Data Literacy: Detecting Data Literacy Gaps within Businesses

Author: Lisa van Disseldorp
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

Nowadays, businesses face new challenges every day while still trying to be more innovative, competitive, and profitable to ensure their future. The rise of data has, on the one hand, complicated this but also enabled businesses to operate differently; data has enabled businesses to make data-driven decisions. To be able to reap the benefits of data, the workforce of any business needs to possess different skills. This is where data literacy plays a significant role. Data literacy is the extent to which an employee can work with data and use it efficiently. By identifying the various levels of data literacy as well as management tools that support businesses to increase the general level of data literacy, this paper shows businesses what data literacy levels exist and how they are put in context. Furthermore, this paper shows managers and project leaders how to assess employees' level of data literacy by offering multiple tools provided by scholars as well as businesses. To complement the theory, this paper examined how businesses monitor the level of data literacy among the workforce, how businesses determine a strategy to detect data literacy gaps, and what actions businesses take to overcome data literacy gaps found among the workforce. The findings are based on six interviews with businesses from the automotive industry, consultancy, tech-business, and the energy sector. The results show that businesses do not have specific strategies in place to detect any data literacy gaps, but rather rely on regular performance reviews. Additionally, the results indicated that businesses favour personal coaching and offering trainings to overcome obstacles faced regarding data literacy.

Graduation Committee members:
First supervisor: Dr. M. de Visser
Second supervisor: Dr. A.B.J.M. Wijnhoven
Third supervisor: L.J.J. Kayser

Keywords
Data, data literacy, levels of data literacy, data literacy management tools, data literacy gaps, data-based decision-making

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

CC-BY-NC

1. INTRODUCTION

1.1 Problem Statement

Due to the forever changing environment businesses face, businesses must continue to be innovative, competitive, and profitable at all times to ensure their future. This does not limit to the products and services offered to their customers, but also includes the management practices and tools which must be properly embedded within businesses as well.

Since the late 20th century, information stream sources have increased – from cities, intelligent home equipment to cars, information has been, and can be, derived. The amount of available information, or *data*, has greatly influenced the way businesses deem to stay innovative, competitive and profitable. The rise of data has led to many businesses collecting, storing, and analysing it and use these data to make data-driven decisions. By doing so, businesses have stepped away from original operations in which businesses used to produce products or offer services they ought to be relevant for their target group, to a so-called process of data science thinking in which the users of a product or service and their needs are the starting point of developing a specific product or service (Pohlmann, 2016).

To successfully embed data in a business, managers and project leaders should go beyond implementing data facilities such as collecting, storing and analysing data and should see data analysts or data departments as part of the whole rather than one department who is responsible for data-driven decisions. This means that managers and project leaders must raise awareness about data within the entire business, they must create a data-driven mindset among their workforce.

However, not all businesses find it easy to overcome this data transformation. Data is a complex, but relevant concept and consists of various aspects. This makes it hard for managers and project leaders to implement the concept of data literacy, which is the ability of someone reading, working with, analysing, and arguing with data (Barth, 2019), and to overcome the corresponding obstacles faced with data literacy. Up until now, it is not clear how managers and project leaders can detect possible data literacy gaps and how they can solve these gaps.

1.2 Defining Data Literacy

Data literacy can be defined by explaining the two words separately. Data are information streams, either structured or unstructured, which can be used to form ideas or concepts. Literacy, on the other hand, is somewhat more complex; researchers from UNESCO define literacy as "the ability to identify, understand, interpret, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society."

Combining these two definitions leads to the following interpretation of data literacy: "a continuous 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". This term is relevant for businesses because data have influenced the way businesses operate.

1.3 Research Project Motivation

The reason for writing a paper about this topic is because businesses require more guidance when it comes to data. Although data are believed to be valuable assets for businesses, many managers and project leaders do not know how to use it effectively and how to shape their workforce accordingly. In an ever-changing environment, they must critically examine what their employees must be capable of and how they can be trained to perform their job better. As data influences the way businesses operate, employees must be aware of this change and must also be educated on this topic. For me, this is an important aspect because I believe jobs should remain interesting for employees and, although most employees obtained valid degrees, it is the responsibility of a manager and project leader to continuously motivate employees to keep on learning and to make sure employees remain an important asset to the business.

1.4 Research Objectives

Within this research, the importance of data within businesses will be emphasised and ways managers and project leaders can detect data literacy gaps among employees will be identified. The goal of this research is to discover possible strategies or common practices that enable managers and project leaders to detect data literacy gaps early on and to know how to overcome such gaps.

To do so, this paper will answer the following research question: "*How can managers and project leaders detect a data literacy gap among their workforce?"*

1.5 Outline

This paper consists of various parts. The first part covers theoretical knowledge to understand the purpose and relevance of this paper. The following sub-questions will be answered in this part:

- *What is data?*
- *What are the differences between data and information?*
- *Why is data relevant for businesses?*
- *What is data literacy and why is it relevant for businesses?*
- *What are differences between informational-, statistical-, and data literacy, and how do they relate with each other?*

The second part of this paper covers the methodology used to execute the research. The chosen methodology is to identify data literacy levels and management tools related to data literacy in literature, and complementing that, interviews will be conducted to get insights into how managers can detect data literacy gaps. The sub-questions related to this, are:

- *What levels of data literacy exist?*
- *How can businesses determine an individual's level of data literacy?*
- *What are management tools that support data literacy among the workforce?*
- *How can managers determine a gap in data literacy?*

These sub-questions will be covered in the third part, where the data collection and data analysis will be performed to answer the research question.

