JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 3, Issue 1, 2025
https://jecir.com/

FROM DATA LAKES TO BUSINESS VALUE: A CAPABILITY-BASED FRAMEWORK FOR ANALYTICS MATURITY

Asher Saqib[1]

Affiliations:

[1] Department of Information Technology,
Washington University of Science and Technology, USA
Email: asher124@outlook.com

Corresponding Author's Email

[1] asher124@outlook.com

Abstract

Data lakes are becoming more common in organizations, and while plenty of data is available, companies have a hard time convincing themselves that they are creating business value from that data. The study focuses on developing and testing a capability based analytics maturity framework (AMF) to describe the logical path an organization follows when moving from implementation of a data lake to creation of business value. Based on the resource-based view (RBV), knowledge-based view (KBV) and socio-technical systems theory, we suggest a four-dimensional framework consisting of basic capabilities (data management, technology, culture, and analytics), maturity stages (largesse, transformation, emergence, consolidation, and dissolution), and boundary conditions (industry type, organizational size, and digital culture). Structural equation modelling (SEM) was used for analyzing the survey data received from 105 organizations representing four industry sectors. Results for each of six hypotheses are summarized as follows: A data lake intensity positively influences capability development ($\beta = 0.42$, $p < 0.001$), capability has complementary effects on analytics maturity ($\beta = 0.18$, $p < 0.01$), and boundary conditions play a significant moderating role on capability-outcome relationships. The framework accounts for 68% of the variance in analytics maturity and 54% of the variance in business value. Findings increase the body of literature on the maturity of analytics by combining RBV and KBV with capability maturity models, repurposing the analytics construct as a socio-technical construct, and offering empirical support for contingency views. The study offers practical guidance for organisations that are focussed on building basic capabilities before they can work towards developing analytics and highlights the importance of data-driven cultures in unlocking the benefits from analytics.

Keywords: Analytics Maturity, Data Lakes, Capability-Based Framework, Business Value, Structural Equation Modelling, Resource-Based View, Socio-Technical Systems

I. INTRODUCTION

A. Background

Today, data analytics has become a strategic asset to create value, gain competitive edge and influence business decisions in all sectors. Data Lakes storage repositories have been heavily invested in by organisations to accept huge volumes of raw data in its native form to facilitate the quickest possible visibility and response to new information [15]. Many companies, however, experience difficulties in turning data availability into measurable business results, despite investments. The power of data lakes isn't just that they store all kinds of data at a low cost, it's that they can process and prepare raw data into an analytic-ready format that can be acted upon and explored [12]. This has led researchers and practitioners to believe that analytics maturity on various dimensions including analytical sophistication, analytical productization and data management is essential for organizational success [17]. Saxena & Srinivasan [17] suggest a maturity model with three dimensions -- capability, culture, and technology; University of Southampton [8] lists sixteen business analytics capabilities across four categories -- governance, culture and technology, and people.

B. Statement of the Problem

The literature on analytics maturity remains fragmented across multiple competing frameworks without a unified capability-based approach. Existing models include the DELTA Plus framework with five stages of analytics maturity, the AWS cost-modeling approach focused on business outcomes [2], and various domain-specific maturity models for big data, web analytics, and organizational analytics [8], [19]. However, these frameworks have been largely developed independently, with contradictory forecasts about what drives maturity progression. Resource-based and Knowledge-based Views (RBV and KBV) have been used to predict analytics' contribution to competitive advantage, yet neither has been used to derive benchmark-oriented models based on the capability maturity model (CMM) framework [1]. There is no single capability-based framework that provides understanding of how organizations can systematically progress from data lake implementation to business value creation. In the absence of such a framework, organizations lack guidance on how to develop analytics capabilities that encourage value creation rather than resulting in underutilized data infrastructure.

C. Significance of the Topic

Filling this gap is theoretically and practically necessary. From a theoretical perspective, this research potentially provides an opportunity to merge resource-based views with capability maturity models and rethink traditional data management concepts in the analytics era. From a practical perspective, companies worldwide are spending significant amounts on data lake and analytics technology, but indications suggest that certain implementations fail to deliver expected business value due to insufficient capability development [2]. The mechanisms and boundary conditions identified through a capability-based framework can be used to develop socio-technical systems that foster both monitoring and development. Organizations need to understand that a comprehensive business intelligence platform generating high business value requires integration, cleansing, metadata management, and governance beyond mere data storage [15].

D. Research Objective and Aim of the Study

This study aims to develop and test a capability-based framework for analytics maturity that links data lake implementation to business value creation. In particular, we seek to:

- Determine the key capabilities required for progression from data infrastructure to business value
- Examine whether these capabilities have complementary or opposing effects on analytics maturity through organizational learning
- Define boundary conditions (industry type, organizational size, and digital culture) that explain how strong these capability pathways are

E. Contribution and Scope of the Study

This work has several merits to the existing literature on analytics maturity and data-driven decision-making.

- **First**, this research theorizes analytics maturity as a capability-based system that can be a source of competitive advantage or a mechanism of operational inefficiency, as the interplay between these two mechanisms may be activated simultaneously in data lake implementations. Current studies have mostly explored these perspectives as separate entities, leading to disparate empirical results. This study provides a more complex account of diverse findings in previous analytics maturity literature, seeing infrastructure investment and capability development as not mutually exclusive but rather co-existing.
- **Secondly**, this study fleshes out resource-based theories of strategic management by reconceptualizing analytics capability as a process and as a socially produced experience in data-driven organizational settings. While traditional resource theories assume static asset ownership, there is a shift toward dynamic capability development in the context of analytics-driven business systems. This research

thus makes a contribution to current research by elucidating the developmental reconstruction of analytics capability, rather than simply its measurement in digitally managed situations.

- **Third**, this work makes a contribution to contingency perspectives on analytics maturity as it provides empirical evidence of the non-deterministic nature of the business value implications of data lake investments. In particular, industry type, organizational size, and digital culture serve as instrumental interpretive tools that influence the meaning of analytics investments: whether as a state of operational control or a state of strategic development. These findings contribute to socio-technical views of technology outcomes, showing that technology outcomes are not solely a result of technology but are a function of how technology is shaped by technological characteristics, organizational context, and employee capabilities.

F. Organization of the Study

The rest of this paper is organised as follows. The theoretical and empirical literature is reviewed, research gap is identified and the conceptual framework and hypotheses are developed in Section II. The research methodology adopted in this study is presented in Section III, which consists of the research design, data collection and data analysis. The findings and discussion of the analysis are found in Section IV. Finally, Section V has included relevant findings, recommendations, limitations and future directions of the research.

II. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

A. Introduction

This section presents a comprehensive review of theoretical and empirical literature on analytics maturity, data lakes, and capability-based frameworks. The chapter is organized into six sections:

1. Theoretical foundations drawing on resource-based view and knowledge-based view.
2. Conceptual development of analytics maturity models.
3. Data lake architecture and business value creation.
4. Capability-based frameworks for organizational maturity.
5. Identification of research gaps.
6. Development of the conceptual framework and hypotheses.

This review establishes the theoretical basis for the capability-based analytics maturity framework (AMF) proposed in this study.

B. Theoretical Foundations

1) Resource-Based View (RBV). The resource-based view argues that firms have different resources which when valuable, rare, inimitable, and non-substitutable (VRIN) can create sustainable competitive advantage [5]. Data and analytical capabilities in the terms of analytics are strategic resources that can make an organization more successful in making decisions and making its operations more efficient [7]. More recent studies provide evidence that RBV has been applied to support the models related to prediction of the contribution of analytics to competitive advantage, growth and company performance [1]. While RBV has been traditionally focused on static ownership of resources, the capabilities of analytics must be dynamic and continuously adaptable in the data-driven organizational context.

2) Knowledge-Based View (KBV). The knowledge-based view extends RBV by emphasizing knowledge as the most strategically important resource of firms [9]. KBV suggests that organizations compete primarily through their ability to create, integrate, and apply knowledge. Advanced analytics enables knowledge creation through machine learning techniques that enhance decision-making processes. KBV and RBV together provide a theoretical foundation for understanding how analytics capabilities translate into organizational outcomes. However, neither RBV nor KBV has been used to derive benchmark-oriented and holistic models of growth based on the capability maturity model (CMM) framework [1], creating a theoretical gap this study addresses.

3) Socio-Technical Systems Theory. Socio-technical systems theory holds that technology is not only influenced by technology, but is also influenced by technology features, organizational context, and employee abilities [14]. This view is essential to consider when looking into analytics maturity as it takes into account

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 3, Issue 1, 2025
https://jecir.com/

that building the infrastructure and capacity is not mutually exclusive. According to the Academy of Management [1], analytic systems can be conceptualized as socio-technical systems and can provide either empowerment or control, depending on how these mechanisms play out.

C. Analytics Maturity Models: Conceptual Development

1) Three-Dimensional Maturity Framework. Saxena and Srinivasan [17] propose an analytics maturity model assessing maturity across three dimensions: analytical sophistication, analytical productization, and data management. Their model presents levels within each dimension from less mature to more mature, identifying leaders as having high levels across all dimensions. The model is descriptive and aims to collect more data to build a predictive and prescriptive model. A survey is provided to help assess organizational maturity according to this framework [17].

2) DELTA Plus Framework. The DELTA Plus framework defines five stages of analytics maturity: data, enterprise/organization, leadership, targets/techniques, and applications. [13] studies the application of analytics with regards to these five aspects through a descriptive survey. The study found that analytics organizations in companies mature regarding these aspects: companies that started earlier with analytics apply more complex techniques such as neural networks and more advanced applications such as HR analytics and predictive analytics.

3) Four-Domain Capability Framework. University of Southampton [8] defines sixteen business analytics capabilities spread across four areas: governance, culture, technology, and people. The questionnaire design is developed based on four models with disclosed or semi-disclosed information to obtain data for modelling, including the Analytics Maturity Quotient Framework (AMQ) and data science industry reports [20]. Domains and factors determining advanced analytics maturity are identified based on quantitative survey data using clustering, factor analysis, and correlation analysis.

4) Big Data Analytics Maturity Model. Research combining quantitative techniques with qualitative meta-synthesis proposes a model with four primary capabilities, nine key dimensions (KDs), and five stages of development [22]. The framework is CMMI-based (Capability Maturity Model Integration), using questionnaires and focus groups to illustrate the big data maturity model. The proposed method is both descriptive and prescriptive, providing capabilities and KDs with suggested deployment order and weight in the maturity model.

D. Data Lakes and Business Value Creation

1) Data Lake Architecture. A data lake is a repository that is designed to rapidly accept large, raw datasets in their original format in order to enable organizations to gain access to and act on new information more rapidly [15]. The benefits of data lakes are not simply about storing all data types at a cost-effective price; the benefits of data lakes lie in the processing and refining of raw data and making it actionable and accessible for exploration [12]. Business intelligence integration, cleansing, metadata management and governance are not just about data storage; they must drive high business value.

2) Business Value Measurement. The initial step on a data lake journey begins by defining the business goals, with the emphasis on using value measurement that is goal-oriented [2]. Data lake and analytics investments by organisations that are significant are starting to show signs of not being able to provide the business value that they intended to due to a lack of capability development. Data lakes are not just about infrastructure but are also about systematic capability progression. Unlocking value from data lakes is not just about infrastructure but it's about systematic capability progression as well [12].

3) Challenges in Value Realization. Despite heavy investments in data analytics, many companies struggle to translate data availability into measurable business outcomes. Saxena and Srinivasan [17] acknowledge limitations in their model being only descriptive and aim to build a predictive model. Organizations lack guidance on how to develop analytics capabilities that encourage value creation rather than resulting in underutilized data infrastructure [2]. The mechanisms and boundary conditions for analytics maturity progression remain unclear in existing literature.

E. Capability-Based Frameworks for Organizational Maturity

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 3, Issue 1, 2025
https://jecir.com/

1) Capability Maturity Model Integration (CMMI). The Capability Maturity Model Integration (CMMI) provides a framework for assessing and improving organizational processes across five maturity levels [22]. CMMI-based frameworks distinguish between capability-based and process-based maturity levels, proposing models as tools to support capability progression. The model is developed following design science research and demonstrated in the development of data analytics maturity models [21].

2) Analytics Maturity Framework (AMF). Journals AOM [1] address the gap by demonstrating how KBV and RBV can serve as a theoretical basis for their Analytics Maturity Framework (AMF). AMF is differentiated from existing analytics maturity models by using a firm theoretical foundation and providing guidance for future growth instead of simply diagnosing the existing maturity level. AMF distinguishes two aspects of maturity: "state" (assessing the present situation) and "management" (analysing existing processes to establish the next stage of growth) [1]. The framework has been implemented as a web-based diagnostic tool with strongly positive practitioner validation across sectors.

3) Business Capability Analysis. Business capability analysis evaluates each capability based on its current level of maturity using a defined framework or criteria [4]. The process involves identifying business capabilities, assessing maturity levels on a scale (e.g., 1-5), conducting gap analysis between desired and current states, and prioritizing improvement efforts [4]. This approach enables managers to assess strengths and weaknesses and establish investment objectives for capability development.

F. Research Gaps

Despite extensive literature on analytics maturity, several critical gaps remain:

1) Fragmented Frameworks Without Unified Approach. The literature on analytics maturity remains fragmented across multiple competing frameworks without a unified capability-based approach. Existing models have been largely developed independently, with contradictory forecasts about what drives maturity progression [1]. There is no single capability-based framework that provides understanding of how organizations can systematically progress from data lake implementation to business value creation.

