacts that establish some commonly agreed way of describing a 'system under study,' [which] implies that all stakeholders work on the same level of abstraction and specificity."

Alter and Bork (2019) presented WSMM as a way to expand WSM through a modeling method that accommodates multiple stakeholders who may have different purposes. WSMM encourages use of multiple modeling techniques; thus, relaxing Karagiannis and Kühn's (2002) assumption that all stakeholders and purposes must be accommodated by a single modeling technique. That approach avoids cognitive overload from increasing metamodel complexity due to adding concepts to separate modeling techniques to suit diverse stakeholder needs and purposes.

Examples within the scope of WSMM can be organized based on Figure 1, which represents a two-dimensional design space for modeling methods and modeling techniques related to a core metaphor (work system in this case). Figure 1 identifies the range of stakeholder purposes as P1 through P7. Technique specificity is the extent to which a technique defines exactly what to include, what to ignore, and how to proceed. Techniques with low specificity tend to be flexible but provide relatively little conceptual or procedural guidance. The reverse applies as well. The shaded area positions most of the modeling techniques that have been applied in using WSM. Most of those techniques focus on system scope and operation, and activity/resource dependencies. Those techniques are relatively low in specificity compared to techniques that might be used for high precision description, system simulation, or code generation. Simpler metamodels apply to informal work system visualization while more expressive metamodels help in identifying and selecting among possible changes and in producing or improving software.

Figure 1. Design space for modeling methods and modeling techniques (Alter & Bork, 2019)



Limits of WSMM

The current research addresses two important limitations of WSMM. The first limitation stems from the fact that most business professionals who want to improve work systems need more than a modeling method. They need understandable, easily used, and easily explained ways to visualize and analyze the structure, operation, and performance of work systems. That analysis focuses on problems and opportunities and considers issues such as performance gaps, risks, customer concerns, and other topics that abstract models do not address directly. This paper proposes a toolkit that addresses those considerations for different modeling purposes. A second important limitation is the assumption (implied in Figure 1) that WSMM covers a specific set of purposes that line up sequentially along a dimension from highly informal to highly formal. The sequence of purposes in Figure 1 was useful for introducing the idea of WSMM, but there is no reason to believe that a particular level of technique specificity is associated with only one purpose. A broader view of WSMM assumes that a particular level of technique specificity could be appropriate for many different purposes.

A SYSTEMS ANALYSIS AND DESIGN TOOLKIT THAT BUILDS ON WSMM

The proposed analysis and design toolkit addresses both WSMM limitations. The toolkit consists of a series of modules each of which is directed at a different stakeholder purpose related to understanding or analyzing an aspect of a work system such as dealing with value capture, shared responsibility, and visibility for providers and customers. Informal versions of some of the analysis modules appeared in most of the WSM outlines used in coursework. Other modules are based on extensions of WST that were not included or had not yet been developed. The overall vision is that these modules will be implemented in an interactive toolkit that will make it easy for users to identify, select, and use individual modules or pre-packaged or standardized groups of modules that are especially relevant to the types of situations that they want to analyze. Whereas some of these modules ask for formal modeling support, others require simpler user interaction e.g., by means of list selection or by filling in pre-defined templates. This section identifies examples but cannot cover all conceivably relevant analysis modules due to length limitations and also, more fundamentally, because initial versions of almost any analysis and design toolkit will surely lead to improvements and extensions.

Typical modules in the toolkit will be illustrated by using the Object Management Group's "EU-Rent" case, aspects of which have been used by other researchers. Similar test cases could be constructed for work systems in production, sales, accounting, software development, and so on. The "EU-Rent" case summarizes the operation of a car rental company, including renting the car, picking up the car, dropping off the car, ending the rental, and accepting payment. The nature of the EU-Rent scenario is apparent from the following excerpt: "EU-Rent is a company that rents cars to persons, operating from geographically dispersed branches. The cars of EU-Rent are divided in car

types (brands and models); for every car type there is a particular rental tariff per day. A car may be rented by a reservation in advance or by a 'walk-in' customer on the day of renting. A rental contract specifies the start and end dates of the rental, the cartype one wishes, the branch where the rental starts" (Op't Land and Dietz, 2017)

This paper's use of this example assumes that examination of the EU-Rent situation was motivated by problems including lengthy waits at rental locations, delayed availability of cars, rental reservations that were not honored by customers, and so on. Consistent with enterprise modeling issues mentioned or implied in Sandkuhl et al. (2018) the exercise of modeling this situation would not generate meaningful business results unless the modeling was integrated in some way with an analysis of the business issues at hand.

This section identifies a representative set of modeling, analysis, and design modules based on various aspects of the EU-Rent case. Some modules emphasize structure, whereas others emphasize performance gaps, risks, workarounds, and other useful evidence for analyzing and understanding a situation. Most illustrations involve entering text or numbers into simple or complex tables.

Table 1 identifies modules in three categories: modeling, analysis, and design. Planning modules related to creating an improvement plan (e.g., project resource requirements, rough schedule, project

Table 1. Examples of modeling, analysis, and design modules

Modeling Modules	Analysis Modules	Design Modules
• Identification • Capabilities • Operation and scope of the work system • Value capture • Responsibilities • Visibility • Activity/resource dependencies • System interactions • Diagrammatic specifications	• Problems and opportunities • Performance gaps • Strengths and weaknesses • Exceptions • Workarounds or noncompliance • Key incidents • Risks • Issues for elements of the work system framework	• Proposed changes in the work system • Rationale for proposed changes • Likely improvements in work system performance

risks, etc.) are not included in Table 1. Any output of a modeling, analysis, or design module that is archived for future use would include a timestamp because data collected using those modules may no longer be accurate or relevant after changes have been implemented.

The remainder of this section presents brief comments and mockups of different modeling, analysis, or design modules with highly abbreviated versions of hypothetical responses related to EU-Rent. A paper produced after toolkit implementation and usage would provide more complete responses and might be organized around a typical order in which the issues would be considered. Figure 2 consolidates modules that call for a short answer or a list plus two modules that require 3 or 4 columns. Other modules are introduced separately. This approach is sufficient for visualizing the types of modules that might be included. There is no assumption that a user would need to use all of the modules or that certain modules would have to be used in a specific order even though reuse of previously entered data, e.g. a list of activities from a work system snapshot (Figure 3), would facilitate subsequent data entry related to the same topics.

Figure 2 starts with modules that call for simple answers or lists. Identification of the work system describes a work system using a verb phrase that summarizes what is being done for whom. Use of a verb phrase prevents confusing the work system with software that it uses, such as CRM or SAP. Naming the main problems and opportunities is important because the relevant work system is assumed to be the smallest work system that exhibits problems and opportunities that launched the analysis. The capabilities of the work system can be summarized as a simple list of

Figure 2. Illustrative analysis modules calling for short answers or lists

Name of the work system	• Renting autos at EU-Rent
Problems and opportunities	• Lengthy waits at rental locations • Delayed availability of cars • Rental reservations that are not honored by customers
Capabilities	• Taking reservations for future rentals • Renting to walk-up customers • Completing rentals with car returns
Strengths	• Excellent conformance to corporate standards • Favorable ratings for the customer experience
Weaknesses	• Rental counters look sloppy and undistinguished • Customer dissatisfaction with confused rental agents
Exceptions	• Car that should have been returned was not returned, making it impossible to satisfy a customer request. • Customer's drivers license seems to be damaged in some way that requires discussion with headquarters.
Workarounds and noncompliance	• Employee overrides corporate policies by giving a luxury car to a renter who rented an unavailable car. • Employee at the exit gate does not fully check a newly rented car for scratches, leading to disputes.
Key incidents	• An electricity outage at a major airport prevented rentals for three hours on a busy morning. • The corporate IT system suffered break-ins that exposed customer records and disrupted operations.
Risks	• Employee treats customer inappropriately • Customer bribes employee to disregard car damage • Accident resulting from improper servicing of a car

Aspect of performance	*Metric (measure of performance)*	*Current numerical value of metric*	*Realistic desired value of metric*
Cost	Employee cost per rental	48.35 euro	40.00 euro
Efficiency	Employee minutes per rental	14.3 minutes	12.6 minutes
Consistency	Nonstandard processing percentage	17%	10%

Activity	Provider responsibility	Customer responsibility
* Renting agent and renter perform initial rental transaction	* Provide a comfortable interaction with the customer * Collect all needed information	* Arrive on time * Provide all necessary information
* Driver picks up the car	* Direct the customer to the car	* Pick up the car
* Driver drops off the car	* Convenient drop off location	* Drop off the car

capabilities or can be described with higher specificity, e.g., in relation to service level expectations. Ideally, a work system's strengths should be maintained and should not be diminished by proposed changes. Most work systems also have identifiable weaknesses that should be overcome. Dealing with significant exceptions often absorbs a great deal of time and sometimes reveals significant work system shortcomings. The simple list of exceptions in Figure 2 could be expanded to identify impacts, outcomes, or responses that affect performance or customer satisfaction. Workarounds or noncompliance often reveals important issues. Employees instructed to follow a process may not comply for reasons including encountering obstacles, finding that a process is extremely awkward in certain situations, and sometimes deviating for personal reasons that are unrelated to corporate goals or customer satisfaction. Significant workarounds or noncompliance often point to significant problems with the work system. Many work systems usually operate adequately but occasionally suffer key incidents such as major disruptions that should be avoided in the future. The analysis of important work systems should always consider risks related to conceivable events that could cause

significant problems for the enterprise or for its employees or customers. Two analysis modules at the bottom of Figure 2 are a bit more complex, one for performance gaps and one for provider and customer responsibilities. Similar modules for value capture and mutual visibility of providers and customers can be envisioned.

Operation and Scope of the Work System

A "work system snapshot" (Figure 3) has been used in all versions of WSM since 2003. This module clarifies the scope of the work system without going into great detail. This snapshot is a formatted one-page work system summary based on six central elements of the work system framework. That easily used description helps in defining the boundaries and contents of the work system and is complete enough to support an initial discussion about the work system's scope and operation. Figure 3 is a

Figure 3. Abbreviated example of a work system snapshot

<table>
<tr><td>Customers</td><td colspan="2">Product/Services</td></tr>
<tr><td>• Renter
• Driver</td><td colspan="2">For customers: Rental of car consistent with rental contract
For providers: Payment for rental</td></tr>
<tr><td colspan="3">Major Processes and Activities</td></tr>
<tr><td colspan="3">• Renting agent performs initial rental transaction through interaction with renter.
• Driver picks up the car.
• Driver drops off the car.
• Drop-off agent ends the rental.
• Renter pays for rental.</td></tr>
<tr><td>Participants</td><td>Information</td><td>Technologies</td></tr>
<tr><td>• Renting agent
• Renter
• Driver
• Drop-off agent</td><td>• Reservation
• Availability of cars at pick-up location
• Rental contract (arrangement for payment, pick-up branch, drop-off branch, start date, end date, type of car, tariff, driver's license, etc.
• Condition of car upon drop-off</td><td>• Corporate software controlling the rental agreement</td></tr>
</table>

highly summarized example that contains only five activities and four participants. Genuinely useful work system snapshots usually contain much more information, even in a single page. As noted earlier, customers may be work system participants, as happens in custom software development. Other forms of work system snapshots have been used. For example, rows for environment, infrastructure, and strategies can be included at the bottom of the work system snapshot if those topics bring key issues.

Many other modeling, analysis, and design modules can be included. The work system snapshot above does not indicate which activities use which information, technology, and other resources. Tables showing activity/resource dependencies are useful when analyzing a work system since every activity in a work system uses resources and produces product/services that constitute its outputs or results. By default, the first column in those tables is the list of activities from the work system snapshot, although users can change that. The other columns are resources used or produced by each activity. One of many versions of that module identifies preconditions, triggers, post conditions, and business rules for each activity.

Other Modules Related to Work System Elements or the Work System as a Whole

Work system attributes and principles have been included in various ways in MS Word outlines for work system analysis and design. An analysis module focusing on conformance to principles lists 24 work system principles and asks the user to express the extent to which the work system conforms with each principle. An experimental four-page checklist for identifying problems listed over 100

attributes of the work system as a whole or of specific elements of the work system framework and asked the user to estimate whether a work system or work system element had 1) no problem, 2) minor issues, or 3) major problems related to that attribute. MBA and Executive MBA students who used that module reported that it was easy to use and useful for identifying topics that they might have overlooked. Treating design dimensions as sliding scales is another approach that has been used. Users of that type of module apply a series of design dimensions by characterizing the current position (C) and desired position (D) of a work system or work system element along multiple continuous dimensions. One example is characterizing product/services along dimensions such as tangible vs. intangible, produced vs. co-produced, produced and transferred vs. consumed as produced, occurring through transactions vs. through relationships, and so on.

Other Theory-Based Analysis Ideas

Extensions of WST and WSM potentially support other analysis modules that have not yet been included in WSM analysis documents. The underlying ideas are related to topics such as a proposed workaround design system that could be used to anticipate likely workarounds, a series of work system axioms, a set of system interaction patterns, a theory of system interactions, ideas from activity theory, value co-creation ideas from service-dominant logic, and so on.

Generic Types of Modules

Histograms, Pareto diagrams, scatter plots, and control charts are examples of generic modules that can be applied to any appropriate dataset regardless of whether the data describes operation or performance results of work systems. The systems analysis toolkit should link conveniently to a range of generic modules because potential users of the toolkit may find generic modules useful if their work system data fits with the generic modules.

High Precision Description of the Work System

Well-known diagrammatic modeling techniques address needs for detailed understanding of how work system components are structured and how the work system operates. For example, it is possible to represent activity sequence and branching logic using BPMN diagrams with activities in swimlanes for different participant roles. Similarly, entity-relationship diagrams and ArchiMate's application and technology layers could be used.

DISCUSSION AND CONCLUSION

This paper extends a stream of research related to WST and WSM by linking WSMM with modeling, analysis, and design modules that can be used by business and IT professionals for personal understanding and for collaboration when analyzing and designing systems. Much of the previous research in that stream was inspired by experience with management briefings by employed students. The initial development of WSMM placed WSM in a broader and more generative context related to modeling. The current research goes a step further by linking WSMM with various types of modeling, analysis, and design modules that individuals and groups can use to attain full benefit from the development and use of models.

General Implications of a Vision That Links Modeling and Analysis

Previous research showed how WSMM potentially facilitates visualization of complex systems such as work systems by relaxing common assumptions about modeling methods. This paper's toolkit proposal assumes that aspirations outlined in Sandkuhl et al. (2018) and elsewhere, such as "grass roots modeling" and "modeling for the masses," are more likely to be achieved, at least partially, if modeling is combined with analysis and design modules that help in identifying, describing,

and analyzing business problems and opportunities. It is worthwhile to explore and question that assumption. The exploration could start by trying to find examples of modeling methods that achieve the goals expressed by Sandkuhl et al. (2018) without strong links to analysis and design modules that address topics other than system content and structure.

Implementation of a Widely Accessible Systems Analysis Toolkit for Understanding and Analyzing Work Systems

This paper focused on presenting ideas rather than illustrating technical implementations. Initial steps in a parallel effort have shown the technical feasibility of implementing WSMM by using ADOxx, a widely used metamodeling development and configuration platform for implementing modeling methods (see Karagiannis et al. (2016) for a selection of methods). ADOxx is available through OMiLAB (the Open Models Laboratory), an open community for the conceptualization of modeling methods (Bork et al., 2019). ADOxx capabilities can be used to implement all of the metamodels needed to support the modeling, analysis, and design modules mentioned in this paper. For example, a blank work system snapshot was created in ADOxx and was used to describe a work system. Further work and experience with ADOxx is required to identify the most convenient ways to use its modeling and analysis capabilities in real situations. For example, this paper identified a set of analysis modules but did not specify how a user interface would make it as easy as possible to find and use individual modules and to combine modules into standardized tools that organizations or individuals can use in standard routines that fit their purposes.

Further Development of the Research Stream Related to WSM, WST, and Extensions

The development of WSM started several decades ago with a focus on issues related to informal but useful description of work systems and with little or no attempt at rigor other than trying to define terms and encourage organized thinking about work systems that involved IT.

This paper can be viewed as a proposal for deconstruction and reconstruction. It used the ideas in WSMM to visualize the deconstruction of WSM Word outlines into modeling, analysis, and design modules that might be applied individually, but whose main value potentially occurs through their use in combination when trying to understand and improve work systems. The most important extensions of this research stream probably will involve packaging the modeling, analysis, and design modules to make them readily accessible, looking for the most effective usage patterns, and building on those patterns with new developments that make this approach broadly available. Realizing this paper's modeling, analysis, and design modules as part of a modularized, flexible work system modeling environment might prove to be a major step toward the vision of "modeling for the masses" that inspired ideas presented here.

REFERENCES

Alter, S. (1995). How should business professionals analyze information systems for themselves? In *Information System Concepts* (pp. 284–299). Boston, MA: Springer. doi:10.1007/978-0-387-34870-4_29

Alter, S. (2006). *The work system method: Connecting people, processes, and IT for business results*. Larkspur: Work System Press.

Alter, S. (2013). Work system theory: Overview of core concepts, extensions, and challenges for the future. *Journal of the Association for Information Systems*, *14*(2), 72–121. doi:10.17705/1jais.00323

Alter, S., & Bork, D. (2019). Work System Modeling Method with Different Levels of Specificity and Rigor for Different Stakeholder Purposes. In *Proceedings of the 14th International Conference on Wirtschaftsinformatik*. Academic Press. pp. 124-138

Alter, S., & Recker, J. C. (2017). Using a work system perspective to expand BPM use cases for research. *Journal of Information Technology Theory and Application*, *18*(1), 47–71.

Bjeković, M., Proper, H. A., & Sottet, J. S. (2014). Embracing pragmatics. In *Proceedings of the International Conference on Conceptual Modeling* (pp. 431-444). Springer. doi:10.1007/978-3-319-12206-9_37

Bork, D., Buchmann, R., Karagiannis, D., Lee, M., & Miron, E. T. (2019). An open platform for modeling method conceptualization: The OMiLAB digital ecosystem. *Communications of the Association for Information Systems*, *44*(1), 32. doi: 10.17705/1CAIS.04432

Burton-Jones, A., Green, P., Parsons, J., & Siau, K. (2017b). Special issue on ontological analysis in conceptual modeling, part 1. *Journal of Database Management*, *28*(1), vii–x.

Burton-Jones, A., Recker, J., Indulska, M., Green, P. F., & Weber, R. (2017a). Assessing Representation Theory with a Framework for Pursuing Success and Failure. *Management Information Systems Quarterly*, *41*(4), 1307–1333. doi:10.25300/MISQ/2017/41.4.13

Correia, F. F., & Aguiar, A. (2013). Patterns of flexible modeling tools. In *Proceedings of the 20th Conference on Pattern Languages of Programs* (pp. 1–9). The Hillside Group.

Engeström, Y., & Miettinen, R. (1999). Activity theory: a wellkept secret. In *Perspectives on Activity Theory*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511812774

Ferstl, O. K., & Sinz, E. J. (2013). *Grundlagen der Wirtschaftsinformatik*. München: Oldenbourg.

Fettke, P. (2009). How conceptual modeling is used. *Communications of the Association for Information Systems*, *25*(43), 571–592.

Figl, K. (2017). Comprehension of procedural visual business process models. *Business & Information Systems Engineering*, *59*(1), 41–67. doi:10.1007/s12599-016-0460-2

Haisjackl, C., Soffer, P., Lim, S. Y., & Weber, B. (2018). How do humans inspect BPMN models: An exploratory study. *Software & Systems Modeling*, *17*(2), 655–673. doi:10.1007/s10270-016-0563-8 PMID:29706861

Hinkel, G., Kramer, M., Burger, E., Strittmatter, M., & Happe, L. (2016). An empirical study on the perception of metamodel quality. In *Proceedings of the 2016 4th International Conference on Model-Driven Engineering*. Academic Press. doi:10.5220/0005632001450152

Johannsen, F., Leist, S., & Braunnagel, D. (2014) Testing the impact of Wand and Weber's decomposition model on process model understandability. In *Proceedings of the International Conference on Information Systems (ICIS)*. Academic Press.

Karagiannis, D. (2015). Agile modeling method engineering. In *Proceedings of the 19th Panhellenic Conference on Informatics* (pp. 5-10). ACM. doi:10.1145/2801948.2802040

Karagiannis, D., & Kühn, H. (2002, September). Metamodelling platforms. In EC-Web (p. 182). Academic Press.

Karagiannis, D., Mayr, H. C., & Mylopoulos, J. (2016). *Domain-specific conceptual modeling*. Springer International Publishing. doi:10.1007/978-3-319-39417-6

Mendling, J., Recker, J., Reijers, H. A., & Leopold, H. (2019). An empirical review of the connection between model viewer characteristics and the comprehension of conceptual process models. *Information Systems Frontiers*, *21*(5), 1111–1135. doi:10.1007/s10796-017-9823-6

Mendling, J., Reijers, H. A., & van der Aalst, W. M. (2010). Seven process modeling guidelines (7PMG). *Information and Software Technology*, *52*(2), 127–136. doi:10.1016/j.infsof.2009.08.004

Op't Land, M., & Dietz, J. L. (2012). Benefits of enterprise ontology in governing complex enterprise transformations. In *Proceedings of the Enterprise Engineering Working Conference* (pp. 77-92). Springer. doi:10.1007/978-3-642-29903-2_6

Rossi, M., Henfridsson, O., Lyytinen, K., & Siau, K. (2013). Design science research: The road traveled and the road that lies ahead. *Journal of Database Management*, *24*(3), 1–8. doi:10.4018/jdm.2013070101

Sandkuhl, K., Fill, H. G., Hoppenbrouwers, S., Krogstie, J., Matthes, F., Opdahl, A., & Winter, R. et al. (2018). From expert discipline to common practice: A vision and research agenda for extending the reach of enterprise modeling. *Business & Information Systems Engineering*, *60*(1), 69–80. doi:10.1007/s12599-017-0516-y

Simões, D., Antunes, P., & Carriço, L. (2018). Eliciting and modeling business process stories. *Business & Information Systems Engineering*, *60*(2), 115–132. doi:10.1007/s12599-017-0475-3

Skyttner, L. (1996). General systems theory: Origin and hallmarks. *Kybernetes*, *25*(6), 16–22. doi:10.1108/03684929610126283

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. doi:10.1016/0959-4752(94)90003-5

Truex, D., Alter, S., & Long, C. (2010). Systems analysis for everyone else: Empowering business professionals through a systems analysis method that fits their needs. In *Proceedings of the European Conference on Information Systems 2010*. Academic Press.

Truex, D., Lakew, N., Alter, S., & Sarkar, S. (2011). Extending a systems analysis method for business professionals. In *Proceedings of the European Design Science Symposium* (pp. 15-26). Springer.

van der Aalst, W. M. (2012). What makes a good process model? *Software & Systems Modeling*, *11*(4), 557–569. doi:10.1007/s10270-012-0265-9

Wand, Y., & Weber, R. (1990). Toward a theory of the deep structure of information systems. In *Proceedings of the International Conference on Information Systems 1999. Academic Press.*

Wand, Y., & Weber, R. (2017). Thirty Years Later: Some Reflections on Ontological Analysis in Conceptual Modeling. *Journal of Database Management*, *28*(1), 1–17. doi:10.4018/JDM.2017010101

Wüest, D., Seyff, N., & Glinz, M. (2019). FlexiSketch: A lightweight sketching and metamodeling approach for end-users. *Software & Systems Modeling*, *18*(2), 1513–1541. doi:10.1007/s10270-017-0623-8

Zur Muehlen, M., & Recker, J. (2013). How much language is enough? Theoretical and practical use of the business process modeling notation. In *Seminal Contributions to Information Systems Engineering* (pp. 429–443). Berlin: Springer. doi:10.1007/978-3-642-36926-1_35

Steven Alter is Professor Emeritus at the University of San Francisco. His experience as vice president of a software start-up and author of four editions of a major IS textbook led to research focused on developing systems analysis and design methods that business professionals could use for their own understanding and to help them collaborate more effectively with IT professionals, consultants, and vendors. Most of his publications are related to the "work system method (WSM)," work system theory (WST), service systems, and extensions of WST such as a theory of workarounds and a service value chain framework.

Dominik Bork is currently working as a postdoctoral researcher at the Research Group Knowledge Engineering of the University of Vienna. Graduated with a diploma and a doctor rerum politicarum in information science from the University of Bamberg, Germany, he moved to Vienna in 2013. Dominik is heavily involved in the OMiLAB organization with a focus on international collaborations and metamodeling. Since 2018, he was elected domain expert of the Special Interest Group on Modelling Business Information Systems of the German Informatics Society. His research works focus on the foundations of metamodeling, the specification of modeling methods, and the conceptualization of domain-specific modeling methods. He was visiting researcher at the University of Technology, Sydney, the Instituto Tecnologico Autonomo de Mexico, and the University of Pretoria.

Journal of Database Management
Volume 31 • Issue 3 • July-September 2020

A Framework for Building Mature Business Intelligence and Analytics in Organizations

Amrita George, Marquette University, USA
https://orcid.org/0000-0002-6441-251X

Kurt Schmitz, Georgia State University, USA
https://orcid.org/0000-0003-1187-9068

Veda C. Storey, Georgia State University, USA

ABSTRACT

As activities are increasingly being digitalized in business and society, organizations have sought ways to effectively and competitively, use data. Business intelligence and analytics (BI&A) systems which support managerial decision-making continue to be developed and used. Given the importance of these systems, it would be useful to have a comprehensive and mature guide to support their development and improvement. This research proposes a BI&A Competitive Advantage Maturity Model to identify the main technical and non-technical dimensions of a system to support business intelligence and analysis. The model is based on work systems theory and related research. It maps descriptive characteristics of its main dimensions across analytic adoption stages of aspirational, experienced, and transformed. The development of the model employed a modified Delphi study technique, design science research, and citation analysis.

KEYWORDS

Analytics, Analytics Adoption Stages, Aspirational, BI&A Competitive Advantage Maturity Model, Business Intelligence, Citation Analysis, Content Analysis, Delphi Study, Descriptive, Design Science Research, Experienced, Maturity Model, Predictive, Prescriptive, Transformed, Work Systems Theory

INTRODUCTION

Business decisions, once the realm of intuition and anecdotal experience, are being reinvented by modern organizations using data. Business Intelligence and Analytics (BI&A) encompasses the technologies, systems, practices, techniques, and methods that analyze data to help organizations understand themselves, their markets and their customers, in order to make business decisions (Chen, Chiang, and Storey, 2012). Mature BI&A systems can be the basis for making effective decisions, improving performance, and exploiting new opportunities (Olszak, 2016). Organizations attempt to improve and expand their BI&A capabilities to advance their competitive position. There has been much press coverage, with exemplars such as Google and Amazon, as well as significant investment

DOI: 10.4018/JDM.2020070102

across industries. However, the majority of organizations (over 87% in one study) are considered as having "low maturity" (Chien 2018), and failing to reach their strategic goals (Grover et al., 2018).

Progressing from stand-alone analytics projects to a BI&A capability that drives decisions across an organization, requires more than big databases and powerful processing capabilities. Competitive advantage requires maturity across many dimensions: collection and analysis of data (Delen and Demirkan, 2013); sophistication of data consumers (Kiron and Shockley, 2011); alignment with organization strategies (Hribar Rajterič, 2010); and cultural commitment of an organization (Davenport, Harris, De Long, and Jacobson, 2001). However, current maturity models focus on specific components (Lahrmann et al., 2011; Foshay et al., 2015; Chen and Nath, 2018). Although various BI&A maturity models have been proposed (Larhmann et al., 2010), they often lack theoretical foundations, empirical validation, and the ability to operationalize maturity measurements (Lasadro, Vatrapu, and Andersen, 2015). Instead, they typically focus on improving technological, business-technical alignment, or analytical capabilities, with little emphasis on the main value proposition of competitive advantage (Chen and Nath, 2018).

Organizations use BI&A system to sense and respond to market needs by focusing on automation and cost reduction, optimization, profitability and exploring newer avenues (LaValle et al., 2011). While the aim is to be competitive, the purchase or development of a BI&A tool alone is not sufficient (Siow, Tiropanis, and Hall, 2018). Many questions arise about the organizational resources and capabilities that influence successful deployment of analytical capabilities (Kohli and Tan, 2016; Abbasi, Sarker, and Chiang, 2016). These include the: role of executive leadership and information technology (IT) departments; data needs and how they can be made available to stakeholders; organizational transformation from an intuition-based decision-making culture to a data-driven one; potential instrumental and humanistic outcomes; and main factors influencing business value for real-time decision making using big data. Organizational resources and capabilities have an immense influence on the success of using a BI&A system to gain competitive advantage. Therefore, to derive value from a BI&A initiative, an organization needs to ensure that its initiative is aligned with its organizational context.