2. RELATED LITERATURE

2.1 Data

The goal of this paper is to identify ways that will help managers and project leaders to discover data literacy gaps among their workforce as well as management tools that will help managers and project leaders to overcome these gaps. As data is the main principle of data literacy, this concept needs to be clarified. Furthermore, this chapter will compare data to information and will determine the effects of data in today's business world.

2.1.1 Definition of (Big) Data

Businesses have always tried to manage their assets – such as property, plants, cash, and inventories (Opher et al., 2016). However, due to digitalisation, businesses must be able to manage another, new type of asset – data. A very plain explanation of data, given by the Merriam-Webster Dictionary, is that data is "factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation", but also addresses that "information output from a sensing device or organ that includes useful and irrelevant or redundant information must be processed to be meaningful". For this paper, it must be understood that data are the outputs of measurements and are truly raw at first (Tien, 2013).

Over the past years, the amount of data has increasingly grown. Not only do smartphones collect data about people's location, the weather, and how mobile phone users touch their screen, but the products people buy are designed in such a way that they can collect data based on usage of the product (Marr, 2016). As data can be found anywhere, the introduction of Big Data has greatly influenced the way businesses operate.

The term Big Data first appeared in the mid-1990s, when John Mashey referred to "handling and analysis of massive datasets" (as cited in Kitchin and McArdle, 2016, p. 1). A study by Doug Laney (2001) focusing on e-commerce channels highlights three characteristics that make for solid data management solutions and led to the so-called concept of V's; volume, velocity, and variety. Volume refers to the amount of data and lead to businesses receiving "ever-growing data of all types, easily amassing terabytes (even petabytes) of information" (as cited in Ebbers et al., 2013, p. 4). Ebbers et al. (2013) refer to velocity as the time to respond to data; not all data can be used forever, which means businesses should be aware of the perishability of data. The last characteristic of Big Data, variety, refers to the type of data. A very generic distinction between data is whether they are structured or unstructured, in which unstructured data cannot easily be managed or analysed with traditional technologies.

Apart from these basic characteristics, scholars have identified many more traits that describe Big Data. Characteristics such as exhaustivity, fine-grained and uniquely indexical, relationality, extensionality and scalability, veracity, value, and variability (Kitchin and McArdle, 2016) have been acknowledged to describe Big Data.

However, in the same study of Kitchin and McArdle (2016), it becomes evident that there is no one definition for Big Data. Big Data comes in all types of forms and can have different traits. In their research, they point out that for data to be Big Data, it must contain most of the seven traits mentioned before, with velocity and exhaustivity being the most important characteristics.

As there is not one single definition of (Big) Data, the following subsection will compare data to information. By doing so, the following subsection aims to give a better insight into what data is.

2.1.2 Data versus Information

Data and information are closely connected and are often used interchangeably. The concepts, however, are not the same. Whereas data are simply raw facts such as an employee number, sales orders or hours worked per week, information is a "collection of facts organised so that they have additional value beyond the value of the individual facts" (as cited in Stair and Reynolds, 2009, p. 5). Therefore, the way data are interpreted, organised, and presented leads to the existence of various types of information.

Once data has been collected, stored, and analysed, it must first be processed to become (relevant) information. It is, however, required to use knowledge when transforming data into information; after the data has been obtained, knowledge must be used to be aware of and get an understanding of a set of information and subsequently must be used to identify ways information "can be made useful to support a specific task or reach a decision" (as cited in Stair and Reynolds, 2009, p. 6). This process is called the transformation process, as shown in Figure 1.

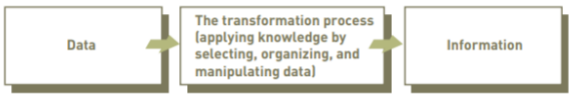

Figure 1. The Process of Transforming Data into Information (Stair and Reynolds, 2009)

2.1.3 Relevance of Data

In a business world that rapidly changes, businesses must remain innovative, competitive, and profitable to stay ahead of their competitors. Data are one of the enablers for businesses to do so, and therefore the relevance of data will be highlighted in this section.

As scholars and business leaders have acknowledged the advantages of data, they are of no use unless they are processed. Tools such as "analytics, machine learning, and AI running on advanced hardware platforms are empowering companies to look at data as a strategic enabler" (as cited in Greer, 2019, p. 3). Data on their own, therefore, only become relevant once they have been processed. Business practices such as business analytics, artificial intelligence, and machine learning, but not limited to, have enabled businesses to use their data effectively.

According to Belissent et al. (2020), 90% of those in charge of making (data-based)decisions recommend data-insights highly when business decisions have to be made. In their paper, Belissent et al. (2020) highlight the importance of an insights-driven mentality. This argument is backed up by Marr (2016), whose research shows that, with the help of data, businesses can predict trends, track their consumers' behaviour, and improve performances. The data gathered from within the organisation, help businesses go through their decision-making process. Instead of relying "on experience, gut feeling, or opinions when making decisions" (as cited in Belissent et al., 2020, p. 2), businesses can now shift toward a data-driven decision-making culture.