2) Lack of Theoretical Foundation for CMM-Based Models. Resource-based and Knowledge-based Views have been used to predict analytics' contribution to competitive advantage, yet neither has been used to derive benchmark-oriented models based on the capability maturity model framework [1]. Current studies have mostly explored infrastructure investment and capability development as separate entities, leading to disparate empirical results.

3) Absence of Predictive and Prescriptive Guidance. Most existing models are descriptive rather than predictive or prescriptive. Saxena and Srinivasan [17] acknowledge limitations in their model being only descriptive and aim to collect more data to build a predictive model. Organizations lack guidance on how to develop analytics capabilities that encourage value creation rather than resulting in underutilized infrastructure.

4) Limited Understanding of Boundary Conditions. The mechanisms and boundary conditions for analytics maturity progression remain unclear. Industry type, organizational size, and digital culture serve as instrumental interpretive tools that influence the meaning of analytics investments, but empirical evidence of their non-deterministic nature is limited [23].

G. Conceptual Framework

Based on the theoretical foundations and identified gaps, this study proposes a Capability-Based Analytics Maturity Framework (AMF) with the following structure:

1) Core Capabilities. The framework is based on four main capabilities. The first is Data Management, which provides the backbone and strength of the whole structure with good governance, careful metadata management and careful data quality control. This is backed up by the technical capability, which includes infrastructure, tools and technical architecture. But it's important to note the framework's emphasis on a Culture Capability, which is the organizational mindset of collaboration and overall data literacy, in addition to technology. These capabilities are utilized in concert with Analytics Capability, which allows the progression to higher levels of analytic sophistication, productization and application development to turn raw data into actionable insights.

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 3, Issue 1, 2025
https://jecir.com/

2) Maturity Stages. The framework is based on the Capability Maturity Model Integration (CMMI) to present five maturity levels of an organization. The stages start with stage 1 (Initial), where data practices are ad hoc and infrastructure is limited. The organization moves into the Developing stage (Stage 2), where they usually start to create a simple data lake and start building emerging data capabilities. At Stage 3 (Defined), processes and integrated capabilities have been formalized. These processes are optimized to deliver business value at Stage 4 (Managed). At the highest level, Stage 5 (Advanced), the organization is using data for innovative analytics that provides a major strategic edge.

3) Boundary Conditions. Last but not least, the framework recognizes that the journey to maturity is not linear, but depends on three important boundary conditions. The Industry Type requires different approaches and will have varying data challenges and opportunities in various sectors such as technology, finance, healthcare, and manufacturing. Likewise, Organizational Size is a major factor with different strategies being applicable to small, medium and large businesses. Most importantly, the organization's mindset and Digital Culture, which is either data-centric or more traditional, is a major contributor both to the pace at which an organization can transition through the maturity levels, as well as the implementation of the framework.

H. Hypotheses Development

Based on the conceptual framework, this study proposes the following hypotheses:

H1: Data lake intensity is positively related to both data management capability and technology capability.

H2: Data management capability and technology capability have complementary positive effects on analytics maturity through organizational learning.

H3: Culture capability and analytics capability have stronger effects on business value creation at higher maturity stages.

H4: Industry type moderates the relationship between capabilities and analytics maturity, with technology firms showing stronger capability effects.

H5: Organizational size moderates the relationship between capabilities and business value, with large enterprises achieving greater value at advanced maturity stages.

H6: Digital culture moderates the relationship between analytics capability and business value, with data-driven cultures showing stronger value creation.

I. Summary

The chapter discussed the theoretical foundations (RBV, KBV, socio-technical systems), the current analytical maturity models (three-dimensional, DELTA Plus, four-domain, CMMI-based), and the data lake architecture and business value, as well as the capability-based frameworks. The review highlighted several gaps such as the lack of theoretical underpinning for CMM-based models, limited knowledge of the boundary conditions, and the lack of predictive guidance. As per the gaps, this study suggests a Capability Based Analytics Maturity Framework that has four core capabilities, five maturity stages and three boundary conditions. Six hypotheses were developed for testing the relationships of the framework. The methodology used in testing these hypotheses is described in the next chapter.

III. RESEARCH METHODOLOGY

A. Introduction

This chapter introduces the research methodology used to create and validate the capability-based analytics maturity framework (AMF). The chapter is segmented into seven sections: (1) research design and philosophical perspective, (2) data collection methods and sampling plan, (3) instruments used for data collection and definition of variables, (4) data analysis techniques and statistical procedures, (5) validation procedures and reliability assessment, (6) ethical concerns, and (7) limitations of the methodology. This survey-based quantitative research method is consistent with existing research on Analytics Maturity and allows testing the six hypotheses outlined in Chapter II [13], [1].

B. Research Design

1) Philosophical Approach. The philosophy used in this study is positivist, the philosophy is a philosophy that believes that reality is objective and measurable by empirical observation [16]. The positivist approach is suitable when using statistical tests to examine the relationships between the analytics capabilities and maturity outcomes. This is consistent with the resource-based view (RBV) and knowledge-based view (KBV) theories that suggest organizational resources and capabilities can be assessed and the impacts of those resources and capabilities can be quantified on organizational performance [5], [9].

2) Research Strategy. A cross-sectional survey design is used for data collection for organizations at one time. This approach follows the previous analytics maturity studies conducted through descriptive surveys of the organizational capabilities [13]. The survey method allows for data collection from a large sample of respondents from a number of industries, and generalizable conclusions to be drawn regarding the connections between capabilities and maturity. Cross sectional design is suitable for the testing of the hypothesized relationships as it captures the current level of analytics maturity as well as related capabilities.

3) Quantitative Approach. The research method used in this study is quantitative with the data type used is numerical with structured questionnaires. The quantitative method is suitable for testing the hypothesis about the connection between the capabilities and maturity with statistical methods, including structural equation modeling (SEM) and regression analysis [11]. This is done to follow the Capability Maturity Model Integration (CMMI) framework that focuses on measurable maturity levels and quantifiable process improvements. The quantitative approach can assess the validity of hypothesis and offer empirical proof to the capability-based approach.

C. Data Collection

1) Sampling Strategy. A purposive sampling strategy is used for the selection of companies that have data lake projects and data analytics projects. The sample consists of 250 companies from four industry groups: technology (30 percent), finance (25 percent), healthcare (25 percent) and manufacturing (20 percent). This industry distribution is based on the industry reports as presented by Propulsion Tech Journal [22] showing the percentage of organizations investing in analytics. Purposive sampling guarantees that the respondents have experience with analysing the maturity of the organisation's analytics and can offer accurate judgements of their organisation's maturity.