Alignment involves understanding under what conditions organizational actors are able to generate insights using BI& A systems, although this topic has not been well-explored (Günther et al., 2017). It necessitates the need to understand the specific contributions of human and algorithmic intelligence to gain insights for different situations (Günther et al., 2017). Examples are cost reduction, operational efficiencies, or using BI&A for emergent or temporal situations. Although many capabilities have been proposed that enable accessing, tracking, collecting, managing, governing, processing, and analyzing big data for data-driven decision making and implementation, organizations still need to identify how to effectively develop, mobilize and use the technical and human resources related to big data in a structured fashion (Günther et al. 2017). Organizations face questions regarding how to acquire or develop technical and human resources, and how to structure teams or departments (Günther et al., 2017). Deploying and structuring organizational resources alone will not enable value creation using BI&A. An understanding of the capabilities of an existing BI&A system for achieving a given objective is needed, as is an understanding of the actions required to achieve a high level of effectiveness (Rajterič, 2010).

The objectives of this research are to: develop a maturity model for BI&A that: defines the capability dimensions needed for data-based decision making; and identifies the characteristics needed at different maturity capability levels to achieve competitive advantage. To do so, relevant literature on BI&A is reviewed, and a Delphi Study conducted from which a Business Intelligence and Analytics (BI&A) Competitive Advantage Maturity Model is derived. Then, a content analysis is applied to improve the validity and robustness of the proposed model. The contribution is the development of a BI&A Competitive Advantage Maturity Model that can help organizations recognize and measure their maturity profile, while providing prescriptive advice for progressing BI&A to advance a firm's competitive position.

RELATED RESEARCH

Maturity models help organizations understand the effectiveness of their current information systems, expose gaps, and guide improvements (Paulk et al., 1993). This section reviews research on identifying different levels of business intelligence and analysis maturity within organizations.

Levels of Maturity

The concept of maturity recognizes that proficiency exists in degrees across a continuum from low to high. Considering infinite granularity across a continuum is unwieldy and led to maturity models that stratify a proficiency spectrum into a manageable number of stages. A recognized example is the Capability Maturity Model Integration (CMMI) for software development, which stratifies maturity into five levels: Initial, Managed, Defined, Quantitatively Managed and Optimizing. For BI&A, Larhmann et al. (2010) document 10 distinct business intelligence maturity models, ranging from three to six levels, with five being the most popular.

Table 1 shows representative BI&A maturity models that have emerging since 2010, revealing that the five levels of maturity remain popular. One criticism of information systems maturity models

Table 1. BI&A maturity models since 2010

Model	Levels	Dimensions/ Components	Source
Three Stages of Analytics Adoption	1. Aspirational 2. Experienced 3. Transformed	• Motive • Functional Proficiency • Business Challenges • Key Obstacles • Data Management • Analytics in Action	LaValle et al, SMR 2011
TDWI: Analytics/Big Data BI Maturity Model	1. Nascent 2. Pre-Adoption 3. Early Adopters 4. Corporate adoption 5. Mature visionary	• Organization • Infrastructure • Data Management • Analytics • Governance	Halper & Stodder, 2015
SAS: Analytic Maturity Scorecard	1. Unaware 2. Aware 3. Astute 4. Empowered 5. Explorative	• Culture • Internal Process Readiness • Analytical Capabilities • Data Environment, Infrastructure & Software	SAS Whitepaper, 2014
Gartner: Maturity Model for Data & Analytics	1. Basic 2. Opportunistic 3. Systemic 4. Differentiating 5. Transformational	• Strategy • People • Governance • Technology	Chien, 2018
HP: Enterprise BIMM	1. Operation 2. Improvement 3. Alignment 4. Empowerment 5. Transformation	• Business enablement (business need and problem) • Information technology • Strategy and program management (nature of management skill)	Hewlett Packard, 2009

is their inability to operationalize maturity measurement (Lasadro et al., 2015). Five or more levels may provide a useful impetus for the sale of the next wave of business intelligence products and services. However, the three-level "analytic adoption stages" stratification (Aspirational, Experienced, Transformed) proposed by LaValle et al. (2011) has conceptual clarity, can rapidly be applied for

self-assessment, and has been applied successfully in thousands of cases by analysts, managers and executives (Kiron and Shockley, 2011).

LaValle et al.'s (2011) three-stage approach empirically demarcates Business Intelligence & Analytics effectiveness with respect to achieving competitive advantage (cost leadership, product differentiation, customer focus/market segmentation), providing an appropriate stratification for our proposed BI&A Competitive Advantage Maturity Model. These stages segment organizations based upon their analytic prowess into three levels: aspirational (farthest from achieving their analytical goal with focus on efficiency and automation of existing processes while searching for cost reduction opportunities); experienced (closer to achieving their analytical goal and use their initial analytical experience to go beyond cost reduction opportunities and emphasize ways to dynamically optimize their organization as circumstances evolve); and transformed (leaders in their domains with substantial experience in analytics and focus on driving profitability and making targeted investments for future growth).

Dimensions of Maturity

In addition to classifying levels, or stages, of maturity, a prescriptive maturity model identifies domains that can be characterized and benchmarked (Lasrado et al., 2015). Commonly labeled dimensions, these domains represent aspects of capabilities that develop and operate as an entity. Each dimension may stagnate, regress, or advance as a collective entity, distinct from other dimensions. Common among business intelligence maturity models is the identification of dimensions for technology (or tools), data (or data management), and processes. A wide variety of other dimensions exist. Foshay et al.'s (2015) study of business intelligence maturity models identified dimensions such as organizational (analytical processes, organization structure, governance and cost-benefit), human (skills, training, sponsorship, culture) and technical (infrastructure/tools, data architecture). Larhmann et al. (2010) discovered a somewhat uniform spread between zero and nine dimensions among models developed prior to 2010.

The collection of dimensions is diverse, but none capture all the relevant characteristics for assessing competitive advantage. The proper identification of dimensions, however, is critical for assessing competitive advantage positioning, because a weakness in one (possibly hidden) area can undermine the overall value of a BI&A system. Delen and Demirkan (2013) found that BI&A effectiveness depends on the capability of analytical tools, the quantity and quality of data, and the organization's procedures. Moving from one BI&A maturity level to another requires changes in all characteristics that establish categorization within a particular level (Olszak 2016). Despite the importance of broad coverage to expose areas that would otherwise be easily overlooked, existing models are incomplete, focusing on a specific point of view or problem domain (Chen and Nath, 2018; Hribar Rajterič, 2010).

Instead of anecdotal insights, building on a theoretical foundation provides a better way to capture the full breadth of BI&A competency domains (Lasrado et al., 2015). The Work Systems Theory (Alter, 2006), which was derived in collaboration with practice over the course of a decade, identifies a comprehensive set of key elements of an effective work system. Work System Theory provides a theoretical framework to understand sociotechnical work systems where humans interact with information systems to perform work and produce value. The framework provides a comprehensive cataloging, from core (information, technologies, participants and processes and activities) to peripheral, of all elements applicable to any work system. When applied to BI&A systems, these elements provide a rigorous accounting of BI&A dimensions, subsuming those identified in previous work. Work Systems Theory elements map to an organization's BI&A dimensions as follows.

Information (Core)

Data is considered the raw material of the 21st Century (Elgendy and Elragal, 2016). Delen and Demirkan (2013) note that BI&A effectiveness depends on data quality, including its accuracy,

integrity and timeliness. There are many challenges related to characteristics of the data itself, including volume, variety, velocity, veracity, volatility, quality, discovery and dogmatism (Sivarajah et al., 2017; Storey and Song, 2017). These evolve as organizations mature (LaValle et al., 2011). Low maturity organizations have limited ability to capture, aggregate, and share information, whereas high maturity organizations are capable of collecting large volumes of data in real time and sharing that data across the organization. Mature organizations collect, manage and use both structured and unstructured data.

In this research, Information encompasses both codified and non-codified data that is captured, created, used, transmitted, stored, retrieved, and manipulated within the core BI&A system to create BI&A products and services. Examples include structured data in isolated form or collected in tables, databases and data-marts, and unstructured data needing codification or interpretation.

Product/Services

BI&A products and services are the transformed output of the core BI&A system, suitable for decision-making. Prominent examples take the form of monthly and yearly reports associated with marketing and finance departments (Trkmann et al., 2010). Delivery often takes place through a visual representation of information in the form of one or more charts, graphs, or diagrams. These representations may be further enhanced by allowing the user to dynamically manipulate the presentation to focus on granular ideas, observe broad generalizations, or specify temporal scope. While the capability to produce an information representation stems from the maturity of the BI&A technologies, the use case and utility of the information product characterizes the maturity of BI&A products and services. Mature BI&A products and services are dynamically interpreted as they empower the BI&A customer to control and manipulate the artifact using techniques such as drill-down and what-if scenarios. The dynamic nature of some BI&A products and services include real-time response to environmental changes (e.g., a digital map on a smartphone updating as the user drives). Limited maturity BI&A systems are suitable for individual departments addressing structured and operational decisions. More mature BI&A systems produce products and services supporting continuous use for day-to-day decisions within multiple business functions, predictive and prescriptive decisions that span across functions, and increasingly focus on the future (Kiron and Shockley 2011).

For our research, Products and Services include information and knowledge produced by the BI&A system that can be used by its customers to perform business tasks and decision making. Examples include charts, diagrams, dashboards, interactive GIS maps and search engines that recognize and respond to past preferences.

Customers

BI&A Customers are the recipients and users of BI&A products and services. In their survey of 105 firms, Lismont et al. (2017) found that analytics is mainly used by marketing (85%), finance (77%) and operations (74%) functions. Whereas some self-service BI&A systems place individuals in the dual role of Participants and Customers, these roles are distinct. An individual can perform the role of a Participant when involved in the generation and creating a BI&A product. That same individual can perform the role of a Customer while using a BI&A product or service as part of the business decision making activity. Characteristics of a customer relevant to BI&A maturity involve the customer's competency to understand and use BI&A products, appreciating the power and limitations of various statistical and analytical techniques that underly specific BI&A products, the span of decision making authority of the customer (operational within a single function, spanning multiple functions, or broad across the organization), and their competence translating BI&A information into actionable decisions.

In this research, Customers are the users and decision-making beneficiaries of BI&A systems. Included are internal and external stakeholders who apply information and knowledge produced by the BI&A system to business tasks and decision making. Examples include operational staff performing routine tasks, management addressing semi-structured problems, and senior management grappling with unstructured problems.

Participants (Core)

Data does not easily transform itself into information and knowledge. Information Technology and business intelligence specialists play a central role in implementing BI&A systems, and then collecting, managing, and shepherding the transformation of data into BI&A products and services. In their study of 159 IT managers, Fink et al. (2017) recognized that an effective Business Intelligence team needs, not only technical capabilities and knowledge of BI&A systems, but also an understanding of the business and its requirements. They further conclude that highly skilled and knowledgeable participants seldom exist in organizations with low BI&A maturity. Sivarajah et al. (2017) note that human expertise and talents required to leverage big data lag behind advances in computing technologies. Big data depends on a range of roles brought together in a fluid, network relationship that requires organizations to adapt dynamically (Braganza et al. 2017).

In this research, Participants include all stakeholders who are directly involved in creation and operationalization of the BI&A system to produce BI&A products and services used by BI&A consumers. Examples include the core IT team such as programmers, database managers, report writers, as well as non-IT participants who contribute to collecting data, turning data into information and delivering that information to customers and decision makers.

Technologies (Core)

Technologies, as tools and techniques, are perhaps the most studied dimension of BI&A. In their basic form, BI&A technologies include a software component, a hardware component and the platform that facilitates data flow (Chaudhuri, Dayal, and Narasayya, 2011; Kromer et al. (2014). Gandomi and Haider (2015) identify two relevant subsets for BI&A tools: data management and analytics. Data management technologies deal with the acquisition, recording, extraction, cleaning, annotation, integration, aggregation, representation and management of data in both structured and unstructured forms. Analytic tools include modeling and analysis, visualization, interpretation, and the expanding area of prediction.

In this research, Technologies are specific technical tools and techniques instantiated in those tools to perform data management and analytic tasks. Examples include database management systems, data warehouses, structured query language, extract-transform-load tools, statistical analysis tools, charting and visualization tools, portals for delivering data and advanced techniques such as machine learning (Hyder, Siau, and Nah, 2019).

Processes and Activities (Core)

Data undergoes a series of transformational steps to create information and knowledge products suitable for business tasks and decision making. Historically, these processes were performed by hand or in the mind of an analyst or decision maker. Although information systems now perform many of the transformational steps, these steps still must be imagined, documented and performed. Process challenges are described in terms of *how* techniques: how to capture data, how to integrate data, how to transform data, how to select the right model for analysis, and how to provide the results (Sivarajah et al., 2017). The data transformation process is an important step in the dissemination of knowledge from data (Davenport et al. 2001). Gandomi and Haider (2015) document basic concepts related to big data, showing that organizations need efficient processes to turn high volumes of fast-moving and diverse data into meaningful insights. Braganza et al. (2017) discovered that processes need to change dynamically in response to, or in anticipation of, both internal and external influences. At the transformational end of value creation, Shollo and Galliers' (2016) case study of a Scandinavian financial institution reveals the pivotal role of articulation and data selection processes in organizational knowledge creation and the emergence of new insights from data. The effectiveness of an organization's BI&A system depends on capabilities of the organization's procedures (Delen and Demirkan, 2013).

In this research, Processes and Activities encompass the series of actions or steps occurring in the BI&A system to transform data into information products and services. Examples include algorithms, access policies, security mechanisms, as well as the steps needed to deliver and disseminate business intelligence to decision makers.

Infrastructure

Infrastructure is technological in nature, but represents a different dimension of capability than the tools and techniques discussed above. The distinction is rooted in the scope of purpose of the technologies. Broad general-purpose technologies, such as electricity, networks and cloud platforms, are available for all types of systems, including those unrelated to BI&A. Although these technologies may be essential components of a BI&A system, they are not managed for the focused purpose of creating and supporting a BI&A system. Rather a BI&A system is one of many subscribers to technical infrastructures. From this perspective, Infrastructure includes certain platform services used to host BI&A technologies (including those involved in collecting, integrating, sharing, processing, storing and managing BI&A data), many sources of data (e.g., social media and external database), and major open-source data management frameworks (e.g., Apache Hadoop, Mahout, Spark, and Storm) (Grover et al., 2018). In a broad sense, infrastructure also includes the training and education systems that prepare BI&A participants for their careers, conferences that educate strategic leaders to employ BI&A, and even a journal that facilitates the distribution of BI&A studies. Although similar infrastructure is available to most firms, individual organizations select combinations of infrastructure unique to themselves as they make decisions between vendors (AT&T vs Verizon; Azure vs. AWS; Oracle vs. Teradata). The portfolio of infrastructure employed by a BI&A system may be more or less mature than that employed by a competitor.

In this research, Infrastructure includes shared human, informational, and instances of technology used by the BI&A system, but managed by entities outside of the core BI&A work system. Examples include network providers, education providers, cloud platform providers, social network platforms, human resource staffing providers, and even the power company.

Environment

Many studies recognize the importance of the environment in which BI&A systems are employed. Davenport et al. (2001) found that contextual elements, including organization and culture, can influence the success of analytical capabilities. LaValle et al. (2001) asserts that many organizations struggle to derive value from their BI&A systems due to misalignment between the BI&A system and organizational context. Kiron and Shockley (2011) found that the most advanced users of analytics typically have a strong data-oriented culture that supports and guides analytics use. Without strong cultural commitments, the success of an analytics program can be shortchanged or derailed.

In this research, Environment is defined as the organizational, cultural, competitive, technical, regulatory, and demographic aspects of the surroundings in which the BI&A system operates. Examples include the decision-making culture (intuition-based or evidence-based), the growth and investment climate of the firm and its industry, and the business cycle (e.g., recession or expansion) affecting customer decisions.

Strategies/Strategic Alignment

Several studies highlight the role of strategic alignment. O'Leary (2013) notes that the success of BI&A depends on how information is utilized for the benefit of the organization. In their case study of a UK retail bank, Audzeyeva and Hudson (2016) found that organizational links that embedding the Business Intelligence into a deep structure, if not adequately aligned, will obstruct the delivery of BI&A benefits. They conclude that alignment is more than an achievement within an organization; it is a dynamic capability to detect and respond to changes in an evolving business environment. Grover et al. (2018) observe that successful big data analytics depends on its inclusion in a firm's long-term

business strategy, with mechanisms to facilitate alignment with this strategy, where alignment extends beyond business strategy, to processes, policies, procedures, organizational structure/governance, and corporate culture.

Here, Strategic Alignment is defined as the dynamic BI&A capability alignment across dimensions and with organizational, business, and operational level strategies. Alignment functions as an overarching capability that ties the other individual dimensions together and magnifies their competitive advantage influence (or lack thereof).

Juxtaposing Alter's work system elements (Alter, 2006) with the analytic adoption stages proposed by LaValle et al. (2011) establishes a comprehensive structure that covers all dimensions of BI&A maturity. These levels and dimensions provide the foundation for addressing our objective of identifying defining characteristics at different maturity capability levels that are needed to achieve competitive advantage. At the intersection of dimension and level are specific indicators that allow an organization to be categorized. These characteristics also expose prescriptive changes needed to progress to higher BI&A maturity.

RESEARCH METHODOLOGY

The proposed Business Intelligence and Analytics (BI&A) Competitive Advantage Maturity Model is a research artifact (Baskerville et al., 2018; Baskerville et al., 2015; Rossi et al., 2013; Prat et al., 2015) developed following a design science approach (Hevner et al., 2004). Table 4 in Appendix A summarizes the development of this artifact.

DERIVATION OF BI&A MATURITY MODEL

The structure of the proposed model (levels and dimensions) was derived from the literature on the Work System Theory and Analytic Adoption Stages. To identify the characteristics at the intersection of each level and dimension, an iterative definition and evaluation approach was used. Each iteration identifies weaknesses, areas for improvements, side effects, efficacy, and utility until a suitable artifact emerges. From this, the ex-ante BI&A structure emerged as an operationalizable model in a naturalistic setting (Venable, Pries-Heje, and Baskerville, 2014).

The development and validation of the model were carried out in two phases. Phase 1 employed a Nominal Group Technique engaging a focus group of BI&A practitioners and experts. Specifically, a Delphi study approach was applied using multiple rounds of input, documentation, review, feedback and revision. This phase established characteristics and indicators for each dimension at each level. An initial BI&A Maturity Model emerged during this process as the focus group opinions converged. The draft model then progressed to Phase 2 for validation using content analysis. Content analysis of appropriate articles and relevant research infuses the collective attitude and intentions (Duriau, Reger, and Pfarrer, 2007), adding validity and robustness to the model. Recent articles were selected, coded and analyzed to validate and, where appropriate, adjust the characteristics applicable to each dimension and level. Since the maturity model should be easily operationalized by practitioners, publications that intentionally embrace practitioners were targeted.

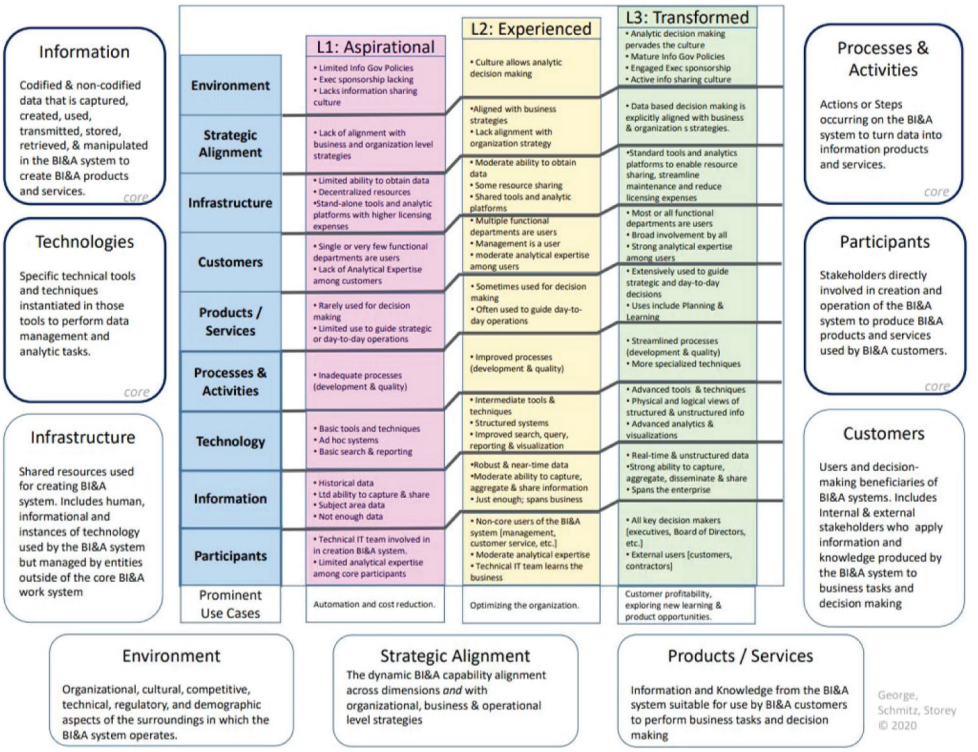
The resulting BI&A Competitive Advantage Maturity Model is presented in Figure 1.

Phase 1 – Establish Characteristics and Indicators Using Delphi

Focused Group Approach

Characteristics and indicators were introduced and refined using a modified Delphi technique (Hasson and Keeney, 2011). This is a structured, multi-pass group discovery process that solicits opinions from experts. The goal is to use expert opinion to ratify the level and dimension structure, and then define appropriate maturity characteristics.

Figure 1. BI&A competitive advantage maturity model



Members of an expert panel were selected from both the researcher and practitioner communities, in an effort to obtain theoretical and practical insights. Participant recruitment took place in a series of evening professional association meetings in a large city. Researcher participants with both practice and scholarly backgrounds were recruited. Nine participants (5 researchers & 4 practitioners) participated (Table 2).

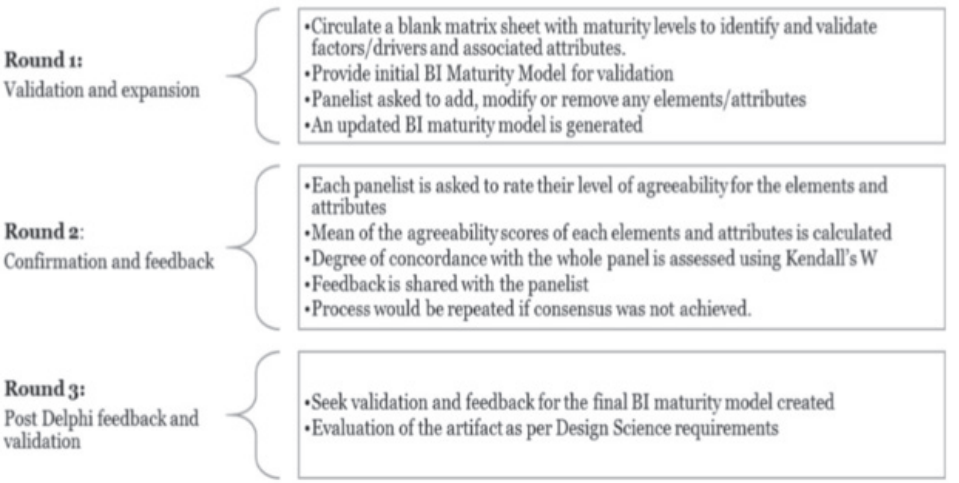
Table 2. Participant profile

Degree of highest education	Business (5 participants), engineering (4 participants)
Total experience in years (range)	5 years to 25 years
Experience with Business Intelligence in years (range)	2 years to 16 years
Industry	Academic Institution, Aviation, IT Service, Marketing Analytics, Retail, Telecommunication

A three-round iterative process was used. The purpose of the first round was to ensure that all dimensions and attributes were identified and correctly assigned. A blank sheet of paper was circulated in which participants were requested to identify the key factors they perceived would affect the stated level of maturity (aspirational, experienced and transformed). They were also asked to assign attributes/values to the factors identified. This round contributed in two ways. First, it helped validate the list of dimensions and attributes given in the proposed model. Second, it helped rule out the presence of bias in the scoring process of the proposed model.

In round 2, the matrix derived from a synthesis of the literature (Table 5 in Appendix B) was circulated for scoring each element and its associated attributes/indicators. Following the Delphi process, the participants scored the matrix privately, with the results aggregated for dissemination and subsequent review by the panel members. Feedback from each round serves as input to successive rounds to achieve concordance amongst the expert group. The final round involved validation of the model components and the artifact as a whole. The participants were encouraged to actively provide recommendations for improving the artifact. The three rounds of the modified Delphi approach are shown in Figure 2.

Figure 2. Modified Delphi technique undertaken



Focused Group Results

A word cloud was created to understand the importance of various factors identified in round 1. Factors influencing each maturity level differed across levels (Figure 3). For example, Training and Top management support emerged as the top two factors for the aspirational level. Critical success factors for organizations at this level align with the common use cases (using analytics to justify actions such as cost reduction/leadership and automation). At the experienced level, the use cases begin to embrace dynamic optimization of the organization as analytics is applied to guide actions and meet market demand in a timely fashion. The top factors were competitive pressure, training, and strategic planning. At the transformed level, organizations focus on customer value, new products, and competitive advantage from market segmentation and/or customer focus, with use cases that are transitioning to prescriptive decision making and organizational learning. The main factors identified at this level were data driven decision-making and strategic planning. The findings from round 1 supported the need for different levels in the proposed model and ratified the three-level structure as suitable for assessing competitive advantage.

In round 2, the maturity level and dimension structures were formally rated along with assignment of characteristics and indicators to specific dimensions. SPSS (Statistical Package for Social Sciences) was used to compute Kendall's coefficient of concordance (W) as a measure of agreeability amongst participants' ratings. The average agreeability scores for each Maturity Dimension with respect to a particular Maturity Level was moderate to high (3.5 to 4.5 out of 5). Kendall's coefficient of concordance was computed for each level of maturity, and as a whole (Table 3). With a score of .7 or above for W signifying strong agreement (Skinner, 2015), a moderate level of concordance was

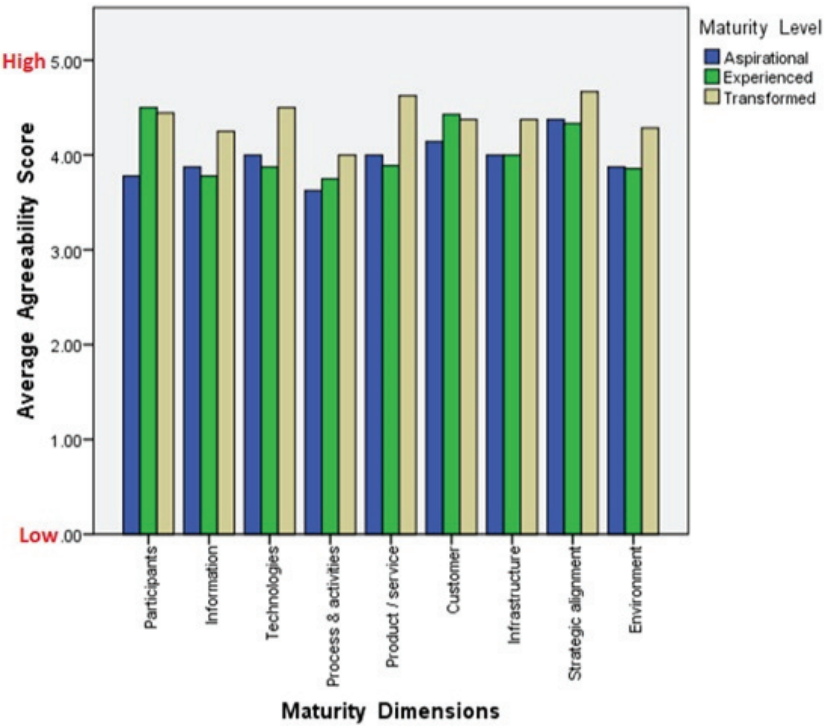
Figure 3. Word cloud of factors identified in round 1

Table 3. Inter-rater agreement (Kendall's coefficient of concordance) in round 2

Maturity Level	Kendall's Coefficient of Concordance *W*	Chi-Square	Df
Aspirational	0.679	10.857	8
Experienced	0.411	9.869	8
Transformed	0.555	8.873	8

observed for aspirational and transformed maturity levels, whereas a low level of concordance was observed for the experienced maturity level. The average agreeability scores reported by participants for each maturity dimensions with respect to the maturity level was also moderate to high (Figure 4), signaling concordance with the dimensions identified in the proposed model.

Figure 4. Average agreeability scores reported by experts for each maturity dimension against the maturity levels in round 2



In round 3, participants reviewed a draft of the BI&A Competitive Advantage Maturity Model to assess the alignment of characteristics and revise the attributes and indicators of each dimension and level. This round also gathered recommendations for improving the model. The feedback included:

- Levels need to be independent of each other and the progression non-linear. However, maintaining a sequential fashion is practical as organizations move from one stage to the next (Kart 2015) building on existing systems, rather than entering at a high level (Kart 2015);
- A base level is necessary to capture organizations entering an initial phase;
- Certain terminology in the maturity characteristics (e.g. structured and unstructured) needs to be stated using common vocabulary for use by technical and non-technical staff.