This new way of decision-making is still relatively new to most businesses, but those who practise it, experience that their organisational performances improve. Due to the processed data, businesses can "identify new opportunities, highlighting potential threats and revealing new business insights" (as cited in Ram, 2016, p. 221).

As a result of these new insights and new ways of decision-making, there has been a shift from the original way of deriving to new products and processes; as businesses can use data during their design process of products and services, they can put the customer first (Pohlmann, 2016). The two main questions asked during this process, which is called design thinking, are: "for whom are we designing and what is the problem they are experiencing?" and "to what end are we modelling the design – to boost consumption and engagement, improve performance, or to achieve scale?" (as cited in Pohlmann, 2016, p. 1). In his paper, Pohlmann (2016) highlights that this process does not only apply to the actual designers of products or services, but these two questions can be of added value to anyone who uses analytics.

As an example, Google will be used (Utica College, 2019) to demonstrate how data-based decision-making works. Google is known for its data-based decision-making, but especially one case stands out. Google's Human Relations department bases their decisions on data, hence they have a so-called People Analytics Department. At one point in time, Google questioned the added value of managers and rose the question "Do managers actually matter?" (Marr, 2012). First of all, Google collected existing data of performance reviews (top-down review of managers) and employee surveys (bottom-up review of managers). After analysing the data, Google first found out that there were not many differences within their analysed data. However, once they analysed the data separately, Google was able to tell which managers did better based on team productivity, employee happiness, and employee turnover. Yet this did not lead to any decisions made by Google; they first wanted to determine what makes a good manager. By collecting more data through the "Great Managers Award" (people could nominate good managers by identifying favourable behaviours from the manager's side) and through interviewing managers to get insights into their operations, Google was able to pinpoint eight favourable behaviours that make for a good manager. These new insights made Google decide to communicate the findings with managers internally but also led to the introduction of a twice-yearly feedback survey that measures whether people fit the prescribed behaviours, to remain the Great Managers Award as well as revising the management training Google used to host.

To summarise, data is essential for businesses to remain competitive, innovative, and profitable. Due to data, businesses can get insights into their operations and can make justified decisions when using data rather than experiences, gut feelings, or opinions. When using data to serve the end-user of a product or service, businesses will be able to outperform businesses that have not incorporated data yet as well as remain competitive with the already-data-driven businesses. The latter relates most to this paper, as businesses who struggle with using data to their benefit should focus on data literacy within their business. For data to be of use, it requires people who understand how to use data, and therefore businesses must have data-literate people within their workforce.

2.2 Data Literacy

As digitalisation has changed the operations of businesses, an internal change is required to be able to adapt to the new way of doing business. As data is becoming more important over time, many employees will experience that "analysing large, messy, unstructured data is going to increasingly form part of everyone's work" (as cited in Harris, 2012, p. 1).

However, as research by Belissent et al. (2020) shows, the majority of organisations struggle to use data effectively when making decisions. To overcome this struggle, not only business leaders and business analysts, but also employees who do not (yet) work with data should improve this so-called data literacy skill. As Wolff et al. (2016) highlight in their paper, data literacy "is considered to be a life skill, as daily interactions with data become ever more commonplace" (as cited in Wolff et al., 2016, p. 2).

The following subsections will clarify what data literacy entails, and how it distinguishes from information- and statistical literacy.

2.2.1 Definition & Relevance of Data Literacy

Data literacy is a combination of two words; data and literacy. As described in section 2.1.1, data are streams of information, structured or unstructured, that need to be processed to become informative value. Data are important for businesses as they enable the workforce to make decisions based on real-time information rather than opinions, experiences, or gut feelings.

Literacy, on the other hand, can simply be defined as "the ability to read and write" (as cited in Zamel and Spack, 2012, p. 55). Despite this generic and basic definition, complexities around literacy do exist. In 1966, the United Nations Educational, Scientific and Cultural Organisation (UNESCO) announced that literacy was a fundamental human right (Zamel and Spack, 2012). It was only in 2004 that UNESCO revised the definition of literacy more broadly: "literacy is the ability to identify, understand, interpret, create, communicate, and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society" (as cited in Shelley et al., 2008, p. 66). Complexities around literacy arise, due to it being a fundamental human right as well as the learning aspect of it; how does one ensure all individuals have access to literacy and who is responsible for educating these individuals? Moreover, the aspect of "learning" implies that one individual might be more successful than others and this may lead to major intellectual gaps between people.

Although there are numerous definitions of data literacy, the core of the definitions largely resembles the definition of literacy. Various definitions of data literacy have been gathered to show what the most important aspects of data literacy are and how it is similar to literacy as described earlier. These definitions can be found in Table 1.

