IO4 Research results for Data Literacy Assessment

Author:
Dr. Guido Ongena –
Hogeschool Utrecht

Co-funded by the Erasmus+ Programme of the European Union

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Data literacy for improving governmental performance

PROJECT INFORMATION

PROJECT TITLE:
Data Literacy in Context

PROJECT ACRONYM:
DaLiCo

PROJECT NUMBER:
2019-1-DE01-KA203-005066

ERASMUS+ PROGRAMME:
KA2 Partnerships for Innovation and the Exchange of Good Practices

WEBSITE:
www.dalico.info

PARTNERS:

HAW HAMBURG | HOGESCHOOL UTRECHT UNIVERSITY OF APPLIED SCIENCES | UNIVERSITY of DEBRECEN | UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Co-funded by the Erasmus+ Programme of the European Union

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Data literacy for improving governmental performance

AUTHOR

Dr. Guido Ongena (Hogeschool Utrecht University of Applied Sciences)

ABSTRACT

Big data analytics received much attention in the last decade and is viewed as one of the next most important strategic resources for organizations. Yet, the role of employees' data literacy seems to be neglected in current literature. The aim of this study is twofold: (1) it develops data literacy as an organization competency by identifying its dimensions and measurement, and (2) it examines the relationship between data literacy and governmental performance (internal and external). Using data from a survey of 120 Dutch governmental agencies, the proposed model was tested using PLS-SEM. The results empirically support the suggested theoretical framework and corresponding measurement instrument. The results partially support the relationship of data literacy with performance as a significant effect of data literacy on internal performance. However, counter-intuitively, this significant effect is not found in relation to external performance. This paper is published under the title "Data literacy for improving governmental performance: A competence-based approach and multidimensional operationalization". https://doi.org/10.1016/j.digbus.2022.100050

DATE: 2022-10-31

LICENCE

Co-funded by the Erasmus+ Programme of the European Union

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Contents

1 Introduction

Increased availability of big data fuels the proliferation of and attention to big data analytics. Big data is considered one of the most valuable strategic business resources in the coming years (Batistic & van der Laken, 2019). As the volume, variety, and velocity of data are expanding, a growing number of organizations initiate and deploy big data analytics initiatives to develop critical insights. The reason behind these initiatives and accompanying investments is that they can provide them with a competitive advantage. Scholars stipulate this potential value creation in a myriad of industries, including the hospitality (Horng, Lio, Chou, Yu, & Hu, 2022), the healthcare (Yu, Zhao, Liu, & Song, 2021), retail (Santoro, Fiano, Bertoldi, & Ciampi, 2019) and circular economy (Kristoffersen, Mikalef, Blomsma, & Li, 2021). To reap the benefits of big data analytics, the organization's workforce needs to interpret this vast amount of data to turn it into a business value (Pothier & Condon, 2019; Tabesh, Mousavidin, & Hasani, 2019; Giudice da Silva Cezar & Maçada, 2021). A workforce that is data-literate and that values information as an asset to enhance the data-driven culture and subsequently achieve fruitful big data analytics projects. More specifically for the public sector, it is suggested that data skills support the improvement of public services as well as decision and policy-making processes by government employees (Gascó-Hernández, Martin, Reggi, Pyo, & Luna-Reyes, 2018; Zhang, Porter, Cunningham, Chiavetta, & Newman, 2021).

In practice, reports - most of which are written by technology consultants and vendors - highlight that a vast percentage of the workforce needs to enhance their data literacy (Bersin & Zao-Sanders, 2020). Several business intelligence vendors like Tableau (Bryla, 2018) and Qlik (Morrow, 2018) stipulate the importance of data literacy. For the public sector, the OECD (2017) enlists data literacy as an essential skill for innovation. One of the organization's critical priorities is thus to foster data literacy across their organization (Hippold, 2019) which is advocated by policymakers (Rayna & Striukova, 2021).

Whilst heralded by practitioners and policymakers, extant academic literature is scarce on the role of data skills in the workforce to enhance the business value of big data analytics initiatives. Recently, Murawski and Bick (2017) underscored the need for research on workforce competences. Despite recent efforts by Sternkopf and Mueller (2018), there is little theoretically grounded knowledge about the concept, dimensions, and measurement, and the role of data literacy in governmental organizations in relation to its performance. To address these concerns a competence-based approach is adopted, and data literacy literature is used to answer two fundamental questions:

RQ1. What are the dimensions of data literacy and how can it be measured?

RQ2. How does data literacy affect the performance of governmental organizations?

This study takes a competence-based approach and seeks to examine the competences that are required to enhance data literacy in the organization. The competence-based view states that the competences of its employees are inimitable and unique and are more adequate than the organization's resources or structures, to explain achievement or competitive success (Rumelt, 1994). The data literacy literature supports understanding the essential conditions organizations need to enhance their achievements or success. To answer the main research questions, a survey is held within public organizations in the Netherlands, collecting 120 responses that supported the examination of the psychometric properties of all measures and empirically assessed the relationships in the proposed research model. The findings of this study help researchers and managers to better understand the conditions required to improve governmental performance.

In the section that follows, theoretical grounding is provided, and data literacy is conceptualized based on relevant academic and practice-based literature by defining it and setting out the dimensions of data literacy. Then, the developed data literacy measurement instrument is presented, followed by the empirical analysis and the outcomes that include an assessment of the measurement model and the structural model. The paper is concluded by discussing the findings from a research and practical standpoint and outlining some key limitations of this study.

2 Theory and conceptual background

This research uses the theoretical lens of the competence-based management (Sanchez, Heene, & Thomas, 1996; Hamel & Heene, 1994) which stems from strategic management literature to develop data literacy as a strategic and operational competence for the organization.

2.1 Competence-based theory

Competence-based theory (Sanchez & Heene, 1997) is based upon the underpinnings of theories such as the resource-based view (Barney, 1991), the knowledge-based view (Kogut & Zander, 1992; Erickson & Rothberg, 2014), strategic assets (Amit & Schoemaker, 1993), and competitive heterogeneity (Hoopes, Madsen, & Walker, 2003) that emphasize the importance of organizational resources and capabilities in creating value and competitive advantage for firms. The resource-based view (RBV) is the most popular of these theories and has seen tremendous growth in the strategic management and information systems literature. The theory has supported the development of new conceptualizations of capabilities in the domain of, for instance, big data analytics (Gupta & George, 2016).

RBV posits that organizations compete based on unique firm resources that are rare, difficult to imitate, and valuable. This suggests that an organization is more successful when it controls more effective and/or efficient resources than its competitors (Barney, 1991). The competence-based view, instead, goes one step further. An organization can only be more successful than its peers if it can make use of the available resources more effectively and/or efficiently (Freiling, 2004). This requires action-related competences that utilize these rather static resources in value-added activities for the organization.

Within a competence-based approach, competence models are used to unify individual capabilities with organizational core competences (Van Der Heijde & Van Der Heijden, 2006). Competences and skills are often used interchangeably. Sanchez (2004), for instance, refers to skills and defines them as 'special forms of capability, usually embedded in individuals or teams. Whereas Freiling and Fichtner (2010) use the term competences and refer to 'a repeatable, knowledge-based, rule-based and therefore non-random ability to render competitive output and to remain competitive'. A competency is generally defined as a set of observable performance dimensions, including individual knowledge, skills, attitudes, and behavior (Bratianu, Hadad, & Bejinaru, 2020). Competence is thus a broader concept encompassing knowledge and attitude next to skills. In the conceptualization that is central in this article, competence is defined at a team level. This is in line with the thought that data literacy must be viewed as a compound competence not just in relation to the individual domains of skill but also in relation to the heterogeneous social composition of the group that as a whole may possess data literacy (Pedersen & Caviglia, 2019).

This study contributes to the discussion on competence-based theory to help explain how data literacy adds value to the organization. Competences refer to the capacity of an organization to deploy resources, data literacy is therefore posited as a strategic and operational compound competency, respectively. This approach extends prior works in this space by not simply viewing big data analytics as a competitive advantage yielding resources, but rather viewing data analytics as a competency that can evoke performance in governmental bodies. Data literacy and its dimensions are further conceptualized in the following subsections.

2.2 Data literacy

Although data literacy is a new kid on the block compared with other literacies (Stordy, 2015), it is an emerging field of study in academic literature from various disciplines (Yousef, Walker, & Leon-Urrutia, 2021) other than the information systems and strategic management discipline. Over the last decade, the number of publications increased exponentially due to increased attention. Fig. 1 presents the total number of publications within the information systems (IS) domain on data literacy per year.