Since the feedback arrived during Phase 1, the authors adjusted the model and prepared information for subsequent iterations. For example, from early feedback, the authors collected a set of candidate characteristics and indicators described in the practitioner literature (Table 5, Appendix B). Seeding the process with these ideas, and then collecting feedback from practitioners helped ensure that the vocabulary and interpretation are suitable for operationalization in real organizations.

Phase 2 – Validation Using Content Analysis

Validation of the model proceeded using Content Analysis. Content analysis of practitioner articles and relevant research improves the robustness of the model by identifying the collective attitude and intentions (Duriau et al. 2007) of practitioners and researchers for each elements of a successful BI&A system.

Literature Search

We searched multiple databases, including Google scholar and the Gartner library. The terms employed were: "descriptive analytics" AND "predictive analytics" AND "prescriptive analytics" as well as "aspirational" AND "experienced" AND "transformed" AND "analytics".

Inclusion Criteria and Exclusion Criteria

Included articles, first, had to contain some aspect of our target analytical levels (i.e. descriptive / prescriptive / predictive OR aspirational / experienced / transformed). Second, articles needed to include elements in the Work Systems Theory. This embraced both technical and non-technical dimensions. Third, the study had to describe one or more dimensions ratified during the Delphi phase to verify dimension characteristics that are distinctly suitable for each maturity level. Fourth, the studies had to be written in English. Older studies were excluded to ensure the characteristics were appropriate for contemporary competitive advantage.

Our search narrowed to 5 articles, which helped to diversify the characteristics vocabulary established during phase 1. For example, Nguyen et al. (2018) was selected because it captured a broad cross-section of 88 supply chain related studies. Siow et al. (2018) captured characteristic dimensions from 311 Internet of Thing related studies. The remaining articles were review articles with characteristic dimensions from a large number of Business Intelligence studies. Despite the modest quantity of articles, this content broadly captures the collective attitude of practitioners and researchers during this particular era of competitive advantage.

Content Analysis Coding

Selected articles were coded by dimension (from the BI&A model, Figure 1) to reflect the degree of alignment of characteristic properties to each target level. For example, when the literature confirmed that customers included Senior Managers for a particular strategic decision use cases at high maturity organizations, this received a score of positive alignment with our model. When the literature contradicted the model (for example, by suggesting that customers included Senior Managers for extensive use cases at low maturity organizations), this received a negative alignment score.

The authors, who had training on the coding practices and constructs of this study, coded the articles. The derived model with dimensions and characteristics provided the coding categories. Each article was coded by two authors. To prevent errors, the coders reviewed all the data and computed the Inter-rater reliability for each article. When the coding diverged, the coders reviewed the original article until they agreed (Orwin and Vevea, 2009).

Content Analysis Results

The agreement between coders was 0.70 – 0.81 (Cohen Kappa). Based on two pilot studies where the coders did not agree, the coding process was clarified by analyzing representative quotes. No new themes emerged from the phase 2 analysis. This high inter-rater reliability score confirmed that the coders agreed on the mapping of article content to the constituent pieces of the proposed model.

This was an iterative exercise with the team progressing through the articles by domain (Supply Chain, Internet of Things) and then the Literature Review articles. Before an article was added to the summary table, the inter-rater reliability among two coders was confirmed and disagreements resolved. As articles were added, characteristics and indicator phrases were adjusted to more effectively

represent attributes in a language that is operationally meaningful across all settings. For example, language applicable to Supply Chain differed from that used for the Internet of Things. As the content analysis progressed iteratively, the vocabulary adapted to being inclusive.

The content analysis confirmed that the 9 dimensions from Work Systems Theory were comprehensive. No new dimensions were suggested. Furthermore, by systematically mapping findings in the literature to each dimension/level combination, there was specific evidence supporting each indicator attribute identified by the experts in the Delphi study.

DISCUSSION

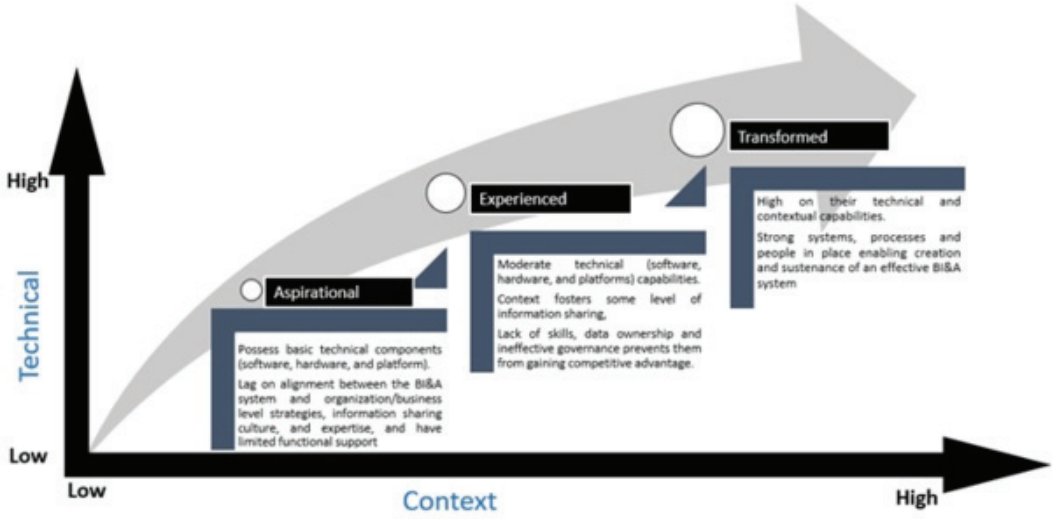
This research has developed a comprehensive Business Intelligence and Analytics Competitive Advantage Maturity Model. The model was derived by systematically examining the Business Intelligence & Analytics literature. A modified Delphi technique (Hasson and Keeney 2011) was used to identify weaknesses, areas for improvements, efficacy, and utility of the model. A content analysis integrated the collective attitude and intentions (Duriau et al. 2007) of practitioners and researchers for each element of a successful BI&A system, to validate the model's robustness.

This study contributes to the BI&A literature by proposing a model that identifies the main technical and non-technical dimensions with defining characteristics for analytical capability levels. Although many maturity models for Business Intelligence & Analytics systems have been devised, each focus on a specific aspect of maturity, rather than embracing the full landscape of important dimensions (Chen and Nath 2018). Moreover, an organization's use of Business Intelligence and Analytics to achieve competitive advantage has been overlooked. The proposed BI&A Competitive Advantage Maturity Model seeks to provide prescriptive advice for advancing maturity capabilities to lead to improved decision making that supports a competitive position.

The BI&A Competitive Advantage Maturity Model is intended to: (1) provide a comprehensive understanding of the technical and non-technical dimensions of an effective BI&A system; and (2) develop a Competitive Advantage focused Business Intelligence Maturity Model artifact that organizations can use to assess the effectiveness of their existing BI&A system(s). The model identifies and defines maturity characteristics for varying analytical levels that influences the effectiveness of business intelligence. Analysis of the characteristics of each dimension shows that aspirational organizations (focused on cost reduction and automation) rate low with respect to their technical and contextual capabilities. They lag with alignment between the BI&A system and the organization/business level strategies, an information sharing culture, and expertise, and have limited expertise and functional support. Experienced organizations (focused on dynamic optimization) achieve improvements across each work system dimension. Their context fosters some level of information sharing, but yet has limitations across one or more dimensions. Transformed organizations (focused on customer profitability, learning and exploring newer avenues) are high on their technical and contextual capabilities. They have systems, processes, and people with strong abilities, thus enabling the creation, sustenance, and use of an effective, mature BI&A system. Figure 5 depicts how organizations, at various adoption stages, perform against the technical and contextual elements of the BI&A system.

This study contributes to the competitive advantage literature by describing the contemporary influence of Information Systems to achieve a competitive position at various levels of sophistication. The proposed BI&A Maturity Model focuses on competitive advantage, generally considered a temporary phenomenon. As such, the characterizations for each maturity level established and verified by this study reflect a maturity assessment during a particular era. For example, there was a time when the use of spreadsheets would place an organization at the cutting edge of BI&A technical capability; spreadsheets now represent the lowest level of BI&A technical maturity. As technology advances, capabilities among competing companies in an industry advance across all dimensions. Whereas the levels and dimensions of the BI&A maturity model are robust across time and setting,

Figure 5. Influence of context and technology on analytic adoption stages



the characteristics that define positioning in one level or another inherently drift as paradigm shifts transform a dimension.

This study contributes to the maturity model literature by proposing a comprehensive framework for assessing the maturity of any given information system. It is appropriate to consider the resilience of the proposed model against the implications of paradigm shifts. For the technology dimension, Cloud Computing is an example of a paradigm shift in the Infrastructure domain. During the early years of the internet boom, organizations had limited choices to acquire computing resources. Although third-party hosting locations existed, network speed and bandwidth limitations drove many organizations to establish their own data centers for their BI&A systems. As the Internet matured and cloud computing became economically competitive, many organizations shifted their BI&A computing from dedicated in-house BI&A technologies, to shared service infrastructures. This led to a shift in skill required by BI&A participants, who no longer had the same need for data center expertise. Instead participants required skills in contract management and properly engaging the redefined infrastructure domain.

Similarly, Blockchain represents a potential technology paradigm shift. One aspect of Blockchain is the democratization of transactional data along with an environment of transparency and a culture of trust. The decentralization of transactional data may influence the characteristics identified for various dimensions in the proposed model. Most likely, new infrastructure platforms and participant skills will emerge as Blockchain matures and diffuses. These, in turn, will affect the characteristics that define cultural maturity within the environment dimension. The wide availability of transparent data distributed on nodes worldwide will alter the types of BI&A products and services, customers, and use-cases to which mature BI&A products are applied.

Another similar paradigm shift is underway in the Information dimension as the Internet of Things propagates data generation from a massively expanding universe of data sources. Effectively harnessing Internet of Things data volumes will propel the lessons from "big data" initiatives to an entirely different level. Infrastructures will need to manage orders of magnitude more data. Processing demands may move from central cloud infrastructures, to where data is generated, for decision-making in real time.

The dimensions that provide the structure for the model, are grounded in Work System Theory, and the levels in the maturity model literature. They are resilient to the change from paradigm shifts. As validated in our multi-phase study, these provide an exhaustive accounting of the dimensions that

constitute a BI&A work system. These are the levels suitable for an organization to understand its competitive advantage position and identify guiding steps toward greater maturity. Characteristics that define low versus high maturity must adapt to reflect a new era of competitive advantage associated with paradigm shifts. Our model should be useful for defining maturity in any era, recognizing the inherent need for periodic updates to the indicator characteristics.

LIMITATIONS AND FUTURE RESEARCH

The effectiveness of the proposed model for assessing areas of improvements will vary across domains. The model should be able to provide a guideline for organizations that undertake a BI&A initiative to identify gaps in their existing implementations. The model could be applied, for example, using the topic-based solution recommender system (Muller et al. 2016) in conjunction with McLaren et al.'s (2011) multilevel strategic fit measurement model. However, every business domain is unique. In healthcare, information access is restricted to care-takers as governed by various statutes and regulations. The applicability of information sharing characteristics to achieve a transformed level is therefore constrained by regulatory and ethical considerations. Further work should establish industry-specific filters to account for practical and regulatory limitations. Such filters would allow situational specific understanding of gaps and pathways to higher maturity.

This research focuses on providing a comprehensive overview of maturity, but not a concise instrument for objective self-assessment, thus constraining the use of this model for self-assessment to organizations that gather data across many dimensions. The result is a daunting challenge because no single department championing such an assessment will possess the expertise and situational awareness needed across all work system domains. Future research should focus on developing a toolkit to enable organizations to use the model for internal assessment and industry certification.

New entrant organizations that have simple or no business intelligence systems are difficult to rate on the BI&A Competitive Advantage Maturity Model. Other maturity models provide a basic level for organizations with no systems. However, our model is intended to aid organizations in understanding the inherent weakness in their existing systems that hindered their progress to gain competitive advantage. It provides a prescriptive approach to progress across established levels of analytical capabilities but lacks the ability to provide domain-specific recommendations for new entrants.

The model is limited when addressing questions from prior research. Our model is intended to be used to address questions such as: How can BI resource allocation and resource orchestration improve BI effective use? What role does c-level leadership play in firms' abilities to leverage and 'compete on big data analytics'? What is the role of IT departments in supporting big data analytics? What data needs to be captured and how can it be made available to stakeholders across the organization? What types of organizational resources are required for successful analytics capabilities (Trieu 2017; Kohli and Tan 2016; Abbasi et al. 2016)? Other questions remain, such as: How do organizations transform from an intuition-based decision-making culture to a data-driven decision-making culture? How does analytical leadership emerge? What are the characteristics of analytical executives? What factors influence business value within the context of real-time decision making using big data? How do we assess big data initiatives when considering relevant stakeholders and potential instrumental and humanistic outcomes (Trieu 2017; Kohli and Tan 2016; Abbasi et al. 2016)? These provide topics for future research.

CONCLUSION

This research has derived a Business Intelligence and Analytics (BI&A) Competitive Advantage Maturity Model. The model is intended to provide a complete accounting of technical and

non-technical dimensions suitable for assessing the BI&A competitive advantage potential of a firm. Its dimensions map to analytic adoption stages. The model was validated using a modified Delphi technique and content analysis of the literature at the intersection of scholarship and practice. Further work is needed to apply the model to a variety of real-world case studies and applications.

REFERENCES

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), I–XXXII. doi:10.17705/1jais.00423

Abzaltynova, Z. & Williams, J. (2013). Developments in Business Intelligence Software. *Journal of Intelligence Studies in Business,* 3.

Alter, S. (2006). *The work system method: connecting people, processes, and IT for business results*. Work System Method.

Alter, S. (2013). Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future. *Journal of the Association for Information Systems*, *14*(2), 72–121. doi:10.17705/1jais.00323

Arnott, D., Lizama, F., & Song, Y. (2017). Patterns of business intelligence systems use in organizations. *Decision Support Systems*, *97*, 58–68. doi:10.1016/j.dss.2017.03.005

Audzeyeva, A., & Hudson, R. (2016). How to get the most from a business intelligence application during the post implementation phase? Deep structure transformation at a UK retail bank. *European Journal of Information Systems*, *25*(1), 29–46. doi:10.1057/ejis.2014.44

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. *Journal of the Association for Information Systems*, *19*(5), 3. doi:10.17705/1jais.00495

Baskerville, R. L., Kaul, M., & Storey, V. C. (2015). Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production. *Management Information Systems Quarterly*, *39*(3), 541–564. doi:10.25300/MISQ/2015/39.3.02

Batra, D. (2017). Adapting agile practices for data warehousing, business intelligence, and analytics. *Journal of Database Management*, *28*(4), 1–23. doi:10.4018/JDM.2017100101

Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, *70*, 328–337. doi:10.1016/j.jbusres.2016.08.006

Chau, M., & Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *Management Information Systems Quarterly*, *36*(4), 1189. doi:10.2307/41703504

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An Overview of Business Intelligence Technology. *Communications of the ACM*, *54*(8), 88–98. doi:10.1145/1978542.1978562

Chee, C. H., Yeoh, W., Gao, S., & Richards, G. (2014). Improving business intelligence traceability and accountability: An integrated framework of BI product and metacontent map. *Journal of Database Management*, *25*(3), 28–47. doi:10.4018/jdm.2014070102

Chen, H. C., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly*, *36*(4), 1165–1188. doi:10.2307/41703503

Chen, L., & Nath, R. (2018). Business analytics maturity of firms: An examination of the relationships between managerial perception of IT, business analytics maturity and success. *Information Systems Management*, *35*(1), 62–77. doi:10.1080/10580530.2017.1416948

Chien, M. (2018, October). Four steps to improve data and analytics capabilities when business intelligence maturity is low. *Gartner.*

Davenport, T. H., Harris, J. G., De Long, D. W., & Jacobson, A. L. (2001). Data to knowledge to results: Building an analytic capability. *California Management Review*, *43*(2), 117–138. doi:10.2307/41166078

Delen, D., & Demirkan, H. (2013). *Data, information and analytics as services*. Elsevier. doi:10.1016/j.dss.2012.05.044

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A content analysis of the content analysis literature in organization studies - Research themes, data sources, and methodological refinements. *Organizational Research Methods*, *10*(1), 5–34. doi:10.1177/1094428106289252

Elgendy, N., & Elragal, A. (2016). Big data analytics in support of the decision making process. *Procedia Computer Science*, *100*, 1071–1084. doi:10.1016/j.procs.2016.09.251

Fink, L., Yogev, N., & Even, A. (2017). Business intelligence and organizational learning: An empirical investigation of value creation processes. *Information & Management*, *54*(1), 38–56. doi:10.1016/j.im.2016.03.009

Foshay, N., Yeoh, W., Boo, Y. L., Ong, K. L., & Mattie, D. (2015). A Comprehensive Diagnostic Framework for Evaluating Business Intelligence and Analytics Effectiveness. *AJIS. Australasian Journal of Information Systems*, *19*, S37–S54. doi:10.3127/ajis.v19i0.1178

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007

Grover, V., Chiang, R. H., Liang, T.-P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, *35*(2), 388–423. doi:10.1080/07421222.2018.1451951

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191–209. doi:10.1016/j.jsis.2017.07.003

Halper, F., & Stodder, D. (2014). TDWI analytics maturity model guide. *TDWI Research*.

Hancu, L. (2012). Acquiring business ecosystem's intelligence through Large-Scale Collaborative Systems. *Paper presented at ITI 2012 34th International Conference on Information Technology Interfaces*.

Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change*, *78*(9), 1695–1704. doi:10.1016/j.techfore.2011.04.005

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in Information Systems research. *Management Information Systems Quarterly*, *28*(1), 75–105. doi:10.2307/25148625

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, *64*, 130–141. doi:10.1016/j.dss.2014.05.013

Hribar Rajterič, I. (2010). Overview of business intelligence maturity models. *Management: Journal of Contemporary Management Issues*, *15*(1), 47–67.

Hribar Rajterič, I. (2010). Overview of business intelligence maturity models. *Management: Journal of Contemporary Management Issues*, *15*, 47–67.

Hyder, Z., Siau, K., & Nah, F. (2019). Artificial Intelligence, Machine Learning, and Autonomous Technologies in Mining Industry. *Journal of Database Management*, *30*(2), 67–79. doi:10.4018/JDM.2019040104

Işı, K. (2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management*, *50*(1), 13–23. doi:10.1016/j.im.2012.12.001

Isik, O., Jones, M. C., & Sidorova, A. (2011). Business intelligence (BI) success and the role of BI capabilities. *Intelligent Systems in Accounting, Finance & Management*, *18*(4), 161–176. doi:10.1002/isaf.329

Khan, R. A., & Quadri, S. (2012). Business intelligence: An integrated approach. *Business Intelligence Journal*, *5*, 64–70.

Kiron, D., & Shockley, R. (2011). Creating Business Value with Analytics. *MIT Sloan Management Review*, *53*, 57–63.

Kohli, R., & Tan, S. S.-L. (2016). Electronic Health Records: How Can IS Researchers Contribute to Transforming Healthcare?. *Management Information Systems Quarterly*, *40*(3), 553–573. doi:10.25300/MISQ/2016/40.3.02

Kromer, M. (2014). Modern Hybrid Big Data Warehouse Architectures. *BI Training Solutions: As Close as Your*, *19*(4), 48.

Lahrmann, G., Marx, F., Winter, R., & Wortmann, F. (2010). Business intelligence maturity models: an overview. *Paper presented at the VII conference of the Italian chapter of AIS (itAIS 2010)*. Academic Press.

Lahrmann, G., Marx, F., Winter, R., & Wortmann, F. (2011). Business intelligence maturity: Development and evaluation of a theoretical model. *Paper presented at the 2011 44th Hawaii International Conference on System Sciences (HICSS)*. Academic Press. doi:10.1109/HICSS.2011.90

Lasrado, L. A., Vatrapu, R., & Andersen, K. N. (2015). Maturity models development in is research: a literature review. *Paper presented at the IRIS Selected Papers of the Information Systems Research Seminar in Scandinavia 2015*. Academic Press.

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, *52*, 21–32.

Lismont, J., Vanthienen, J., Baesens, B., & Lemahieu, W. (2017). Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, *37*(3), 114–124. doi:10.1016/j.ijinfomgt.2016.12.003

Lu, Y., Wang, F., & Maciejewski, R. (2014). Business intelligence from social media: A study from the vast box office challenge. *IEEE Computer Graphics and Applications*, *34*(5), 58–69. doi:10.1109/MCG.2014.61 PMID:25248200

McLaren, T. S., Head, M. M., Yuan, Y. F., & Chan, Y. E. (2011). A Multilevel Model for Measuring Fit between a Firm's Competitive Strategies and Information Systems Capabilities. *Management Information Systems Quarterly*, *35*(4), 909–929. doi:10.2307/41409966

Müller, O., Junglas, I., Debortoli, S., & vom Brocke, J. (2016). Using text analytics to derive customer service management benefits from unstructured data. *MIS Quarterly Executive*, *15*, 243–258.

Nguyen, T., Li, Z., Spiegler, V., Ieromonachou, P., & Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*, *98*, 254–264. doi:10.1016/j.cor.2017.07.004

O'Leary, D. E. (2011). Building and evolving data warehousing and business intelligence artefacts: The case of Sysco. *Intelligent Systems in Accounting, Finance & Management*, *18*(4), 195–213. doi:10.1002/isaf.330

O'Leary, D. E. (2013). Artificial intelligence and big data. *IEEE Intelligent Systems*, *28*(2), 96–99. doi:10.1109/MIS.2013.39 PMID:25505373

Oestreich, T. W. (2016). Magic Quadrant for Business Intelligence and Analytics Platforms. *Analyst*, *501*, G00275847.

Olszak, C. M. (2014). Dynamic business intelligence and analytical capabilities in organizations. *Paper presented at the e-skills for knowledge production and innovation conference*. Academic Press.

Olszak, C. M. (2016). Toward better understanding and use of Business Intelligence in organizations. *Information Systems Management*, *33*(2), 105–123. doi:10.1080/10580530.2016.1155946

Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In The handbook of research synthesis and meta-analysis (Vol. 2, pp. 177-203). Academic Press.

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). Capability Maturity Model, Version 1.1. *IEEE Software*, *10*(4), 18–27. doi:10.1109/52.219617

Pirttimäki, V., Lönnqvist, A., & Karjaluoto, A. (2006). Measurement of business intelligence in a Finnish telecommunications company. *Electronic Journal of Knowledge Management*, *4*, 83–90.

Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems*, *54*(1), 729–739. doi:10.1016/j.dss.2012.08.017

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A taxonomy of evaluation methods for information systems artifacts. *Journal of Management Information Systems*, *32*(3), 229–267. doi:10.1080/07421222.2015.1099390

Ransbotham, S., Kiron, D., & Prentice, P. K. (2015). Minding the analytics gap. *MIT Sloan Management Review*.

Rossi, M., Henfridsson, O., Lyytinen, K., & Siau, K. (2013). Design Science Research: The Road Traveled and the Road That Lies Ahead. *Journal of Database Management*, *24*(3), 1–8. doi:10.4018/jdm.2013070101

Rusaneanu, A. (2013). Comparative analysis of the main business intelligence solutions. *Informações Econômicas*, *17*(2/2013), 148–156. doi:10.12948/issn14531305/17.2.2013.12

Shankaranarayan, G., Ziad, M., & Wang, R. Y. (2003). Managing data quality in dynamic decision environments: An information product approach. *Journal of Database Management*, *14*(4), 14–32. doi:10.4018/jdm.2003100102

Shollo, A., & Galliers, R. D. (2016). Towards an understanding of the role of business intelligence systems in organisational knowing. *Information Systems Journal*, *26*(4), 339–367. doi:10.1111/isj.12071

Siow, E., Tiropanis, T., & Hall, W. (2018). Analytics for the Internet of Things: A survey. *ACM Computing Surveys*, *51*(4), 74. doi:10.1145/3204947

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. doi:10.1016/j.jbusres.2016.08.001

Skinner, R., Nelson, R. R., Chin, W. W., & Land, L. (2015). The Delphi Method Research Strategy in Studies of Information Systems. *Communications of the Association for Information Systems*, *37*, 31–63. doi:10.17705/1CAIS.03702

Storey, V. C., & Song, I. Y. (2017). Big data technologies and management: What conceptual modeling can do. *Data & Knowledge Engineering*, *108*, 50–67. doi:10.1016/j.datak.2017.01.001

Trieu, V.-H. (2017). Getting value from Business Intelligence systems: A review and research agenda. *Decision Support Systems*, *93*, 111–124. doi:10.1016/j.dss.2016.09.019

Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, *49*(3), 318–327. doi:10.1016/j.dss.2010.03.007

Van De Graaff, J., & Cameron, A. (2013). Quest for business intelligence in health care. *Healthcare Financial Management*, *67*, 44–48. PMID:23413668

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. doi:10.1057/ejis.2014.36

Wendler, R. (2012). The maturity of maturity model research: A systematic mapping study. *Information and Software Technology*, *54*(12), 1317–1339. doi:10.1016/j.infsof.2012.07.007

Yeoh, W., & Popovič, A. (2016). Extending the understanding of critical success factors for implementing business intelligence systems. *Journal of the Association for Information Science and Technology*, *67*(1), 134–147. doi:10.1002/asi.23366

Zeng, L., Li, L., & Duan, L. (2012). Business intelligence in enterprise computing environment. *Information Technology Management*, *13*(4), 297–310. doi:10.1007/s10799-012-0123-z

APPENDIX A

Table 4. Design science research guidelines

Guidelines	Description
Guideline 1: Design as an Artifact Design	This research identifies the technical and non-technical dimensions of the BI&A system and their levels, mapping them to the analytical adoption stages proposed by LaValle et al. (2013) with appropriate characteristics that defines each dimension for each level.
Guideline 2: Problem Relevance	The amount of interest amongst practitioners for using BI&A system to compete testifies to its practical relevance. Researchers also continue to raise questions about the complementary resources and capabilities that are essential for building effective analytics to compete in a dynamic environment (Kohli and Tan, 2018; Günther et al., 2017). The BI&A literature have explored various facets of a BI&A environment and have even proposed various maturity models. A review of existing models reveal they lack theoretical foundations, empirical validation, and the ability to operationalize maturity measurement (Lasadro et al., 2015). Competitive advantage requires maturity across a full range of dimensions that span the collection and analysis of data (Delen and Demirkan, 2013), sophistication of data consumers (Kiron and Shockley, 2011), alignment with organization strategies (Hribar Rajterič, 2010), and cultural commitment (Davenport et al., 2001).
Guideline 3: Design Evaluation	This study uses a two phased approach. Phase one employs a Delphi approach to derive a comprehensive maturity model with competitive advantage focus. Weaknesses and areas for improvements are identified. Participants scored the derived maturity model to ensure efficacy, and utility of the model. Phase two is a content analysis of relevant articles that infuses collective attitude and intentions (Duriau et al., 2007), adding validity and robustness.
Guideline 4: Research Contributions	BI&A Competitive Advantage Maturity Model provides: (1) A comprehensive understanding of technical and non-technical dimensions with defining characteristics at different maturity capability levels; and (2) Competitive Advantage focused artifact that organizations can use to assess the effectiveness of their existing BI&A system(s). Framework derived using Work System Theory can be used for creation of other IT-related maturity models associated with paradigm shifts.
Guideline 5: Research Rigor Design	Kernel theories have been used to develop the proposed framework (e.g. Work Systems Theory). Employed a Delphi study to derive and validate the proposed model is followed by content analysis of literature
Guideline 6: Design as a Search Process	BI& A system is defined by two components: core and peripheral elements. Two sets of elements are defined and described. Study identifies different levels for each element when using BI&A to achieve competitive advantage. Defining characteristics, derived from the literature and practitioner insights, are assigned for each element.
Guideline 7: Communication of Research	Communications with researchers has been achieved by reviewing and coding key publications. Communications with practitioner community in the Delphi stage represents engaged scholarship.