Scholar	*Definition*
Forrester/Belissent et al. (2020, p. 4)	The ability to recognise, evaluate, work with, communicate, and apply data in

	the context of business priorities and outcomes
Gartner/Panetta (2019, p. 4)	The ability to read, write and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied – the ability to describe the use case, application and resulting value
Grillenberg (2018, p. 1)	Competencies regarding gathering, storing, processing and visualising data
Sternkopf (2018, p. 1)	The ability to use data productively and to think about it in a critically reflective way
Wolff et al. (2016, p. 23)	Ask and answer real-world questions through an inquiry process (ethical use of data); practical and creative skills: select, clean, analyse, visualise, critique, interpret, and communicate
Bhargava (2015, p. 1)	The ability to read, work with, analyse, and argue with data
Harris (2012, p. 2)	The competence in finding, manipulating, managing and interpreting data, and it does not only include numbers but texts and images as well
Schield (2004, p. 7)	How to obtain and manipulate data

Table 1 Definitions of Data Literacy

The definitions of data literacy given above, show that data literacy is a process; various tasks are required to be data literate. In their research, Wolff et al. (2016) performed a qualitative analysis of various definitions of data literacy as well as statistical literacy to find commonalities between somehow unique definitions. Based on their findings, they propose the following definition of data literacy: "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 goal. These include the abilities to select, clean, analyse, visualise, critique, and interpret data, as well as to communicate stories from data and to use data as part of a design process" (as cited in Wolff et al., 2016, p. 23). Their definition includes the abilities someone should possess to be data literate, but also includes aspects such as ethical behaviour, being able to present the findings of data to others as well as how to use this in a design process. For that reason, this definition is leading in this paper as it applies to businesses as well.

Research by Censuswide (Morrow, 2018) shows that only 24% of the business decision-makers are fully confident when using data. These data highlight the importance of data literacy since the process of transforming data into business value is crucial for businesses (Greer, 2019). According to Greer (2019), data literacy is a necessity to be able to successfully go through this process. More scholars believe data literacy is crucial for businesses: first of all, data literacy is important to "solve problems and gain knowledge" (as cited in Hippold, 2019, p. 1). Without the skills to use data effectively, businesses cannot rely on their data when trying to solve issues or getting insight into their business. Additionally, Lee (2020) points out that data literacy enables businesses to ensure success as well as scalability for their business. Last but not least, in research performed by Gartner, 80% of organisations in 2020 will try to overcome the data literacy gap by creating development programs and 50% of the organisations in 2020 will not have access to proper artificial intelligence and data literacy skills. These statistics show that many organisations do not see data literacy as an enabler of (digital) business and do not treat their information as their way to speak. As a result, many businesses will not be able to thrive on their data to gain business value.

2.2.2 Comparing Information-, Statistical-, and Data Literacy

Data literacy is closely connected to information- and statistical literacy. Although they are not the same, the three types of literacy are interrelated. In his research, Schield (2004) defines the literacies and gives insights into how they relate to each other.

Since 1989, the need for critical evaluation of information rose. In 1998, the American Library Association and the Association of College and Research Libraries defined information literacy as the ability to "recognise when information is needed and have the ability to locate, evaluate, and use effectively the needed information" (as cited in Schield, 2004, p. 6).

With information comes statistics; without statistics, it is hard to put (social) issues into perspective. Therefore, statistical literacy covers the "use of statistics as evidence in arguments" (as cited in Schield, 2004, p. 6). Two main elements of statistical literacy are that 1) statistical literacy focuses on assembly ("how the statistics are defined, selected and presented" (as cited in Schield, 2004, p. 7)) and 2) the aspect of context and confounding is important within statistical literacy.

Now that information- and statistical literacy have been clarified, the relationship between the three literacies can be established. According to the research of Schield (2004), information literacy needs statistical literacy as well as data literacy. First of all, people must be information literate to be able to "think critically about concepts, claims and arguments" (as cited in Schield, 2004, p. 8). This means one should possess abilities such as reading, interpreting, and evaluating information. Next to that, statistical literacy is an important aspect of information literacy; statistical literacy requires someone to "be able to think critically about basic descriptive statistics" and includes "analysing, interpreting, and evaluating statistics as evidence in a special skill" (as cited in Schield, 2004, p. 8). Last but not least, Schield explains that to be data literate, one should be able to "access, assess, manipulate, summarise, and present data" (as cited in Schield, 2004, p. 8). Schield (2004) concludes by pointing out that information- and statistical literacy both need data literacy as this is an essential component.

The reason why information- and statistical literacy is important for this paper, is because data literacy alone may not be enough for a business to focus on; to be fully confident using data in any decision-making process, employees should first know how to handle information properly and then they should know how to abstract the right information to back up their arguments using

statistics. Possessing these two abilities may lead to an increase in data literacy among the workforce.

3. RESEARCH METHODOLOGY

Up until this far, this paper used literature gathered from sources such as Google Scholar and Web of Science to explain the following concepts: data, data literacy, and the comparison between information-, statistical-, and data literacy. These concepts are important for businesses to understand how to turn data into business value.

However, little to no concepts can be found that explain how to detect a data literacy gap within a business. As a consequence of data becoming ever more important to businesses, a slight shift of who's responsibility it is to educate people, can be seen. Whereas data literacy is being taught at schools, it is as much of a responsibility of a business to strive for a data literate workforce.

To better understand how businesses can detect a potential data literacy gap, this paper will first focus on what levels of data literacy exist and how the level of data literacy of an individual can be determined. This is important because one cannot detect a gap without any classifications. This also means that businesses should know what they expect from their employees data-literacy wise beforehand and have proper tools in place to evaluate an employee's level of data literacy.

Secondly, this paper will look at various management tools related to data literacy and will analyse whether the tools can be implemented by businesses. Management tools help managers achieve their goals, and therefore can add value when a business aims to improve the overall data literacy level of the workforce.

Thirdly, interviews will be conducted with managers and project leaders from various businesses. In this part of the research, new insights will be generated about the current practices of managers and project leaders regarding how they detect a data literacy gap among their employees.