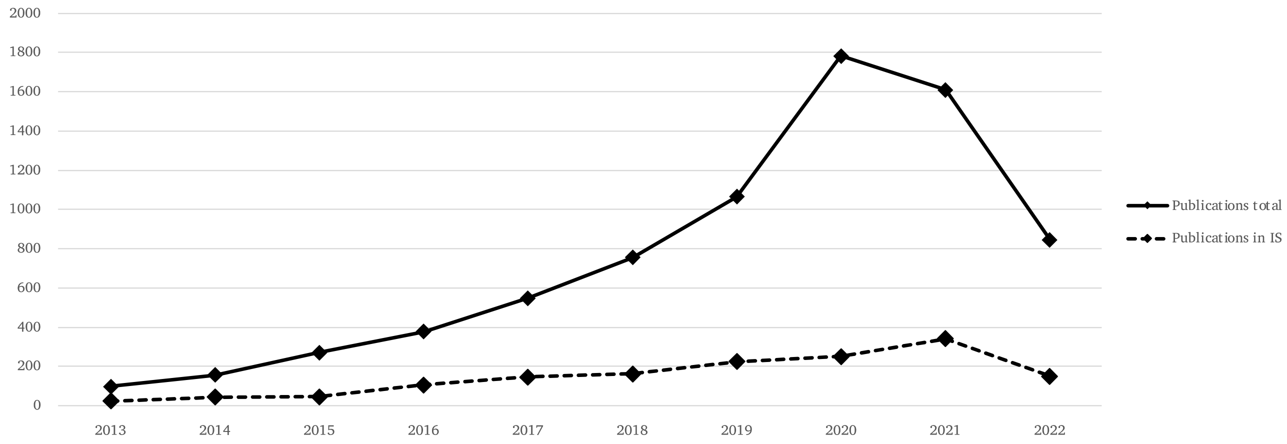

Fig. 1. Number of publications on data of literacy per year (source: https://app.dimensions.ai/)

Despite the increased attention in the last decade, to date, there is no unifying and accepted definition throughout the literature. Data literacy is an evolving concept, and it is in the need of unified terminology and definition (Koltay, 2015). This can support the grounding of empirical studies on data literacy and a better understanding in its entirety. To build a coherent understanding of data literacy, it is necessary to first explore the notion of "literacy", before ascribing this concept to data and defining the compound term "data literacy". The Oxford dictionary defines literacy first as “the ability to read and write”, the second meaning provided is: “knowledge or skills in a specific area”. This second meaning is embraced by UNESCO that, due to a rapidly evolving digital environment, details the concept beyond its conventional set of reading, writing, and counting skills, as literacy is now understood as a means of identification, understanding, interpretation, creation, and communication (UNESCO, 2021). In combination with the above, the notion of "data" pertains to that specific area where these means can be applied. Building on the meaning of these two core notions, it is crucial that a more sophisticated understanding of the term data literacy is developed. To enable a more holistic and comprehensive understanding of what data literacy is, five definitions of data literacy are identified and selected from relevant (practice-based) articles, which are presented in Table 1.

These definitions for the most part address the reading and writing of data as an analogy to the earliest definition of literacy. Interestingly, the definitions of Wolff et al. (2016) and Ridsdale et al. (2015) explicitly emphasize critiquing the data. In contrast to knowledge and skills, this relates to the attitude toward the data. Attitude in the form of critical thinking and dealing with data ethically is increasingly recognized in building 21st-century citizenry (Ridsdale, et al., 2015). Building on these definitions, as well as on the delineation of the two comprising terms that form the overall notion, an integrative definition of data literacy is provided that is used throughout this article. The goal in doing so is not to provide yet another definition of data literacy, but one that is relevant in the context of information systems research and its role in the organizational setting. Hence, the following definition is formulated: *Data literacy is the ability of an employee to identify, understand, use, reflect, and communicate data to achieve predetermined organizational and societal goals.*

This definition views data literacy as a competence[1] and is most in line with the definition of Ridsdale et al. (2015), with the exception that it is limited in scope as it is formulated toward the study of management and information systems–related phenomena. By developing this definition, it is thus easier to identify what does and what does not constitute data literacy within the organizational setting.

[1] In academic literature the term skills is often used. The term competence is favored as this is a broader concept encompassing knowledge and attitude next to skills.

Table 1
Sample definitions of data literacy

Author(s)	Definition
Wolff, Gooch, Cavero Montaner, Rashid, and Kortuem (2016, p. 23)	Data literacy is the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data. It is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyze, visualize, critique and interpret data, as well as to communicate stories from data and to use data as part of a design process.
Ridsdale, et al. (2015, p. 2)	Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner.
Data to the People (2018)	Data literacy is our ability to read, write and comprehend data, just as literacy is our ability to read, write and comprehend our native language.
Gartner (2019)	The ability to read, write and communicate data in context, with an understanding of the data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case application and resulting business value or outcome.
Sternkopf and Mueller (2018)	Data literacy is a continuous learning journey that creates the ability to identify, understand, interpret, create, communicate, and compute pieces of information (data) to develop knowledge and the ability to participate fully in our society.
Mandinach & Gummer (2013, p. 30)	The ability to understand and use data effectively to inform decisions which involves "knowing how to identify, collect, organize, analyze, summarize, and prioritize data" and "how to develop hypotheses, identify problems, interpret the data, and determine, plan, implement, and monitor courses of action."

2.3 Dimensions of data literacy

Although the published research on data literacy in the organizational setting is still quite limited, some studies have identified dimensions of data literacy. A large proportion of these studies have been from the practice-based press. Nevertheless, there is a lack of theoretically grounded research about how organizations can enhance employees' data literacy. This is an important gap for both research and practice, as it can indicate the core areas that organizations should steer their focus toward when deploying data initiatives and provide a notion upon which to gauge the potential organizational value and mechanisms of value creation.

The research framework of this study is built upon the two most comprehensive frameworks on data literacy, from Sternkopf and Mueller (2018) and Ridsdale et al. (2015). To capture the prevalent behavior of employees, verbs are used to formulate the first-order constructs. Not only this is in line with previous research, but it also emphasizes the processing and working with, the data. The framework is enriched by the framework of Pangrazio and Selwyn (2019) to develop a higher-order construct. Higher-order constructs have several advantageous features. One reason is that they help to reduce the number of path model relationships, thereby achieving model parsimony. Another reason pertains to the bandwidth-fidelity tradeoff or the idea that broader constructs are better predictors of criteria that span multiple dimensions (Jenkins & Griffith, 2004; Johnson, Rosen, & Chang, 2011). In addition, from a statistical viewpoint, higher-order constructs provide a means for reducing collinearity among formative indicators by offering a vehicle to rearrange the indicators and/or constructs across different concrete subdimensions of the more abstract construct (Sarstedt, Hair, Jun-Hwa, Becker, & Ringle, 2019).

Eleven skills are proposed that jointly constitute data literacy (Fig. 2). In the sub-sections that follow, each of these resources is discussed in detail.

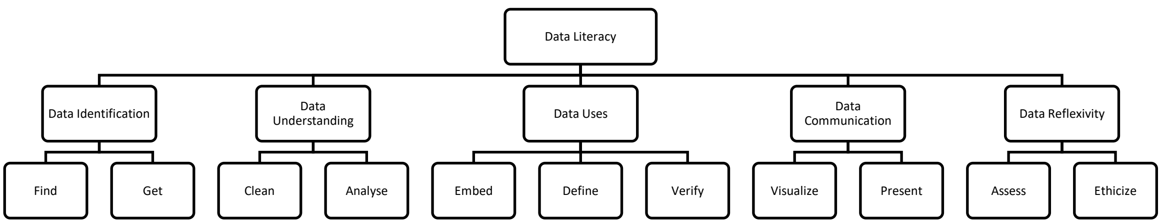

Fig. 2. Data literacy and categorization of dimensions

2.3.1 Data Identification

The first step in turning data into organizational value is to discover and obtain relevant data. One should be able to *find* this data, what it describes, in which systems it is collected, and what type it is. In addition to its discovery, it is imperative to know how to access the data and import it into applications that can explore the data. Data is accessible in various formats. A typical form is a comma-separated values (.csv) file, which is a delimited text file that uses a comma to separate values. This format is often used for internal data. However, data is increasingly stored in the cloud rather than locally. Data exchange between different devices and applications has thus become a necessity nowadays, to support the exchange of data between devices filetypes like XML or JSON. Especially the JSON format is often favored as it is more effective concerning the response time (Breje, Győrödi, Győrödi, Zmaranda, & Pecherle, 2018). These formats are specifically used when the data is made accessible through API's which makes it possible for an organization to *get* this external data. Both external and internal data have expanded in volume in the last decade which poses a great challenge for organizations (Colson, 2019), especially concerning real-time or near-real-time data.

2.3.2 Data Understanding

Assessing the data quality and repairing 'dirty' data is one of the perennial challenges in data analytics. Data quality problems are present in single data collections due to misspellings during data entry, missing information, or other invalid data. This becomes even more complex and present when multiple sources are combined as sources often contain redundant data in different representations (Rahm & Hai Do, 2000). It is thus imperative to *clean* the data as it otherwise can result in inaccurate analytics and unreliable decisions when not done adequately (Chu, Ilyas, & Krishnan, 2016). The importance of data cleaning has increased with the renaissance of big data analytics and also calls for new and more advanced approaches to improve the quality of the data (Gudivada, Apon, & Ding, 2017). Cleaning the data has a symbiotic relation with *analyzing* the data as exploring the data has to be done first to determine actions for cleaning the data (e.g., which records or cases should be deleted or altered). In academia, this is often referred to as exploratory data analysis which is used to maximize insight into the data and to spot anomalies. Surveys show that analyzing and cleaning, thus preparing the data, is the most time-consuming task for employees that work with data (Press, 2016).