APPENDIX B

Table 5. Literature to derive proposed model

Article	Publication	Author(s)
Acquiring Business Ecosystem's Intelligence through Large-Scale Collaborative Systems	International Conference on Information Technology Interfaces	Hancu, 2012
Analytics: The new path to value	Sloan Management Review	LaValle et al., 2011
Best Practices for Driving Successful Analytics Governance	Gartner (ID G00278625)	Oestreich and Tapadinhas, 2015
Best-Practice Insights from Gartner's Business Intelligence and Analytics Summits in Sydney, London and Las Vegas, 2015	Gartner (ID G00278955)	Yeun and Schlegel, 2015
BI Experts' Perspective - BI for manufacturing	Business Intelligence Journal	TWDI
BI Experts Perspective - Infrastructure for Advanced Analytics	Business Intelligence Journal	TWDI
Building and Evolving Datawarehousing and Business Intelligence Artifact: The case of SYSCO	Intelligent System in Accounting, Finance And Management	O`Leary, 2011
Business Intelligence (BI) success and the role of BI capabilities	Intelligent System in Accounting, Finance And Management	Isik et al., 2011
Business Intelligence And Analytics: From Big Data To Big Impact	MIS quarterly	Chen et al., 2012
Business Intelligence from Social Media	Business Intelligence Analytics	Lu et al., 2014
Business Intelligence in blogs: Understanding Consumer Interactions and Communities	MIS quarterly	Chau and Xu, 2012
Business intelligence in enterprise computing environment	Information Technology Management	Zeng et al., 2012
Business intelligence success: The roles of BI capabilities and decision environments	Information & Management	Isik et al., 2013
Business Intelligence Teams Need to Change with the Times	Gartner (ID G00270899)	Duncan, 2016
Business Intelligence: An Integrated Approach	International Journal of Management and Innovation	Khan and Quadri, 2014
Comparative Analysis of the Main Business Intelligence Solutions	Informatica Economică	Rusaneanu, 2013
Critical Capabilities for Data Warehouse and Data Management Solutions for Analytics	Gartner (ID G00277836)	Greenwald et al., 2016
Develop Good Decision Models to Succeed at Decision Management	Gartner (ID G00278480)	Schulte and Kart, 2016
Dynamic Business Intelligence and Analytical Capabilities in Organizations	Proceedings of the e-Skills for Knowledge Production and Innovation Conference	Olszak, 2014
Extend Your Portfolio of Analytics Capabilities	Gartner (ID G00254653)	Kart et al., 2015
Extending the Understanding of Critical Success Factors for Implementing Business Intelligence Systems	Journal of the Association for Information Science and Technology	Yeoh and Popovič, 2016
Facilitators Are Crucial for Stakeholder Engagement in a Data-Driven Culture	Gartner (ID G00308815)	Duncan, 2016
How CDOs Engage with Their Stakeholders to Deliver Real Business Value	Gartner (ID G00304772)	Duncan et al., 2016
How Chief Data Officers Can Succeed by Driving Analytic Value	Gartner (ID G00278412)	Duncan and Buytendijk, 2016
How to Apply Advanced Analytics Capabilities to Social Data	Gartner (ID G00301354)	Sussin, 2016
How to Architect the Bi and Analytics Platform	Gartner (ID G00265003)	Tapadinhas, 2016
How to Establish a Data-Driven Culture in the Digital Workplace	Gartner (ID G00275916)	Kart, 2015
How to Get Started with Prescriptive Analytics	Gartner (ID G00273713)	Duncan, 2015
How to Take a First Step to Advanced Analytics	Gartner (ID G00276728)	Kart, 2015

continued on following page

Table 5. Continued

Article	Publication	Author(s)
Hype Cycle for Business Intelligence and Analytics, 2016	Gartner (ID G00290879)	Kart, 2015
Implications and Directions of Development of Web Business Intelligence Systems for Business Community	Economic Insights – Trends and Challenges	Bucur, 2012
Its core for BI and Analytics	Gartner	Howson and Duncan, 2016
Magic Quadrant for Business Intelligence Platforms	Gartner (ID G00225500)	Hagerty et al., 2012 Sallam et al., 2011
Market Guide for Optimization Solutions	Gartner (ID G00273712)	Kart and Zaidi, 2015
Predictive Algorithms That Healthcare Delivery Organizations Are Using to Improve Outcomes	Gartner (ID G00295509)	Craft, 2016
Predicts 2015_Power Shift in Business Intelligence and Analytics Will Fuel Disruption	Gartner	Paranteau et al., 2014
Quest for business intelligence in health care	Healthcare Financial Management	Van De Graaff and Cameron, 2013
Redefine, Reorganize, Revamp and Rebrand Your Bicc to Shift Focus to Analytics	Gartner (ID G00270091)	Paranteau and Zaidi, 2014

APPENDIX C: SUMMARY OF CODED ARTICLES

Validation of the Model Dimensions

The importance of the core dimensions of Work Systems Theory (i.e. participants, information, technologies, and processes & activities) for building effective BI&A systems in organizations has been documented in the articles reviewed. For example, Braganza et al. (2017) note that human capital (e.g. data analyst, statistician), physical capital (e.g. hardware) and intangible assets (e.g. knowledge, managerial skills, organizational brand/goodwill) are important for any big data initiative. Similarly, Sivarajah et al. (2017) points out that as big data becomes mainstream, more and more organizations are aiming to derive valuable information from the data using appropriate tools, technologies and people. Gandomi and Haider (2015) also suggest that organizations are adopting newer and innovative data management technologies to enable them to integrate data into their business processes to compete, which in turn enables organizations to improve operational efficiency, drive new revenue streams and gain competitive advantages over business rivals (Sivarajah et al., 2017). Large data volumes, widely distributed data sources, and multiple stakeholders characterize typical e-business settings with ubiquitous technologies (e.g., mobile devices) increasing data volumes distributed across data sources (Shankaranarayan, Ziad, and Wang, 2003). Such environments necessitate decision-makers to act/react quickly to decision-tasks including mission-critical decisions with demands for efficient data quality management. In addition, matured analytical methodologies (e.g., artificial intelligence, machine learning) provide many economic benefits for organizations to achieve cost reduction, improve efficiency, improve productivity, reduce exposure of workers to hazardous conditions, and enable continuous production (Shankaranarayan et al., 2003). Yet, the implementation of these technologies can face economic, financial, technological, workforce, and social challenges that needs to be addressed to enable higher levels of BI&A maturity.

Work Systems Theory identifies peripheral components (customers, infrastructure, products/services, strategic alignment and environment) that influence any sociotechnical work systems where humans interact with information systems to perform work and produce value. The articles analyzed also provide support for these peripheral dimensions in enabling effective BI&A systems. For example, Trieu (2017) notes that presenting information to users (e.g. senior management) for analysis, query, and reporting can enable effective decision-making as well as improve the performance of business processes. Business value, project management, shared understanding, technological capability, top management commitment, complexity, and organizational culture are all important for improving the maturity of development processes and activities used for BI&A system implementations within an organization (Batra, 2017). The maturity of the hardware infrastructure within an organization influences the functional affordances of the BI&A system (Popovic et al., 2012; Arnott, Lizama, and Song, 2017). The strategic context (e.g., to improve profitability, increase efficiency or reduce costs), and knowledge of how the information produced is utilized for the benefit of the organization contributed to the success of the BI&A system (Pirttimaki, Lönnqvist, and Karjaluoto 2006). A data-driven BI&A system that concentrates mainly on localized data itself will yield lower levels of efficiency in business decision-making (Chee et al., 2014). The need for comprehensive metadata to enhance the understanding of business intelligence products has grown (Chee et al. 2014) in addition to having analytical expertise among users (Ransbotham, Kiron, and Prentice, 2015). While having quality information would enable effective decision making, an organization's absorptive capacity and culture could hamper value creation (Pirttimaki et al., 2006).

Amrita George is an Assistant Professor in Management at the College of Business at Marquette University. She has 9 years of IT experience spanning business analysis, project management, and consulting across multiple business domains. Her research interests include health information technology, business intelligence and analytics, software development, and emergent technologies. Her research has been published in Journal of Medical Internet Research, International Conference on Information Systems, Americas Conference on Information Systems, and Academy of Management Global Proceedings.

Kurt Schmitz is a Clinical Associate Professor in Computer Information Systems at J. Mack Robinson College of Business at Georgia State University. He is certified by the Project Management Institute as a program management professional (PgMP) with extensive industry experience having served in multiple global IT leadership roles in large enterprises including industrial automation in the 1980s, networking and eCommerce in the 1990s, and life sciences in the 2000s. His research interests involve adaptation and change in the area of IT project management and the IT technologies produced by those projects. His publications appear in MIS Quarterly, IEEE Software, Journal of Information Systems Education and various conference proceedings.

Veda C. Storey is the Tull Professor of Computer Information Systems and Professor of Computer Science at the J. Mack Robinson College of Business, Georgia State University. Her research interests are in intelligent information systems, data management, conceptual modeling, and design science research. Dr. Storey is a member of the AIS College of Senior Scholars, an AIS Fellow, and an advisor to the Workshop on Information Technologies and Systems. She is also a member of the steering committee of the International Conference of Conceptual Modeling, where she is an ER Fellow and a recipient of the Peter P. Chen Award. Dr. Storey received her PhD from the University of British Columbia and holds a degree in flute performance from the Royal Conservatory of Music, University of Toronto.

Journal of Database Management
Volume 31 • Issue 3 • July-September 2020

Creativity of Participants in Crowdsourcing Communities:
The Effects of Promotion Focus and Extrinsic Motivation

Lingfei Zou, Huawei Technologies Co., Ltd, China
Shaobo Wei, University of Science and Technology of China, China
Weiling Ke, Southern University of Science and Technology, China
Kwok Kee Wei, Singapore Institute of Management, Singapore

ABSTRACT

Organizations can tap the wisdom of the crowd through digital platforms of crowdsourcing for ideation. However, we have limited understanding of factors affecting the innovativeness of ideas and solutions submitted by individual participants. Drawing upon self-determination theory and regulatory focus theory, we investigate how participants' regulatory focus and extrinsic motivation aroused by incentivizing mechanisms affect their creativity in crowdsourcing communities. Based on the data collected from 164 participants in a crowdsourcing platform, we find that promotion focus positively influences participants' creativity, and different types of extrinsic motivation have differential effects. Although external, identified, and integrated motivation positively affect participants' creativity, introjected motivation is not significantly related to participants' creativity. In addition, external and identified motivation strengthen the relationship between promotion focus and creativity. The theoretical contributions and managerial implications of this study are discussed.

KEYWORDS

Creativity, Crowdsourcing, Extrinsic Motivation, Promotion Focus, Self-Determination Theory

INTRODUCTION

Advances in information technology (IT) transform the way organizations form relationships with the large flexible labor pool provided by Internet users (Brynjolfsson, McAfee, and Spence, 2014; Cui, Kumar Pm, and Gonçalves, 2019; Jiang and Wang, 2019; Lukyanenko and Parsons, 2018). Crowdsourcing is an innovative practice enabled by IT wherein an organization outsources a task to a "crowd," rather than to a designated "agent" (Howe, 2006; Jeppesen and Lakhani, 2010). Online platforms enable organizations to use crowdsourcing for different tasks, such as product testing, logistics, production, after sales support, and ideation (Durward, Blohm, and Leimeister, 2016). In

DOI: 10.4018/JDM.2020070103

particular, crowdsourcing for ideation provides organizations with opportunities to tap the wisdom of the crowd to acquire innovative ideas for their products, services, and business processes (Howe, 2006; Nambisan and Baron, 2009, 2010). By issuing calls for ideas and solutions, organizations can reach out to individuals outside the organization and provide rewards to the individuals who generate the best solutions (Franke, Keinz, and Klausberger, 2013).

Organizations can reap significant benefits from crowdsourcing for ideation as they can attract a large and diverse group of people to engage in complex innovation (e.g., Afuah and Tucci, 2012; Franke et al., 2013; Huang, Vir Singh, and Srinivasan, 2014; Terwiesch and Xu, 2008). However, ensuring that the ideas generated by crowdsourcing ideators (hereafter participants) are creative enough to meet the organization's expectations is difficult (Blohm, Bretschneider, Leimeister, and Krcmar, 2011; Kevin J Boudreau, 2012; Leimeister, Huber, Bretschneider, and Krcmar, 2009). On the one hand, unoriginal ideas and solutions submitted by the crowd would defeat the organization's purpose of using crowdsourcing to acquire innovative ideas, and I would make crowdsourcing costly as the organization needs to expend resources in identifying the winning solutions (Cui et al., 2019; Wang, Feng, Jiang, and Xie, 2019). On the other hand, poor ideas and solutions do not allow the participants to get the desired reward, thus discouraging future participation in crowdsourcing (Hofstetter, Zhang, and Herrmann, 2018). Therefore, the antecedents of participants' creativity that determines the innovativeness of ideas and solutions submitted in crowdsourcing communities must be investigated (Bayus, 2013; Liu, Yang, Adamic, and Chen, 2014).

A thorough literature review shows that previous studies on crowdsourcing for ideation primarily focus on examining how extrinsic motivation mobilizes individuals to expend effort on innovation-related tasks (Franke et al., 2013; Hossain, 2012; Liu et al., 2014; Lukyanenko and Parsons, 2018; Zheng, Li, and Hou, 2011). For example, investigates how individuals' identification with the community affects participants' continued generation of ideas. study how the number of competitors affect participants' effort on different types of problems in crowdsourcing communities. find that a high financial reward can induce participants to submit high-quality solutions. This stream of research sheds light on our understanding of task design, reward schemes, and community development that effectively enhance the cognitive effort expended by individuals. Yet, cognitive effort may be necessary—but not enough—for participants to come up with innovative ideas and solutions. Indeed, previous creativity studies have mixed findings on how extrinsic motivation affects individuals' creativity (Deci, Koestner, and Ryan, 1999). For example, several researchers suggest that extrinsic motivation crowds out intrinsic motivation and negatively affects creativity (Amabile, Hill, Hennessey, and Tighe, 1994; Edward L Deci et al., 1999), whereas others find that extrinsic motivation enhances creativity (Robert Eisenberger and Selbst, 1994; Slot, 2013). In the view that crowdsourcing platforms implement various mechanisms aimed to arouse participants' extrinsic motivation, we must investigate how extrinsic motivation affects individuals' creativity in crowdsourcing communities. In addition, the creativity literature clearly establishes that the novelty and originality of ideas stem from flexible, associative, and global thinking (T. M. Amabile, Conti, Lazenby, and Herron, 1996; Dreu, W., Baas, Nijstad, and A., 2008; Förster and Dannenberg, 2010). Without the right thinking style, participants are unlikely to turn cognitive effort into original, novel, or uncommon ideas (Dreu et al., 2008; Förster and Dannenberg, 2010). Hence, a study on factors that affect participants' thinking style will help us understand how to increase participants' creativity in crowdsourcing communities.

To this end, we conduct the research reported in this paper. Drawing upon self-determination theory (SDT), regulatory focus theory (RFT), and creativity literature, we develop a model to investigate how extrinsic motivation and regulatory focus jointly affect individuals' creativity in crowdsourcing communities. Creativity or creative performance refers to the development of novel and useful ideas by individuals (T. M. Amabile et al., 1996). Extrinsic motivation is defined as a behavior that aims toward outcomes extrinsic to the behavior itself (Ryan and Deci, 2000a). Specifically, anchoring on RFT (E Tory Higgins, 1997, 1998), we contend that promotion focus—a motivational system behind individuals' strivings to satisfy their desire for growth, advancement,

and achievement—enables individuals to engage in global, inclusive, and flexible thinking and thus positively influences creativity (Förster and Dannenberg, 2010; Friedman and Förster, 2001, 2002; Xu, Wang, and Xue, 2019). In addition, following SDT, we examine the effects of a spectrum of extrinsic motivations and categorize them into four types, namely, external motivation, introjected motivation, identified motivation, and integrated motivation (E. L. Deci and Ryan, 2000; Gagné and Deci, 2005; Ryan and Deci, 2000b). We argue that they positively affect participant's creativity in crowdsourcing communities. Further, while regulatory focus initiates purposeful goal strivings, extrinsic motivations associated with the task can act in concert with these strivings, which allow individuals to experience the psychological state of meaningfulness and thus expend sustained effort on the task (T. M. Amabile et al., 1996; Dreu et al., 2008; Förster and Dannenberg, 2010; Nijstad, De Dreu, Rietzschel, and Baas, 2010). Thus, we propose that extrinsic motivation aroused by the incentivizing mechanisms in crowdsourcing communities would moderate the effect of participants' promotion focus on their creative performance. Our research model is supported by data from 164 participants in Zhubajie, one of the largest crowdsourcing communities in China. As a third-party platform, Zhubajie allows registered organizations to post open calls for ideas and solutions with set deadlines and options of rewards for the tasks.

The current research makes three important contributions to the literature. First, this study is a response to the call for further research on identifying the factors that can improve the contribution quality of individuals in a computer-meditated group collaboration environment like crowdsourcing (Dean, Hender, Rodgers, and Santanen, 2006; Jung, Schneider, and Valacich, 2010). We extend the current understanding of critical factors affecting participants' creativity, which is most valuable for organizations aiming to acquire innovative ideas by using crowdsourcing. To the best of our knowledge, this study is the first to theorize and empirically test how promotion focus, as a determining factor, affects individuals' creativity in crowdsourcing communities. The finding on the positive effects of promotion focus suggests that the crowdsourcing platforms can prime participants and get them into the promotion focus mode when working on the crowdsourcing tasks. Second, by examining the effects of a spectrum of extrinsic motivation in a holistic model, our research provides a nuanced understanding of how different incentivizing mechanisms implemented by the crowdsourcing platform would affect participants' creativity in the communities. Further, our study reveals different types of extrinsic motivation interact differently with the driving force of promotion focus in affecting participants' creativity. Thus, our research provides new insights, which have implications for how the crowdsourcing platforms can better stimulate participants to generate innovative ideas and solutions.

THEORETICAL BACKGROUND AND RESEARCH HYPOTHESES

Self-Determination Theory

Motivation is recognized as a critical factor affecting human behavior and performance (Locke and Latham, 2004). It can be characterized into two categories, i.e., intrinsic and extrinsic motivation. According to SDT, motivation is a spectrum, rather than a unitary construct, varying along the internalization of value underlying the specific behavior. Based on the degrees to which people internalize the related value of a task, their motivation can vary from being externally controlled to self-determined (Ryan and Deci, 2000a). In particular, SDT categorizes extrinsic motivations into four types, namely, external, introjected, identified, and integrated motivation (Deci and Ryan, 2000; Gagné and Deci, 2005; Ryan and Deci, 2000b; Xu et al., 2019). External motivation is the most controlled motivation. It refers to how individuals perform a task to obtain a reward or satisfy an external demand. As such, when externally motivated, people act to obtain a desired consequence and are energized into action that are instrumental to those ends (Gagné and Deci, 2005). Introjected motivation is a moderately controlled form of regulation (Gagné and Deci, 2005). Although the locus of control for introjected motivation is within the individual, it is relatively controlled because it pressures people to behave to feel worthy or to buttress their egos (Gagné and Deci, 2005). By contrast,

identified regulation is a relatively self-determined extrinsic motivation that energizes individuals to take action (Gagné and Deci, 2005). With such a motivation, people identify the value underlying the specific task and thus feel great freedom and volition (Gagné and Deci, 2005). The most autonomous extrinsic motivation is integrated motivation that involves the integration of identification with other aspect of oneself, including values and interests (Gagné and Deci, 2005). With integrated motivation, individuals fully identify with the importance of the related activities and willing to devote their effort to the tasks even though they are uninteresting (Ryan and Deci, 2000b).

Previous research identifies several factors that can motivate individuals to participate in crowdsourcing communities (Afuah and Tucci, 2012; Bayus, 2013; Boudreau and Lakhani, 2013; Brabham, 2010; Hossain, 2012). For example, find that a sense of community membership motivates participants to comment actively on the ideas of others. suggest that participants are motivated by the opportunity to receive financial reward and take up freelance work. Although these previous studies provide great insights into how extrinsic motivation can energize individuals to expend effort on tasks in crowdsourcing communities, they conceptualize extrinsic motivation differently and examine these motivating factors in a piecemeal manner. More importantly, the literature does not help us understand how participants' extrinsic motivation aroused by various incentivizing mechanisms afforded by the platform affect participants' creativity in the crowdsourcing community. In the view that the influencing mechanism varies across different types of motivation (Deci and Ryan, 2000) and the mixed views on the effects of extrinsic motivation on creativity in the literature, as mentioned above, we must categorize extrinsic motivation into a unified framework and examine its effects on participants' creativity in crowdsourcing in a holistic model.

Drawing on SDT, we categorize the motivating factors identified by previous research on crowdsourcing into different motivation types, as shown in Table 1. Specifically, financial rewards

Table 1. Mapping of motivation in crowdsourcing communities to SDT

Motivation Type	Motivations in Crowdsourcing Communities
External motivation	To make money by contributing creative ideas or solutions (Bayus, 2013) To share profit by selling creative ideas to firms (Horton and Chilton, 2010) To improve job prospect (Brabham, 2008, 2010)
Introjected motivation	To gain reputation in a specific field (Brabham, 2008, 2010) To demonstrate their abilities (Battistella and Nonino, 2012; Zheng et al., 2011)
Identified motivation	To identify with the value of sponsored firms (Hossain, 2012) To identify the value of crowdsourcing communities (DiPalantino and Vojnovic, 2009) To believe that they belong to crowdsourcing communities (Brabham, 2008)
Integrated motivation	To believe that customers should contribute to their favorite company (Hossain, 2012) To constantly performing creative tasks is important to keep themselves mentally active (Afuah and Tucci, 2012) To believe that companies should pay attention to customer needs when designing new products (Bayus, 2013)
Intrinsic motivation	Being happy to express creative ideas freely (Battistella and Nonino, 2012) Being interested in firms' new product development process (Brabham, 2008, 2010) Be Enjoying working on the tasks in the community (Estelles-Arolas and Gonzalez-Ladron-de-Guevara, 2012)

and job opportunity provided by crowdsourcing platforms or sponsoring organizations are the most controlled factors and are thus identified as external motivation. The mechanisms enabling participants to gain reputation and demonstrate their abilities in crowdsourcing communities are introjected motivation. Although seeking reputation and demonstrating abilities allow participants to see the

value in participating in crowdsourcing communities, such regulations are controlling as they pressure participants to work on the task to buttress their egos. In addition, crowdsourcing platforms typically implement features and functionalities that allow participants immerse in the communities. As such, participants would develop community identification and affective commitment, which are relatively self-determined and are thus classified as identified motivation. In addition, crowdsourcing platforms may advocate open-innovation and emphasize the importance of individuals' participation in the joint development of products, services, and processes of the sponsoring organizations. Accordingly, participants develop their personal beliefs in crowdsourcing, which are the most autonomous motivation and are thus categorized as integrated motivation. Further, we categorize interests and enjoyment in crowdsourcing communities as intrinsic motivation. In the current research, we focus on the effects of the four types of extrinsic motivation because they are the motivation aroused by the incentivizing mechanisms implemented in crowdsourcing communities. Intrinsic motivation is included as a control variable in our research model.

Regulatory Focus Theory

RFT posits that people regulate their goals, emotion, and behavior based on two fundamental motivational orientations, i.e., promotion and prevention focus (Crowe and Higgins, 1997). With promotion focus, individuals are sensitive to the presence or absence of positive factors. They generally engage in approach-related behaviors toward positive outcomes and try to satisfy their desire for growth, advancement and achievement, and ideal self-pursuit. Moreover, promotion focus is associated with feelings of cheerfulness when the individual is making progress toward the desired end state (Higgins, 1997; Idson, Liberman, and Higgins, 2000). By contrast, with prevention focus, individuals are sensitive to the presence or absence of negative factors. They generally engage in avoidance-related behaviors away from negative outcomes and seek to satisfy their desire for security, duties, obligations, and ought self-pursuit (Higgins, 1997). In addition, prevention focus is associated with fear, tension, and worry before the individual attains the goal of avoiding the undesired end state (Higgins, 1997; Idson et al., 2000).

Promotion and prevention focus are distinct but co-exist in the motivation system of individuals (Crowe and Higgins, 1997; Xu et al., 2019). Previous research regards regulatory focus as a chronic characteristic derived from personality and culture (Higgins, 2000). However, recent studies increasingly regarded regulatory focus as a psychological state that varies across different situations (Lanaj, Chang, and Johnson, 2012). For example, found that a safety climate sparks prevention focus but inhibits promotion focus. Therefore, investigating the influence of regulatory focus can provide implications for community managers to induce the desired orientation of participants that can lead to participants' quality contributions. Considering the special characteristics of crowdsourcing communities, we focus on promotion focus in the current research. Crowdsourcing communities consist of individuals who are self-motivated and voluntarily engage in ideation and solution development activities posted by organizations (Franke et al., 2013). These communities adopt an open structure and value participants' autonomy, with no formal contract to regulate the kinds of behavior or levels of effort from participants. As such, participants can quit from the project or the community any time, and no penalty or punishment is imposed for their immature contributions to the community (Franke et al., 2013). Accordingly, we argue that promotion focus is the driving force that leads individuals to initiate the self-regulatory process aiming at achieving gains and participating in crowdsourcing communities.

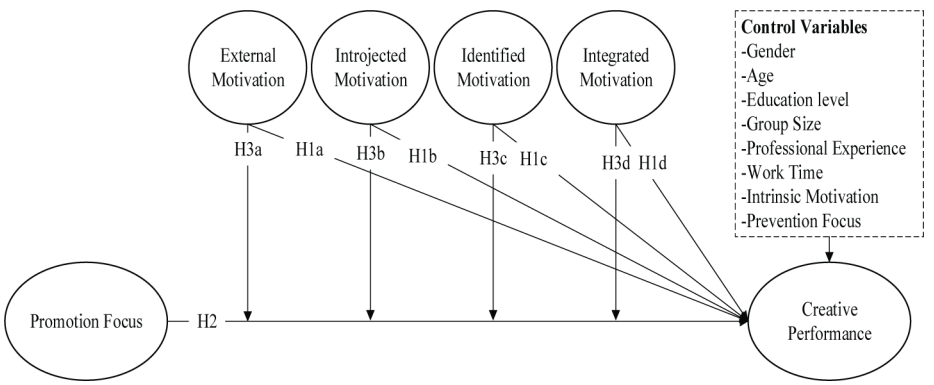
Further, recent studies on creativity found that regulatory focus significantly impacts people's capacity for creativity, i.e., the ability to generate ideas, insights, and solutions that are innovative (Amabile et al., 1996; Runco, 2004). Specifically, with promotion focus, individuals apply eager strategic means to pursue their aspiration goals, that is, to insure hits and to insure against errors of omission (Brendl and Higgins, 1996). By contrast, with prevention focus, individuals apply vigilant strategic means to pursue their defensive goals, that is, to insure correct rejection and to insure

against errors of commission (Brendl and Higgins, 1996). As such, promotion-focused individuals would engage in more global, inclusive, and flexible thinking than their prevention-focused counterparts (Förster and Dannenberg, 2010). With such a thinking style, individuals can access remote informational links and find new connections among categories and concepts (Förster and Dannenberg, 2010). The use of broad, global, and inclusive cognitive categories and the adaptive switching among categories, approaches, and sets increases creative capacity (Knoblich, Ohlsson, Haider, and Rhenius, 1999). In addition, promotion focus makes individuals experience positive affect such as cheerfulness when they are making progress on tasks requiring creativity (Baas, De Dreu, and Nijstad, 2008). Such a positive affect would further activate individuals' promotion system and make them even more creative than their counterparts with prevention focus (Baas et al., 2008; Lyubomirsky, King, and Diener, 2005). Hence, promotion focus is a critical determinant of individuals' creative performance on tasks that require creative ideation and solutions.

Hypothesis Development

Drawing upon RFT and SDT, we develop the research model depicted in Figure 1. As discussed earlier, external motivation refers to the expectation of obtaining tangible rewards that are contingent

Figure 1. Research model



on task performance. In crowdsourcing communities, participants receive financial rewards when their ideas or solutions outperform others' or are adopted by the sponsoring organization. Financial incentives are the most common tangible rewards that crowdsourcing platforms use to motivate people in generating creative ideas (Bayus, 2013; Cai, 2006; Franke et al., 2013). In the literature on creativity, mixed findings exist about the effects of external motivation on creativity. Although several previous studies argued and found that financial incentives would crowd out individuals' intrinsic motivation for the task and thus decrease creativity, other research suggested that financial rewards could enhance individuals' interest in the task and thus increase creativity instead of undermining intrinsic motivation (R. Eisenberger and Rhoades, 2001; Eisenberger & Selbst, 1994; Liu et al., 2014).

To reconcile the differences, researchers come to a consensus that how financial rewards affect creativity depends on the reward orientation (Eisenberger and Rhoades, 2001; Jiang and Wang, 2019). In particular, external rewards that are linked to specific dimension of the task performance will guide individuals to put in effort to meet the expectations (Eisenberger and Rhoades, 2001; Liu et al., 2014; Slot, 2013). In crowdsourcing communities, external rewards are awarded based on the innovativeness of ideas and solutions submitted by the individuals. Only when a participant's submission is selected as the best solution does he or she receive the external rewards. As such, external motivation directs

participants' effort toward generating innovative ideas and thus would promote creativity. Indeed, the positive effects of external motivation gain empirical support (Bayus, 2013; Liu et al., 2014; Slot, 2013). Accordingly, we propose the following.