Interviews aim at collecting data through asking qualitative and/or quantitative questions. Quantitative questions are closed while qualitative questions are open-ended (Doody and Noonan, 2013). Interviews are one of the most common practices to gather data, specifically interviews including qualitative questions. As with any other collecting method, interviews have their advantages as well as disadvantages. These have been displayed in Table 2.

Advantages	**Disadvantages**
- They are useful to gain insight and context; - They help participants describe what is important to them; - They are useful in generating quotes and stories; -They enable the researcher to develop a report; - They give the researcher the opportunity to observe as well as listen; - They enable more complex questions to be asked; - The researcher can explain the purpose of the research and answer any questions the participant may have about the study; - The researcher can probe the participant's responses and seek further clarification; - Participants can seek clarification of a question; - They help the participant to give detailed responses; - Can explore participants' reason for acting in a certain way or their interpretations of events; - They are more appropriate for certain groups, such as those with reading or writing difficulties; - Interviews can be a rewarding for participants as they stimulate self-exploration and discovery; - Personal benefit: the telling of one's story.	- They may seem intrusive to the participant; - They are time-consuming, not only in terms of conducting them but also in relation to arranging them, travelling to the venue, post-interview transcriptions and analysis of the data; - They can be expensive compared with other methods; - Interviews on a personal and/or intimate subject can evoke strong feelings and these feelings need to be handled with great sensitivity; - They are susceptible to bias, which may include: 1. The participant's desire to please the researcher 2. Saying what they think/feel the researcher wishes to hear, such as giving an official point of view rather than their personal view 3. The desire to create a good impression may lead to participants not answering honestly 4. There is a tendency to say something rather than nothing if the participant cannot answer a question or has nothing to say on a topic. 5. The researcher's views can influence the participant's responses by expressing surprise of disapproval.

Table 2 Advantages and Disadvantages of Interviews (Doody and Noonan, 2013)

Furthermore, interviews can be structured, unstructured, or semi-structured. According to Doody and Noonan (2013), structured interviews ask the same, identical question to each participant; unstructured interviews, on the other hand, start with a general, open question that is asked to every participant and then the researcher asks questions based on the participant's answer; and lastly, semi-structured interviews are a mix of the described interviews above: semi-structured interviews make use of predetermined questions but ask follow-up questions based on the answers of the participants.

For this research, the reason for collecting data through an interview is because interviews enable researchers to "further explore responses or findings" (as cited in Doody and Noonan, 2013). Although interviewers are obliged to think critically beforehand about the plan and format of the interview, the direct contact makes for the perfect situation to ask follow-up questions. For this paper, interviews are favoured to discover the experiences of the managers and project leaders, and therefore a semi-structured interview will be conducted.

Before the interview, participating managers or project leaders will get information about the nature of the research. The text and consent form they received can be found in Appendix A and Appendix B.

Before the interview, the participants will be asked to orally agree upon several conditions. These include that they participate voluntarily, the outcomes of the interview will be used anonymously, and the participant has the right to stop at any moment without giving reason. Aside from that, the participants will be

asked whether they would like to receive the outcomes of the research.

The interviews will be conducted online. Depending on the participants, the interviews will be conducted via Skype (if not possible, another app will be proposed). Due to the scope of this research and time limits, six interviews will be conducted. Of these six interviews, five different businesses have been interviewed of which one business operates in the automotive industry, two businesses in consultancy, one business is a tech-business, and one business operates in the energy sector. In Table 3, an overview of each participants' function has been given.

Participant	Role
P1	Advanced Intelligence Data Analyst
P2	Manager/Project Leader
P3	Manager
P4	Chief Data Officer
P5	Founder & Chief Executive Officer
P6	Corporate Data Strategist

Table 3 Overview of participants' functions

All interviews will start with the same questions:

- If you think about data literacy, what are the first things you think of?
- How do you define what data literacy level an employee must have to perform a certain job?
- How do you monitor their performance regarding data literacy?
- Do you have a strategy, or process, in place to detect a data literacy gap?
- When you detect a data literacy gap, what do you do?

As this is a semi-structured interview, follow-up questions may be asked depending on the answers given by the participants. The interviews will take no longer than thirty minutes per interview.

4. DATA COLLECTION

This chapter is omitted intentionally

5. DATA ANALYSIS

This chapter is omitted intentionally

6. CONCLUSION

The objective of this paper is to reveal how managers and project leaders can detect data literacy gaps among their workforce and if detected, how they can solve these data literacy gaps. To do so, this paper answers the following research question: "*How can managers and project leaders detect a data literacy gap among their workforce?"*. To answer this research question, a literature review and semi-structured interviews have been conducted.

Based on a literature review, this paper discussed the importance of data to businesses and how data contribute to the performance of businesses today. With tools such as "analytics, machine learning, and AI running on advanced hardware platforms" (as cited in Greer, 2019, p. 3), businesses now have the option to withstand their competition and have been able to "identify new opportunities, highlighting potential threats and revealing new business insights" (as cited in Ram, 2016, p. 221) with the use of processed data. However, using data as a benefit requires the entire workforce to be data literate. Data literacy, in this paper, is defined as "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,

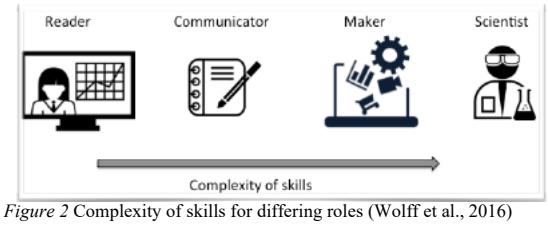

Figure 2 Complexity of skills for differing roles (Wolff et al., 2016)

with the ability to extend knowledge of specialist data handling skills according to goal. These include the abilities to select, clean, analyse, visualise, critique, and interpret data, as well as to communicate stories from data and to use data as part of a design process" (as cited in Wolff et al., 2016, p. 23). To ensure a good understanding of what data literacy is, this paper discussed the differences between information-, statistical- and data literacy and highlighted their interdependence.