2.3.3 Data Uses

Naturally, data initiatives must add value to the business. Value is created by adequately *defining* business questions to the data. Precise and target-oriented questions to find meaningful answers enhance the potential fruitful outcomes of these initiatives. Moreover, evaluating the data analyses outcomes against a business or project goals supports. This also means that the business should be involved to *verify* these outcomes. Data initiatives are no longer solely relevant to IT departments. It is relevant to every single member of a functional organization (Pagador, Huynh, Davis, & Abhari, 2020). Additionally, cultivating data literacy, or *embedding* data within the organizational culture, allows an organization, especially government agencies, to become data literate (Pagador, Huynh, Davis, & Abhari, 2020). This pertains to the idea that the use of data is supported by higher management and is seen as an enabler or opportunity rather than a threat.

2.3.4 Data Communication

Using data to support a larger narrative intended to communicate some message to a particular audience (Bhargava & D'Ignazio, 2015) is often an integral part of data literacy definitions. On the one hand, this means that an employee should be able to *present* data analysis outcomes in a comprehensive way to relevant stakeholders. Turning data insights into a narrative and communicating this narrative to an audience is often linked to processes like storytelling (Lee, Riche, Isenberg, & Carpendale, 2015). Storytelling has a long history during which it has been used in many ways in different domains. As a result, it invokes meaning and nuance without having one single agreed-upon definition. Most definitions of storytelling require some sort of controlled delivery or presentation of information. It thus goes hand in hand with constructive visualizations to convey and support the message (Lee, Riche, Isenberg, & Carpendale, 2015). Moreover, having skills to create and comprehend mapped data, graphs, pie charts, and emerging forms of visualizations, thus *visualizing* the data, will help to effectively navigate visually rich data sets (Fontichiaro & Oehrili, 2016) that support data understanding.

2.3.5 Data Reflexivity

The Oxford dictionary defines reflexivity as “the process of self-consciousness where an individual subject or group becomes the object of its own scrutiny” (Oxford Reference, 2021). Concerning data, this pertains to the question of what the implications are of the utilization of data (Pangrazio & Selwyn, 2019). The development of data-driven technologies can impact the organization as well as (negatively) affect society. Or to paraphrase Pentland (2013): "data can be used for good or ill". Employees thus should be able to *assess* the data on these implications. Moreover, they should develop a critical thinking competence about the larger issues regarding the use of data and must have an understanding and awareness of the *ethics* surrounding data (Ridsdale, et al., 2015; Wolff, Gooch, Cavero Montaner, Rashid, & Kortuem, 2016).

2.4 Government performance

A myriad of researchers examined the black box of governmental performance over the years. Rainey and Steinbauer (1999), for instance, developed a theory that posited that organizational performance was affected by political authorities, agency autonomy in refining and implementing its mission, high mission valence, a strong, mission-oriented culture, and certain leadership behaviours. These determinants were empirically tested which confirmed most of the hypotheses (Brewer & Selden, 2000). Later studies found additional determinants that positively affect organizational performance, especially from a human-resource perspective. For instance, Giauque, Anderfuhren-Biget, and Varone (2013) found several intrinsic motivators (e.g., fairness, job enrichment, individual appraisal, and professional development) that contribute to organizational performance. Similarly, research showed a positive impact of family-friendly work practices (Ko, Hur, & Smith-Walker, 2013).

Although Rainey and Steinbauer (1999) already provided a valuable stepping-stone with their proposed theory of effective government organization two decades ago, much attention was paid to the determinants of governmental performance, but less to the concept itself. The measurement of performance by governmental bodies is thus still a youthful and under-investigated field of research (Andrews, Boyne, Moon, & Walker, 2010). Therefore, there is hitherto no consensus on how to measure governmental performance. The proposed measures of organizational performance based on the perceptions of the organization’s members by Brewer and Selden (2000) are to date the most comprehensive and theoretically founded measurement instrument, that is also successfully used by other scholars (e.g., Kim, 2005). Their measurement comprises two organizational foci dimensions (i.e., internal, and external) and three administrative value dimensions (i.e., efficiency, effectiveness, and fairness). Table 2 shows the different dimensions.

Table 2.
Dimensions of governmental performance (adopted from Brewer and Selden (2000))

			Administrative Values	
		Efficiency	Effectiveness	Fairness
Organizational Focus	Internal	Internal Efficiency	Internal Effectiveness	Internal Fairness
	External	External Efficiency	External Effectiveness	External Fairness

3 Research model

To examine the structure fit in a nomological network this study has employed internal and external performance for evaluating nomological validity. Based on the argumentation above on the role of data literacy in organizations, it is clear that a lot of emphasis has been placed on the role data-driven innovations may play in improving organizations’ performance. As this study views data literacy as compound competence, data literacy is posited as a team competence which is the manifestation of knowledge, skills and abilities of individuals in the teams (Eby & Dobbins, 1997; Potnuru & Sahoo, 2016). In line with prior research (e.g., Salman, Ganie & Saleem, 2020) this study assumes that team-level competence will positively affect organizational performance. The argumentation is developed on this relationship through the conceptual research model presented in Fig. 3.

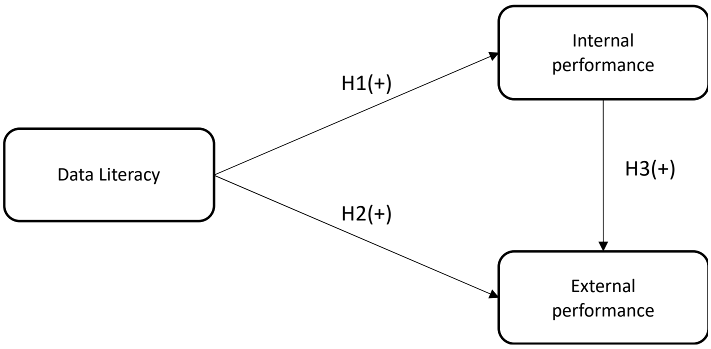

Fig. 3. Conceptual research model

The employment of data is used toward organizational goals so they can enable managers to gain insight that was previously unobtainable by making sense of vast amounts of data and uncovering patterns and relationships. Based on the RBV of an organization, literature shows evidence that enhancing data analytics competency can help organizations improve their internal decision-making performance (Ghasemaghaei, Ebrahimi, & Hassanein, 2018). In addition, more specifically to the governmental context, it is suggested that it will improve policy-making processes by government employees (Gascó-Hernández, Martin, Reggi, Pyo, & Luna-Reyes, 2018). Adequate use of data thus supports this data-driven decision, or policy-making, which leverages the internal performance of a governmental organization.

Yet, enhancing the internal process is not the only way in which data can deliver value to organizations. In line with the competence perspective that can be used to describe internal factors that contribute to creating value in the market (Harmsen & Jensen, 2004), data-driven technologies also cross-organizational boundaries. Such technologies can also be deployed to improve the efficiency, effectiveness, and quality of public service delivery through applications like chatbots and intelligent agents (Zuiderwijk, Chen, & Salem, 2021). Big data analytics literature suggests a strong link between big data analytics capabilities and organization performance, and provides empirical evidence for this (e.g., Rialti, Zollo, Ferrari, & Alond, 2019). Performance is in these studies typically measured by the organizational financial and market position in relation to its competitors. Based on the foregoing argumentation, the following hypotheses are posited:

H1. A higher degree of data literacy will have a positive effect on internal performance
H2. A higher degree of data literacy will have a positive effect on external performance

Early public management literature stipulates the importance of distinguishing between internal and external performance for governmental bodies (Epstein, 1992). Both internal and external performance entails the efficiency and effectiveness-related values which are either within their organization or team or towards customers or the public. To develop a more comprehensive picture of the performance of a public service organization, Kim (2005) added the performance-related value of fairness. This pertains to the provision of fair and equitable treatment of employees (internal) and services to the public (external), with no consideration of their backgrounds. The governmental internal performance on management and operation contributes substantially to the achievement of the goals and mission (Rainey & Steinbauer, 1999). The latter is externally oriented. Thus, this study hypothesizes the following:

H3. A higher degree of internal performance will have a positive effect on external performance

4 Methodology

4.1 High-order construct specification

The data literacy construct is conceptualized as a multidimensional third-order formative construct, which is comprised of the following data literacy dimensions: Data Identification, Data Understanding, Data Uses, Data Communication, and Data Reflexivity. These dimensions are, in turn, conceptualized as second-order formative constructs comprising ten reflective first-order constructs (Table 3). Hence, a reflective-formative type II model typically implies the use of Mode A to estimate specified measurement models (Becker, Klein, & Wetzels, 2012).

The choice of a reflective or formative higher construct is not a trivial one and is important as it may result in a different overall construct (Diamantopoulos & Winklhofer, 2001). Examining whether data literacy and digital literacy have been developed as reflective or formative constructs in prior studies is the first step. To date, there is however no literature that specified either data or digital literacy as a high-order construct. Recent studies such as those of Jang, Aavakare, Nikou, and Kim (2021) and Cetindamar, Abedin, and Shirahada (2021) developed their conceptualization of digital literacy as a first-order construct. More in general on competency constructs, Ghasemaghaei, Ebrahimi, and Hassanein (2018) conceptualized their big data analytics competency as a higher-order formative construct.