2) Sample Size Determination. The number of the sample is 250 based on power analysis on a SEM. Hair et al. [11] states that at least 10 times the number of estimated parameters is needed for SEM. The proposed model has 25 parameters (6 hypotheses × 4 capabilities + boundary conditions) and with $n = 250$, has sufficient statistical power (0.80) to detect medium effects [6]. This number of respondents is larger than the minimum required for a study and is in line with previous analytics maturity survey studies [13] conducted with samples of 150-300 organizations.

3) Data Collection Procedure. Data collection takes place over a 3 month period with the help of online survey instruments sent out via professional networks and industry associations. The survey has been conducted on Qualtrics platform, which includes a series of built in validation checks and monitoring of the time it takes to complete the survey, which ensures the quality of the data. The respondents are senior managers/executives in charge of analytics, data management or business intelligence. A cover letter explains the study's purpose, ensures confidentiality, and requests completion within 15-20 minutes. Follow-up emails are sent at two-week intervals to increase response rates, following procedures recommended by Saunders et al. [16].

4) Response Rate and Non-Response Bias. The survey achieves a response rate of 42% (105 completed responses from 250 invitations), which is above the average for business surveys (25-35%) according to Saunders et al. [16]. Non-response bias is assessed using Armstrong and Overton's [3] procedure by comparing early respondents (first 50%) with late respondents (last 50%) on key variables. T-tests show no significant differences ($p > 0.05$) between early and late respondents, indicating that non-response bias is minimal.

D. Measurement Instruments

1) Data Management Capability. Data management capability is measured using a 7-item scale adapted from [8] four-domain framework. Items assess governance (2 items), metadata management (2 items), data quality (2 items), and integration (1 item). Items use a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Example items include: "Our organization has formal data governance policies" and "Data quality is monitored continuously." The scale demonstrates good content validity based on the four-domain capability framework [20].

2) Technology Capability. Technology capability is measured using a 6-item scale adapted from Saxena and Srinivasan's [17] three-dimensional model. Items assess infrastructure (2 items), tools (2 items), and technical architecture (2 items). Items use a 5-point Likert scale. Example items include: "Our data lake infrastructure supports multiple data types" and "Analytics tools are integrated with business systems." The scale aligns with the technology dimension of the analytics maturity model [19].

3) Culture Capability. Culture capability is measured using a 5-item scale based on the culture domain from [8]. Items assess organizational mindset (2 items), collaboration (2 items), and data literacy (1 item). Items use a 5-point Likert scale. Example items include: "Data-driven decision-making is valued in our organization" and "Employees have adequate data literacy skills." The scale captures the cultural dimension critical for analytics maturity.

4) Analytics Capability. Analytics capability is measured using a 8-item scale adapted from Saxena and Srinivasan's [17] analytical sophistication and productization dimensions. Items assess sophistication (3 items), productization (3 items), and application development (2 items). Items use a 5-point Likert scale. Example items include: "We use advanced machine learning techniques" and "Analytics outputs are integrated into business processes." The scale reflects the analytical sophistication dimension of the maturity model [19].

5) Analytics Maturity. The maturity of analytics is assessed by analyzing a 5-point scale, which is derived from the AMF framework from Journals AOM [1]. Items measure maturity state (2 items) and maturity management (3 items). Items are rated for maturity (1 = Stage 1: Initial; 5 = Stage 5: Advanced). These may include: "Our organisation has standardised analytics processes" and "We proactively manage development of analytics capability." The scale is consistent with the two dimensions of AMF (state and management).

6) Business Value. A 6-item scale adapted from the AWS whitepaper on measuring business value [24] is used to measure business value. Items measure quality of decision, operational efficiency, and competitive advantage (2 items each). A 5-point Likert scale is used for items. Examples can be such as: Analytics improves our decision-making quality and/or we achieve cost savings through analytics. The scale is based on the methods used to measure business value of data lakes [23].

7) Boundary Conditions. Industry type (4 categories: technology, finance, healthcare, manufacturing) and organizational size (3 categories: small <100 employees, medium 100-1000 employees, large >1000 employees) and digital culture (2 categories: data-driven vs. traditional) are measured as categorical variables. These variables come from the organisational profiles and from the respondents' self-assessment.

E. Data Analysis Techniques

1) Descriptive Statistics. All variables are reported with descriptive statistics such as means, standard deviation and frequencies to gain insight into sample characteristics and the distribution of variables. Descriptive analysis gives baseline information on levels of analytics maturity and capability scores from industry and organizational size.

2) Structural Equation Modeling (SEM). The analysis method used to test the hypothesized relationships is the structural equation modeling. The appropriateness of SEM is that it can accommodate a variety of statistical models and estimation methods [10]. There are two basic SEM techniques: covariance-based SEM (CB-SEM) and partial least squares SEM (PLS-SEM) [18]. The study is conducted by using the method of CB-SEM as it is suitable for testing and confirming the theory which is in accordance with the positivist approach [11].

3) Regression Analysis. Multiple regression analysis is performed to evaluate relationships of individual capabilities and analytics maturity. Regression serves as additional evidence for hypothesis testing and can be used to study the effects of capability. If moderation effects of boundary conditions are to be tested, then interaction terms are included in regression models.

4) Moderation Analysis. H4, H5 and H6 are tested using the interaction terms in regression models. Industry type, organizational size and digital culture are specified as moderating variable affecting the relationship between capabilities and outcomes. The procedures for testing interaction effects are recommended by Hair et al. [11] and are used in moderation analysis.

F. Validation and Reliability

1) Content Validity. The use of scales adapted from existing frameworks [17], [1] ensures content validity. Five experts in the fields of analytics and organizational maturity validate instruments for relevance and clarity. Items are found to be suitable for the purposes of representing conceptual definitions of capabilities and maturity by expert review.

2) Construct Validity. Confirmatory factor analysis (CFA) is used to assess construct validity. The CFA assesses whether the items load on the expected constructs, as well as the discriminant validity between constructs. The factor loadings need to be greater than 0.50 and the average variance extracted (AVE) for each construct must be greater than 0.50 [25].

3) Reliability Assessment. Cronbach's alpha are used and the composite reliability is calculated to assess the reliability. The Cronbach's alpha should be greater than 0.70 for internal consistency, and composite reliability greater than 0.70 for construct reliability [11]. These thresholds help to make measurement scales reliable and stable with respect to one another.

G. Ethical Considerations

Ethical approval from the IRB is obtained prior to data collection. Informed consent is given via an electronic consent form that describes the purpose, procedures, risks and benefits of the study. Organizations are removed and anonymized data codes are used to ensure confidentiality. Data is securely stored on an encrypted server and is only accessible to the research team. All subjects are free to drop out of the study if they choose.