To answer the research question, an explorative research, including semi-structured interviews, has been conducted to confirm the literature review done in this paper as well as complement it. Six interviews have been conducted, from which two participants operate in the automotive industry, two participants in consultancy, one participant operates in a tech-business and one participant works in the energy sector. The interviews have given a general understanding of how data literacy is perceived by the participants and how their respective businesses try to foster data literacy. The analysis of the interviews show that the businesses are aware of the definition of data literacy and the importance of statistics that comes with it, as highlighted in the literature review of this paper. However, the participants were unaware of the identified data literacy levels in literature and did not specify the roles they use within their business. Instead of defining various data literacy levels and assigning such levels to specific jobs beforehand, the analysis revealed that businesses use the hiring process to identify what skills someone at least needs for a specific job. Next, the analysis showed that only a few businesses monitored the employees' performance regarding data literacy. Instead of having specific outcomes, or KPIs, in place, most businesses use their regular performance reviews to monitor someone's skills. One exception in this was P3, who explained they have a program in place that is accredited by the Hult University and is aligned with the desired outcomes of the Human Relations department. Regarding the strategies to detect data literacy gaps, no business had a strategy in place to specifically detect data literacy gaps. However, many of the businesses emphasised the use of trainings which give them a good indication of someone's skills, as well as the regular performance reviews where these skills are assessed. Based on the interviews, it became apparent that many businesses use coaching as a tool to overcome data literacy gaps. This coaching implied pairing up employees so they can learn from each other as well as offering internal and/or external trainings to the employee(s) in need.

To answer the research question *"How can managers and project leaders detect a data literacy gap among their workforce?"*, it can be said that the proposed data literacy levels including their assessment tools and the data literacy programs, have not been adopted by the businesses yet. For businesses to outperform competitors, more emphasis should be placed on the importance of data literacy among the entire workforce of any company. However, this paper did show that businesses do value programs/trainings and are looking for ways to improve the overall data literacy level of the workforce. In addition to that, something that has not been covered in literature extensively yet, is the benefits of working together in teams; in this way, employees learn from each other and are exposed to new ways of retrieving, handling, visualising, interpreting and communicating with data.

To conclude, managers and project leaders should develop internal data literacy levels using the context of their business environment, should have assessment tools in place that can monitor employees' level of data literacy, and should stimulate project work to detect data literacy gaps.

7. LIMITATIONS & FUTURE RESEARCH

During this research, some limitations occurred. First of all, only six interviews could be conducted. This is a very small sample size, and so can only be interpreted as general views of businesses regarding data literacy. However, due to time constraints and the scope of this research, it was not feasible to approach more participants. For future research, it might be suggested to interview more participants. In this way, the collected data can be validated.

Another limitation in this research was the type of people interviewed; many of the participants closely worked with data on a daily bases because they are data analysts or own a tech-businesses. As data literacy should be implemented throughout the entire organisation for businesses to reap the benefits from it, it would have been interesting to interview employees from e.g. the Human Relations department, the administration office, or the assembly department. Therefore, for future research, it is suggested to widen the scope of the research and include employees from several departments with various backgrounds. Doing this enables the researcher to get a grasp of how data literacy truly is perceived throughout the entire business.

The literature found while doing the literature review consisted mainly of scholars relating data literacy to education and emphasising data literacy as to be taught in elementary school, high school, or universities. For future research, it would be interesting to see how the findings in these researches can complement the current uses and practices of businesses to increase the level of data literacy. In line with that, in particular, one participant of this research, emphasised the need for a change in education when it comes to data (literacy). According to this participant, the educational system has not adapted to the digitised world we live in today. This participant advocates for earlier involvement of students in data and/or statistics to prepare them better for the jobs of the future. To research this, it would be interesting to set up panel meetings with both the management of educational programs as well as businesses. During such panels, both parties can discuss how they handle data literacy at the moment, what are required skills for students to be able to perform future jobs, and align this with the interpretation and expectations of businesses. By doing so, this future research could pinpoint how educational programs can embed data literacy and can better prepare students for future jobs.

8. CONTRIBUTION

This research complements existing knowledge about data literacy, data literacy levels, and data literacy tools, and on top of that tries to uncover strategies to detect data literacy gaps and practices to solve these gaps.

This paper has developed a framework to understand data literacy and the relevance of data literacy for businesses. Furthermore, this paper gives businesses a good overview of what data literacy levels currently exist and for what context they have been designed, and what management tool may help businesses to improve the data literacy levels of employees.

Finally, this paper has tried to uncover how businesses detect data literacy gaps. In doing so, this paper has advanced the levels of data literacy by emphasising the need for good alignment of data literacy levels.