Hypothesis 1a: External motivation is positively related to participant's creativity.

With introjected motivation, participants aim at demonstrating their ability and gaining reputation in a specific domain. In crowdsourcing communities, reputation systems are an important mechanism to indicate the ability and past performance of participants (Majchrzak and Malhotra, 2013; Y. Zhang and van der Schaar, 2012). When participants' creative ideas are adopted by the sponsoring firm, their scores of performance indicators, such as "number of creative ideas," "accumulated points," and "achievement level" in the system, will increase. With the increased ranking of participants, they are then considered to have high ability and good reputation in crowdsourcing communities (Adler and De Alfaro, 2007).

According to, ability-proving-oriented people seek recognition from peers and thus try to identify environmental cues to learn which kinds of behavior and performance are valued or respected and then act accordingly to meet the expectations. The reputation system in crowdsourcing communities runs on the basis of the innovativeness of ideas and solutions submitted by participants, providing social information to participants that creative ideas are respected in the community. As such, reputation systems can arouse crowdsourcing participants' introjected motivation. To rank high in the reputation system, participants are likely to direct their energy to creative activities. The positive effects of introjected motivation on creativity gain empirical support. For example, found that ability-proving orientation motivates individuals to exchange information with peers and thus increase creativity. Contextual influence on individual creativity shows that ability-proving orientation further generates creative ideas when creativity is valued as an indicator of ability (Hirst et al., 2009). Therefore, following hypothesis is proposed:

Hypothesis 1b: Introjected motivation is positively related to a participant's creativity.

With identified motivation, participants internalize the value and identify with the crowdsourcing community. Community identification reflects that people have a sense of belonging to the community and regard community success or failure as their own (Johnson, Chang, and Yang, 2010). It is a determining factor for members' behavior in the community (Ashforth and Mael, 1989). When identifying the community's value, participants perceive community activities as significant and meaningful, and they align their goals with those of the community. Therefore, they are willing to exert sustained effort to achieve community goals and improve community success (Ashforth and Mael, 1989; Dick, Hirst, Grojean, and Wieseke, 2007). Applying this notion to the current research, we argue that participants with identification motivation will actively engage in the creative process and generate creative ideas because the key value in crowdsourcing communities is fostering creative ideas to help the sponsoring firm solve complex tasks. Previous studies offer empirical support to the positive effects of identification motivation on creativity. For example, confirm that people become creative when they perceive meaningfulness at work. provide evidence that users perform better when they strongly identify with the community. Accordingly, we hypothesize the following.

Hypothesis 1c: Identified motivation is positively related to a participant's creativity.

Integrated motivation manifests the extent to which participants regard their behavior as an important part of themselves (Ryan and Deci, 2000b). When the required behaviors in crowdsourcing communities match their personal beliefs, participants will have a strong self-identity and view their

effort as valuable and meaningful (Stryker, 1987). Self-identity drives people to persist in role-consistent behavior (Hong, Ye, Du, Wang, and Fan, 2019). With integrated motivation, participants will be committed to generating innovative ideas and solutions that are valued by the sponsoring organization in crowdsourcing communities. Accordingly, we hypothesize the following.

Hypothesis 1d: Integrated motivation is positively related to a participant's creativity.

In addition, we expect promotion focus to positively affect participants' creative performance in crowdsourcing communities. According to RFT, promotion focus drives individuals to conduct approach-oriented behavior (Higgins, 1997). When opportunities for gains exist, promotion-focused individuals will proactively take action to strive for the goal of achieving personal growth and advancement (Parker, Williams, and Turner, 2006). Participation in crowdsourcing communities, as a self-regulatory process, is at the individuals' discretion and thus an opportunity for gains rather than possibilities of losses (Bayus, 2013; Franke et al., 2013; Huang et al., 2014). Indeed, individuals are attracted to crowdsourcing communities due to the opportunities to work on innovative tasks that can potentially benefit themselves and the organization (Boudreau and Jeppesen, 2015; Huang et al., 2014). Therefore, we conjecture that promotion focus is the driving force for individuals' participation in crowdsourcing communities.

In addition, previous studies in creativity literature theorize and empirically find that promotion focus enables individuals' global, associative, and flexible thinking, which is required for the generation of innovative ideas and solving of insight problems (Knoblich et al., 1999). Applying this notion to the current research context, we argue that promotion focus positively affects participants' creativity and the innovativeness of their ideas because promotion focus activates the cognitive system for creativity. Specifically, with promotion focus, participants apply eager strategic means and search globally for information and knowledge that help them generate innovative ideas. They also engage in the adaptive combination and association of various categories, approaches, and sets of knowledge related to the specific tasks. In addition, the ambitious strivings driven by promotion focus make participants focus on their ideals, sustain enthusiasm, and be committed to the tenacious goal pursuit (Brendl and Higgins, 1996). Further, promotion focus allows participants to experience positive affect such as cheerfulness when they are working on the innovative task, thereby making them take broad and flexible approaches (Baas et al., 2008; Boudreau and Jeppesen, 2015; Lyubomirsky et al., 2005). With the flexible and global thinking style, participants can access remote informational links and find new connections among different categories and concepts, allowing them to generate innovative ideas and solve insight problems (Förster and Dannenberg, 2010). Hence, promotion focus positively affects participants' creativity in crowdsourcing communities.

Hypothesis 2: Promotion focus is positively related to a participant's creativity.

In addition to appropriate cognitive style, innovative ideas and insights require a cognitive process that involves persistent task-directed effort (T. M. Amabile et al., 1996; Dreu et al., 2008; Förster and Dannenberg, 2010). The generation of creative ideas and solutions usually costs sustained effort and attention (Nijstad et al., 2010; Schooler et al., 1993). As such, a context that arouses individuals' motivation to expend prolonged effort can benefit ideation and solution development (Rietzschel, Nijstad, and Stroebe, 2007). In particular, when extrinsic motivation aroused from incentivizing mechanisms acts in concert with the purposeful goal strivings initiated by promotion focus, individuals will experience a psychological state of meaningfulness that, in turn, will encourage them to expend effort on the task and endeavor to attain the desired outcomes (Barrick et al., 2012). Prior research has consistently found that incentive mechanisms can be applied to entice individuals to conduct effortful exploration of problem space, which facilitates the generation of novel ideas and solving

of insight problems (Rhoades, Eisenberger, and Armeli, 2001; Rietzschel et al., 2007). Therefore, extrinsic motivation can interact with promotion focus in exerting positive influence on individuals' creative performance (Nijstad et al., 2010).

By applying this notion to our research, we expect that extrinsic motivation aroused by the incentive mechanisms in crowdsourcing communities will moderate the effects of promotion focus on participants' creativity. Specifically, when individuals are driven by their promotion focus to work on tasks posted in crowdsourcing communities, extrinsic motivation allows them to feel that their effort is valued and sense a psychological state of meaningfulness (Barrick, Mount, and Li, 2012). In turn, the sense of meaningfulness will further activate individuals' motivational system and make them focus on pursuing the goal of advancing gains and personal development. Therefore, extrinsic motivation stimulates sustained effort and thus would facilitate individuals to turn their creative thinking aspired by promotion focus into fruitful results of innovative ideas and solutions in crowdsourcing communities.

Specifically, external motivation in the form of financial rewards for quality contributions in crowdsourcing communities make participants feel that their effort is appreciated by the crowdsourcing platform or sponsoring organization. Introjected motivation, by enhancing reputation and demonstration of abilities in the communities, allows participants to gain intangible benefits that can possibly help them with career advancement. Identified motivation makes participants feel great freedom because working on tasks in crowdsourcing communities is congruent with their personal goals and identities, thereby making participation important to them. With integrated motivation, participants fully understand that working on tasks in crowdsourcing communities is an integral part of who they are and emanates from their sense of self. All these different types of extrinsic motivation will allow participants to perceive the value in the course of action and be willing to expend sustained effort on the crowdsourcing tasks. We expect that such effort will work in concert with the flexible, associative, and global thinking activated by promotion focus and will lead participants to achieve enhanced creative performance in crowdsourcing communities.

Hypothesis 3: Extrinsic motivation moderates the relationship between a participant's promotion focus and creativity.

Hypothesis 3a: External motivation moderates the relationship between a participant's promotion focus and creativity.

Hypothesis 3b: Introjected motivation moderates the relationship between a participant's promotion focus and creativity.

Hypothesis 3c: Identified motivation moderates the relationship between a participant's promotion focus and creativity.

Hypothesis 3d: Integrated motivation moderates the relationship between a participant's promotion focus and creativity.

RESEARCH METHODOLOGY

Data Collection

To test the research model, an online survey was conducted in Zhubajie (http://www.zhubajie.com), one of the largest crowdsourcing communities in China. Zhubajie is a third-party platform on which registered organizations can post innovative-related tasks in the form of an open call, set a deadline, and fund rewards for the tasks. Participants can contribute their ideas or solutions by working on these tasks. After the deadline, sponsoring organizations evaluate the solutions and adopt what they consider best satisfies their requirements. Finally, rewards are assigned to the corresponding contributors. Zhubajie was selected as the research website for three reasons. First, organizations issued calls for ideas and solutions in this platform, which made it appropriate for our research topic

on creative performance in crowdsourcing communities. Second, a large number of active participants were registered with this platform as ideators who are willing to work on the innovation-related tasks issued by sponsoring organizations, thereby allowing us to invite participants to respond to our survey. Third, as one of the largest crowdsourcing communities in China, Zhubajie represented a successful model in crowdsourcing. Collecting data with this community will reveal relatively general insights that would be applicable to other crowdsourcing communities.

We sent the platform (i.e., Zhubajie's website) a request for a survey and asked them to help us invite 1000 active participants in the crowdsourcing community. We checked the invited participants' respective homepages for their participation record to ensure that they had been participating in crowdsourcing tasks. The questionnaire started with a detailed introduction that describes the purpose and requirement of our survey and assured that all information would be kept strictly confidential. To encourage participation, we offered RMB 20 (approximately USD 3.2) to participants who would complete the questionnaire. As precaution against repeated participation of any respondent, we rejected any multiple questionnaires derived from the same IP address. We received a total of 166 responses. Two responses were disregarded because the participants had never participated in innovation-related tasks. Table 2 shows the participants' demographic data.

Measurement

All items in the questionnaire were adopted from validated scales in previous studies. To ensure the face validity of the items in the context of China, we used standard and back translation methods. The questionnaire was pretested with 10 crowdsourcing participants who filled out the questionnaire and reported some unclear or inappropriate questions. On the basis of their feedback, we revised the questionnaire and then administered the survey by sending invitations to participants of projects hosted in Zhubajie. Participants answered the questions on a seven-point Likert scale ranging from "strongly disagree" to "strongly agree." The instruments for the four types of extrinsic motivation were adopted from Teresa M Amabile (1993); Becker, Billings, Eveleth, and Gilbert (1996); Mullan, Markland, and Ingledew (1997); W. L. Ke and Zhang (2010); and T. M. Amabile et al. (1994). The measurement scales for the promotion focus were adopted from Haws, Dholakia, and Bearden (2010), and the items for creativity were adopted from Tierney, Farmer, and Graen (1999) and Taggar (2001).

We included control variables, i.e., age, gender, education level, group size of the project, the participant's professional experience, amount of time that the participant spends on the project per week, intrinsic motivation, and prevention focus. In particular, intrinsic motivation has been suggested to play an important role in influencing an individual's innovative behavior (Li, Hsieh, and Rai, 2013); thus, we controlled it in our model. Although our study centers on promotion focus, individuals, particularly Asian people (e.g., Chinese), are influenced by prevention focus (Ke, Tan, Sia, and Wei, 2013). To account for alternative explanations, we also controlled individuals' prevention focus in our model. The five items used to measure intrinsic motivation were adopted from T. M. Amabile et al. (1994). The four items measuring prevention focus were adopted from Haws et al. (2010). Table 3 shows the measurement items for each construct. In addition, following prior studies (e.g., Shen, Lee, and Cheung, 2014), we controlled the possible effects of gender, age, and education level on a participant's creative performance.

Common Method Bias

When data are self-reported and collected from a single source, common method bias could be a serious issue because of the halo effect or participants' leniency biases (Podsakoff, MacKenzie, Lee, and Podsakoff, 2003). Similar to most prior research, we adopted the most widely used Harman's single factor method to test for common method bias in our measurement. The result shows eight constructs with eigenvalues greater than 1, which account for 73.87% of the total variance. The first construct accounts for only 11.39% of the total variance. Therefore, the result demonstrates that no serious common method bias exists in this research. Furthermore, we addressed the nonresponse bias

Table 2. Respondents' demographics in crowdsourcing communities

	Category	Frequency (N=164)	Percent (%)
Gender	Male	101	61.6%
	Female	63	38.4%
Age	18 years old and below	1	0.6%
	19–29 years old	134	81.7%
	29–39 years old	26	15.9%
	39–49 years old	2	1.2%
	50 years old and above	1	0.6%
Education	Junior college or below	46	28.0%
	Bachelor's degree	115	70.2%
	Master's degree	3	1.8%
Group size	1 member	102	62.2%
	2–5 members	50	30.6%
	5–10 members	5	3.0%
	10–15 members	2	1.2%
	15 or above members	5	3.0%
Professional experience	Less than 6 months	34	20.7%
	6 months to 1 year	32	19.5%
	1–3 years	57	34.8%
	3–5 years	25	15.2%
	More than 5 years	16	9.8%
Community tenure	Less than 6 months	93	56.7%
	6 months to 2 years	54	32.9%
	2 years to 4 years	16	9.8%
	More than 4 years	1	0.6%
Work time in community per week	Less than 1 hour	6	3.7%
	1–5 hours	47	28.7%
	5–10 hours	39	23.8%
	10–20 hours	29	17.6%
	More than 20 hours	43	26.2%

by comparing the demographics of the first and final 25% of the participants via chi-square test. The results show no significant difference between the two groups with regard to demographic variables, thus indicating the absence of a nonresponse bias in this study.

DATA ANALYSIS AND RESULTS

Our data analysis involves two steps. The first step is the confirmatory factor analysis, which tests the reliability and validity of our measurement. The second step is multiregression analysis using SPSS, which examines our research hypotheses.

Table 3. Measurement items

Construct	Items	Source
Promotion focus	I often think about the person I would ideally like to be in the future.	Haws et al. (2010)
	I frequently imagine how I will achieve my hopes and goals.	
	In general, I am focused on achieving positive outcomes in my life.	
	When I see an opportunity to perform tasks that I like, I get excited right away.	
External motivation	I am strongly motivated by the money I can earn from performing tasks in this community.	Amabile (1993); Mullan et al. (1997)
	I am keenly aware of the income goals I have for myself when I participate in this community.	
	I seldom think about the contest award. (Reverse coded)	
	I am concerned about the award in the innovation contest.	
Introjected motivation	I am strongly motivated by the recognition I can gain from other people in this community.	Ke and Zhang (2010); Amabile et al. (1994)
	I want other people to find out how good I am in generating novelty ideas.	
	For me, successfully participating in this community means doing better than other people.	
	I am concerned about how other people will react to my ideas.	
Identified motivation	This community group has a great deal of personal meaning for me.	Ke and Zhang (2010)
	I have a strong positive feeling toward this community group.	
	I identify with the value of participating in this community.	
Integrated motivation	The reason I participate in this community is because of what it stands for, that is, its values.	Becker et al. (1996)
	My attachment to this community is primarily on the basis of the similarity of my values to those represented by this community.	
	If the values of the community group were different, I would not be attached to it.	
Creativity	My solution provides a new perspective on this kind of task.	Tierney et al. (1999); Taggar (2001)
	My solution is novel but also meets the task requirement.	
	My solution is original.	
Intrinsic Motivation	It is important for me to be able to do what I most enjoy in the crowdsourcing community.	Amabile et al. (1994)
	What matters most to me is enjoying what I do in this crowdsourcing community.	
	Interest is the driving force behind much of what I do in this crowdsourcing community.	
	I enjoy trying to solve complex problems.	
Prevention Focus	I often worry that I will fail to accomplish my goals.	Haws et al. (2010)
	I am anxious that I will fall short of my responsibilities and obligations.	
	I frequently think about how I can prevent failures in my life.	

Measurement Model

We conducted confirmatory factor analysis to assess the reliability and validity of all constructs and test our measurement model. Reliability was assessed using Cronbach's alpha. As suggested, Cronbach's alpha above 0.70 indicates good reliability. Table 4 shows that the Cronbach's alpha values of all constructs were higher than 0.70. This finding indicates that our measurement model has good reliability.

Construct validity has two dimensions: convergent and discriminant. Convergent validity was assessed using (1) composite reliability, (2) average variance extracted (AVE), and (3) item loadings. The composite reliability of each construct should be higher than 0.70 (Fornell and Larcker, 1981); all constructs satisfy this standard, as shown in Table 4. AVE captures the amount of variance explained by the indicators. suggested that AVE above 0.50 indicates good convergent validity. The confirmatory factor analysis results show that the AVE values of all constructs were greater than 0.60. Moreover, item loadings should be above 0.60 (Fornell and Larcker, 1981). Our analysis results also satisfy this requirement. Thus, all three conditions are satisfied by our measurement model, which means that our measurement has good internal consistency and convergent validity.

Discriminant validity was assessed in two ways. First, according to, the square root of the AVE should be greater than the correlation coefficient between the corresponding construct and other constructs. Table 4 shows that the numbers on the diagonal are the square roots of the AVE and the numbers off the diagonal are the correlation coefficients between constructs. The results show that all diagonal numbers are higher than off-diagonal numbers, thus indicating good discriminant validity. Second, the item loading on the corresponding construct should be higher than that on the other constructs. recommended that the difference between the item loadings on the corresponding construct and other constructs should be more than 0.10. Table 5 shows that all item loadings are higher than 0.60 and that item cross loadings are lower than 0.50. Therefore, our measurement has good discriminant validity.

Structural Model

We conducted a series of multiregression analyses to test our hypotheses. In Model 1, we included only the control variables to predict participants' creativity. In Model 2, we included the control variables and four types of extrinsic motivation to predict participants' creative performance. In Model 3, we included the variables in Model 2 and promotion focus. Furthermore, we added the interaction terms of different types of extrinsic motivation and promotion focus in Model 4. Table 6 presents the regression results.

As shown in Model 1, we found that gender, educational level, and intrinsic motivation had positive effects on participants' creativity ($\beta = 0.151$, $p < 0.05$; $\beta = 0.197$, $p < 0.01$; $\beta = 0.304$, $p < 0.001$). All other control variables were insignificantly related to creativity.

In Model 2, we determined that external motivation was significantly related to creative performance ($\beta = 0.200$, $p < 0.01$), which supported H1a. Moreover, identified and integrated motivations positively affected creativity ($\beta = 0.223$, $p < 0.01$; $\beta = 0.225$, $p < 0.01$); thus, H1c and H1d were supported. However, we found a negative yet insignificant relationship between introjected motivation and creativity ($\beta = -0.136$, $p > 0.05$). Hence, H1b was unsupported.

In Model 3, we added the promotion focus to test its direct effect on creativity. As expected, we identified that promotion focus had a significant and positive effect on creativity ($\beta = 0.227$, $p < 0.01$). Hence, H2 was supported.

We added the interaction variables of extrinsic motivation and promotion focus to the regression model to test the moderating effect of different types of extrinsic motivation on the relationship between promotion focus and creativity in Model 4. The results show that the interaction term of external motivation and promotion focus was positively significant ($\beta = 0.157$, $p < 0.05$), thus providing support to H3a. The interaction term of identified motivation and promotion focus was also positively significant ($\beta = 0.169$, $p < 0.05$). This result supported

Table 4. Internal consistency and discriminant validity of constructs

Constructs	Mean	S.D.	Cronbach's Alpha	C.R.	AVE	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Promotion Focus	5.70	0.82	0.843	0.905	0.761	**0.87**													
2. External Motivation	4.71	1.04	0.794	0.868	0.622	0.23	**0.79**												
3. Introjected Motivation	4.72	1.11	0.874	0.914	0.727	0.27	0.35	**0.85**											
4. Identified Motivation	5.16	0.83	0.864	0.917	0.788	0.36	0.25	0.45	**0.89**										
5. Integrated Motivation	4.83	0.97	0.872	0.922	0.797	0.37	0.34	0.50	0.49	**0.89**									
6. Creativity	4.98	0.80	0.834	0.901	0.753	0.43	0.34	0.21	0.41	0.41	**0.87**								
7. Prevention focus	5.35	0.81	0.801	0.872	0.630	0.44	0.22	0.25	0.32	0.30	0.30	**0.79**							
8. Intrinsic motivation	5.01	1.00	0.799	0.872	0.630	0.41	0.21	0.42	0.40	0.49	0.38	0.44	**0.79**						
9. Group size	NA	NA	NA	NA	NA	0.04	0.09	0.10	0.08	0.07	0.04	0.09	0.08	NA					
10. Professional experience	NA	NA	NA	NA	NA	-0.11	0.09	0.09	0.02	0.01	0.14	-0.01	-0.07	0.04	NA				
11. Work time	NA	NA	NA	NA	NA	0.02	0.08	0.10	0.09	0.05	-0.03	-0.11	-0.01	0.12	0.24	NA			
12. Gender	NA	NA	NA	NA	NA	0.05	0.01	-0.15	0.00	-0.17	0.17	-0.01	-0.03	0.03	0.15	0.17	NA		
13. Age	NA	NA	NA	NA	NA	0.05	0.06	0.13	0.11	0.05	0.05	-0.04	-0.09	0.05	0.38	0.19	0.11	NA	
14. Education level	NA	NA	NA	NA	NA	-0.02	-0.02	-0.01	-0.01	0.01	0.19	-0.13	0.02	0.11	0.02	-0.11	0.07	-0.12	NA

Note. The diagonals are the square roots of the AVE; NA = Not applicable

Table 5. Cross loadings of measurement items to latent variables

Items	Promotion Focus	External Motivation	Introjected Motivation	Identified Motivation	Integrated Motivation	Creativity	Prevention Focus	Intrinsic Motivation
PRMF_1	**0.811**	0.078	0.212	-0.028	0.043	0.126	0.081	0.150
PRMF_2	**0.845**	0.059	0.082	0.173	0.088	0.066	0.081	0.079
PRMF_3	**0.767**	0.021	0.005	0.190	0.183	0.186	0.167	0.131
PRMF_4	**0.807**	0.085	0.030	0.081	0.072	0.184	0.049	0.178
EXTE_1	0.110	**0.683**	0.400	0.071	-0.041	0.118	-0.017	0.065
EXTE_2	0.234	**0.668**	0.213	0.075	-0.027	0.243	0.067	0.028
EXTE_3	-0.029	**0.825**	0.067	0.013	0.160	0.000	0.082	0.100
EXTE_4	0.004	**0.815**	-0.036	0.068	0.177	0.108	0.035	-0.023
INTR_1	0.091	0.177	**0.746**	0.119	0.018	0.149	0.027	0.151
INTR_2	0.151	0.054	**0.814**	0.099	0.186	-0.027	0.006	0.162
INTR_3	0.016	0.126	**0.803**	0.277	0.188	0.013	0.103	0.118
INTR_4	0.049	0.092	**0.791**	0.129	0.207	0.013	0.101	0.121
IDEN_1	0.211	0.120	0.221	**0.731**	0.147	0.263	0.060	0.152
IDEN_2	0.118	0.066	0.170	**0.844**	0.116	0.109	0.056	0.090
IDEN_3	0.087	0.034	0.200	**0.838**	0.152	0.100	0.106	0.160
INTE_1	0.171	0.203	0.243	0.020	**0.759**	0.182	0.114	0.205
INTE_2	0.127	0.089	0.198	0.200	**0.806**	0.223	-0.052	0.221
INTE_3	0.133	0.108	0.256	0.346	**0.734**	0.056	0.149	0.155
CRPF_1	0.161	0.272	0.022	0.111	0.087	**0.773**	0.012	0.082
CRPF_2	0.190	0.084	0.042	0.194	0.127	**0.798**	-0.009	0.160
CRPF_3	0.166	0.060	0.071	0.103	0.164	**0.818**	0.188	0.136
PRVF_1	0.166	0.053	0.059	0.073	0.045	0.047	**0.854**	0.005
PRVF_2	-0.040	0.044	0.117	0.022	0.126	-0.042	**0.820**	0.191
PRVF_3	0.298	0.077	0.003	0.143	-0.053	0.265	**0.707**	0.224
IM_1	0.059	0.050	0.020	0.113	0.220	0.040	0.115	**0.801**
IM_2	0.188	0.028	0.119	0.116	-0.001	0.281	0.140	**0.753**
IM_3	0.178	-0.048	0.205	0.121	0.144	0.106	0.128	**0.677**
IM_4	0.147	0.146	0.273	0.069	0.131	0.036	0.033	**0.676**

H3c regarding the positive moderating effect of identified motivation on the relationship between promotion focus and creativity. However, the interaction term of introjected motivation and promotion focus and that of integrated motivation and promotion focus are insignificant ($\beta = -0.070$, $p > 0.05$; $\beta = -0.024$, $p > 0.05$). Therefore, the hypotheses regarding the positive moderate effects of introjected and integrated motivations on the relationship between promotion focus and creativity were unsupported. As such, H3b and H3d were unsupported.

For further analysis, we plotted the moderating effects of external and identified motivations in Figures 2 and 3, respectively. As predicted, at high levels of external motivation (or identified motivation), creativity increased rapidly as promotion focus increased. At low levels of external motivation (or identified motivation), creativity did not increase (increased marginally) as promotion

Table 6. Multi-regression analyses results

Dependent Variable	Creativity			
	Model 1	Model 2	Model 3	Model 4
Control Variables				
Gender	**0.151***	**0.177****	**0.157***	**0.144***
Age	0.053	0.012	-0.010	-0.014
Education Level	**0.197****	**0.180****	**0.172***	**0.161***
Group Size	-0.025	-0.035	-0.028	-0.023
Professional Experience	0.128	0.127	**0.164***	**0.176***
Work Time	-0.049	-0.091	-0.103	-0.097
Intrinsic Motivation	**0.304*****	**0.162***	0.127	0.095
Prevention Focus	0.199	0.103	0.036	0.053
Independent Variables				
External Motivation (EXTE)		**0.200****	**0.187****	**0.147***
Introjected Motivation (INTR)		-0.136	-0.137	-0.111
Identified Motivation (IDEN)		**0.223****	**0.198***	**0.182***
Integrated Motivation (INTE)		**0.225****	**0.194***	**0.191***
Promotion Focus (PMF)			**0.227****	**0.227****
Interaction Terms				
EXTE*PMF				**0.157***
INTR*PMF				-0.070
IDEN*PMF				**0.169***
INTE*PMF				-0.024
R^2	0.258	0.398	0.433	0.478
Adjust R^2	0.219	0.350	0.384	0.417
F	**6.723*****	**8.328*****	**8.815*****	**7.872*****

Note: *p < 0.05; **p < 0.01; ***p < 0.001.

focus increased. This phenomenon confirmed the positive moderating effect of external and identified motivations on the relationship between promotion focus and creativity.