Literature is scarce and widely accepted decision rules and considerations are used to conceptually assess whether the construct should be developed as a higher-order formative or reflective one (Jarvis, MacKenzie, & Podsakoff, 2003; Petter, Straub, & Rai, 2007; Coltman, Devinney, Midgley, & Venaik, 2008). First, from the proposed underlying dimensions, there is no single one that can adequately explain the notion of data literacy. This observation is a strong criterion that the dimensions are core characteristics, rather than manifestations of data literacy. The dimensions that comprise the construct cover complementary facts of overall data literacy. Second, the indicators in the research model do not share a common theme and are thus not interchangeable. Consequently, the omission of an indicator alters the conceptual domain of the construct. Thirdly, the indicators do not necessarily covary with each other. For instance, one can develop competences in using and analyzing data but that does not necessarily entail that the person also mastered the presentation of the data. Based on the above, it is argued that the second-order constructs fit the criteria of a formative model.

The theoretical assumptions are also empirically tested through confirmatory tetrad analysis (CTA-PLS). CTA-PLS provides insights into whether a reflective indicator specification or formative indicator specification is more appropriate. Following the CTA-PLS process (Gudergan, Ringle, Wende, & Will, 2008), first, all vanishing tetrads for the measurement model of each latent variable are formed and computed; then model-implied vanishing tetrads are identified, which is followed by eliminating redundant model-implied vanishing tetrads. Then based on examination of the statistical significance test for each vanishing tetrad, the results are evaluated for all model-implied non-redundant vanishing tetrads per measurement model. The results provided support for the reflective mode for the measurement of the first-order constructs of data literacy.

Table 3
Latent constructs and sub-dimensions

Third order	Type	Second-order (sub-dimensions)	Type	First-order (sub-dimensions)	Type
Data Literacy	Formative	Data Identification	Formative	Find	Reflective
				Get	Reflective
		Data Understanding	Formative	Clean	Reflective
				Analyze	Reflective
		Data Uses	Formative	Embed	Reflective
				Define	Reflective
				Verify	Reflective
		Data Communication	Formative	Visualize	Reflective
				Present	Reflective
		Data Reflexivity	Formative	Assess	Reflective
				Ethicize	Reflective

4.2 Measurements

The measures used to develop the first-order constructs were adapted or created from existing literacy literature (e.g., Sternkopf & Mueller, 2018). As such, a data literacy construct differs significantly from other literacies like digital literacy and information literacy constructs as the dimensions that it is comprised of are data specific. Digital literacy relates to the competences required for working with digital technologies like digital media, information processing, and retrieval (Cetindamar, Abedin, & Shirahada, 2021). Although these digital technologies entail or are based on data, it is the combination with other competences (e.g., interpreting the data) that collectively lead to the emergence of data literacy. This idea is reflected in the proposed theoretical framework (Fig. 2) and in the items used to capture the first-order constructs, which are related specifically to data literacy as in Table 4.

Reflective multi-item measurement was created for the first-order constructs. The use of multi-items is favorable as, from a psychometric perspective, the combination of multiple items in reflective measurement averages out random error in the items, thereby increasing the reliability (Sarstedt & Wilczynski, 2009). Moreover, as multiple items cover a larger number of distinct construct facets, they also offer higher degrees of construct validity (Wanous, Reichers, & Hudy, 1997). Conversely, single items that may perform as well as the multi-item scale in one context may not do so in another (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012) and is thus not without risks. To mitigate this risk and provide a more robust assessment, this study uses multi-item scales as it provides more predictive validity (Cheah, Sarstedt, Ringle, Ramayah, & Ting, 2018).

To ensure the content validity of the indicators, a group of experts on data literacy was asked to provide feedback on the instrument. A group of eight senior academics (15+ years of experience) with expertise in data literacy from different disciplines (e.g., statistics, information science, information systems) were individually asked to provide recommendations on questions that were not comprehensive or aspects of questions that could be improved. The feedback provided resulted in some minor modifications, including some examples. This feedback shows that the content validity of the instrument was established.

Table 4
Constructs and measures of data literacy (adopted from Sternkopf & Mueller, 2018)

Second-order construct	First-order construct	Items
Data Identification	Find	We have a broad understanding of different data sources, and most relevant ones can be chosen from a selection of data sources
		We are aware of and use data portals for specific topics
	Get	We can access data using more complex data formats (e.g., JSON, XML)
		We are able to use APIs to get data
Data Understanding	Clean	We can detect and remove invalid records using programs that support data cleaning (e.g., OpenRefine)
		We have a high awareness of data quality criteria (e.g., machine processable, empty fields, duplicate detection)
	Analyze	We are able to work with advanced statistics (e.g., inferential view of data, linear regression, decision trees)
		We can research and select the most appropriate tool for our analysis needs
Data Uses	Embed	We perceive data as a source of security and see it as an enabler for progress and support for existing and planned activities
		Higher management and leaders support data initiatives
	Define	We are able to formulate questions to data precisely and target-oriented to find meaningful answers in most of the cases
		We are able to identify and describe problems in practical situations using a range of data sources
	Verify	We have implemented multiple layers of data checking in standard procedures across the organization
		We can identify whether the data is trustworthy and locate alternate sources if required
Data Communication	Visualize	We are able to create interactive charts and dashboards
		We are able to visualize uncertainties along with the data
	Present	Our projects are supported by interactive visualizations and more sophisticated narrative in a broader context (e.g., data storytelling, conferences, talks, monthly updates, blog posts)
		We can read and understand a range of tables, charts and graphs
Data Reflexivity	Assess	We can interpret data outputs and results confidently and critically
		We have internalized evaluation criteria
	Ethicize	We are aware of the impacts of data use
		We have defined guidelines for responsible data handling and have incorporated them internally to activities

A nomological network is used to assess the validity of the multidimensional structure as recommended by MacKenzie, Podsakoff and Podsakoff (2011). To this end, this study has employed internal and external performance for evaluating nomological validity. As part of examining the nomological validity of the proposed construct, two performance metrics were introduced to capture the suggested effect that data literacy has at an organizational level. Internal and external governmental performance is therefore included, in addition to the existing constructs of the data literacy scale as introduced earlier. The items of these performance metrics are based on the measures from the study of Kim (2005). This empirical study was developed, based on prior studies, and validated the items. The set of items comprises of measures related to efficiency, effectiveness, and fairness (see Appendix A).

4.3 Data collection

The data used in this research was collected through a self-administered online questionnaire targeted at team leaders within governmental agencies. The data was collected in the Netherlands. Three strategies were used to collect the data. First, municipalities were contacted directly via email. The Netherlands consists of 352 municipalities. Second, LinkedIn was used, a business-oriented social network available worldwide since 2003, which comprised 756 million users as of August 2021, to disseminate the questionnaire. Third, taking a chain-referral sampling approach respondents were asked to share the survey among potentially relevant respondents. The data-collection process lasted for approximately three months (July 2021–September 2021). A total of 318 respondents started to complete the survey, with 120 providing complete responses.

Non-response bias was assessed. Non-response bias refers to a situation in which people who do not respond to a questionnaire may bias the research results. To determine whether there was any non-response bias in the sample, the profile of the respondents was compared with those on the mailing list that was collected for each municipality, such as size. In addition, the nonresponse approach follows the suggested wave analysis by Armstrong and Overton (1977), who suggested that late respondents are more likely to resemble non-respondents than to resemble early respondents. The first and last waves of respondents on all the variables are compared, which treats late respondents as a proxy of non-respondents. No statistically significant differences were found ($p < 0.01$). Hence, one can conclude that there is no critical degree of non-response bias.

The responses received came from a representative group of municipalities in the Netherlands. Table 5 shows the distribution of size classes in terms of number of employees. As can be seen in the same table, the experience of the respondents is somewhat skewed to both ends of the spectrum. Most respondents are experienced with data (more than 4 years) or are inexperienced (less than 1 year).

To assess common method bias, this study first used procedural controls (ex-ante) during the design of the survey (Podsakoff, MacKenzie, & Podsakoff, 2012). Respondents were assured that all the information they provided would remain completely anonymous and confidential and stipulated and that any analysis would be done on an aggregate level solely for research purposes. Clear instructions were also provided, avoiding complex and ambiguous items. The latter was done through pre-testing by survey experts, which subsequently refined the formulation of the questions and further eliminated any repeated or similarly sounding items. In addition, the Harman's one-factor test was employed as a statistical control (ex-post). The results show that 55.2% of the variance was explained by one single factor. Recent literature (Fuller, Simmering, Atinc, Atinc, & Babin, 2016) indicates that Harman's test produced both false positives and false negatives, and the potential for a false positive conclusion is particularly high when scale reliability is high (>0.90). As this was the case in this study (see

Table 6), the explained variance should not exceed 70% or more before substantial concern about inflated relationships can arise. These steps assured that common method bias was not an issue with the collected data.

Table 5
Sample characteristics

	Frequency (N=120)	Percentage (%)
Size-class of the organization (number of employees)		
1-50	7	5.8%
51-100	5	4.2%
101-1,000	70	58.3%
More than 1,000	38	31.7%
Length of data experience (number of years)		
Less than 1 year	38	31.7%
1 - 2 years	16	13.3%
2 - 3 years	10	8.3%
3 - 4 years	11	9.2%
More than 4 years	45	37.5%

5 Results

5.1 Measurement model

A confirmatory factor analysis (CFA) was conducted to validate the instrument. CFA combines ex-ante theoretical expectations with empirical data to validate the factor structure and is therefore a stronger statistical method than its alternatives (e.g., exploratory factor analysis) in the theory-driven instrument development (Bhattacherjee, 2002). CFA enables testing whether the assumed theoretical relationships exist between the observed indicator variables and their underlying latent constructs. The method of partial least squares structural equation modeling (PLS-SEM) is employed with the software SmartPLS (Ringle, Wende, & Becker, 2015). PLS-SEM is the preferred approach when formative constructs are included in the structural model (Hair, Risher, Sarstedt, & Ringle, 2019). Moreover, it is less demanding on measurement scales (Gefen, Straub, & Boudreau, 2000).