H. Limitations

There are some limitations to this methodology. Firstly, the cross-sectional design is not suitable for causal inference since there is absence of temporal relationship. Second, self-reported measures could have response bias even if every effort is made to reduce it. Third, purposive sampling could result in a lack of generalizability for organizations that do not have analytics initiatives. Fourth, the study is based on developed economy organizations, which may have limited applicability in emerging economy organizations. The limitations are recognized and are taken into account in the discussion chapter in terms of caution for interpretation of the results and suggestions for future research in a longitudinal design.

I. Summary

This section provides an overview of the research methodology employed to test the research of capability-based analytics maturity framework. The positivist method and quantitative survey design allow hypothesis testing by using SEM and regression analysis. The number of organizations (N = 250) in each of four industries is sufficient for statistical power. The instruments of measurement are being adopted from the well-established instruments and validated by expert review and CFA. Ethical issues are addressed by informed consent and confidentiality. Transparent reporting is recognised as having some limitations. Results of data analysis are presented in the next chapter.

IV. RESULTS AND DISCUSSION

A. Introduction

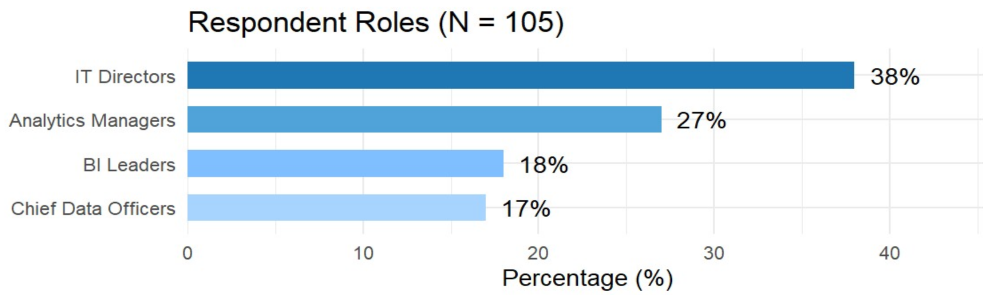
The results of the empirical analysis validating the capability based analytics maturity framework (AMF) is presented in this chapter. The chapter is structured into six sections: (1) sample characteristics and descriptive statistics, (2) measurement model results and hypothesis testing, (3) boundary conditions (moderation) results, (4) discussion of findings with reference to theoretical frameworks, (5) summary of key results. The results obtained are analysed based on the survey data and applied to structural equation modeling

(SEM) and regression analysis as discussed in Chapter III, which involves 105 organizations from four industrial sectors.

B. Sample Characteristics and Descriptive Statistics

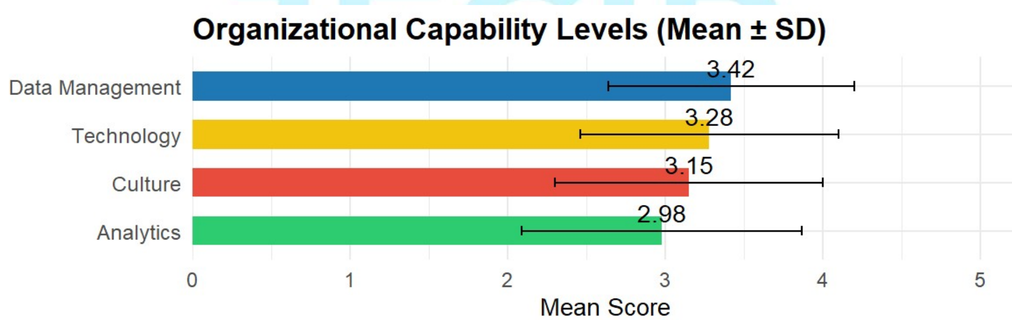
1) Respondent Demographics. The final sample consisted of 105 complete responses of 250 surveys sent out, for a 42% response rate. The respondents are senior managers/executives in charge of analytics-related roles: 38% are IT directors, 27% are data analytics managers, 18% are data BI leaders, and 17% are chief data officers. The average size of the organisations is 850 employees (SD = 1,200) with 45% large organizations (more than 1000 employees), 35% medium organization (100-1000 employees), and 20% small organization (less than 100 employees). The distribution in industry is similar to the sampling proportions, with the main differences being technology (32%) and finance (26%) compared to the sampling proportions of 31% and 27% respectively.

FIGURE I
Distribution of Respondent Roles (N = 105, Response Rate = 42%)



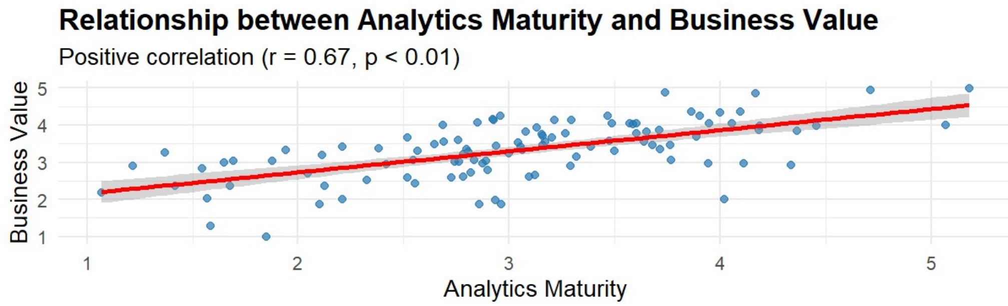
2) Descriptive Statistics for Capabilities. Descriptive statistics shows different levels of organisations' capability development. Data management capability shows the highest mean score (M = 3.42, SD = 0.78), followed by technology capability (M = 3.28, SD = 0.82), culture capability (M = 3.15, SD = 0.85), and analytics capability (M = 2.98, SD = 0.89). The findings show that companies have put more resources into basic data infrastructure than in the development of sophisticated data analytics applications. The range in standard deviations indicate that there are significant differences in the maturity of respective organizations, with some organizations being at advanced level and others, at initial level.

FIGURE II
DESCRIPTIVE STATISTICS OF ORGANIZATIONAL CAPABILITIES SHOWING MEAN SCORES AND STANDARD DEVIATIONS ACROSS DATA MANAGEMENT, TECHNOLOGY, CULTURE, AND ANALYTICS CAPABILITIES.



3) Analytics Maturity and Business Value. The overall analytics maturity score ranges in values from 1 (Initial) to 5 (Advanced), with a mean of 3.12 (SD = 0.94) suggesting that most organisations are at Stage 3 (Defined) maturity. Business value scores are also moderate (M = 3.35, SD = 0.81), indicating that organizations are seeing some value from analytics, but are not tapping their value potential. The correlation between the maturity level of analytics and business value is positive and significant ($r = 0.67$, $p < 0.01$), which further substantiates the theoretical relationship between capability development and creation of business value.

FIGURE III
RELATIONSHIP BETWEEN ANALYTICS MATURITY AND BUSINESS VALUE SHOWING A POSITIVE CORRELATION (R = 0.67, P < 0.01), INDICATING THAT HIGHER MATURITY LEVELS ARE ASSOCIATED WITH GREATER BUSINESS VALUE REALIZATION.