Post Hoc Analyses

We conducted a series of post hoc analyses to provide further insights into our results. First, our results showed that gender was significantly related to creativity ($\beta = 0.151$, $p < 0.05$), which indicates that male and female participants might be affected by different factors when they are generating creative ideas in the crowdsourcing context. Hence, we further conducted the analyses of our model for male and female samples. As shown in Table 7, the direct effects of extrinsic motivation and promotion focus had different influences on creativity between female and male groups. Specifically, external motivation was positively related to creativity only for the female group, whereas identified motivation, integrated motivation, and promotion focus were positively related to creative performance only for the male group. Regarding the interaction effects of extrinsic motivation and promotion focus, the interaction between identified motivation and promotion focus became insignificant for the female

Figure 2. The moderating effect of external motivation on the relationship between promotion focus and creativity

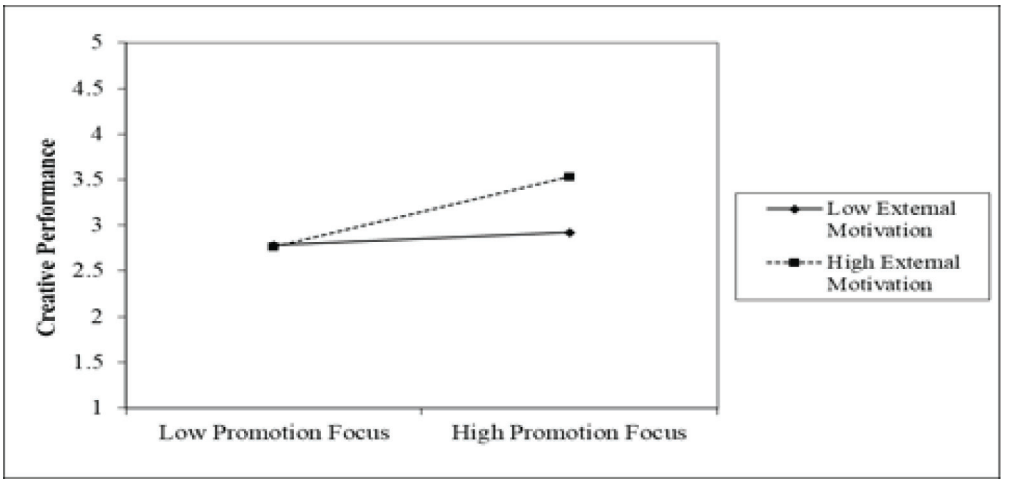
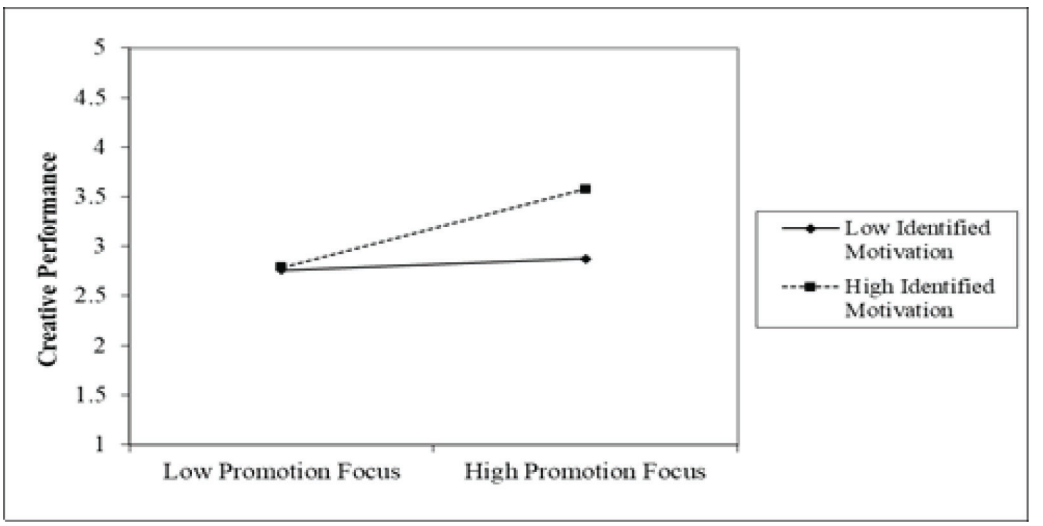


Figure 3. The moderating effect of identified motivation on the relationship between promotion focus and creativity



group ($\beta = 0.056$, $p > 0.05$). The interaction effects of external motivation and promotion focus on creative performance were marginally significant for female and male groups at the 0.10 level.

Second, although we argued that promotion focus is the primary driving force for creativity in the crowdsourcing context, we further tested the possible moderating effect of prevention focus on the relationships among different types of extrinsic motivation and creativity. As shown in Model 2 in Table 8, prevention focus only positively moderated the effect of identified motivation on creativity ($\beta = 0.187$, $p < 0.05$).

Third, as intrinsic motivation can play an important role when participants generate innovative ideas and solutions, we examined whether intrinsic motivation interacts with different types of extrinsic motivation in affecting participants' creativity. Our results (Model 3 in Table 8) indicated no significant interaction effects between intrinsic motivation and the four types of extrinsic motivation.

Table 7. Regression analyses results across female and male samples

Dependent Variable	Creativity			
	Female (N=63)		Male (N=101)	
	Model 1	Model 2	Model 3	Model 4
Control Variables				
Age	0.007	-0.009	-0.057	-0.037
Education Level	**0.166**	**0.172**	**0.155**	**0.1639**
Group Size	-0.073	-0.093	-0.045	-0.017
Professional Experience	-0.050	0.009	**0.291****	**0.281****
Work Time	0.001	0.021	-0.090	-0.106
Intrinsic Motivation	**0.368***	**0.349***	0.011	-0.028
Prevention Focus	-0.207	-0.141	0.155	0.185
Independent Variables				
External Motivation (EXTE)	**0.263***	0.148	0.097	0.087
Introjected Motivation (INTR)	-0.067	-0.035	-0.193	-0.163
Identified Motivation (IDEN)	0.206	0.137	**0.277****	**0.251***
Integrated Motivation (INTE)	0.106	0.109	**0.290****	**0.294****
Promotion Focus (PMF)	0.133	0.215	**0.213***	0.166
Interaction Terms				
EXTE*PMF		**0.254+**		**0.158+**
INTR*PMF		-0.139		-0.094
IDEN*PMF		0.056		**0.274****
INTE*PMF		0.033		-0.083
R^2	0469	0.513	0.482	0.549
Adjust R^2	0.341	0.344	0.412	0.463
F	**3.674****	**3.032****	**6.829*****	**6.391*****

Note: +p<0.10; *p<0.05; **p<0.01; ***p<0.001.

DISCUSSION AND CONCLUSION

Our objective was to investigate the factors affecting participants' creativity in crowdsourcing communities. Drawing upon SDT, we categorized extrinsic motivation into four types and investigated their effects on participants' creativity in a single holistic model. In addition, building upon RFT and recent literature on creativity, we argued that promotion focus is a critical determining factor of creativity and its effects are strengthened by extrinsic motivation aroused by the incentivizing mechanisms. Our results confirmed most of our predictions and underscored the importance of most types of extrinsic motivation and promotion focus in driving creativity in crowdsourcing communities. In particular, we found that the four types of extrinsic motivation do not equally affect participants' creativity. Although external, identified, and integrated motivations have positive effects, introjected motivation insignificantly affects creativity. These findings are consistent with those of previous studies on extrinsic motivation's effect on individuals' participation effort in crowdsourcing (Bayus, 2013; Franke et al., 2013; Liu et al., 2014; Terwiesch and Xu, 2008). The surprising finding regarding

Table 8. Moderating roles prevention focus and intrinsic motivation

Dependent Variable	Creativity		
	Model 1	Model 2	Model 3
Control Variables			
Gender	**0.157***	**0.154***	**0.141***
Age	-0.010	-0.013	0.001
Education Level	**0.172***	**0.178***	**0.160***
Group Size	-0.028	-0.006	-0.016
Professional Experience	**0.164***	**0.170***	**0.146***
Work Time	-0.103	-0.092	-0.072
Intrinsic Motivation (IM)	0.127	**0.158***	0.150
Prevention Focus (PRF)	0.036	0.006	0.059
Independent Variables			
External Motivation (EXTE)	**0.187****	**0.171***	**0.173***
Introjected Motivation (INTR)	-0.137	-0.128	-0.158*
Identified Motivation (IDEN)	**0.198***	**0.201****	**0.237****
Integrated Motivation (INTE)	**0.194***	**0.185***	**0.167***
Promotion Focus (PMF)	**0.227****	**0.247****	**0.169***
Interaction Terms			
PRF*EXTE		-0.038	
PRF*INTR		-0.039	
PRF*IDEN		**0.187***	
PRF*INTE		-0.125	
IM*EXTE			0.063
IM*INTR			0.139
IM*IDEN			0.061
IM*INTE			-0.046
R^2	0.433	0.463	0.461
Adjust R^2	0.384	0.400	0.398
F	**8.815*****	**7.390*****	**7.341*****

Note: *p < 0.05; **p < 0.01; ***p < 0.001.

the insignificant effects of introjected motivation can be attributed to evaluation apprehension (Diehl and Stroebe, 1987; Gallupe et al., 1992). With introjected motivation, individuals aim at an improved reputation to demonstrate their ability, implying that negative evaluation by others would cause anxiety and thus may harm their creative performance (Hildebrand, Haubl, Herrmann, and Landwehr, 2013). In addition, promotion focus explains a significant unique variance beyond the effects of extrinsic motivation, which highlights the importance of considering regulatory focus when studying the outcomes of self-initiated and self-regulated behavior in open innovation communities. This finding is consistent with the recent literature on creativity (Förster and Dannenberg, 2010). Furthermore, external and identified motivations reinforce the positive effects of promotion focus as they energize

participants to expend sustained effort on the related tasks and thus facilitate individuals to turn their creative thinking into the generation of ideas and solutions.

Limitations and Future Research

Recognizing the limitations of a research project is important in identifying future directions for research. This study has several limitations. First, all data were collected from a single crowdsourcing community and thus could limit the generalizability of the results. Although conducting the survey in a single context can alleviate the influence of confounding variables and improve research rigor, the relationship between extrinsic motivation and creativity may not hold in other crowdsourcing communities. In future studies, a cross-crowdsourcing community research can be conducted to increase the generalizability of our findings. Second, although all items were adopted from validated scales, the data were self-reported by participants at one point in time. Thus, such a design is vulnerable to common method bias, although Harman's one-factor analysis shows that it is not a serious problem with the current research. In future studies, objective data can be collected to measure participants' creativity for avoiding common method bias. Furthermore, a longitudinal design of future study can offer information that allows us to unveil the causal effects of promotion focus and extrinsic motion on participants' creativity in the crowdsourcing communities. Third, the sample size of the current research was small. The platform had more than 10 million registered participants at the time of data collection. However, we were unable to invite additional participants to respond to our survey due to limited resources. Although the crowdsourcing platform randomly sent invitations to active participants at our request, future research should launch an increased scale of study and collect data from an increased number of participants to enhance sample representativeness. Finally, although the current study focused on promotion focus and extrinsic motivation, future study can explore other factors that may influence the relationships among promotion focus, extrinsic motivation, and creative performance. For example, in the post hoc analyses, our results varied between male and female groups.

Theoretical Implications

Our work offers several major contributions to the crowdsourcing and creativity literature. First, this study advances our understanding of participants' contribution to crowdsourcing communities by investigating the factors leading to creativity. Most previous studies have identified many antecedents of individuals' participation in crowdsourcing (Cai, 2006; C. Zhang, Hahn, and De, 2013; Zheng et al., 2011). However, several scholars have recently criticized this stream of research because they found that an enormous number of ideas exceed the processing ability of sponsoring organizations and even most ideas and solutions submitted by participants are not innovative enough and do not meet the expectations (Bloodgood, 2013). The present study, with a focus on participants' creativity, provides a new perspective on enhancing the quality of individual ideators' contributions, which determines the extent to which organizations can materialize the purported benefits of crowd ideation.

Second, we examine the effects of a set of comprehensive motivational factors related to the context of crowdsourcing communities. In particular, we provide a framework for categorizing these factors on the basis of SDT. Previous studies have regarded extrinsic motivation as a unitary construct and thus ignored how different types of extrinsic motivations affect participants' performance. In this view, our examination on the effects of the entire spectrum of extrinsic motivation in a single model provides a nuanced understanding of the effective incentivizing mechanisms in stimulating participants' development of creative ideas. As such, the current research complements prior research that has primarily focused on the amount of motivation, rather than the types of motivation, and thus neglected the differences in the effects of different incentivizing mechanisms implemented by crowdsourcing platforms.

Third, we provide new insights into the critical role of promotion focus in affecting individuals' creative performance in crowdsourcing communities. Prior research has primarily focused on

how incentivizing mechanisms motivate individuals to expend cognitive effort in crowdsourcing communities. In this view, our work extends the literature and underscores the importance of considering factors that directly determine individuals' cognitive styles. Without an appropriate thinking style and information processing approach, effort energized by incentivizing mechanisms will unlikely lead to innovative ideation and solutions (Baas et al., 2008; Förster and Dannenberg, 2010). Therefore, the current research complements previous studies that have ignored the possible effects of regulatory focus on creativity.

Fourth, the findings of this research enrich our understanding by revealing that promotion focus and extrinsic motivation can generate synergistic effects on creativity in crowdsourcing communities. Although promotion focus enables global, flexible, and associative thinking, external and identified motivations help keep participants engaged in the related tasks and expending sustained effort. This situation allows participants to turn their thinking into ideation and solutions. Thus, our research shows that the effectiveness of providing financial rewards and developing participant identification with the sponsoring organizations exceeds their direct effects on creativity in crowdsourcing communities.

Practical Implications

This study provides some important practical implications for practitioners, especially the sponsoring organizations in crowdsourcing communities. We found that different types of extrinsic motivation have differential effects on participants' creativity, indicating that various incentivizing mechanisms implemented would differently affect creativity. Particularly, sponsoring organizations must be aware that external, identified, and integrated motivations promote creativity, whereas introjected motivation has an insignificant effect. To facilitate participants' generation of innovative ideas, firms can increase the financial incentives or the number of winners to understand participants' strong external motivation. Firms may remove the reputation system to diminish the influence of ranking in the community because introjected motivation does not directly improve participants' creativity. Moreover, organizations can implement mechanisms to foster a strong sense of community among participants and help them internalize the value of crowdsourcing, which will arouse identified and integrated motivations and enhance participants' creativity.

Our findings suggest that participants' promotion focus enhances their creativity. Sponsoring organizations aiming to capture innovative ideas by using crowdsourcing can ensure that promotion focus becomes the dominant motivational system among participants in the communities. Although we operationalize promotion focus as a chronic variable, recent research suggests that it can be a state variable (Lanaj et al., 2012). As such, administrators of the digital platforms can prime participants to be promotion focus-oriented. For example, platform administrators can provide additional positive feedback and encouragement when they are communicating with participants in the community. In addition, an open and high-risk-tolerant atmosphere in the community can help participants strive for their ideal selves and generate additional creative ideas and solutions.

ACKNOWLEDGMENT

This research is funded by Singapore Ministry of Education (Grant No. R-253-000-127-133). We thank editor-in-chief Dr. Keng Siau and anonymous reviewers for providing us with insightful and constructive comments.

Weiling Ke is a corresponding author on this paper. They can be reached at: wke@clarkson.edu.

REFERENCES

Adler, B. T., & De Alfaro, L. (2007). A content-driven reputation system for the Wikipedia. *Paper presented at the 16th international conference on World Wide Web*. Academic Press. doi:10.1145/1242572.1242608

Afuah, A., & Tucci, C. L. (2012). Crowdsourcing as a Solution to Distant Search. *Academy of Management Review*, *37*(3), 355–375. doi:10.5465/amr.2010.0146

Amabile, T. M. (1993). Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human Resource Management Review*, *3*(3), 185–201. doi:10.1016/1053-4822(93)90012-S

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal*, *39*(5), 1154–1184.

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology*, *66*(5), 950–967. doi:10.1037/0022-3514.66.5.950 PMID:8014837

Ashforth, B. E., & Mael, F. (1989). Social identity theory and the organization. *Academy of Management Review*, *14*(1), 20–39. doi:10.5465/amr.1989.4278999

Baas, M., De Dreu, C. K., & Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus? *Psychological Bulletin*, *134*(6), 779–806. doi:10.1037/a0012815 PMID:18954157

Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology studies, 2*(2), 285-309.

Barrick, M. R., Mount, M. K., & Li, N. (2012). The Theory of Purposeful Work Behavior: The Role of Personality, Higher-Order Goals, and Job Characteristics. *Academy of Management Review*, *38*(1), 132–153. doi:10.5465/amr.2010.0479

Battistella, C., & Nonino, F. (2012). Open innovation web-based platforms: The impact of different forms of motivation on collaboration. *Innovation-Management Policy & Practice*, *14*(4), 557–575. doi:10.5172/impp.2012.14.4.557

Bayus, B. L. (2013). Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community. *Management Science*, *59*(1), 226–244. doi:10.1287/mnsc.1120.1599

Becker, T. E., Billings, R. S., Eveleth, D. M., & Gilbert, N. L. (1996). Foci and bases of employee commitment: Implications for job performance. *Academy of Management Journal*, *39*(2), 464–482.

Blohm, I., Bretschneider, U., Leimeister, J. M., & Krcmar, H. (2011). Does collaboration among participants lead to better ideas in IT-based idea competitions? An empirical investigation. *International Journal of Networking and Virtual Organisations*, *9*(2), 106–122. doi:10.1504/IJNVO.2011.042413

Bloodgood, J. (2013). Crowdsourcing: Useful for Problem Solving, But What About Value Capture? *Academy of Management Review*, *38*(3), 455–457. doi:10.5465/amr.2012.0318

Boudreau, K. J. (2012). Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. *Organization Science*, *23*(5), 1409–1427. doi:10.1287/orsc.1110.0678

Boudreau, K. J., & Jeppesen, L. B. (2015). Unpaid crowd complementors: The platform network effect mirage. *Strategic Management Journal*, *36*(12), 1761–1777. doi:10.1002/smj.2324

Boudreau, K. J., Lacetera, N., & Lakhani, K. R. (2011). Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis. *Management Science*, *57*(5), 843–863. doi:10.1287/mnsc.1110.1322

Boudreau, K. J., & Lakhani, K. R. (2013). Using the Crowd as an Innovation Partner. *Harvard Business Review*, *91*(4), 60–69. PMID:23593768

Brabham, D. C. (2008). Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. *First Monday*, *13*(6). doi:10.5210/fm.v13i6.2159

Brabham, D. C. (2010). Moving the Crowd at Threadless. *Information Communication and Society*, *13*(8), 1122–1145. doi:10.1080/13691181003624090

Brendl, C. M., & Higgins, E. T. (1996). Principles of Judging Valence. *What Makes Events Positive or Negative*, *28*, 95–160. doi:10.1016/s0065-2601(08)60237-3

Brynjolfsson, E., McAfee, A., & Spence, M. (2014). New world order: Labor, capital, and ideas in the power law economy. *Foreign Affairs*, *93*(4), 44–53.

Cai, J. (2006). Knowledge Management Within Collaboration Processes. *Journal of Database Management*, *17*(1), 33–48. doi:10.4018/jdm.2006010103

Cohen-Meitar, R., Carmeli, A., & Waldman, D. A. (2009). Linking meaningfulness in the workplace to employee creativity: The intervening role of organizational identification and positive psychological experiences. *Creativity Research Journal*, *21*(4), 361–375. doi:10.1080/10400410902969910

Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes*, *69*(2), 117–132. doi:10.1006/obhd.1996.2675

Cui, Z., Kumar Pm, S., & Gonçalves, D. (2019). Scoring vs. Ranking: An Experimental Study of Idea Evaluation Processes. *Production and Operations Management*, *28*(1), 176–188. doi:10.1111/poms.12910

Dean, D. L., Hender, J. M., Rodgers, T. L., & Santanen, E. L. (2006). Identifying quality, novel, and creative ideas:construct and scales for idea evaluation. *Journal of the Association for Information Systems*, *7*(10), 646–699. doi:10.17705/1jais.00106

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668. doi:10.1037/0033-2909.125.6.627 PMID:10589297

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. doi:10.1207/S15327965PLI1104_01

Dick, R., Hirst, G., Grojean, M. W., & Wieseke, J. (2007). Relationships between leader and follower organizational identification and implications for follower attitudes and behaviour. *Journal of Occupational and Organizational Psychology*, *80*(1), 133–150. doi:10.1348/096317905X71831

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups - toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497–509. doi:10.1037/0022-3514.53.3.497

DiPalantino, D., & Vojnovic, M. (2009). Crowdsourcing and all-pay auctions. *Paper presented at the 10th ACM conference on electronic commerce*.

Dreu, D. (2008). Hedonic tone and activation level in the mood-creativity link: Toward a dual pathway to creativity model. *Journal of Personality and Social Psychology*, *94*(5), 739–756. doi:10.1037/0022-3514.94.5.739 PMID:18444736

Durward, D., Blohm, I., & Leimeister, J. M. (2016). Crowd work. *Business & Information Systems Engineering*, *58*(4), 281–286. doi:10.1007/s12599-016-0438-0

Eisenberger, R., & Rhoades, L. (2001). Incremental effects of reward on creativity. *Journal of Personality and Social Psychology*, *81*(4), 728–741. doi:10.1037/0022-3514.81.4.728 PMID:11642357

Eisenberger, R., & Selbst, M. (1994). Does reward increase or decrease creativity? *Journal of Personality and Social Psychology*, *66*(6), 1116–1127. doi:10.1037/0022-3514.66.6.1116

Estelles-Arolas, E., & Gonzalez-Ladron-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200. doi:10.1177/0165551512437638

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *JMR, Journal of Marketing Research*, *18*(1), 39–50. doi:10.1177/002224378101800104

Förster, J., & Dannenberg, L. (2010). GLOMOsys: A Systems Account of Global Versus Local Processing. *Psychological Inquiry*, *21*(3), 175–197. doi:10.1080/1047840X.2010.487849

Franke, N., Keinz, P., & Klausberger, K. (2013). "Does This Sound Like a Fair Deal?": Antecedents and Consequences of Fairness Expectations in the Individual's Decision to Participate in Firm Innovation. *Organization Science*, *24*(5), 1495–1516. doi:10.1287/orsc.1120.0794

Friedman, R. S., & Förster, J. (2001). The effects of promotion and prevention cues on creativity. *Journal of Personality and Social Psychology*, *81*(6), 1001–1013. doi:10.1037/0022-3514.81.6.1001 PMID:11761303

Friedman, R. S., & Förster, J. (2002). The Influence of Approach and Avoidance Motor Actions on Creative Cognition. *Journal of Experimental Social Psychology*, *38*(1), 41–55. doi:10.1006/jesp.2001.1488

Fuller, J., Matzler, K., & Hoppe, M. (2008). Brand community members as a source of innovation. *Journal of Product Innovation Management*, *25*(6), 608–619. doi:10.1111/j.1540-5885.2008.00325.x

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, *26*(4), 331–362. doi:10.1002/job.322

Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, L. M., & Nunamaker, J. F. (1992). Electronic brainstorming and group size. *Academy of Management Journal*, *35*(2), 350–369.

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, *16*(1), 91–109.

Gong, Y., Kim, T. Y., Lee, D. R., & Zhu, J. (2012). A Multilevel Model of Team Goal Orientation, Information Exchange, and Creativity. *Academy of Management Journal*, *56*(3), 827–851. doi:10.5465/amj.2011.0177

Haws, K. L., Dholakia, U. M., & Bearden, W. O. (2010). An Assessment of Chronic Regulatory Focus Measures. *JMR, Journal of Marketing Research*, *47*(5), 967–982. doi:10.1509/jmkr.47.5.967

Higgins, E. T. (1997). Beyond pleasure and pain. *The American Psychologist*, *52*(12), 1280–1300. doi:10.1037/0003-066X.52.12.1280 PMID:9414606

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology*, *30*, 1–46. doi:10.1016/S0065-2601(08)60381-0

Higgins, E. T. (2000). Making a good decision: Value from fit. *The American Psychologist*, *55*(11), 1217–1230. doi:10.1037/0003-066X.55.11.1217 PMID:11280936

Hildebrand, C., Haubl, G., Herrmann, A., & Landwehr, J. R. (2013). When Social Media Can Be Bad for You: Community Feedback Stifles Consumer Creativity and Reduces Satisfaction with Self-Designed Products. *Information Systems Research*, *24*(1), 14–29. doi:10.1287/isre.1120.0455

Hirst, G., van Knippenberg, D., & Zhou, J. (2009). A Cross-Level Perspective on Employee Creativity: Goal Orientation, Team Learning Behavior, and Individual Creativity. *Academy of Management Journal*, *52*(2), 280–293. doi:10.5465/amj.2009.37308035

Hofstetter, R., Zhang, J. Z., & Herrmann, A. (2018). Successive Open Innovation Contests and Incentives: Winner-Take-All or Multiple Prizes? *Journal of Product Innovation Management*, *35*(4), 492–517. doi:10.1111/jpim.12424

Hong, H., Ye, Q., Du, Q., Wang, G. A., & Fan, W. (2019). Crowd characteristics and crowd wisdom: Evidence from an online investment community. *Journal of the Association for Information Science and Technology*, asi.24255. doi:10.1002/asi.24255

Horton, J. J., & Chilton, L. B. (2010). The labor economics of paid crowdsourcing. *Paper presented at the 11th ACM conference on Electronic commerce*. Academic Press. doi:10.1145/1807342.1807376

Hossain, M. (2012). Users' motivation to participate in online crowdsourcing platforms. *Paper presented at the 2012 International Conference on Innovation, Management and Technology Research*. Academic Press. doi:10.1109/ICIMTR.2012.6236409

Howe, J. (2006). The rise of crowdsourcing. *Wired magazine, 14*(6), 1-4.

Huang, Y., Vir Singh, P., & Srinivasan, K. (2014). Crowdsourcing New Product Ideas Under Consumer Learning. *Management Science*, *60*(9), 2138–2159. doi:10.1287/mnsc.2013.1879

Idson, L. C., Liberman, N., & Higgins, E. T. (2000). Distinguishing Gains from Nonlosses and Losses from Nongains: A Regulatory Focus Perspective on Hedonic Intensity. *Journal of Experimental Social Psychology*, *36*(3), 252–274. doi:10.1006/jesp.1999.1402

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and Problem-Solving Effectiveness in Broadcast Search. *Organization Science*, *21*(5), 1016–1033. doi:10.1287/orsc.1090.0491

Jiang, J., & Wang, Y. (2019). A Theoretical and Empirical Investigation of Feedback in Ideation Contests. *Production and Operations Management*, poms.13127. doi:10.1111/poms.13127

Johnson, R. E., Chang, C. H., & Yang, L. Q. (2010). Commitment and motivation at work: The relevance of employee identity and regulatory focus. *Academy of Management Review*, *35*(2), 226–245.

Jung, J. H., Schneider, C., & Valacich, J. (2010). Enhancing the Motivational Affordance of Information Systems: The Effects of Real-Time Performance Feedback and Goal Setting in Group Collaboration Environments. *Management Science*, *56*(4), 724–742. doi:10.1287/mnsc.1090.1129

Ke, W., Tan, C.-H., Sia, C. L., & Wei, K.-K. (2013). Inducing Intrinsic Motivation to Explore the Enterprise System: The Supremacy of Organizational Levers. *Journal of Management Information Systems*, *29*(3), 257–290. doi:10.2753/MIS0742-1222290308

Ke, W. L., & Zhang, P. (2010). The Effects of Extrinsic Motivations and Satisfaction in Open Source Software Development. *Journal of the Association for Information Systems*, *11*(12), 784–808. doi:10.17705/1jais.00251

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(6), 1534–1555. doi:10.1037/0278-7393.25.6.1534

Lanaj, K., Chang, C. H., & Johnson, R. E. (2012). Regulatory focus and work-related outcomes: A review and meta-analysis. *Psychological Bulletin*, *138*(5), 998–1034. doi:10.1037/a0027723 PMID:22468880

Leimeister, J. M., Huber, M., Bretschneider, U., & Krcmar, H. (2009). Leveraging crowdsourcing: Activation-supporting components for IT-based ideas competition. *Journal of Management Information Systems*, *26*(1), 197–224. doi:10.2753/MIS0742-1222260108

Li, X., Hsieh, J. J. P.-A., & Rai, A. (2013). Motivational Differences Across Post-Acceptance Information System Usage Behaviors: An Investigation in the Business Intelligence Systems Context. *Information Systems Research*, *24*(3), 659–682. doi:10.1287/isre.1120.0456

Liu, T. X., Yang, J., Adamic, L. A., & Chen, Y. (2014). Crowdsourcing with All-Pay Auctions: A Field Experiment on Taskcn. *Management Science*, *60*(8), 2020–2037. doi:10.1287/mnsc.2013.1845

Locke, E. A., & Latham, G. P. (2004). What should we do about motivation theory? Six recommendations for the twenty-first century. *Academy of Management Review*, *29*(3), 388–403. doi:10.5465/amr.2004.13670974

Lukyanenko, R., & Parsons, J. (2018). Beyond Micro-Tasks. *Journal of Database Management*, *29*(1), 1–22. doi:10.4018/JDM.2018010101

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, *131*(6), 803–855. doi:10.1037/0033-2909.131.6.803 PMID:16351326

Majchrzak, A., & Malhotra, A. (2013). Towards an information systems perspective and research agenda on crowdsourcing for innovation. *The Journal of Strategic Information Systems*, *22*(4), 257–268. doi:10.1016/j.jsis.2013.07.004

Mullan, E., Markland, D., & Ingledew, D. K. (1997). A graded conceptualisation of self-determination in the regulation of exercise behaviour: Development of a measure using confirmatory factor analytic procedures. *Personality and Individual Differences*, *23*(5), 745–752. doi:10.1016/S0191-8869(97)00107-4

Nambisan, S., & Baron, R. A. (2009). Virtual Customer Environments: Testing a Model of Voluntary Participation in Value Co-creation Activities. *Journal of Product Innovation Management*, *26*(4), 388–406. doi:10.1111/j.1540-5885.2009.00667.x

Nambisan, S., & Baron, R. A. (2010). Different Roles, Different Strokes: Organizing Virtual Customer Environments to Promote Two Types of Customer Contributions. *Organization Science*, *21*(2), 554–572. doi:10.1287/orsc.1090.0460

Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, *21*(1), 34–77. doi:10.1080/10463281003765323

Parker, S. K., Williams, H. M., & Turner, N. (2006). Modeling the antecedents of proactive behavior at work. *The Journal of Applied Psychology*, *91*(3), 636–652. doi:10.1037/0021-9010.91.3.636 PMID:16737360

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, *88*(5), 879–903. doi:10.1037/0021-9010.88.5.879 PMID:14516251

Rhoades, L., Eisenberger, R., & Armeli, S. (2001). Affective commitment to the organization: The contribution of perceived organizational support. *The Journal of Applied Psychology*, *86*(5), 825–836. doi:10.1037/0021-9010.86.5.825 PMID:11596800

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2007). Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology*, *43*(6), 933–946. doi:10.1016/j.jesp.2006.10.014

Runco, M. A. (2004). Everyone has creative potential. doi:10.1037/10692-002

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. doi:10.1006/ceps.1999.1020 PMID:10620381

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *50*(1), 68–78. doi:10.1037/0003-066X.55.1.68 PMID:11392867

Schooler, W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology. General*, *122*(2), 166–183. doi:10.1037/0096-3445.122.2.166

Shen, X.-L., Lee, M. K., & Cheung, C. M. (2014). Exploring online social behavior in crowdsourcing communities: A relationship management perspective. *Computers in Human Behavior*, *40*, 144–151. doi:10.1016/j.chb.2014.08.006

Slot, J. H. (2013). *Crossing boundaries: Involving external parties in innovation: CentER*. Tilburg University.