The first-order constructs in this study are specified as reflective constructs. Typical metrics applied to assess the measurement model with reflective constructs include reliability, convergent validity, and discriminant validity (Sarstedt, Hair, Jun-Hwa, Becker, & Ringle, 2019). Cronbach's alphas and the composite reliabilities were calculated to examine the reliability of the instrument. The reliability met the conventional standard of internal consistency, with both Cronbach's alphas composite reliabilities falling between 0.703 and 0.958. The values of the average variance extracted (AVE) are greater than 0.5 (see

Table 6). Hence, it is assumed that convergent validity exists in the theoretical model (Fornell & Larcker, 1981). Further, the discriminant validity is examined in three ways. First by comparing the square root of AVEs of the constructs concerning the correlation between two constructs, also known as the Fornell–Larcker criterion. None of the correlations between the latent constructs were found to be higher than the square root AVE for each individual construct. Next, each indicator loadings show greater values than its cross-loadings with other constructs (see Appendix B). In addition, the heterotrait-monotrait ratio of correlations (HTMT) is used to assess discriminant validity (Henseler, Ringle, & Sarstedt, 2015). To clearly discriminate between two factors, the HTMT should be smaller than 0.90 (Henseler, Hubona, & Ray, 2016). The obtained correlations (see Appendix C) show that not all of the values comply with that criterion. A few are above this threshold. As they comply with the other two criteria (Fornell–Larcker and cross-loadings), this is thus not a major issue. Hence, it is confirmed that the constructs of the research model possess sufficient discriminant validity.

Table 6
Assessment of reliability, convergent, and discriminant validity of reflective first-order constructs

	Construct	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Find	*0.931*												
2	Get	0.713	*0.936*											
3	Clean	0.510	0.573	*0.878*										
4	Analyze	0.668	0.690	0.697	*0.959*									
5	Embed	0.538	0.516	0.365	0.487	*0.879*								
6	Define	0.645	0.595	0.575	0.651	0.534	*0.947*							
7	Verify	0.639	0.623	0.687	0.665	0.515	0.632	*0.936*						
8	Visualize	0.762	0.723	0.559	0.728	0.592	0.710	0.654	*0.936*					
9	Present	0.750	0.685	0.550	0.750	0.532	0.647	0.653	0.789	*0.882*				
10	Assess	0.679	0.643	0.536	0.730	0.541	0.700	0.690	0.756	0.824	*0.898*			
11	Ethicize	0.546	0.490	0.381	0.417	0.509	0.550	0.567	0.511	0.460	0.541	*0.918*		
12	Internal perf.	0.483	0.407	0.153	0.303	0.460	0.345	0.338	0.436	0.476	0.412	0.520	*0.737*	
13	External perf.	0.548	0.495	0.367	0.519	0.520	0.580	0.493	0.589	0.614	0.587	0.556	0.733	*0.707*
AVE		0.867	0.876	0.771	0.919	0.772	0.896	0.875	0.876	0.778	0.807	0.842	0.500	0.543
Cronbach's Alpha		0.846	0.858	0.703	0.912	0.705	0.884	0.858	0.859	0.715	0.761	0.813	0.797	0.831
Composite Reliability		0.929	0.934	0.871	0.958	0.871	0.945	0.934	0.934	0.875	0.893	0.914	0.855	0.877

AVE = Average Variance Extracted

The relevant criteria for formative constructs differ from reflective constructs because formative constructs are aggregate constructs that do not assume the covariation of indicators. A two-stage approach is taken by calculating the latent variables scores from the first-order constructs (first-stage) and then additionally utilizing these in the model instead of the reflective items (stage-two) which is a recommended approach with regard to high-order constructs (Sarstedt, Hair, Jun-Hwa, Becker, & Ringle, 2019). Indicator collinearity, statistical significance, and relevance of the indicator weights are examined. To evaluate collinearity the variance inflation factors (VIFs) are calculated. VIF values of 5 or above indicate critical collinearity issues among the indicators of formatively measured constructs. All VIF values, as shown in Table 7, are below the threshold and thus do not indicate problems with collinearity. Even more, the values are below the more conservative threshold of 3.3. Weights measure the relative contribution of a dimension to its construct. Whether the weights of dimensions were significant is examined. As presented in Table 7, all the weights are significant at the 0.01 level. The validity of different dimensions as significant parts of the higher-order constructs is therefore warranted.

Table 7
Assessment of VIF, significance, and relevance of the indicator weights of formative constructs

Construct	Measures	VIF	Weight	*t*-value	Significance
Data Identification	Find	2.037	0.586	9.62	p<0.01
	Get	2.037	0.493	7.73	p<0.01
Data Understanding	Clean	1.945	0.288	4.30	p<0.01
	Analyze	1.945	0.777	12.72	p<0.01
Data Uses	Embed	1.510	0.247	4.60	p<0.01
	Define	1.847	0.445	7.68	p<0.01
	Verify	1.797	0.481	8.45	p<0.01
Data Communication	Visualize	2.654	0.562	10.11	p<0.01
	Present	2.654	0.495	8.87	p<0.01
Data Reflexivity	Assess	1.415	0.809	16.44	p<0.01
	Ethicize	1.415	0.295	4.73	p<0.01

5.2 Structural model

A nomological network is used to assess the validity of the multidimensional structure as recommended by MacKenzie, Podsakoff and Podsakoff (2011). To this end, this study has employed internal and external performance for evaluating nomological validity.

The potential existence of collinearity issues among the exogeneous latent variables was first assessed. This analysis revealed that the VIF values were all below 2.701 which is lower than a suggested threshold of 5 or, ideally, 3.3 (Hair, Risher, Sarstedt, & Ringle, 2019), thus indicating that collinearity was not an issue in this study.

To test the hypotheses, the structural model is examined using a bootstrapping technique specifying 5,000 subsamples. In structural model analysis, it is important to determine the significance and association of each hypothesized path and the variance explained. The results are shown in Figure 4. Data literacy has significantly positive effects on internal performance ($\beta = 0.855$, $p < 0.001$), however not on external performance ($\beta = 0.147$, *n.s.*). Internal performance has significantly positive effects on external performance ($\beta = 0.641$, $p < 0.001$). Therefore, the hypotheses are partially supported. The size of the organization and years of experience with data was included to control for. The results show that it did not have significant effects in this model.

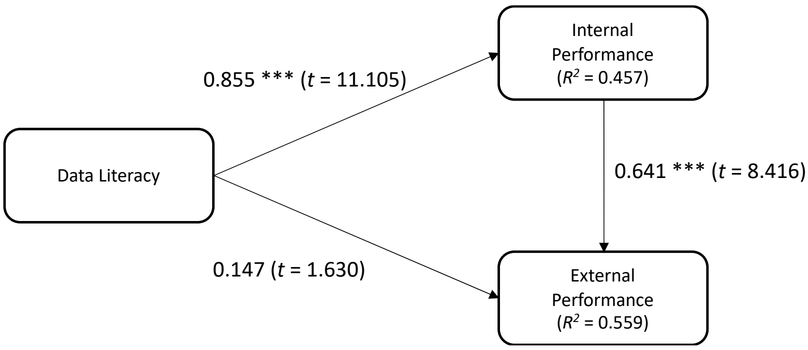

Fig. 4. Results of the structural model

The model accounted for 43.2% of the variance with regard to internal performance ($R^2 = 0.432$), and 52.8% of the variance for external performance ($R^2 = 0.528$). To assess the model's out-of-sample predictive power, the holdout-sample-based procedure of Shmueli et al. (2016) is used. This procedure generates case-level predictions on an item or a construct level to reap the benefits of predictive model assessment in PLS-SEM. To assess the predictive power of the model a *k*-fold cross-validation is executed with PLSpredict. Comparing the root mean squared error (RMSE) against the naïve linear regression model (LM) benchmark, some PLS-SEM RMSE's values are higher than LM RMSE's. However, the *majority* of PLS-SEM RMSE's values are lower. This indicates that the model has a medium predictive power (Shmueli, et al., 2019).

In summary, the results from the nomological model provide evidence for a strong, positive relationship between data literacy, and internal and external performance, as well as a highly significant positive effect of internal performance on external performance.

6 Discussion and conclusion

The purpose of this study was to operationalize data literacy, develop and validate a scale for it, and uncover the multifaceted structure that underlay the construct. Data literacy was defined as a competency encompassing five dimensions as a second-order construct: Data Identification, Data Understanding, Data Uses, Data Communication, and Data Reflexivity. Eleven first-order constructs were formulated. A content validity assessment was performed to ensure that there was sufficient rigor in the measures. Testing the instrument through a confirmatory factor confirmed the reliability and construct validity of the instrument. According to the results of these tests, all the criteria used for assessing the characteristics of data literacy were confirmed. To the best of the author's knowledge, no study has been done so far to develop a holistic high-order instrument for measuring data literacy.