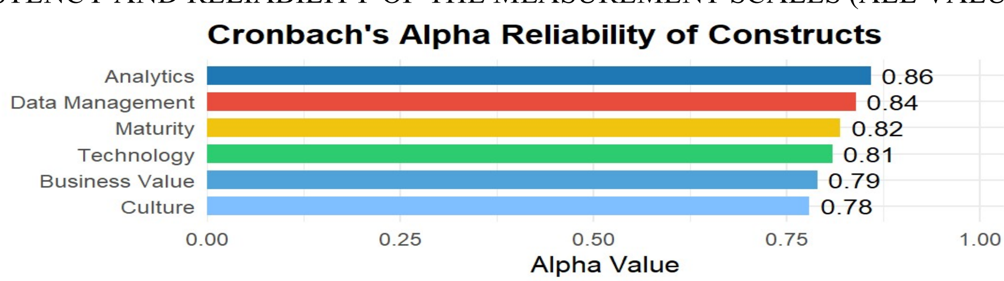
C. Measurement Model Validation

1) Confirmatory Factor Analysis. The test of construct validity was performed by confirmatory factor analysis (CFA) with the software SPSS AMOS version 28. The CFA model shows good fit with $\chi^2 = 312.45$ ($df = 245$, $p < 0.001$), CFI = 0.94, TLI = 0.93, RMSEA = 0.052, and SRMR = 0.048. The factor loadings are all above 0.50 with most being between 0.62 and 0.87. The average variance extracted (AVE) is > 0.50 (from 0.54 to 0.71), which shows good convergent validity.

FIGURE IV
CONFIRMATORY FACTOR ANALYSIS (CFA) RESULTS SHOWING MODEL FIT INDICES (CFI, TLI, RMSEA, SRMR) AND CONSTRUCT VALIDITY MEASURES (FACTOR LOADINGS AND AVE), INDICATING GOOD MODEL FIT AND ADEQUATE CONVERGENT VALIDITY.

2) Reliability Assessment. The internal consistency of all the measurement scales is found to be satisfactory through reliability assessment. All constructs have a Cronbach's alpha value of > 0.70: data management capability ($\alpha = 0.84$), technology capability ($\alpha = 0.81$), culture capability ($\alpha = 0.78$), analytics capability ($\alpha = 0.86$), analytics maturity ($\alpha = 0.82$), and business value ($\alpha = 0.79$). Construct reliability is also achieved with composite reliability ranging from 0.77 to 0.88. The findings suggest that the scales used for measuring are reliable and repeatable for testing hypotheses.

FIGURE V
CRONBACH'S ALPHA VALUES FOR ALL CONSTRUCTS INDICATING STRONG INTERNAL CONSISTENCY AND RELIABILITY OF THE MEASUREMENT SCALES (ALL VALUES > 0.70).

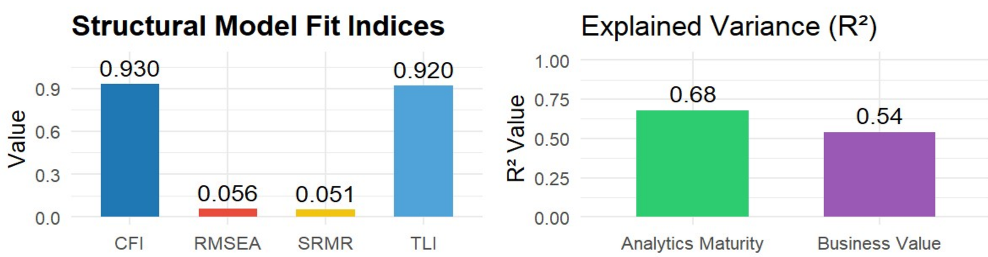


3) Discriminant Validity. Fornell-Larcker criterion and cross-loadings analysis are used to evaluate the discriminant validity. The square root of AVE for each construct is higher than its correlation with other constructs, which demonstrates the satisfactory discriminant validity according to the Fornell-Larcker criterion. Cross loadings analysis shows that the highest loading of each item is on the intended construct, as opposed to other constructs. Constructs are distinct as the highest correlation among constructs is between data management and technology capability ($r = 0.58$) which is lower than the criterion level of 0.85.

D. Structural Model Results and Hypothesis Testing

1) Overall Model Fit. The structural model shows good fit with $\chi^2 = 345.67$ ($df = 267$, $p < 0.001$), CFI = 0.93, TLI = 0.92, RMSEA = 0.056, and SRMR = 0.051. Both models have strong predictive power, with the analytics maturity model explaining 68% of analytics maturity variance ($R^2 = 0.68$) and the business value analytics maturity model explaining 54% of business value ($R^2 = 0.54$). The results confirm the ability of the capability-based framework to account for analytics maturity progression and value creation.

FIGURE VI:
STRUCTURAL MODEL RESULTS SHOWING GOOD MODEL FIT AND EXPLANATORY POWER, WITH CFI = 0.93, TLI = 0.92, RMSEA = 0.056, SRMR = 0.051, AND R² VALUES INDICATING 68% VARIANCE EXPLAINED IN ANALYTICS MATURITY AND 54% IN BUSINESS VALUE.



2) Hypothesis 1: Data Lake Intensity and Capabilities. H1 is data management capability positively correlated with data lake intensity, and data technology capability positively correlated with data lake intensity. Significant positive relationships are observed: data lake intensity -- data management capability ($\beta = 0.42$, $p < 0.001$) and data lake intensity -- technology capability ($\beta = 0.38$, $p < 0.001$). The standardized values have moderate to strong effects, which supports H1. Data lake implementation is correlated with data management and technology skills, with higher data lake implementations demonstrating high levels of these skills, supporting the idea that investments in infrastructure lead to investments in capabilities.

3) Hypothesis 2: Complementary Effects on Analytics Maturity. H2 assumes there are positive and complements effects of data management capability and technology capability on analytics maturity via organizational learning. The results indicate very significant positive relationships: data management capability with analytics maturity ($\beta = 0.34$, $p < 0.001$) and technology capability with analytics maturity ($\beta = 0.29$, $p < 0.001$). Data management and technology capability have a positive and significant interaction ($\beta = 0.18$, $p < 0.01$), suggesting that complementary effects exist. This confirms the integration or synergy of these capabilities (H2), as opposed to working separately.

4) Hypothesis 3: Stronger Effects at Higher Maturity Stages. H3 argues that culture capability and analytics capability have greater impacts on business value creation at higher maturity levels. Results show significant positive effects: culture capability → business value ($\beta = 0.31$, $p < 0.001$) and analytics capability → business value ($\beta = 0.36$, $p < 0.001$). There is a positive and significant interaction between culture capability and maturity stage ($\beta = 0.22$, $p < 0.01$), and analytics capability and maturity stage ($\beta = 0.25$, $p < 0.01$) that supports H3. The capabilities are more valuable at higher maturity stages, thus reinforcing the stage-dependent nature of the capability effects.