Stryker, S. (1987). Identity theory: Developments and extensions. *Bulletin of the British Psychological Society*, *37*, 123.

Taggar, S. (2001). Group composition, creative synergy, and group performance. *The Journal of Creative Behavior*, *35*(4), 261–286. doi:10.1002/j.2162-6057.2001.tb01050.x

Terwiesch, C., & Xu, Y. (2008). Innovation Contests, Open Innovation, and Multiagent Problem Solving. *Management Science*, *54*(9), 1529–1543. doi:10.1287/mnsc.1080.0884

Tierney, P., Farmer, S. M., & Graen, G. B. (1999). An examination of leadership and employee creativity: The relevance of traits and relationships. *Personnel Psychology*, *52*(3), 591–620. doi:10.1111/j.1744-6570.1999.tb00173.x

Wallace, C., & Chen, G. (2006). A multilevel integration of personality, climate, self-regulation, and performance. *Personnel Psychology*, *59*(3), 529–557. doi:10.1111/j.1744-6570.2006.00046.x

Wang, Q., Feng, J., Jiang, X., & Xie, J. (2019). Multiple-Winner Award Rules in Online Procurement Auctions. *Production and Operations Management*, *28*(10), 2533–2551. doi:10.1111/poms.13061

Xu, F., Wang, X., & Xue, B. (2019). The Differential Effects of Interpersonal Justice and Injustice on Computer Abuse. *Journal of Database Management*, *30*(3), 1–17. doi:10.4018/JDM.2019070101

Zhang, C., Hahn, J., & De, P. (2013). Continued Participation in Online Innovation Communities: Does Community Response Matter Equally for Everyone? *Information Systems Research*, *24*(4), 1112–1130. doi:10.1287/isre.2013.0485

Zhang, Y., & van der Schaar, M. (2012). Reputation-based incentive protocols in crowdsourcing applications. *Paper presented at the INFOCOM 2012*. IEEE. doi:10.1109/INFCOM.2012.6195597

Zheng, H. C., Li, D. H., & Hou, W. H. (2011). Task Design, Motivation, and Participation in Crowdsourcing Contests. *International Journal of Electronic Commerce*, *15*(4), 57–88. doi:10.2753/JEC1086-4415150402

Lingfei Zou is currently a product manager at Huawei Technologies Co. Ltd. He obtained his PhD degree in Information Systems from City University of Hong Kong. Lingfei is interested in research areas such as corporate IT architecture, supply chain IT solutions, and data warehouse design.

Shaobo Wei is an Assistant Professor in the School of Management at the University of Science and Technology of China (USTC). He obtained his Ph.D. from USTC and City University of Hong Kong. His research focuses on supply chain integration and IT business value creation. He has published papers in journals such as Journal of Operations Management, Journal of Business Ethics, Decision Sciences, and International Journal of Logistics Management, as well as in conference proceedings, including those of the International Conference on Information Systems.

Weiling Ke is a professor in College of Business at Southern University of Science and Technology. She holds a Ph.D. from the National University of Singapore. Her primary research activities address the management of IT-enabled innovations. Specifically, her research seeks to understand issues related to supply chain integration, enterprise systems and open innovations. She has published with a variety of journals, including MIS Quarterly, Journal of Operations Management, Journal of Management Information Systems, Journal of the Association for Information Systems, Journal of Business Ethics and Personnel Psychology.

Kwok Kee Wei is Provost at Singapore Institute of Management. He is a Fellow of the Association of Information Systems (AIS) and was President of that Association in 2003/4. He was awarded the AIS LEO Award for Lifetime Exceptional Achievement in Information Systems in 2015. Professor Wei has served as Senior Editor of MIS Quarterly. He has also played major roles in ICIS and PACIS. He has published more than 200 papers. Based on Scopus, his works have been cited over 9280 times and his h-index is 44 (Dec 2019).

Journal of Database Management
Volume 31 • Issue 3 • July-September 2020

The Rise of NoSQL Systems:
Research and Pedagogy

Akhilesh Bajaj, University of Tulsa, USA
Wade Bick, University of Tulsa, USA

ABSTRACT

Transaction processing systems are primarily based on the relational model of data and offer the advantages of decades of research and experience in enforcing data quality through integrity constraints, allowing concurrent access and supporting recoverability. From a performance standpoint, they offer joins-based query optimization and data structures to promote fast reads and writes, but are usually vertically scalable from a hardware standpoint. NoSQL (Not Only SQL) systems follow different data representation formats than relations, such as key-value pairs, graphs, documents or column-families. They offer a flexible data representation format as well as horizontal hardware scalability so that Big Data can be processed in real time. In this review article, we review recent research on each type of system, and then discuss how teaching of NoSQL may be incorporated into traditional undergraduate database courses in information systems curricula.

KEYWORDS

Column Family Database, Database Consistency, Database Course, Database Partitioning, Graph Database, Information Systems Curriculum, JSON Database, Key Hashing, Key-Value Database, NoSQL, Relational Database, XML Database

INTRODUCTION

Starting from the 1980-s, the relational model of data has dominated the storage and retrieval of commercial data in the business world. Relational database management systems (DBMS) by vendors such as Oracle, IBM, Microsoft and open source offerings such as MySQL have evolved into stable platforms that use the SQL (Structured Query Language) standard for developing online transaction processing applications (OLTP), which constitute the bulk of information systems used by businesses.

However, the explosive growth of Web 2.0 and the Internet of Things (IoT) has led to a proliferation of massive amounts of data, often termed big data, that needs to be processed in real time (van der Aalst, 2018). The structuring of information into normalized tables with subsequent joins for querying, the overhead required for checking integrity constraints, and the enablement of data recoverability has led to performance bottlenecks when big data is stored on relational DBMSs. The first NoSQL (Not only SQL) applications were developed by the initial users of big data, such as Google that began development of Bigtable in 2004 (Whitchcock, 2005). Over time, several different NoSQL data models have emerged based on the needs of different domains. In each case, they have emerged because of

DOI: 10.4018/JDM.2020070104

faster performance and support of evolving data schemas. As of now, a plethora of NoSQL systems exist, though relational DBMSs are still by far the most popular (see usage data for various DBMSs at https://db-engines.com/en/ranking). As each of these systems has evolved, they have raised their own set of research questions.

From a teaching standpoint, the undergraduate curricula in Information Systems have contained a course on Database Management starting from the 1980s, and the most recent ACM IS 2010 curriculum specifies that this course largely deal with designing a normalized relational schema and building applications using the structured query language (SQL) standard (Topi et al., 2010). However, recent industry trends indicate that the adoption of NoSQL systems is increasing by organizations who are increasingly willing to try out these new data stores in a polyglot approach to managing their data storage (Simone, 2019). At the same time, many of the cloud vendors that offer NoSQL stores also offer relational stores which are increasing in popularity. The CIO of Amazon describes how their relational Aurora product is the fastest growing service in the history of Amazon Web Services (AWS) in (Vogel, 2019).

Given this changing landscape, a review of open research questions in SQL and NoSQL systems is needed. Another question that needs exploration is: should IS curricula adopt NoSQL as part of their database course, perhaps at the expense of teaching relational concepts? In this work we describe current and evolving research questions for each type of NoSQL system. Next, we contrast relational versus NoSQL systems, based on their application development methods, their application domains and the level of standards available. Using this review, we analyze how NoSQL can fit into an IS database curriculum over the near to midterm horizon, and if it is appropriate to replace relational concepts with NoSQL concepts in a core database course.

The rest of this paper is organized as follows. In Section 2, we describe the different categories of data models available today and describe open research questions for each model. In Section 3, we describe the application development methods commonly used for each type, as well as the domains of common application. Section 4 contains an analysis of appropriate curricular content for each type in an undergraduate IS (Information Systems) curriculum. We conclude in Section 5 with a discussion on the future importance of each type of system.

REVIEW OF DIFFERENT DATA MODELS AVAILABLE

First, we describe research related to consistency and availability issues, followed by data partitioning research. Next, we look at the following data models that are in wide use today: the relational model, the key-value (KV) model, the column family model, the graph model, the JSON model and the XML (eXtensible Markup Language) model.

ACID, BASE and the CAP Theorem

The consistency, availability, and partition tolerance (CAP) theorem was introduced in the early 2000-s as the size of the Internet was exploding (Gilbert & Lynch, 2002). In its simplest form, it asserts that a database implementation can only maintain two of the three concepts listed in the CAP acronym. For example, if a database system prioritizes consistency and partition tolerance, it will sacrifice some amount of availability. Trying to maintain high availability and partition tolerance would cause consistency issues. Network partitioning is a requirement for database systems being accessed today over networks, so partition tolerance is a basic requirement. Hence the tradeoff when building or selecting a system is considering the continuum with guaranteed consistency on one end, and guaranteed availability on the other end (Brewer, 2012).

The well-known ACID (Atomicity, Consistency, Isolation, and Durability) model signifies the four requirements for transactions in more traditional relational systems (Gray & Reuter, 1992). Systems that follow a pure ACID model are at one end of the CAP continuum, where consistency is always guaranteed. However, in a big data context, this occurs at the expense of availability. At the

other end of the CAP continuum is BASE (Basically Available, Soft-state, Eventually consistent) (Pritchett, 2008). This is a set of design principles where availability is paramount, as in big data real time systems where data partitions ensure that locally consistent data is always available, but global consistency can be accomplished at a later time (V. Gudivada, Apon, & Rao, 2018). Collectively, NoSQL systems tend to be on the side of BASE to different degrees.

The highest level of consistency is strict serializability of transactions. This assures that all transactions can be ordered strictly, so that lost updates and dirty reads are eliminated. For non-transactional systems, this is called linearizability or strong consistency (Davoudian, Chen, & Liu, 2018). Sequential consistency is the next lower level where eventually a total ordering is created, but a read operation may see a dirty value. Since providing consistency over a distributed system is very expensive, many NoSQL systems use atomic aggregates where denormalized data is stored together on the same node. Eventual consistency is the next level of relaxation of consistency, where ordering is not enforced, but ultimately all the copies of the data converge (Vogels, 2009). This allows for real time speeds of big data stores, but dirty writes and reads are a possibility here. Strategies used here include the last write wins rule where every write is timestamped. In some variants of this, the values of each attribute are timestamped, for example in Apache Cassandra and Amazon Dynamo (Lakshman & Malik, 2010). In addition, data can also be repaired while reading inconsistent values. In this strategy, the read operation for a data item gets to see all the data values with their timestamps on different nodes, and then sends back updates to each node with the latest data value (DeCandia et al., 2007). This strategy is used in several systems, including Dynamo, Cassandra and Voldemort.

There are different approaches to data replication, for example, master-slave (Couchbase, MongoDB, Espresso), master-master (BDR for PostgreSQL, GoldenGate for Oracle) and masterless (Dynamo, Cassandra). In a master-slave replication architecture, database clients send the data to one known master partition, and the system in turn updates the slave partitions at a later time. As pointed out recently in (Gonzalez-Aparicio, Younas, Tuya, & Casado, 2019), pushing the tradeoff envelope between availability and consistency for NoSQL stored continues to be an ongoing research area.

Data Partitioning Schemes

Data in distributed systems can be split horizontally, vertically or functionally. Horizontal splits occur when the data schema is preserved across each partition, but only a small subset of data is on each segment or *shard.* Vertical partitioning occurs when a subset of fields in the schema are in each partition, usually based on field access performance of the system. Functional partitioning occurs when the entire schema is split based on some functionality, for example all accounts payable records may be in one partition.

When data is partitioned, the performance scaling is usually horizontal, so that subsequent shards or partitions are loaded onto low cost hardware, and scaling up is indefinite. The design goals for partitioning strategies include minimizing inter-partition requests, load balancing of read and write requests across nodes after normalizing for node capacity and finally facilitating the insertion or removal of new nodes in the system with minimum disruption to partitions (Huang, Wang, Zhong, Song, & Yu, 2015; Schall & Härder, 2015; Stonebraker, Brown, Zhang, & Becla, 2013).

Key based sharding is the most common method of horizontal data partitioning, except in the case of graph databases, where graph vertices that are strongly connected are grouped onto one node. Research in key based sharding may be divided into workload-unaware and workload-aware sharding. Workload-unaware sharding is static in nature, and optimizes storage. Example strategies include range-based sharding where key ranges determine the destination partition for a data item as in MongoDb (Kookarinrat & Temtanapat, 2015). The disadvantages here are that some partitions may have more data since distribution of data items amongst range buckets may not be uniform, and the maintenance of a central mapping structure is required that maps range values to actual nodes.

Hashing is another method, where data item keys are hashed and the hash value is used to determine the storage node. A disadvantage here is the increase in multi partition queries, since data

items that are queried together often may have very different hash values for their keys. Another disadvantage of simple hashing is the removal and insertion of nodes. A solution to this is consistent hashing, where nodes are also given several virtual positions based on a hash function and data items are allocated based on proximity of their hashed key value to the hashed value of the node. This distributes the load so that the removal or addition of nodes is easier. Hashing schemes to meet the design goals of partitioning, described above, continue to be an ongoing area of research. For example, a consistent hashing scheme with bounded loads is proposed in (Mirrokni, Thorup, & Zadimoghaddam, 2018) for dynamic load balancing, where the goal is to minimize the maximum load on any node, and to minimize the number of data items to be moved if any node is added or removed.

Relational Model

Proposed in 1970 (Codd, 1970), the relational model is the most popular model in existence for the storage and access of alpha-numeric data. Decades of research have created secure and reliable systems that allow concurrent access to users in an organization, secure recovery in case of software or hardware failure and declarative access to data via SQL. Role based access is also a feature of relational models. While earlier iterations were optimized for more writes to the database, also called online transaction processing (OLTP), newer iterations provide storage that is optimized for online analytic processing (Chaudhuri & Dayal, 1997), where roll-up and drill-down queries are optimized across measures of interest such as sales figures along different dimensions.

The relational schema needs to be usually defined at the onset of deploying an application, and is not amenable to frequent evolution (V. N. Gudivada, Rao, & Raghavan, 2016). Read and write performance can be scaled up in relational systems via vertical scaling (adding more expensive hardware) though this is expensive and reaches a ceiling quickly (Vaquero, Rodero-Merino, & Buyya, 2011). The relational database market is over $40 billion and expected to constitute over 80% 0f the entire database market till 2022(Doherty, 2013; Wells, 2019). Over 100 relational database management system vendors exist, including Oracle, Microsoft SQL server, IBM DB2 and MySQL.

From a research standpoint, relational systems are still an active area of research in many areas. Hardware advances, such as NVMe (Non Volatile Memory Express) drives that deliver significantly lower latency and higher bandwidth for I/O applications continue to appear, and exploring how these advances impact relational database design going forward is an important research question. For example, it was found in (Xu et al., 2015) that NVMe backed systems delivered up to eight times faster client-side performance over enterprise-class SATA (serial advanced technology attachment) solid state drives. Indexing schemes to get better read and write performance from relational stores also continue to be an area of research. A survey of 48 indexing techniques was presented in (Gani, Siddiqa, Shamshirband, & Hanum, 2016), with a view to managing big data on different data models. For example, a compact steiner index was presented in (Li, Feng, Zhou, & Wang, 2011) to facilitate keyword searches in relational databases.

Another ongoing area of research is associating ontologies with relational data. This has become easier since OWL (One World Language or Web Ontology Language) became the de-facto standard for creating class-based ontologies. A survey of tools to map ontologies to relational data is presented in (Moldovan et al., 2015). Ontologies can be used to create more natural query interfaces. For example, usage of an OWL schema to create fuzzy queries on scalar data in a relational database is presented in (Martínez-Cruz, Noguera, & Vila, 2016). A natural language query interface using ontologies is proposed in (Lei et al., 2018). Mapping semantic constraints in OWL to relational database implementation mechanisms such as triggers is another current area of research (Achpal, Kumar, & Mahesh, 2016). A useful summary of ontologies for knowledge modeling and information retrieval is available in (Munir & Anjum, 2018).

A third area of research is the emergence of NewSQL databases that combine the throughput performance of NoSQL systems with the ACID properties of relational systems (Duggirala, 2018). Many of these use some form of in-memory storage to enhance I/O performance. Application domains

of NewSQL databases are described in (Almassabi, Bawazeer, & Adam, 2018), and include archival data warehouses, real-time and streaming systems and systems that support time stamped information.

Key Value Model

Data in the key value model is stored as simple key-value pairs, where a key is generated using any partitioning algorithm, such as a hash algorithm, and the value is stored as an opaque binary BLOB (binary large object). Once the application has the key, the BLOB can be made to appear as a list, or a dictionary, or another structure, depending on the system and the application. The performance of key based querying is independent of the amount of data stored. The KV data model is very extensible in a deployed application, so that the value in the next KV pair does not have to match the format of the value of an earlier KV pair. KV systems lend themselves well to horizontal scaling (adding cheaper machines to a pool of resources) since data can be sharded more easily and shards can be stored on different servers. Aggregate querying is also facilitated using a MapReduce program that runs queries in parallel across multiple shards and combines the result (Dean & Ghemawat, 2008). Big data applications that require real time reading and writing performance in milliseconds benefit from the KV model. Over 50 vendors exist for KV applications, including Aerospike, Redis, Memcached, Riak and DynamoDB.

While applications such as web session logs and shopping carts need primarily key based querying, most applications require some form of value-based querying. Rather than require the application to parse the value object and write query code, recent KV systems support indexing and querying values of certain data types, such as lists (redis and Aerospike) or table rows (HyperDex and Spinnaker).

In-memory KV systems (such as Memcached and Voldemort) offer very fast access, and are used for cloud caching or web session information. Persistent systems (such as Riak and Oracle NoSQL) offer solid state disk storage, while hybrid systems (such as Aerospike and Redis) offer in memory performance with persistent storage if certain conditions are met (Davoudian et al., 2018). Ongoing research in KV systems focuses on developing access control models for data access (Moreno, Fernandez, Fernandez-Medina, & Serrano, 2018) and teasing out data semantics in the value portion (Rudnicki, Cox, Donohue, & Jensen, 2018). While KV systems offer fast query performance, the issue of data quality in the value portion is an issue since integrity constraints have to be written at the application level.

Column Family Model

Unlike relational systems that are relatively immutable when it comes to schema evolution, column family systems model data as multiple families of columns, where usually intra-column mutability is facilitated. Column families are not easily mutable once created for an application, but columns within a family can be easily added or taken away. This allows for the data schema to change easily as the application evolves, so that each "row" within a column family can have a flexible number of columns. Columns within a family are physically stored together and easily accessed quickly. In essence, a nested KV model can be used to represent a column family schema, where each "row" has one key, and then nested key-value pairs to store the data. New values do not replace old ones, but are instead time stamped, so a triple <*row-key, column-key, timestamp*> can be used to access data (Davoudian et al., 2018; Wiese, Waage, & Brenner, 2019). Computing aggregate metrics on large databases that have temporal data, such as financial applications, is a classic application of column family databases. Optimizing aggregate queries is an ongoing research question for column family systems (Storey & Song, 2017). For example, storing the results of expected searches on big data, so that, say, all in-box mail messages related to 'picnics' are in one row and all messages related to user 'johnDoe21' are in another row is another example of the application of column family databases. An aggregate search can then be run very fast for retrieving messages based on search criteria. Notable systems include Cassandra, HBase, BigTable and HyperTable.

The partitioning of column family systems can be by column families (vertical) or by rows (horizontal). Storage is usually in-memory, using a data structure such as an LSM (log-structured merge) tree for each column family (Wu, Xu, Shao, & Jiang, 2015). Copies of these trees are kept in persistent storage, usually immutably, so that the most recent copy is used to access the data. Updates become faster because of the immutability of old values in older copies. The MapReduce programming model pioneered by Google works well with the column family model, and there is ongoing research to alter the model to accommodate different types of domains. For example, processing Geospatial data in HBase using MapReduce is covered in (Gao, Yue, Wu, & Zhang, 2017). A two-phased MapReduce algorithm to facilitate replication of data in data warehouses is presented in (Barkhordari & Niamanesh, 2018). A summary of the different MapReduce variants used in Column Family systems is presented in (Seera & Taruna, 2018).

Graph Model

The native graph model stores data as vertices with edges. Usually vertices represent things or events, and edges represent relationships between these things or events. Attributes can be used to further describe the vertices and the edges. Ongoing research in this model involves two broad areas: storage of graph data, and optimizing deep queries.

Systems such as early versions of twitter FlockDB stored their graph data in adjacency lists on a relational system such as MYSQL. Adjacency lists store a single list of neighbor vertices for each vertices in a non-directed graph, and two lists (in and out) for each vertex for a directed graph. This storage causes data redundancy, since each link information is stored once for each vertex in the link. Updates are faster with this kind of storage over a matrix representation, since a vertex can be easily added by concatenation of a list for the new vertex. A variation of this is called an edge list, where the source and destination vertice IDs of each edge are stored in a two column relational database table, that may be indexed using, say, a B-Tree (Comer, 1979). In this case, data redundancy is lower since each piece of information is retained only once in the database.

The most popular graph storage format is a compressed sparse row (CSR) where two integer arrays are used (Qian, Childers, Huang, Guo, & Wang, 2018). The first is an edge array that maps an edge ID to the ID of its destination vertex, and the second is a vertex array that maps each vertex to the ID of its first outgoing edge. The main advantage of CSR is the contiguous storage of outgoing edges of each vertex in main memory, thereby leading to a reduced need for secondary storage access. The cost of edge insertion or removal is O(*m)* where *m* is the number of edges. CSR is commonly found in many graph engines such as GraphChi (Kyrola, Blelloch, & Guestrin, 2012) where billions of edges can be managed. Recent advances in the area include G-store where trillion edge graphs are optimized (Kumar & Huang, 2016).

The second area of research in graph databases is the querying of large graphs. A good survey of modern query languages for graph databases is presented in (Angles et al., 2017). Graph queries can be real time queries, relating to a small portion of a large graph, or off line computations that optimize queries covering over large portions of the graph. The storage mechanisms for these two needs are different, in that the former uses an online model, where updates and queries over small portions of the graph are optimized, while the latter may actually take the graph offline and optimize for quick traversal over the graph. Systems like GraphCHI and Microsoft Trinity support both aspects of graph querying. The three most popular query languages for graphs are SPARQL (Harris & Seaborne, 2013), Neo4J's Cypher and Apache's Gremlin. All three employ graph pattern matching, where graph patterns with variables for edges and nodes are matched with actual data in a graph schema. Both homomorphic (SPARQL) and isomorphic (Neo4j) match strategies may be supported when it comes to criteria for finding matches between patterns and actual data. A second aspect of queries is finding paths between nodes that satisfy certain criteria such as existence of the path, satisfaction of a regular expression or more complex queries (Libkin, Martens, & Vrgoč, 2016).

The selection of proper standardized semantics for these aspects of graph queries, across languages, is an ongoing research issue.

Document Model (JSON and XML)

The data model used here is a variant of the key-value model, where the value is not an opaque BLOB, but is instead a structured file format, usually JSON (JavaScript Object Notation) or XML (eXtensible Markup Language). Each value is a file with its own attributes or fields. The data schema can evolve since subsequent files need not have the same fields as prior ones. Documents that are semi structured with widely varying fields whose field structure and data types can change are suitable for the document model. Indices can be built for different attributes. Aggregate querying across document collections is optimized in these systems (V. N. Gudivada et al., 2016). Over 40 vendors for document systems exist, including MongoDB, CouchDB, CouchBase, DynamoDB. Since the difference between the KV model and the document data model lies in how the value is stored, many systems offer a choice of both. An example query language in this domain is N1QL (non first normal form query language) in CouchBase that supports declarative querying across documents.

Database management systems supporting the XML standard go beyond just document model systems. XML schemas can be validated using namespaces or document type definitions (DTD), so a schema can be enforced on data. At the same time, because of the extensibility, a new DTD can be defined easily, thereby supporting the evolution of the schema as the domain changes. Rule-based data validation can also be done using the Schematron standard (Van der Vlist, 2007). Unlike add-on modules to some systems that store XML files as Character Large Object (CLOB) files, data is stored in these systems natively as an XML file (V. N. Gudivada et al., 2016).

XPath and XQuery extensions are standards that can support reads and writes to the data. Other accompanying standards such as XForms and XProc also may be implemented in these systems, thereby supporting an evolving standard in data processing and application development. Standards based storage, access and processing allows applications to be more easily moved across different native XML systems and allows for better recovery and security policies. Examples of native XML systems include MarkLogic, BaseX, and Sedna. An ongoing area of research in document-based systems is polyglot storage, so that the relationships between documents may be stored in, say, a graph representation, while the data itself is stored in an underlying JSON or XML format (Oliveira & del Val Cura, 2016; Sadalage & Fowler, 2013).

Table 1 summarizes the ongoing research questions in the different NoSQL data models discussed in this section.

Next, we briefly describe the application development methods that may be used for the different models, and how this may impact pedagogical content in IS.

APPLICATION DEVELOPMENT AND DOMAINS

Relational Model

There is a great deal of mathematical theory associated with the design of relational schemas, primarily based on the notion of functional dependencies between attributes in the schema (Kent, 1983). Reducing redundancy of the data that is stored with appropriate forms of normalization, the relational schema is widely taught in undergraduate database courses in both IS and computer science curricula. Semantic models like the extended entity relationship model (EER) have also been developed to bridge the gap between natural language descriptions of domain requirements and a relational schema design (Engels et al., 1992; Teorey, Yang, & Fry, 1986). Thus, the notion of designing a schema for improved data quality in the application is well understood when using the relational model. Inbuilt constraint enforcement mechanisms such as primary key and referential integrity constraints and triggers are built in to most relational OLTP systems, further enhancing their

Table 1. Ongoing research areas in NoSQL systems

NoSQL Area of Research	Ongoing Research Questions
Consistency/Availability	Algorithms to push the consistency-availability envelope
Partitioning Schemes	Schemes to reduce inter-partition requests, load balance read/write requests across nodes and reduce disruption to partitions when nodes are inserted or removed.
Relational Model	Memory management changes because of advances in secondary storage latency and bandwidth. Indexing schemes for improved read/write performance. Mapping of ontologies to relational schemas and integrity constraints such as referential integrity and triggers. Increase throughput and read/write latency to NoSQL levels while still allowing for relational schemas.
Key Value Model	Providing data access security to roles. Capturing application semantics and integrity constraints in the value portion.
Column Family Model	Balance latency of aggregate queries with storage requirements. Apply column family data model to different domains. Modify MapReduce algorithms for increased performance.
Graph Model	Formulate graph storage methods so that storage space is lower, link insertion and removal is faster and access is faster. Standardize and optimize query language constructs for small segment queries and larger traversal queries.
Document Model	Develop systems with polyglot storage to capture links between document segments. Evolving standards in JSON to match those of XML for document schema validation and querying.

abilities to facilitate data quality for transaction processing applications with alpha numeric data, as is currently found in most functional business areas (Reiter, 1988). Role based security mechanisms with a fine-grained level of control, the ability to support concurrent users and recovery from system faults are also built into these systems.

Typically, the performance scaling for these systems is vertical, in that faster speed is usually obtained through increasing the main memory, disk access speeds and processing speeds of existing hardware. This limitation implies that relational systems typically are not suitable for very high throughput applications where input data is generated from programs, such as data from Internet of Things (IoT) appliances, or webserver logs. Instead, they are more suitable for applications where data is being input by several humans or machines in a typical business setting, such as customer relationship management, banking, etc.

Column Family Systems

Column family systems like Cassandra do allow the creation of a primary key. The key is used to ensure uniqueness of records, and also to determine the location of the partition that will house that record. This facilitates searches by key. However, there is no support for referential integrity, triggers or other data quality checks. This means that data is often stored in duplicate locations, leading to redundancy but very fast reads since no joins are needed.