The findings indicate that data literacy has a direct positive impact on internal performance. Making employees data literate thus improves the effectiveness of governmental bodies. This is in line with previous studies which found similar results in the private sector (Ghasemaghaei, Ebrahimi, & Hassanein, 2018). Our results extend this finding not only by confirming this in the public sector, but also by including fairness in the equation of internal performance. Enhancing one's data literacy thus also affects the fairness of the organization supporting equitable treatment for employees. Interestingly, the findings revealed that improving employees' data literacy is not sufficient to enhance external performance. That means that data literacy alone is not sufficient to provide the public with a worthwhile return on their tax or increase customer satisfaction. This could be caused by the fact that employees' perspective is limited and that tangible (e.g., data) and intangible resources (e.g., data-driven culture) should be taken into account to develop adequate services that improve customer satisfaction as stipulated by Gupta and George (2016). Conversely, as suggested by literature (Rainey & Steinbauer, 1999), our results show empirical confirmation that internal performance does provoke external performance. This suggests that data literacy indirectly improves external performance.

This study has several theoretical implications for data literacy research. First, a comprehensive theoretical framework was developed which entails the dimensions of data literacy. These dimensions were derived through a rigorous review of practice-based literature and scientific research. The dimensions were evaluated by experts. This process also supported the formulation of a definition that is relevant in the context of the information systems discipline. By doing so it is then concluded that a complete set of dimensions jointly comprises a data literacy competency. Second, a measurement instrument was developed that can be empirically applied to assess the data literacy of organizations. By using the latest PLS-SEM guidelines the measures were empirically validated. This study thus provided new directions for measuring data literacy in future studies. In essence, this study has provided a measurement that scholars could use to assess the higher-order construct(s) in their empirical studies. Third, the impact was demonstrated by arranging data literacy in a nomological net to validate its effect on key governmental performance. More specifically, this study assessed the degree to which data literacy affects internal and external performance. Although this relationship is suggested by practice-based literature (Bersin & Zao-Sanders, 2020; Hippold, 2019) and academics (Gascó-Hernández, Martin, Reggi, Pyo, & Luna-Reyes, 2018), this is the first empirical large-scale study linking a theoretically grounded conceptualization of data literacy with key performance indicators.

The results of this study also have important implications for managers engaged in using data analytics to improve decision-making (internal performance) and service delivery (external performance). This is the main motivation for organizations making significant investments in big data analytics technologies. Hence, having explained a large amount of the variance in both internal and external performance, managers need to pay particular attention to improving employees' data literacy. Without data-literate employees, the implementation of big data analytics technologies may fail to improve this governmental performance. The 22-item data literacy instrument is a comprehensive yet easily administered measure of data literacy. Practitioners (e.g., educators and human resource employees) can administer the measurement instrument to employees to evaluate their degree of data literacy. The data literacy instrument could be administered over successive periods to assess the changes in the degree of the different dimensions and to track the development of data literacy within the organization. By examining the relative importance and contribution of each dimension in determining data literacy, corresponding dimension-specific measures could be taken to leverage data literacy. For example, to then improve employees' data literacy, organizations could embark on training on different dimensions.

This study is not without limitations. First, the use of big data technologies can be affected by factors other than data literacy competences. Prior research also showed the importance of tangible (e.g., basic resources) assets (Gupta & George, 2016). Second, this study was conducted among Dutch governmental agencies of different sizes. Given the fundamentally different nature of the public sector to the private sector (Rainey, Backoff, & Levine, 1976), future research should examine the associations studied here, in a commercial context. Third, while this study relied on senior professionals within the governmental body, the choice of a single respondent could potentially include some bias in the results. A way to overcome this would be to opt for survey designs that collect data from multiple respondents within one organization. Finally, one must caution that ongoing debates about the issues and challenges related to formative versus reflective modeling and their interpretations (Hardin, 2017), corresponding criteria in the use of partial least squares modeling (Hair, Sarstedt, & Ringle, 2019), and particularly the assessment of complex higher-order constructs (Sarstedt, Hair, Jun-Hwa, Becker, & Ringle, 2019) as developed in this study. This study considered and applied guidelines that were current at the time of writing, but one must recognize that related debates are active and ultimate verdicts remain elusive.

This research can be considered a stepping stone for an avenue of future studies in this area. This present study contributes significantly to the presentation of a comprehensive instrument with a scientific design in the field of management and information systems. Considering the appropriate validity and reliability of the developed instrument, it can be introduced as a valid tool for managers and educators to assess the degree of their data literacy and manage them more efficiently.

7 Acknowledgments

This research was financially supported by the European Union, under the Erasmus+ Program Strategic Partnerships (grant number: 2019-1-DE01-KA203-005066). The author would like to thank [left blank for review purposes] for supporting the collection of data.

Appendix A: Governmental performance measures

Below is the measurement that is used for governmental performance. It is based on the measurement as developed by Kim (2005). It was translated into Dutch for use in this study.

Construct	Items
Internal performance	
- efficiency	My organization has made good use of my knowledge and skills in looking for ways to become more efficient.
	My organization is trying to reduce cost in managing organization and performing works.
- effectiveness	In the past two years, the productivity of my work unit has improved.
	Overall, the quality of work performed by my current coworkers in my immediate work group is high.
- fairness	My organization provides fair and equitable treatment for employees and applicants in all aspects of personnel management without regard to their political affiliation, sex, hometown, marital status, age, or handicapping condition.
	In general, all are treated with respect in my organization, with no regard to status and grade
External performance	
- efficiency	My organization has conducted business relations with outside customers very promptly.
	It is rare to make big mistakes in my organization when conducting work
- effectiveness	The work performed by my work unit provides the public a worthwhile return on their tax dollars.
	The occurrence of goal attainment is very high in my organization.
- fairness	My organization provides fair and equitable services to the public, with no considering of their individual backgrounds.
	The customer satisfaction toward my organization is very high.

Appendix B: Cross-loadings

	Find	Get	Clean	Analyze	Embed	Define	Verify	Visualize	Present	Assess	Ethicize	Internal perf.	External perf.
Find1	**0.932**	0.675	0.461	0.626	0.547	0.68	0.587	0.739	0.738	0.63	0.535	0.549	0.472
Find2	**0.930**	0.654	0.488	0.618	0.454	0.52	0.602	0.679	0.658	0.635	0.480	0.470	0.426
Get1	0.690	**0.938**	0.510	0.646	0.503	0.584	0.561	0.715	0.647	0.629	0.482	0.450	0.407
Get2	0.645	**0.934**	0.564	0.644	0.462	0.529	0.605	0.638	0.636	0.573	0.434	0.478	0.354
Clean1	0.395	0.442	**0.867**	0.567	0.297	0.429	0.49	0.447	0.427	0.398	0.275	0.249	0.096
Clean2	0.496	0.560	**0.889**	0.654	0.342	0.574	0.707	0.531	0.534	0.537	0.389	0.390	0.170
Analyze1	0.591	0.646	0.694	**0.960**	0.458	0.580	0.64	0.668	0.711	0.709	0.402	0.487	0.265
Analyze2	0.692	0.677	0.642	**0.958**	0.476	0.669	0.636	0.730	0.728	0.691	0.397	0.509	0.315
Embed1	0.492	0.431	0.293	0.405	**0.881**	0.465	0.470	0.475	0.444	0.415	0.441	0.425	0.424
Embed2	0.453	0.476	0.348	0.451	**0.877**	0.474	0.435	0.566	0.490	0.538	0.453	0.489	0.384
Define1	0.523	0.514	0.572	0.605	0.468	**0.947**	0.631	0.628	0.575	0.624	0.508	0.551	0.276
Define2	0.698	0.614	0.516	0.627	0.543	**0.947**	0.565	0.717	0.650	0.701	0.533	0.548	0.377
Verify1	0.572	0.607	0.668	0.656	0.493	0.568	**0.935**	0.631	0.611	0.675	0.532	0.474	0.351
Verify2	0.623	0.559	0.617	0.589	0.471	0.615	**0.937**	0.594	0.612	0.616	0.53	0.449	0.281
Visualize1	0.678	0.674	0.502	0.64	0.556	0.644	0.574	**0.938**	0.76	0.664	0.486	0.559	0.436
Visualize2	0.749	0.68	0.544	0.724	0.553	0.686	0.652	**0.935**	0.718	0.753	0.471	0.543	0.381
Present1	0.692	0.642	0.514	0.742	0.456	0.584	0.592	0.709	**0.885**	0.743	0.418	0.541	0.353
Present2	0.631	0.567	0.456	0.580	0.483	0.557	0.561	0.684	**0.879**	0.711	0.393	0.543	0.489
Assess1	0.668	0.667	0.475	0.651	0.515	0.691	0.617	0.738	0.815	**0.908**	0.537	0.559	0.470
Assess2	0.547	0.479	0.488	0.662	0.455	0.56	0.623	0.615	0.659	**0.888**	0.432	0.493	0.262
Ethicize1	0.523	0.426	0.313	0.329	0.478	0.516	0.459	0.479	0.413	0.459	**0.913**	0.579	0.582
Ethicize2	0.480	0.472	0.384	0.433	0.457	0.495	0.579	0.46	0.431	0.532	**0.923**	0.445	0.379
Int.perf.1	0.462	0.368	0.300	0.406	0.380	0.500	0.416	0.504	0.521	0.563	0.318	**0.747**	0.456
Int.perf.2	0.517	0.382	0.372	0.509	0.428	0.531	0.475	0.508	0.573	0.535	0.469	**0.776**	0.641
Int.perf.3	0.548	0.472	0.479	0.591	0.463	0.544	0.527	0.557	0.599	0.601	0.473	**0.806**	0.439
Int.perf.4	0.330	0.345	0.310	0.379	0.372	0.417	0.346	0.403	0.345	0.394	0.393	**0.683**	0.354
Int.perf.5	0.185	0.237	0.023	0.110	0.261	0.206	0.135	0.204	0.212	0.101	0.360	**0.599**	0.627
Int.perf.6	0.186	0.268	-0.024	0.099	0.269	0.182	0.098	0.250	0.254	0.191	0.319	**0.603**	0.583
Ext.perf.1	0.343	0.248	0.114	0.179	0.356	0.246	0.258	0.264	0.282	0.306	0.471	0.519	**0.777**
Ext.perf.2	0.436	0.331	0.113	0.312	0.454	0.393	0.252	0.407	0.442	0.380	0.378	0.638	**0.777**
Ext.perf.3	0.308	0.329	0.141	0.193	0.283	0.206	0.255	0.363	0.364	0.268	0.280	0.503	**0.663**
Ext.perf.4	0.434	0.373	0.258	0.354	0.302	0.284	0.379	0.389	0.459	0.383	0.391	0.512	**0.696**
Ext.perf.5	0.272	0.299	0.000	0.063	0.280	0.108	0.134	0.196	0.204	0.164	0.383	0.526	**0.760**
Ext.perf.6	0.326	0.218	0.058	0.219	0.333	0.259	0.218	0.296	0.338	0.308	0.398	0.522	**0.743**