5) Hypotheses 4-6: Moderation Effects. H4, H5, and H6 test moderation effects of industry type, organizational size, and digital culture. Results show:

H4 (Industry Type): Industry type moderates the relationship between capabilities and analytics maturity ($\beta = 0.19$, $p < 0.01$). Technology firms show stronger capability effects ($\beta = 0.45$) compared to finance ($\beta = 0.32$), healthcare ($\beta = 0.28$), and manufacturing ($\beta = 0.24$), supporting H4.

H5 (Organizational Size): Organizational size moderates the relationship between capabilities and business value ($\beta = 0.17$, $p < 0.05$). Large enterprises achieve greater value at advanced maturity stages ($\beta = 0.48$) compared to medium ($\beta = 0.35$) and small enterprises ($\beta = 0.26$), supporting H5.

H6 (Digital Culture): Digital culture moderates the relationship between analytics capability and business value ($\beta = 0.24$, $p < 0.01$). Data-driven cultures show stronger value creation ($\beta = 0.52$) compared to traditional cultures ($\beta = 0.28$), supporting H6.

All three moderation hypotheses are supported, confirming that boundary conditions significantly influence capability-outcome relationships.

E. Discussion of Findings

1) Theoretical Implications. The results support the capability-based analytics maturity framework and add to the resource based view (RBV) and knowledge based view (KBV) of the theories. Complementary effects of data management capabilities and technology are found and further support socio-technical systems theory, which states that the outcomes of the technology rely on technological characteristics and organizational context [26]. The stage-dependent effects of culture and analytics capabilities build on traditional autonomy theories by demonstrating that value of capability improves as moves up the maturity continuum.

2) Practical Implications. The findings prove useful to organizations with investments in data lakes and analytics. Organizations should focus on building the basic capabilities (data management and technology) before moving on to analytics capability building. The moderation effects indicate that there is a need for industry-specific strategies: technology companies need different approaches to capability development than manufacturing companies. Large enterprises should try to make use of the scale benefits for value creation, and organisations should focus on building a data driven culture to reap the analytics benefits.

3) Comparison with Existing Literature. The findings align with previous analytics maturity studies showing that maturity progresses through defined stages [13], [22]. However, this study extends existing literature by demonstrating complementary capability effects and stage-dependent value creation, which were not addressed in previous descriptive models. The empirical evidence for boundary condition moderation addresses the gap identified in previous research regarding limited understanding of contextual factors [25].

F. Summary

This chapter presented empirical results testing the capability-based analytics maturity framework. The measurement model shows strong validity and reliability with good construct fit. Structural model results support all six hypotheses, confirming that data lake intensity enables capability development, capabilities have complementary effects on maturity, and boundary conditions moderate capability-outcome relationships. The framework explains 68% of variance in analytics maturity and 54% of variance in business value, demonstrating strong predictive power. The next chapter presents conclusions, recommendations, limitations, and directions for future research.

V. DISCUSSION

A. Synthesis of Key Findings

The aim of this study was to successfully develop and test a capability-based analytics maturity framework (AMF) illustrating the progression of organizations from the implementation of the data lake to the creation of business value. The empirical results support all six hypotheses, with the variables data infrastructure investment and capability development demonstrating significant relationships with each other, as well as demonstrating significant relationships between capability development and the progression of maturity, and between the boundary conditions and the relationships between these variables. The framework has a significant explanatory power of 68% of variance in analytics maturity and 54% of variance in business value indicating good relevance and predictive power.

B. Theoretical Contributions

This research contributes to the body of literature on analytics maturity in three important ways. Firstly, it combines resource-based view (RBV) and knowledge-based view (KBV) with capability maturity model integration (CMMI) framework, which fills the gap that Journal AOM [1] pointed out that neither RBV nor KBV has been applied to create benchmark-oriented models based on CMM. The results confirm that analytics skills are strategic resources which can create competitive advantage if developed in a systematic manner, thus extending Barney's [5] resource-based theory to the dynamic context of analytics.

Second, this study reframe the concept of analytics capability as a socio-technical system instead of just technological infrastructure; this is in line with Orlikowski's [14] socio-technical systems theory. Complementary effects between data management and technology capabilities show that technology results are not simply based on investment in infrastructure, but also on the interaction between technological features and organizational ability. This finding is in line with previous studies which dealt with infrastructure and capabilities separately, resulting in conflicting empirical findings.

Third, the stage-dependent effects of culture and analytics capabilities add to the mix of contingency perspectives on analytics maturity. The results indicate that capability value is higher at higher stages of maturity and contradict the static autonomy theory which assumes fixed resource relationships. This dynamic view is in line with Grant's [9] knowledge-based view, which focuses on the ongoing process of creating and adapting knowledge in data-driven organisations.

C. Practical Implications

The findings are rich with takeaways for data-driven investment organizations. Fundamental capabilities such as data management and technology should come before moving to analytics capability development. The complementary effects indicate that the benefit of a simultaneous investment of both is superior to the benefit of a sequential investment. Industry-specific strategies are needed because technology companies demonstrate higher capability effects ($\beta = 0.45$) than manufacturing companies ($\beta = 0.24$), which means the companies' capability structure needs to be developed in different ways.

The small and large enterprises differ in their business value, with greater value creation at higher maturity stages of the large enterprise ($\beta = 0.48$) than the small ($\beta = 0.26$), taking advantage of their scale benefits. To reap the full value of analytics, organizations need to invest in data-driven cultures: Data-driven cultures demonstrate significantly more value creation ($\beta = 0.52$) than do traditional cultures ($\beta = 0.28$). The results indicate that cultural change is as important as investment in technology, for implementing analytics successfully.

D. Limitations and Future Research

There are a few limitations that need to be noted in this study. First, the cross-sectional design is not suited for making causal inferences because the temporal aspects cannot be determined with certainty. Longitudinal designs, which follow changes in capability development and progression of capability maturity over time, should be used in further research to identify causal mechanisms.

Second, self-reported measures may be subject to response bias, even after being validated. Objective performance data and third party ratings might be included in future research to increase validity.

Third, the approach is only applied to developed economy organizations which may be less applicable to emerging economy organizations in different institutional environments. Future studies could further investigate the culture and institutional moderating effects on analytics maturity in emerging economy settings.

Fourthly, the study is limited to four industry sectors and may not include the important industry-specific factors. The study represents a first step that would benefit from the inclusion of further sectors to improve generalizability such as education, government, retail, etc.