9. ACKNOWLEDGEMENTS

This paper could not have been realised without the support of various people. First of all, I would like to thank all the participants for taking part in my interviews. Thank you for making time for my bachelor thesis and for the valuable insights and experiences you have given me.

Secondly, I would like to thank my first supervisor, Matthias de Visser, for being enthusiastic along the entire way and for constructing valuable feedback. I would like to thank my second supervisor, Fons Wijnhoven, for reading my thesis at the end of this process and still being able to provide me with useful feedback. Last but not least, I would like to thank my third supervisor, Liza Kayser, for helping me find participants for my interviews and giving me so much hope throughout these, somewhat turbulent, times. I am grateful for all my supervisors' input.

Lastly, I would like to thank all my friends and family for being so supportive of me. Without their genuine love as well as encouragement, I would not have been able to finalise this bachelor thesis.

10. REFERENCES

1. Barth, P. (2019, November 19). Data Literacy and the Colin Powell Rule. Retrieved from https://blog.qlik.com/data-literacy-and-the-colin-powell-rule#:%7E:text=MIT%20defines%20data%20literacy%20as,analyze%20and%20argue%E2%80%9D%20with%20data.

2. Belissent, ., Leganza, G., Arcand, K., Bennett, M., Little, C., Hoerman, E., & Vale, J. (2020). *Data Literacy Matters: The Writing's on the Wall*. Retrieved from file:///C:/Users/lisav/Downloads/Data%20Literacy%20Matters_%20The%20Writing's%20On%20The%20Wall%20(1).pdf.

3. Bhargava, R., & D'Ignazio, C. (2015). Designing Tools and Activities for Data Literacy Learners. In Wed Science: Data Literacy Workshop. Oxford, UK.

4. Databilities. (2020). A data literacy competency framework [Slides]. Retrieved from https://docs.wixstatic.com/ugd/1ff4ae_14805e0c8ef14b54bdafd38e44d5de23.pdf

5. Doody, O., Noonan, M., 2013. Preparing and conducting interviews to collect data. Nurse Researcher.. doi:10.7748/nr2013.05.20.5.28.e327

6. Ebbers, M., Abdel-Gayed, A., Budhi, V. B., Dolot, F., Kamat, V., Picone, R., … Redbooks, I. (2013). *Addressing Data Volume, Velocity, and Variety with IBM InfoSphere Streams V3.0* (1st ed.). Retrieved from https://books.google.nl/books?hl=en&lr=&id=cnXJAgAAQBAJ&oi=fnd&pg=PP1&dq=data+volume+velocity+variety&ots=OJo03pmYET&sig=E1XreCKILCePcdpU2eq1jgTsgmk&redir_esc=y#v=snippet&q=volume&f=false.

7. Greer, M. (2019, February 7). Data: The Fuel Powering AI & Digital Transformation. Retrieved from https://www.forbes.com/sites/cognitiveworld/2019/02/06/data-the-fuel-powering-ai-digital-transformation/#20965fce578b.

8. Grillenberger, A., & Romeike, R. (2018). Developing a theoretically founded data literacy competency model. *Proceedings of the 13th Workshop in Primary and Secondary Computing Education on - WiPSCE '18*, 1. https://doi.org/10.1145/3265757.3265766

9. Harris, J. (2012, September 13). Data Is Useless Without the Skills to Analyze It. Retrieved from https://hbr.org/2012/09/data-is-useless-without-the-skills

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

11. Hornick, M. (2018, June 8). Data Science Maturity Model - Data Awareness Dimension (Part 6). Retrieved from https://blogs.oracle.com/r/data-science-maturity-model-data-awareness-dimension-part-6

12. Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 1–10. https://doi.org/10.1177/2053951716631130.

13. Laney D (2001) 3D data management: Controlling datavolume, velocity and variety. In: Meta Group. Available at:http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

14. Lambers, E. E., Goedhart, B., & Madlener, J. J. (2019, January 8). How to become data literate and support a data-driven culture. Retrieved from https://www.compact.nl/articles/how-to-become-data-literate-and-support-a-data-driven-culture/#Levels_of_Data_Literacy

15. Lee, B. (2020, January 28). How to Close the Data-Literacy Gap At Your Company. Retrieved from https://blog.kloud.io/how-to-close-the-data-literacy-gap-at-your-startup

16. Marr, B. (2016). *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results* (1st ed.). Retrieved from http://www.bdbanalytics.ir/media/1169/bernard-marr-big-data-in-practice_-how-45-successful-companies-used-big-data-analytics-to-deliver-extraordinary-results-wiley-2016.pdf.

17. Merriam-Webster Dictionary. Data. Retrieved from https://www.merriam-webster.com/dictionary/data.\

18. Morrow, J. (2018, July). Developing a data literate workforce: A strategy and framework for the enterprise. Retrieved from https://www.qlik.com/us/-/media/files/training/global-us/qlik-education-data-literacy-program-strategy-and-framework.pdf

19. Open Data Institute (2020). Data Skills Framework – The ODI. Retrieved from http://theodi.org/wp-content/uploads/2020/05/2020-05-Data-Skills-Framework.pdf

20. Opher, A., Chou, A., Onda, A., & Sounderrajan, K. (2016). *The Rise of the Data Economy: Driving Value through Internet of Things Data Monetization*. Retrieved from https://www.ibm.com/downloads/cas/4JROLDQ.