Columns within a family can be dynamically or statically typed, and families are extensible, which supports applications with an evolving data schema. An application development framework for column family systems is more involved than KV systems, since there is some structure to data (columns), though normalization is not an option. Since time stamps are assigned to each update/delete/insert mutation with micro second granularity, big data applications such as financial transactions and streaming multimedia where time stamping is important are suitable for column family systems.

Graph Systems

The logical model for graph systems is vertex-edge information which is often physically implemented in different formats. Support for transactions or data partitioning varies based on the system, and there seems to be no converging standard at present for how to develop applications. Graph systems are very useful where large datasets of vertex-edge information is needed, as in recommender systems, security access systems, social networks, etc.

Document Systems

Document systems that are non-XML usually store information as Key-JSON file pairs. Systems like CouchDB offer both a key-binary value pair and a key-JSON file storage model. Systems like MongoDB and CouchDB also allow system specific document validation features, though no standard exists. JSON has advantages over XML in that it can be parsed with JavaScript functions versus using an XML parser and it supports arrays, so less coding is required to move data to and from a JSON file. For AJAX style applications, JSON is easier to use than XML since no looping through the XML document object model is required. JSON files also tend to be smaller and faster to transmit and retrieve in the case of big data applications. Applications that require flexible schemas, have mainly alpha numeric data and require fast throughput on the web are good candidates for JSON based document systems.

XML systems store data in native XML format, which means they can also store data with varying degrees of structure, similar to JSON based systems. However, in XML, attributes can be put into the opening tags in XML, thereby delineating metadata from actual data. Application development is the most mature here amongst all the NoSQL systems. First, data schema design can be undertaken using constraints such as primary key and referential integrity constraints, with automated checking using Schematron. Every new documents' schema can also be checked against a DTD or a namespace (Bray, Hollander, & Layman, 1999).

Second, query-based applications can be written using the Xpath/XQuery standards, so applications are portable across systems. Third, front end client applications can be built using XForms, that supports common graphic user interface (GUI) architectures. Standards-based development permits the usual advantages of greater availability of technical personnel, portability across different XML platforms and easier outsourcing of development. XML based systems are slower than JSON based systems, but provide better support for document heavy applications. Typical applications for XML systems include workflow systems with very long transactions and evolving schemas, automated supply chain systems and any application with large amounts of semi-structured data that is mainly alphanumeric.

Key Value Systems

KV systems are applicable in very high throughput, big data environments where real time response is essential. Examples include web session data, auction systems, dynamic customized user interface generation for a user, with thousands of users concurrently logged in, recommendation engines, etc. The design of data in KV systems is very flexible, since the storage structure is an opaque binary object. The selection of what type of key to use is somewhat important, though often a hash key may be sufficient. Thus, in terms of data quality enforcement, all mechanisms must be coded in the application code, unlike the relational model with built in constraints and triggers. Applications where the primary lookup is via keys are especially well suited for KV systems, especially if the data schema evolves constantly.

Table 2 summarizes the discussion above and contrasts the application development in he different data models.

Next, we explore the extent to which NoSQL needs to be incorporated into a database course in Information Systems.

Table 2. Application domains and development for relational and NoSQL data models

Data Model	Logical Representation	Major Characteristics	Application Development	Application Areas
Relational Model	Relations, Attributes	Integrity constraints, fine grained security, SQL standard, strong ACID support	Strong normal form design theory, SQL standard to write code, vertical scaling	OLTP and OLAP systems, multiuser systems usually within an organization. Query speed and throughput not as critical as data quality. Typically not used for big data applications.
Column Family Model	Columns which can be dynamically added to a family	Primary key, but no normalization support	Lots of duplication of data, since each query often becomes its own family. Leads to very fast reads.	Big data applications where time stamping is important, and data still needs to be organized into columns. Includes financial applications and data warehouses.
Graph Model	Nodes and edges, with attributes for both	No normalization or other standardized integrity constraint support. Supports graph oriented operations such as shortest-path, minimum spanning tree, cluster analysis, etc.	Identifying what to model as nodes versus edges, what attributes to capture. Developing queries using the language provided by the particular system.	Big data application where richness of relationships is important. Examples include recommender systems and social network analysis.
JSON Systems	Key-Value pairs with a JSON file as each value	Use JSON syntax to capture data structure, which is lightweight, though lacks meta data specification support.	Structure the data as JSON files, and Uses JavaScript to process quickly. Data integrity constraint checking is unstandardized and needs to be coded.	Web based big data applications where there is lots of network traffic and the data model evolves, but is mainly alphanumeric. Examples include web based sharing of large numbers of documents, text searches on large databases, etc.
XML Systems	Key-Value pairs with an XML file as each value	Use XML syntax to model data structure, structure can be validated against DTD or namespace, and values can be validated using a standard like Schematron.	Structure the data fields, and list integrity constraints. Develop applications using Schematron and XForms. Set up queries using XQuery and XPath.	Data applications where data sharing is essential and the data model evolves, but where real time speed is not critical and data quality and integrity checking is more important. Examples include multi organizational supply chain systems, and other OLTP software that spans multiple databases.
Key-Value Systems	Key-value pairs where the value is an opaque binary value	Keys are partitioned and usually hash values. Maximum flexibility since each value can be anything, including multimedia. No checking for integrity constraints	Decide on the key partitioning scheme and which objects to implement as values. Decide on data partition mechanisms so local data is always consistent for application.	Very fast real time performance in big data sets for writes and reads, independent of size of data. Works well for distributed applications, as long as localized data partitions are consistent. Any kind of application that requires non alpha numeric data with very flexible structure. Examples include large scale video databases, music databases, image databases, etc.

HOW SHOULD NOSQL DATABASES BE INCORPORATED INTO AN IS CURRICULUM?

The overall market for big data software and services is estimated to grow from $ 42 billion to $103 billion from 2017 to 2027 (Columbus, 2018). The market for NoSQL databases is relatively small, and though growing relatively fast, likely to remain at under 10% of the overall database market till 2022, as per a report in (Wells, 2019). An analysis of data published by IDC, the market intelligence firm, indicates that the total database market in 2022 will be approximately $40.4 billion with relational databases accounting for about $33 billion (Mullins, 2019). The NoSQL market in the same analysis is estimated to be approximately $3.7 billion in 2022. This implies that storage and usage of transactional data, stored in a relational database, in the business world will dominate information generation and usage in the near future.

The ACM/AIS curriculum for IS schools recommends a course termed "Data and Information Management" (IS 2010.2) with an intent to provide "an introduction to the core concepts in data and information management" (Topi et al., 2010). The recommended learning goals incorporate conceptual modeling (*e.g.* entity-relationship model), relational database design with normalization levels, with a primary focus on OLTP systems. Understanding OLAP, semi-structured and unstructured data management is listed in goals 19 and 21 (out of a total of 21 learning goals). One of the authors of this paper has been teaching an IS database class at various US based Universities since 1995, and this is typical of the current syllabi used in IS curricula, where ER or EER (extended ER) modeling, normalization, SQL and application development on a relational platform are the bulk of the course, with one or two lectures regarding warehousing and semi-structured or unstructured data. These components conform to the well understood requirement that University level courses offer knowledge with a "long shelf life" that is independent of any particular IS artifact or system.

The discussion of NoSQL systems in Sections 2 and 3 of this work indicates that NOSQL has undergone a bottom-up evolution, where solutions were put together to solve performance and scaling issues in emerging web and big data scenarios. There is a lack of underlying theory guiding database design and standards similar to SQL, except to some degree in the case of XML systems. Application development on NoSQL systems also seems to be largely based on the features provided by each vendor, versus a set of standards. Finally, there are many different types of NoSQL systems, depending on the big data application domain.

An early work on incorporating NoSQL into a traditional database course argued that NoSQL needs to be added because of "industry changes" and suggested that students be asked to do the same project on relational and a NoSQL platform (Stanier, 2012). More recently, a case was proposed that could be used as a NoSQL project in a traditional database course, using CouchDB (Fowler, Godin, & Geddy, 2016). The goal behind this project was to show students how an evolving data schema can be captured in CouchDB, and sample JavaScript queries were provided. Pre- and post- test measures showed that students had a better understanding of why NoSQL is helpful in a social networking context and how it supports an evolving data schema. However, as shown in (Topi et al., 2010), the current AIS curriculum for a database course has little room to add new content, without taking away content. Further, the challenge is to teach "long shelf life" intellectual content, which can be a challenge in a landscape that evolved bottom-up and has few standards.

Should a new course for NoSQL be created? A recent work described the creation of such a course in the curriculum at Rose-Hulman Institute of Technology (Mohan, 2018). It was argued that in computer science, the emphasis on course creation is based on realistic use case scenarios and problem-based learning (Walker & Slotterbeck, 2002), and not just on theory-based learning. The course exposed students to the underlying motivations behind NoSQL, and a variety of NoSQL databases (document, graph, KV and Columnar). Students had to work on an application on each system and reflect on the differences and applicability of each system. A final project required

students to develop an information system using a NoSQL system they select. Student feedback was positive on the course.

While concepts such as scaling, big-data performance and multiple system application development can be covered in a computer science course, are they relevant in an IS curriculum? As per the 2010 IS Curriculum in IS, the goal of IS programs is to produce individuals who can use technology to make the enterprise work better (Topi et al., 2010) pp. 374. Given the fluid landscape of NoSQL models, and the continued preponderance of relational systems, we recommend that IS courses in the area of databases do not significantly change their content at present. However, the fast growth of NoSQL in big data applications implies that at the very least, IS student should be aware of the contrast between relational and NoSQL systems, and how application development would occur in a NoSQL scenario. One option here is to utilize a lecture in the traditional IS database course to discuss these issues, perhaps with example scenarios. A second choice would be to introduce a NoSQL course as a more technical elective in the curriculum, where students build applications in a variety of NoSQL systems, along the lines of the course described in (Mohan, 2018).

CONCLUSION

NoSQL systems evolved from the ground-up to help manage latency and throughput in big data applications where much of the data comes from automated sources, such as a database of web pages maintained by a search engine, or web server logs to customize user experience. Several data models, such as column families, key-value pairs, graphs and document-based systems have emerged to represent data in different domains. They share common traits such as emphasizing low latency and high throughput at the expense of data consistency and management of data integrity. In this work, we highlighted ongoing research questions for each of these data models, and summarized current techniques used to develop applications on the different data models' platforms.

The historic evolution of relational databases systems offers some guidance on the future of application development in the NoSQL landscape. The relational model was proposed in (Codd, 1970), and the relational calculus shortly thereafter (Codd, 1971). The emergence of SQL (Chamberlin, 1980; Pirotte, 1979) allowed the implementation of relational systems on the university campus and eventually into industry. At the same time, the rise of conceptual modeling (Chen, 1976) allowed the creation of an application development methodology that bridged the gap between end-user requirements in an enterprise and the design and implementation of transaction processing systems. Database courses in computer science curricula dealt with learning the architecture of relational DBMSs, from the point of view of the algorithms related to enforcing the ACID requirement. Courses in IS curricula evolved differently and dealt with translating business requirements into a normalized design, using an intermediate conceptual model such as the E-R model, and applying SQL to create applications on an existing DBMS platform.

In the area of NoSQL systems, there has been an absence of a top-down theory that has driven the evolution of the different data models. Bottom-up principles such as horizontal scaling, data partitioning and aggregating denormalized data have been the driving factors, motivated by real world data handling requirements that have mushroomed for specific big data applications. However, it is important to note that while big data application usage grows, the market for NoSQL systems is still projected to be a small fraction of the overall DBMS market, in the future. While University courses in computer science are starting to appear that discuss architectural issues, there has not been an application development methodology that has emerged, specific to NoSQL, that would warrant a course in IS curricula at present. IS students need to be aware of NoSQL principles and where NoSQL can be applied, but relational design and SQL continues to remain a core skill for IS majors that cannot be replaced.

REFERENCES

Achpal, A., Kumar, V. B., & Mahesh, K. (2016). Modeling Ontology Semantic Constraints in Relational Database Management System. *Paper presented at the International MultiConference of Engineers and Computer Scientists*. Academic Press.

Almassabi, A., Bawazeer, O., & Adam, S. (2018). Top NewSQL Databases and Features Classification. *International Journal of Database Management Systems*, *10*(2), 11–31. doi:10.5121/ijdms.2018.10202

Angles, R., Arenas, M., Barceló, P., Hogan, A., Reutter, J., & Vrgoč, D. (2017). Foundations of modern query languages for graph databases. *ACM Computing Surveys*, *50*(5), 68. doi:10.1145/3104031

Barkhordari, M., & Niamanesh, M. (2018). Chabok: A Map-Reduce based method to solve data warehouse problems. *Journal of Big Data*, *5*(1), 40. doi:10.1186/s40537-018-0144-5

Bray, T., Hollander, D., & Layman, A. (1999). Namespaces in XML. *World Wide Web Consortium.* Retrieved from http://www.w3.org/TR/1999/REC-xml-names-19990114

Brewer, E. (2012). CAP Twelve years Later: How the. *Computer*, *45*(2), 23–29. doi:10.1109/MC.2012.37

Chamberlin, D. D. (1980). *A summary of user experience with the SQL data sublanguage. IBM Thomas J*. Watson Research Division.

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *SIGMOD Record*, *26*(1), 65–74. doi:10.1145/248603.248616

Chen, P. P.-S. (1976). The entity-relationship model—Toward a unified view of data. *ACM Transactions on Database Systems*, *1*(1), 9–36. doi:10.1145/320434.320440

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, *13*(6), 377–387. doi:10.1145/362384.362685

Codd, E. F. (1971). A data base sublanguage founded on the relational calculus. *Paper presented at the 1971 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control*. Academic Press. doi:10.1145/1734714.1734718

Columbus, L. (2018). 10 Charts That Will Change Your Perspective Of Big Data's Growth. *Forbes*.

Comer, D. (1979). Ubiquitous B-tree. *ACM Computing Surveys*, *11*(2), 121–137. doi:10.1145/356770.356776

Davoudian, A., Chen, L., & Liu, M. (2018). A survey on NoSQL stores. *ACM Computing Surveys*, *51*(2), 40. doi:10.1145/3158661

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107–113. doi:10.1145/1327452.1327492

DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., . . . Vogels, W. (2007). Dynamo: amazon's highly available key-value store. *Paper presented at the ACM SIGOPS Operating Systems Review*. Academic Press. doi:10.1145/1294261.1294281

Doherty, S. (2013). *The Future Of Enterprise Data: RDBMS will be there*. Wired.

Duggirala, S. (2018). NewSQL Databases and Scalable In-Memory Analytics. *Advances in Computers*, *109*, 49–76. doi:10.1016/bs.adcom.2018.01.004

Engels, G., Gogolla, M., Hohenstein, U., Hülsmann, K., Löhr-Richter, P., Saake, G., & Ehrich, H.-D. (1992). Conceptual modelling of database applications using an extended ER model. *Data & Knowledge Engineering*, *9*(2), 157–204. doi:10.1016/0169-023X(92)90008-Y

Fowler, B., Godin, J., & Geddy, M. (2016). Teaching Case Introduction to NoSQL in a Traditional Database Course. *Journal of Information Systems Education*, *27*(2), 99–103.

Gani, A., Siddiqa, A., Shamshirband, S., & Hanum, F. (2016). A survey on indexing techniques for big data: Taxonomy and performance evaluation. *Knowledge and Information Systems*, *46*(2), 241–284. doi:10.1007/s10115-015-0830-y

Gao, F., Yue, P., Wu, Z., & Zhang, M. (2017). Geospatial data storage based on HBase and MapReduce. *Paper presented at the 2017 6th International Conference on Agro-Geoinformatics*. Academic Press. doi:10.1109/Agro-Geoinformatics.2017.8047040

Gilbert, S., & Lynch, N. (2002). Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. *Acm Sigact News*, *33*(2), 51–59. doi:10.1145/564585.564601

Gonzalez-Aparicio, M. T., Younas, M., Tuya, J., & Casado, R. (2019). Evaluation of ACE properties of traditional SQL and NoSQL big data systems. *Paper presented at the Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. Academic Press. doi:10.1145/3297280.3297474

Gray, J., & Reuter, A. (1992). *Transaction processing: concepts and techniques*. Elsevier.

Gudivada, V., Apon, A., & Rao, D. L. (2018). Database systems for big data storage and retrieval. In Handbook of Research on Big Data Storage and Visualization Techniques (pp. 76–100). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-3142-5.ch003

Gudivada, V. N., Rao, D., & Raghavan, V. V. (2016). Renaissance in database management: Navigating the landscape of candidate systems. *Computer*, *49*(4), 31–42. doi:10.1109/MC.2016.115

Harris, S., & Seaborne, A. (2013). sparql 1.1 query language. Recommendation. *World Wide Web Consortium*.

Huang, X., Wang, J., Zhong, Y., Song, S., & Yu, P. S. (2015). Optimizing data partition for scaling out NoSQL cluster. *Concurrency and Computation*, *27*(18), 5793–5809. doi:10.1002/cpe.3643

Kent, W. (1983). A simple guide to five normal forms in relational database theory. *Communications of the ACM*, *26*(2), 120–125. doi:10.1145/358024.358054

Kookarinrat, P., & Temtanapat, Y. (2015). Analysis of range-based key properties for sharded cluster of mongodb. *Paper presented at the 2015 2nd International Conference on Information Science and Security (ICISS)*. Academic Press. doi:10.1109/ICISSEC.2015.7370983

Kumar, P., & Huang, H. H. (2016). G-store: high-performance graph store for trillion-edge processing. *Paper presented at the SC'16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. Academic Press. doi:10.1109/SC.2016.70

Kyrola, A., Blelloch, G., & Guestrin, C. (2012). GraphChi: Large-Scale Graph Computation on Just a PC. *Paper presented at the 10th USENIX Symposium on Operating Systems Design and Implementation OSDI 12*. Academic Press.

Lakshman, A., & Malik, P. (2010). Cassandra: A decentralized structured storage system. *Operating Systems Review*, *44*(2), 35–40. doi:10.1145/1773912.1773922

Lei, C., Özcan, F., Quamar, A., Mittal, A. R., Sen, J., Saha, D., & Sankaranarayanan, K. (2018). Ontology-Based Natural Language Query Interfaces for Data Exploration. *IEEE Data Eng. Bull.*, *41*(3), 52–63.

Li, G., Feng, J., Zhou, X., & Wang, J. (2011). Providing built-in keyword search capabilities in RDBMS. *The VLDB Journal*, *20*(1), 1–19. doi:10.1007/s00778-010-0188-4

Libkin, L., Martens, W., & Vrgoč, D. (2016). Querying graphs with data. [JACM]. *Journal of the Association for Computing Machinery*, *63*(2), 14. doi:10.1145/2850413

Martínez-Cruz, C., Noguera, J. M., & Vila, M. A. (2016). Flexible queries on relational databases using fuzzy logic and ontologies. *Information Sciences*, *366*, 150–164. doi:10.1016/j.ins.2016.05.022

Mirrokni, V., Thorup, M., & Zadimoghaddam, M. (2018). Consistent hashing with bounded loads. *Paper presented at the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. Academic Press. doi:10.1137/1.9781611975031.39

Mohan, S. (2018). Teaching NoSQL Databases to Undergraduate Students: A Novel Approach. *Paper presented at the 49th ACM Technical Symposium on Computer Science Education*. Academic Press. doi:10.1145/3159450.3159554

Moldovan, D., Antal, M., Valea, D., Pop, C., Cioara, T., Anghel, I., & Salomie, I. (2015). Tools for mapping ontologies to relational databases: A comparative evaluation. *Paper presented at the 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*. Academic Press. doi:10.1109/ICCP.2015.7312609

Moreno, J., Fernandez, E. B., Fernandez-Medina, E., & Serrano, M. A. (2018). A Security Pattern for Key-Value NoSQL Database Authorization. *Paper presented at the 23rd European Conference on Pattern Languages of Programs*. Academic Press. doi:10.1145/3282308.3282321

Mullins, C. S. (2019). The New World Of Database Technologies. *Paper presented at the Data Summit*. Retrieved from http://conferences.infotoday.com/documents/331/A101_Sherman.pdf

Munir, K., & Anjum, M. S. (2018). The use of ontologies for effective knowledge modelling and information retrieval. *Applied Computing and Informatics*, *14*(2), 116–126. doi:10.1016/j.aci.2017.07.003

Oliveira, F. R., & del Val Cura, L. (2016). Performance evaluation of NoSQL multi-model data stores in polyglot persistence applications. *Paper presented at the Proceedings of the 20th International Database Engineering & Applications Symposium*. Academic Press. doi:10.1145/2938503.2938518

Pirotte, A. (1979). Fundamental and secondary issues in the design of non-procedural relational languages. *Paper presented at the Fifth International Conference on Very Large Data Bases*. Academic Press. doi:10.1109/VLDB.1979.718139

Pritchett, D. (2008). Base: An acid alternative. *Queue*, *6*(3), 48–55. doi:10.1145/1394127.1394128

Qian, C., Childers, B., Huang, L., Guo, H., & Wang, Z. (2018). CGAcc: A Compressed Sparse Row Representation-Based BFS Graph Traversal Accelerator on Hybrid Memory Cube. *Electronics (Basel)*, *7*(11), 307. doi:10.3390/electronics7110307

Reiter, R. (1988). On integrity constraints. *Paper presented at the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*. Academic Press.

Rudnicki, R., Cox, A. P., Donohue, B., & Jensen, M. (2018). Towards a methodology for lossless data exchange between NoSQL data structures. *Paper presented at the Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX*. Academic Press. doi:10.1117/12.2307717

Sadalage, P. J., & Fowler, M. (2013). *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. Pearson Education.

Schall, D., & Härder, T. (2015). Dynamic physiological partitioning on a shared-nothing database cluster. *Paper presented at the 2015 IEEE 31st International Conference on Data Engineering*. Academic Press. doi:10.1109/ICDE.2015.7113359

Seera, N. K., & Taruna, S. (2018). Leveraging MapReduce with Column-Oriented Stores: Study of Solutions and Benefits. In Big Data Analytics (pp. 39-46): Springer.

Simone, S. (2019). Experts Predict Trends for Data Architecture in 2019.

Stanier, C. (2012). Introducing NoSQL into the database curriculum. *Paper presented at the 10th International Workshop on the Teaching, Learning and Assessment of Databases*. Academic Press.

Stonebraker, M., Brown, P., Zhang, D., & Becla, J. (2013). SciDB: A database management system for applications with complex analytics. *Computing in Science & Engineering*, *15*(3), 54–62. doi:10.1109/MCSE.2013.19

Storey, V. C., & Song, I.-Y. (2017). Big data technologies and management: What conceptual modeling can do. *Data & Knowledge Engineering*, *108*, 50–67. doi:10.1016/j.datak.2017.01.001

Teorey, T. J., Yang, D., & Fry, J. P. (1986). A logical design methodology for relational databases using the extended entity-relationship model. *ACM Computing Surveys*, *18*(2), 197–222. doi:10.1145/7474.7475

Topi, H., Valacich, J. S., Wright, R. T., Kaiser, K., Nunamaker, J. F. Jr, Sipior, J. C., & de Vreede, G.-J. (2010). IS 2010: Curriculum guidelines for undergraduate degree programs in information systems. *Communications of the Association for Information Systems*, *26*(1), 18.

van der Aalst, W. M. (2018). Responsible Data Science in a Dynamic World. *Paper presented at the IFIP International Internet of Things Conference*. Academic Press.

Van der Vlist, E. (2007). *Schematron*. O'Reilly Media, Inc.

Vaquero, L. M., Rodero-Merino, L., & Buyya, R. (2011). Dynamically scaling applications in the cloud. *Computer Communication Review*, *41*(1), 45–52. doi:10.1145/1925861.1925869

Vogel, W. (2019). Amazon Aurora ascendant: How we designed a cloud-native relational database. All Things Distributed. Retrieved from https://www.allthingsdistributed.com/2019/03/Amazon-Aurora-design-cloud-native-relational-database.html

Vogels, W. (2009). Eventually consistent. *Communications of the ACM*, *52*(1), 40–44. doi:10.1145/1435417.1435432

Walker, E. L., & Slotterbeck, O. A. (2002). Incorporating realistic teamwork into a small college software engineering curriculum. *Journal of Computing Sciences in Colleges*, *17*(6), 115–123.

Wells, J. (2019). Key Database Trends Now and for the Future.

Whitchcock, A. (2005). Google's Big Table. Retrieved from https://web.archive.org/web/20060616203323/http://andrewhitchcock.org/?post=214

Wiese, L., Waage, T., & Brenner, M. (2019). CloudDBGuard: A framework for encrypted data storage in NoSQL wide column stores. *Data & Knowledge Engineering*. doi:10.1016/j.datak.2019.101732

Wu, X., Xu, Y., Shao, Z., & Jiang, S. (2015). LSM-trie: An LSM-tree-based ultra-large key-value store for small data items. *Paper presented at the 2015 USENIX Annual Technical Conference (USENIX 15)*. Academic Press.

Xu, Q., Siyamwala, H., Ghosh, M., Suri, T., Awasthi, M., Guz, Z., . . . Balakrishnan, V. (2015). Performance analysis of NVMe SSDs and their implication on real world databases. *Paper presented at the 8th ACM International Systems and Storage Conference*. ACM Press. doi:10.1145/2757667.2757684

Akhilesh Bajaj is a Chapman Professor of Computer Information Systems at the University of Tulsa. He has a B Tech (Chem. E.) from the Indian Institute of Technology, MBA from Cornell University and a PhD from the University of Arizona. He has published in several academic journals including Management Science, Journal of Association of Information Systems, IEEE Transactions on Knowledge and Data Engineering, Journal of Information Systems and Information Systems. He is on the editorial board of several IS journals and has taught courses from the undergraduate to the Executive level on database design and development, web programming, and IS strategy.

Wade Bick earned a degree in electrical engineering at the University of Southern California and has worked at Teradyne for over 25 years. Positions included memory applications, ASIC test, systems engineering, and software verification. He is currently enrolled in the Master's of Data Analytics Program at the University of Tulsa. Research interests include databases, analytics, and neural networks. Outside of academics, he spends time with his family, enjoys wood working and officiates swim meets for USA Swimming.

Journal of Database Management

Volume 31 • Issue 3 • July-September 2020 • ISSN: 1063-8016 • eISSN: 1533-8010

MISSION

The **Journal of Database Management (JDM)** publishes original research on all aspects of database management, design science, systems analysis and design, and software engineering. The primary mission of JDM is to be instrumental in the improvement and development of theory and practice related to information technology, information systems, and management of knowledge resources. The journal is targeted at both academic researchers and practicing IT professionals.

SUBSCRIPTION INFORMATION

The Journal of Database Management (JDM) is available in print and electronic formats and offers individual or institution-level pricing. Full subscription information can be found at www.igi-global.com/JDM.
JDM is also included in IGI Global's InfoSci-Journals Database which contains all of IGI Global's peer-reviewed journals and offers unlimited simultaneous access, full-text PDF and XML viewing, with no DRM. Subscriptions to the InfoSci-Journals Database are available for institutions. For more information, please visit www.igi-global.com/infosci-journals or contact E-Resources at eresources@igi-global.com.

CORRESPONDENCE AND QUESTIONS

EDITORIAL

Keng Siau, Editor-in-Chief • JDM@igi-global.com

SUBSCRIBER INFO

IGI Global • Customer Service
701 East Chocolate Avenue • Hershey PA 17033-1240, USA
Telephone: 717/533-8845 x100 • **E-Mail:** cust@igi-global.com

The *Journal of Database Management* is indexed or listed in the following.

ABI/Inform; ACM Digital Library; Aluminium Industry Abstracts; Australian Business Deans Council (ABDC); Bacon's Media Directory; Burrelle's Media Directory; Cabell's Directories; Ceramic Abstracts; Compendex (Elsevier Engineering Index); Computer & Information Systems Abstracts; Corrosion Abstracts; CSA Civil Engineering Abstracts; CSA Illumina; CSA Mechanical & Transportation Engineering Abstracts; Current Contents®/Engineering, Computing, & Technology; DBLP; DEST Register of Refereed Journals; EBSCOhost's Business Source; EBSCOhost's Computer & Applied Sciences Complete; EBSCOhost's Computer Science Index; EBSCOhost's Computer Source; EBSCOhost's Current Abstracts; EBSCOhost's STM Abstracts; Electronics & Communications Abstracts; Emerald Abstracts; Engineered Materials Abstracts; Gale Directory of Publications & Broadcast Media; GetCited; Google Scholar; INSPEC; Internet & Personal Computing Abstracts; ISBIB; JournalTOCs; Library & Information Science Abstracts (LISA); Library Literature & Information Sciences; Materials Business File - Steels Alerts; MediaFinder; Norwegian Social Science Data Services (NSD); PubList.com; SCOPUS; Solid State & Superconductivity Abstracts; The Index of Information Systems Journals; The Standard Periodical Directory; Ulrich's Periodicals Directory; Web of Science; Web of Science Science Citation Index Expanded (SCIE)