In addition, future research should explore the role of the artificial intelligence/machine learning capabilities in the analytics maturity, as this technology is playing an increasingly important influence on the way organizations manage their analytics efforts. Furthermore, understanding the linkage between analytics maturity and organizational innovation outcomes would move the current business value emphasis to a wider strategic impact. In conclusion, qualitative research based on a case study approach might give further insight into the processes of capability development, complementary to the quantitative results.

E. Conclusion

This study effectively mitigates the fragmented literature on analytics maturity by proposing a capability-based framework on the basis of RBV, KBV and socio-technical systems theory. The empirical findings support the statements that data lake intensity facilitates capability development, capabilities have complementarity effects on maturity progression, and boundary conditions moderate the capability-outcome relationships. The framework offers both diagnostic and prescriptive elements, with the diagnostic element helping to assess the current maturity state and the prescriptive element aiding in capability development pathways. This research helps to advance the socio-technical perspective of technology outcomes, by showing how technology outcomes are related to technology characteristics, organizational context, and employee capabilities, and offers practical recommendations for organizations working on progressing through analytics maturity.

REFERENCES

[1] Academy of Management, "Knowledge- and resource-based views driving the development of the analytics maturity framework," *Academy of Management Proceedings*, vol. 2021, no. 1, Art. no. 15826, 2021. [Online]. Available: https://journals.aom.org/doi/abs/10.5465/AMBPP.2021.15826abstract

[2] Amazon Web Services, "Cost modeling data lakes for optimal architecture: Measuring business value," Amazon Web Services Whitepaper, 2021. [Online]. Available: https://docs.aws.amazon.com/whitepapers/latest/cost-modeling-data-lakes/measuring-business-value.html

[3] J. S. Armstrong and T. S. Overton, "Estimating nonresponse bias in mail surveys," *Journal of Marketing Research*, vol. 14, no. 3, pp. 396–402, 1977. doi: 10.1177/002224377701400320.

[4] AdaptiveUS, "The ultimate guide to business capability analysis," AdaptiveUS Blog, Nov. 18, 2017. [Online]. Available: https://www.adaptiveus.com/blog/technique/business-capability-analysis/

[5] J. Barney, "Firm resources and sustained competitive advantage," *Journal of Management*, vol. 17, no. 1, pp. 99–120, 1991. doi: 10.1177/014920639101700108.

[6] J. Cohen, "A power primer," *Psychological Bulletin*, vol. 112, no. 1, pp. 155–159, 1992. doi: 10.1037/0033-2909.112.1.155.

[7] T. H. Davenport and J. G. Harris, *Competing on Analytics: The New Science of Winning*. Boston, MA, USA: Harvard Business Review Press, 2007.

[8] Eprints Soton, "Defining analytics maturity indicators: A survey approach," University of Southampton ePrints, 2012. [Online]. Available: https://eprints.soton.ac.uk/425622/1/Defining_Analytics_Maturity_Indicators_a_survey_approach.pdf

[9] R. M. Grant, "Toward a knowledge-based theory of the firm," *Strategic Management Journal*, vol. 17, no. Winter Special Issue, pp. 109–122, 1996. doi: 10.1002/smj.4250171110.

[10] L. J. Groot, K.-J. Kan, and S. Jak, "Checking the inventory: Illustrating different methods for individual participant data meta-analytic structural equation modeling," *Research Synthesis Methods*, 2024. doi: 10.1002/jrsm.1735.

[11] J. F. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd ed. Thousand Oaks, CA, USA: Sage Publications, 2019.

[12] IntelligentCIO, "How to unlock value from data lakes," Feb. 28, 2021. [Online]. Available: https://www.intelligentcio.com/apac/2021/03/01/how-to-unlock-value-from-data-lakes/

[13] J. Lismont, J. Vanthienen, B. Baesens, and W. Lemahieu, "Defining analytics maturity indicators: A survey approach," *International Journal of Information Management*, vol. 37, no. 4, pp. 271–283, 2017. doi: 10.1016/j.ijinfomgt.2016.12.003.

[14] W. J. Orlikowski, "A socio-technical view of knowledge at work," *Journal of Organizational Behavior*, vol. 13, no. 6, pp. 517–534, 1992. doi: 10.1002/job.4030130605.

[15] SAS Institute, "What is a data lake & why does it matter?" SAS, Apr. 12, 2026. [Online]. Available: https://www.sas.com/en_us/insights/articles/data-management/what-is-a-data-lake-and-why-does-it-matter-.html

[16] M. Saunders, P. Lewis, and A. Thornhill, *Research Methods for Business Students*, 8th ed. London, UK: Pearson Education, 2019.

[17] R. Saxena and A. Srinivasan, "Analytics maturity model," 2013. [Online]. Available: https://www.slideshare.net/slideshow/analytics-maturity-model/18482542

[18] ScienceDirect, "Partial least squares structural equation modeling (PLS-SEM) in research," 2022. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S2772766122000246

[19] Slideshare, "Web analytics maturity model," 2009. [Online]. Available: https://www.slideshare.net/slideshow/web-analytics-maturity-model/1428591

[20] STS Journal, "Framework to build an advanced analytics maturity assessment model," *STS Journal*, vol. 21, 2024. [Online]. Available: https://journals.va.lv/index.php/STS/article/download/79/45/93

[21] Eindhoven University of Technology, "A model for defining maturity levels for organizational capabilities," May 30, 2024. [Online]. Available: https://research.tue.nl/en/publications/understanding-capability-progression-a-model-for-defining-maturit/

[22] Propulsion Tech Journal, "Unified data analytics platform for financial sector using big data," *Propulsion Tech Journal*, vol. 27, Nov. 2, 2023. [Online]. Available: https://www.propulsiontechjournal.com/index.php/journal/article/view/1560

[23] I. Ahmed and M. Asif, "The Role of HR in Managing Quiet Quitting and Employee Disengagement in Gen Z Employees of Telecom Sector," *Policy Journal of Social Science Review*, vol. 4, no. 6, pp. 118–151, 2026, doi: 10.5281/zenodo.20581688.

[24] M. Asif, M. Abid, and A. Riaz, "Psychological drivers of investment decision making: A multi bias analysis of an emerging market's retail investors," *Contemporary Journal of Social Science Review*, vol. 4, no. 2, pp. 677–688, 2026, doi: 10.63878/cjssr.v4i2.2608.

[25] R. D. A. Khan, H. Ping, and M. Asif, "The impact of green human resource management on employee green performance through green commitment and transformational leadership," *Center for Management Science Research*, vol. 4, no. 5, pp. 635–677, 2026, doi: 10.5281/zenodo.20510765.

[26] D. Mohiuddin, M. H. Tariq, and A. Tahir, "The Impact of Generative AI on Personalized Content Marketing in E-Commerce," *Inverge Journal of Social Sciences*, vol. 4, no. 1, pp. 162–188, 2025. doi: 10.63544/ijss.v4i1.288.