21. Panetta, K. (2019, February 6). A Data and Analytics Leader's Guide to Data Literacy - Smarter With Gartner. Retrieved from https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guide-to-data-literacy/

22. Pohlmann, T. (2016, February 16). What Design Thinking Means for Data and Analytics. Retrieved from https://www.datanami.com/2016/02/16/what-design-thinking-means-for-data-analytics/.

23. Ram, J., Zhang, Changyu., & Koronios, A. (2016). The Implications of Big Data Analytics on Business Intelligence: A Qualitative Study in China. *Procedia Computer Science*, *87*, 221–226. https://doi.org/10.1016/j.procs.2016.05.152.

24. Schield, M. (2004). Information Literacy, Statistical Literacy, Data Literacy. *IASSIST Quarterly*, *28*(2), 6–11. https://doi.org/10.29173/iq790

25. Shelley, M. C., Yore, L. D., & Hand, B. B. (2008). *Quality Research in Literacy and Science Education*. Retrieved from https://books.google.nl/books?id=HVwt7pUKP5MC&pg=PA66&dq=the+ability+to+identify,+understand,+interpret,+communicate+and+compute,+using+printed+and+written+materials+associated+with+varying+contexts.+Literacy+involves+a+continuum+of+learning+in+enabling+individuals+to+achieve+their+goals,+to+develop+their+knowledge+and+potential,+and+to+participate+fully+in+their+community+and+wider+society&hl=nl&sa=X&ved=0ahUKEwjUsMfGjb3pAhUNKewKHbu-CP0Q6AEIJzAA#v=onepage&q=the%20ability%20to%20identify%2C%20understand%2C%20interpret%2C%20communicate%20and%20compute%2C%20using%20printed%20and%20written%20materials%20associated%20with%20varying%20contexts.%20Literacy%20involves%20a%20continuum%20of%20learning%20in%20enabli

ng%20individuals%20to%20achieve%20their%20goals%2C%20to%20develop%20their%20knowledge%20and%20potential%2C%20and%20to%20participate%20fully%20in%20their%20community%20and%20wider%20society&f=false

26. Stair, R., & Reynolds, G. (2009). *Principles of Information Systems* (9th ed.). Retrieved from https://drive.uqu.edu.sa/_/fbshareef/files/principles%20of%20information%20systems%209th%20-stair,%20reynolds.pdf.

27. Sternkopf, H., & Mueller, R. M. (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*, 5045–5054. https://doi.org/10.24251/hicss.2018.630

28. Tarrant, D., & Mezeklieva, V. (2019, May 8). Data Skills Framework – The ODI. Retrieved from https://theodi.org/article/open-data-skills-framework/

29. Tien, J. M. (2013). Big Data: Unleashing information. *Journal of Systems Science and Systems Engineering*, *22*(2), 127–151. https://doi.org/10.1007/s11518-013-5219-4.

30. Utica College. (2019, December 5). Three Examples of How Companies Make Data-Driven Decisions. Retrieved from https://programs.online.utica.edu/articles/data-driven-decisions

31. Wolff, A., Gooch, D., Cavero Montaner, J.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.

32. Zamel, V., & Spack, R. (2012). *Negotiating Academic Literacies*. Retrieved from https://books.google.nl/books?hl=nl&lr=&id=NKWdDqyJQgIC&oi=fnd&pg=PA51&dq=what+is+literacy&ots=iTNmqDrNTT&sig=JRDF2m-ypeIPcUTv0sIIBWmI5IE&redir_esc=y#v=onepage&q=what%20is%20literacy&f=false

11. APPENDICES
11.1 Appendix A

Data Literacy: Detecting Data Literacy within Businesses

Date: 02.06.2020
Researcher: Lisa van Disseldorp

Thank you very much for your interest to participate in this interview. The interview is part of my graduation project at the University of Twente and will help me to explore how managers and project leaders detect a data literacy gap among their employees.

The duration of the interview will be approximately 30 minutes and participation is voluntary. This means that you can refuse to answer questions and/or stop it at any time without consequences. There will not be any processing or storing of personal data, and your name and company will not be mentioned within the study.

Before we will start the interview, I will ask you to sign a consent form. Do you agree to this interview being recorded? The recording will be deleted after the graduation project is completed.

Interview questions:
1. If you think about data literacy, what are the first things you think of?
2. How do you define what data literacy level an employee must have to perform a certain job?
3. How do you monitor their performance regarding data literacy?
4. Do you have a strategy, or process, in place to detect a data literacy gap?
5. When you detect a data literacy gap, what do you do?

Contact details Lisa van Disseldorp
E-mail: e.m.vandisseldorp@student.utwente.nl
Mobile: +31 6 14 28 43 49

Contact details of the BMS Ethics Committee to file:
E-mail: ethicscommittee-bms@utwente.nl

11.2 Appendix B

Consent Form for Data Literacy: Determining the right level

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated 02/06/2020, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	□	□
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	□	□
I understand that taking part in the study involves an audio-recorded interview, which will be transcribed as text and deleted after the graduation project is completed.	□	□
Use of the information in the study		

I understand that information I provide will be used for a bachelor graduation project. □ □

I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team. □ □

Signatures

_____ _____ _____

Name of participant Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

Lisa van Disseldorp _____ _____

Researcher name Signature Date

Study contact details for further information:

Lisa van Disseldorp, e.m.vandisseldorp@student.twente.nl