Appendix C: Heterotrait-monotrait Ratio of Correlations (HTMT)

	Find	Get	Clean	Analyze	Embed	Define	Verify	Visualize	Present	Assess	Ethicize	Internal performance	External performance
Find													
Get	0.837												
Clean	0.658	0.735											
Analyze	0.761	0.780	0.867										
Embed	0.696	0.663	0.517	0.607									
Define	0.745	0.683	0.725	0.725	0.677								
Verify	0.750	0.727	0.878	0.752	0.662	0.726							
Visualize	0.894	0.842	0.717	0.824	0.762	0.815	0.763						
Present	0.963	0.874	0.771	0.927	0.750	0.814	0.834	1.007					
Assess	0.843	0.789	0.728	0.877	0.738	0.849	0.855	0.933	1.112				
Ethicize	0.659	0.585	0.499	0.482	0.673	0.649	0.677	0.612	0.603	0.684			
Internal performance	0.641	0.593	0.484	0.582	0.686	0.671	0.572	0.694	0.785	0.734	0.689		
External performance	0.572	0.481	0.203	0.344	0.594	0.395	0.402	0.513	0.614	0.509	0.638	0.897	

8 Resources

Amit, R., & Schoemaker, P. J. (1993). Strategic assets and organizational rent. *Strategic Management Journal, 14*(1), 33–46.

Andrews, R., Boyne, G., Moon, M., & Walker, R. (2010). Assessing Organizational Performance: Exploring Differences Between Internal and External Measures. *International Public Management Journal, 13*(2), 105-129.

Armstrong, J., & Overton, T. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research, 14*(3), 396-402.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management,, 17*(1), 99-120.

Batistic, S., & van der Laken, P. (2019). History, Evolution and Future of Big Data and Analytics: A Bibliometric Analysis of Its Relationship to Performance in Organizations. *British Journal of Management, 30*, 229–251.

Becker, J., Klein, K., & Wetzels, M. (2012). Hierarchical latent variable models in pls-sem: Guidelines for using reflective-formative type models. *Long Range Planning, 45*, 259-394.

Bersin, J., & Zao-Sanders, M. (2020). *Boost Your Team's Data Literacy.* Retrieved from https://hbr.org/2020/02/boost-your-teams-data-literacy

Bhargava, R., & D'Ignazio, C. (2015). Designing tools and activities for data literacy learners. *Wed Science: Data Literacy Workshop.* Oxford, UK.

Bhattacherjee, A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of Management Information Systems, 19*(1), 211-241.

Bratianu, C., Hadad, S., & Bejinaru, R. (2020). Paradigm shift in business education: A competence-based approach. *Sustainability, 12*.

Breje, A., Győrödi, R., Győrödi, C., Zmaranda, D., & Pecherle, G. (2018). Comparative Study of Data Sending Methods for XML and JSON Models. *International Journal of Advanced Computer Science and Applications, 9*(12), 198-204.

Brewer, G., & Selden, S. (2000). Why Elephants Gallop: Assessing and Predicting Organizational Performance in Federal Agencies. *Journal of Public Administration Research and Theory, 10*(4), 685-712.

Bryla, M. (2018). *Data literacy: A critical skill for the 21st century*. Retrieved from Tableau: https://www.tableau.com/about/blog/2018/9/data-literacy-critical-skill-21st-century-94221

Cetindamar, D., Abedin, B., & Shirahada, K. (2021). The role of employees in digital transformation: A preliminary study on how employees' digital literacy impacts use of digital technologies. *IEEE Transactions on Engineering Management*.

Cheah, J., Sarstedt, M., Ringle, C., Ramayah, T., & Ting, H. (2018). Convergent validity assessment of formatively measured constructs in PLS-SEM: On using single-item versus multi-item measures in redundancy analyses. *International Journal of Contemporary Hospitality Management, 30*(11), 3192-3210.

Chu, X., Ilyas, I., & Krishnan, W. J. (2016). Data cleaning: Overview and emerging challenges. *SIGMOD '16: Proceedings of the 2016 International Conference on Management of Data* (pp. 2201-2206). San Francisco, USA: ACM.

Colson, E. (2019). *What AI-driven decision making looks like.* Retrieved from Harvard Business Review: https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like

Coltman, T., Devinney, T., Midgley, D., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, 61*, 1250-1262.

Data to the People. (2018). *Databilities: A Data Literacy Competency Framework*. Retrieved from https://www.datatothepeople.org/databilities

Diamantopoulos, A., & Winklhofer, H. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269-277.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science, 40*, 434-449.

Eby, L. T., & Dobbins, G. H. (1997). Collectivistic orientation in teams: An individual and group level analysis. *Journal of Organizational Behaviour, 18*(3), 275–295

Epstein, P. D. (1992). Measuring the performance of public services. In M. Holzer, *Public productivity handbook* (pp. 161-195). New York: Marcel Dekker.

Erickson, S., & Rothberg, H. (2014). Big data and knowledge management: establishing a conceptual foundation. *Electronic Journal of Knowledge Management, 12*(2).

Fontichiaro, K., & Oehrili, J. (2016). Why data literacy matters. *Knowledge Quest, 44*(5), 20-27.

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

Freiling, J. (2004). A Competence-based Theory of the Firm. *Management Revue, 15*(1), 27–52.

Freiling, J., & Fichtner, H. (2010). Organizational Culture as the Glue between People and Organization: A Competence-based View on Learning and Competence Building. *German Journal of Human Resource Management: Zeitschrift für Personalforschung, 24*(2), 152-172.

Fuller, C., Simmering, M., Atinc, G., Atinc, Y., & Babin, B. (2016). Common methods variance detection in business research. *Journal of Business Research, 69*, 3192–3198.

Gartner. (2019). *Data Literacy*. Retrieved from https://www.gartner.com/en/information-technology/glossary/data-literacy

Gascó-Hernández, M., Martin, E., Reggi, L., Pyo, S., & Luna-Reyes, L. (2018). Promoting the use of open government data: Cases of training and engagement. *Government Information Quarterly, 35*(2), 233-242.

Gefen, D., Straub, D., & Boudreau, M. (2000). Structural equation modeling techniques and regression: Guidelines for research practice. *Communications of AIS, 4*(7), 1-77.

Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *Journal of Strategic Information Systems, 27*, 101-113.

Giauque, D., Anderfuhren-Biget, S., & Varone, F. (2013). HRM Practices, Intrinsic Motivators, and Organizational Performance in the Public Sector. *Public Personnel Management, 42*(2), 123-150.

Giudice da Silva Cezar, B., & Maçada, A. (2021). Data literacy and the cognitive challenges of a data-rich business environment: an analysis of perceived data overload, technostress and their relationship to individual performance. *Aslib Journal of Information Management, 73*(5), 618-638.

Gudergan, S., Ringle, C., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research, 61*(12), 1238-1249.

Gudivada, V., Apon, A., & Ding, J. (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software, 10*(1), 1-20.

Gupta, M., & George, J. (2016). Toward the development of a big data analytics capability. *Information & Management, 53*, 1049-1064.

Hair, J., Risher, J., Sarstedt, M., & Ringle, C. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2-24.

Hair, J., Sarstedt, M., & Ringle, C. (2019). Rethinking some of the rethinking of partial least squares. *European Journal of Marketing, 53*(4), 566-584.

Hamel, G., & Heene, A. (1994). *Competence-based competition.* New York: Wiley.

Hardin, A. (2017). A call for theory to support the use of causal-formative indicators: A commentary on Bollen and Diamantopoulos. *Psychological Methods, 22*(3), 597-604.

Harmsen, H., & Jensen, B. (2004). Identifying the determinants of value creation in the market: A competence-based approach. *Journal of Business Research, 57*(5), 533-547.

Henseler, J., Hubona, G., & Ray, P. (2016). Using PLS path modeling in new technology research: updated guidelines. *Industrial Management & Data Systems, 116*(1), 2-20.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135.

Hippold, S. (2019, 9 11). *CDOs Must Take the Lead to Improve Data Literacy*. Retrieved from Gartner: https://www.gartner.com/smarterwithgartner/cdos-must-take-the-lead-to-improve-data-literacy/

Hoopes, D., Madsen, T. L., & Walker, G. (2003). Guest editors' introduction to the special issue: why is there a resource-based view? Toward a theory of competitive heterogeneity. *Strategic Management Journal, 24*(10), 889-902.

Horng, J.-S., Lio, C.-H., Chou, S.-F., Yu, T.-Y., & Hu, D.-C. (2022). Role of big data capabilities in enhancing competitive advantage and performance in the hospitality sector: Knowledge-based dynamic capabilities view. *Journal of Hospitality and Tourism Management, 51*, 22-38.

Jang, M., Aavakare, M., Nikou, S., & Kim, S. (2021). The impact of literacy on intention to use digital technology for learning: A comparative study of Korea and Finland. *Telecommunications Policy, 45*(7).

Jarvis, C., MacKenzie, S., & Podsakoff, P. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*(2), 199-218.

Jenkins, M., & Griffith, R. (2004). Using personality constructs to predict performance: Narrow or broad bandwidth. *Journal of Business and Psychology, 19*, 255–269.

Johnson, R., Rosen, C., & Chang, C. (2011). To aggregate or not to aggregate: Steps for developing and validating higher-order multidimensional constructs. *Journal of Business and Psychology, 26*, 241-248.

Kim, S. (2005). Individual-level factors and organizational performance in government organizations. *Journal of Public Administration Research and Theory, 15*(2), 245–261.

Ko, J., Hur, S., & Smith-Walker, A. (2013). Family-friendlywork practices and job satisfaction and organizational per-formance: moderating effects of managerial support andperformance-oriented management. *Public Personnel Management, 42*(4), 545-565.
Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science, 3*(3), 383–397.
Koltay, T. (2015). Data literacy: in search of a name and identity. *Journal of Documentation, 71*(2), 401 - 415.
Kristoffersen, E., Mikalef, P., Blomsma, F., & Li, J. (2021). Towards a business analytics capability for the circular economy. *Technological Forecasting and Social Change, 171*.
Lee, B., Riche, N., Isenberg, P., & Carpendale, S. (2015). More than telling a story: A closer look at the process of transforming data into visually shared stories. *IEEE Computer Graphics and Applications, 35*(5), 84–90.
MacKenzie, S., Podsakoff, P., & Podsakoff, N. (2011). Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. *MIS Quarterly, 35*(2), 293–334.
Mandinach, E., & Gummer, E. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30-37.
Morrow, J. (2018). *The Data Literacy Project*. Retrieved from Qlik: https://www.qlik.com/blog/the-data-literacy-project
Murawski, M., & Bick, M. (2017). Digital competences of the workforce – a research topic? *Business Process Management Journal, 23*(3), 721-734.
OECD. (2017). *Core skills for public sector innovation.* Paris: OECD Publishing.
Oxford Reference. (2021, 09 08). *Reflexivity*. Retrieved from Oxford Reference: https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100410506
Pagador, J., Huynh, N., Davis, D., & Abhari, K. (2020). Data swagger: Cultivating data literacy to drive digital transformation. *41st International Conference on Information Systems, ICIS2020.*
Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society, 21*(2), 419-437.
Pedersen, A. Y., & Caviglia, F. (2019). Data literacy as a compound competence. In T. Antipova, & A. Rocha, *Digital Science* (Vol. 850, pp. 166–173). Springer International Publishing.
Pentland, A. (2013). The data-driven society. *Scientific American, 309*(4), 78-83.
Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly, 31*(4), 623-656.
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*(1), 539–569.
Pothier, W., & Condon, P. (2019). Towards data literacy competencies: business students, workforce needs, and the role of the librarian. *Business and Finance Librarianship, 24*(3), 1-24.
Potnuru, R. K. G., & Sahoo, C. K. (2016). HRD interventions, employee competencies and organizational effectiveness: An empirical study. *European Journal of Training and Development, 40*(5), 345–365.
Press, G. (2016, 3 23). *Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says*. Retrieved from Forbes: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/
Rahm, E., & Hai Do, H. (2000). Data cleaning: Problems and current approaches. *Bulletin of the Technical Committee on Data Engineering, 23*(4), 3-13.
Rainey, H. G., & Steinbauer, P. (1999). Galloping elephants: Developing elements of a theory of effective government organizations. *Journal of Public Administration Research and Theory, 9*(1), 1–32.
Rainey, H., Backoff, R., & Levine, C. (1976). Comparing public and private organizations. *Public Administration Review, 36*(2), 233–244.
Rayna, T., & Striukova, L. (2021). Fostering skills for the 21st century: The role of Fab labs and makerspacesmaker spaces. *Technological Forecasting and Social Change, 164*.
Rialti, R., Zollo, L., Ferrari, A., & Alond, I. (2019). Big data analytics capabilities and performance: Evidence from a moderated multi-mediation model. *Technological Forecasting and Social Change, 149*.
Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., . . . Matwin, S. W. (2015). *Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report.* Halifax: Dalhouse University.
Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. Boenningstedt: SmartPLS GmbH. Retrieved from http://www.smartpls.com
Rumelt, R. (1994). Foreword. In G. Hamel, & A. Heene, *Competence-based competition.* New York: Wiley.
Salman, M., Ganie, S. A., & Saleem, I. (2020). Employee Competencies as Predictors of Organizational Performance: A Study of Public and Private Sector Banks. *Management and Labour Studies, 45*(4), 416–432.

Sanchez, R. (2004). Understanding competence-based management: Identifying and managing five modes of competence. *Journal of Business Research, 57*, 518–532.
Sanchez, R., & Heene, A. (1997). Reinventing strategic management: New theory and practice for competence-based competition. *European Management Journal, 15*(3), 303-317.
Sanchez, R., Heene, A., & Thomas, H. (1996). *Dynamics of competence-based competition: Theory and practice in the new strategic management.* Oxford (UK): Elsevier Science.
Santoro, G., Fiano, F., Bertoldi, B., & Ciampi, F. (2019). Big data for business management in the retail industry. *Management Decision, 57*(8), 1980-1992.
Sarstedt, M., & Wilczynski, P. (2009). More for less? A comparison of single-item and multi-item measures. *Business Administration Review, 69*(2), 211-227.
Sarstedt, M., Hair, J., Jun-Hwa, C., Becker, J., & Ringle, C. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal, 27*(3), 197-211.
Shmueli, G., Ray, S., Velasquez Estrada, J., & Shatla, S. (2016). The elephant in the room: Evaluating the predictive performance of PLS models. *Journal of Business Research, 69*(10), 4552-4564.
Shmueli, G., Sarstedt, M. H., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *European Journal of Marketing, 53*(11), 2322-2347.
Sternkopf, H., & Mueller, R. (2018). Doing good with data: Development of a maturity model for data literacy in non-governmental organizations. *Proceedings of the 51st Hawaii International Conference on System Sciences*, (pp. 5045–5054). Waikoloa Village, Hawaii.
Stordy, P. (2015). Taxonomy of literacies. *Journal of Documentation, 71*(3), 456-476.
Tabesh, P., Mousavidin, E., & Hasani, S. (2019). Implementing big data strategies: A managerial perspective. *Business Horizons, 62*(3), 347-358.
UNESCO. (2021). *Literacy*. Retrieved from https://en.unesco.org/themes/literacy
Van Der Heijde, C., & Van Der Heijden, B. (2006). A competence-based and multidimensional operationalization and measurement of employability. *Human Resource Management Journal, 45*(3), 449-476.
Wanous, J., Reichers, A., & Hudy, M. (1997). Overall job satisfaction: how good are single-item measures? *The Journal of Applied Psychology, 82*(2), 247-252.
Wolff, A., Gooch, D., Cavero Montaner, J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3), 9–26.
Yousef, A., Walker, J., & Leon-Urrutia, M. (2021). Defining data literacy communities by their objectives: A text mining analysis. *WebSci '21: 13th ACM Web Science Conference 2021* (pp. 26–33). Virtual event: United Kingdom.
Yu, W., Zhao, G., Liu, Q., & Song, Y. (2021). Role of big data analytics capability in developing integrated hospital supply chains and operational flexibility: An organizational information processing theory perspective. *Technological Forecasting and Social Change, 163*.
Zhang, Y., Porter, A., Cunningham, S., Chiavetta, D., & Newman, N. (2021). Parallel or intersecting lines? Intelligent bibliometrics for investigating the involvement of data science in policy analysis. *IEEE Transactions on Engineering Management, 68*(5), 1259 - 1271.
Zuiderwijk, A., Chen, Y., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research ag