Toward a Curriculum for Data Literacy in Enterprises

Hippolyte Lefebvre
University of Lausanne, Faculty of Business and Economics
hippolyte.lefebvre@unil.ch

Christine Legner
University of Lausanne, Faculty of Business and Economics
christine.legner@unil.ch

Abstract

To create business value from data, firms need a data literate workforce capable of reading, working, analyzing, and arguing with data. Prior studies on data literacy have mostly focused on educational settings and identified data-related skills. However, the suggested generic skill catalogs do not account for the highly situated nature of data practices. In this paper, we delve into five data literacy programs at multinational companies and examine their unique scope and characteristics. We leverage curriculum theory to analyze the different curriculum components and how they foster workplace data practices. As a contribution to data literacy research, we propose a theory-inspired and situated curriculum for data literacy in enterprises built upon five learning blocks, namely generic skills, disciplinary content, disciplinary skills, workplace awareness, and workplace experience. We also disclose each block's target audience, scope, and delivery mode and thereby inform practitioners on how to build their own curricula.

Keywords: Data literacy, Data competences, Data skills, Data practices, Curriculum

1. Introduction

Firms increasingly recognize the strategic potential data has and they seek to bring an increasing number of employees on board to engage in data-related activities. Expanding data practices to beyond the data expert domain requires a data literate workforce, i.e., employees that are able to read, work, analyze, and argue with data (D'Ignazio & Bhargava, 2015). For instance, business managers should be able autonomously to tackle basic tasks such as defining the requirements of a simple dashboard, while also accessing and analyzing data on it (Lennerholt et al., 2021). However, most companies still lack a workforce with the requisite data skills that would allow them to draw meaningful business insights out of data (Grover et al., 2018). Besides, managers tend to overestimate their workforce's capabilities and their readiness to work with data (Vohra & Morrow, 2020). Also, they struggle to establish working relationships between business and data experts (Redman, 2022).

Data literacy has mostly been studied as an educational theme that equips students with a list of core skills to prepare them for the job market (Carlson et al., 2013; Koltay, 2017). In the enterprise context, data literacy is often embedded in digital literacy research (Cordes & Weber, 2021; Goel et al., 2021). Thereby, it does not account for the distinctive nature of data (Paparova, 2023), nor for the fact that data use is highly situated (Alaimo & Kallinikos, 2022). Since data mainly gains value when it is put to use as a part of local actors' sense-making processes (Aaltonen et al., 2021), data literacy must be taught in context (Jones, 2019; Micheli et al., 2020). Although some studies have focused on listing skills for certain roles such as for data scientists (Demchenko et al., 2016; Saltz et al., 2018), they overlook the majority of employees who are less technically skilled but should still have a key role in creating value by making business sense out of the data they work with.

In contrast to IT literature that has investigated upskilling to address the surge in demand for IT workers at the beginning of the millennium (Ho & Frampton, 2010), very little research has looked into today's need to train a larger emerging community of employees to fulfil data roles in context. Considering this lacuna, we ask the following research question:

RQ: *How do companies develop data literacy programs to upscale their data practices?*

We opted for multiple case studies to capture rich and diverse insights directly from practitioners' working contexts (Paré, 2004). Based on Bennet et al.'s (1999) curriculum model, we analyzed data literacy programs with different scopes and target audiences from five multinational companies. This prism helped us to examine the cases using a common framework, to compare them, and to identify recurring patterns (Miles et al., 2014). As a contribution to data literacy research, we propose a theory-inspired and situated curriculum for large-scale data literacy programs that comprises five learning blocks covering

URI: https://hdl.handle.net/10125/107123
978-0-9981331-7-1

generic skills, disciplinary content, disciplinary skills, workplace awareness, and workplace experience. For each block, we indicate the target audience, the scope, and the delivery mode. Besides contributing to data literacy research, our cases and findings inform practitioners on how to build their data literacy curriculum.

In this paper, we first review the literature on data literacy and competence development, and we identify the research gap. Second, we explain our case study methodology and the research process. Third, we present our findings related to the five learning blocks. Finally, we discuss our findings and provide an outlook on future research.

2. Background

2.1. Data literacy

Data literacy refers to the ability to read, work, analyze, and argue with data (D'Ignazio & Bhargava, 2015). Interestingly, the existing body of knowledge on data literacy has mainly built upon concepts from educational research (e.g., high school, university) and library studies to define and investigate data literacy as a bundle of skills (Calzada Prado & Marzal, 2013; Carlson et al., 2011; Ridsdale et al., 2015). To identify them, researchers have mainly analyzed data experts' profiles and derived a set of generalizable and context-independent skills. Across these studies, data literacy is traditionally associated with a generic set of key skills such as data analysis, data curation, data visualization, data ethics and data security (see **Table 1**). More recent studies have extended these sets of data literacy skills for work and society (Schüller, 2020; Sternkopf & Mueller, 2018; Wolff et al., 2016). The resulting set of skills also reflects their application in a more specific context e.g., in developing hypotheses, identifying related sources of data that could support an investigation, accessing data, analyzing and creating explanations from data, or communicating with data. Additionally, a recent understanding of data literacy not only encompasses skills, but also includes behavioral dimensions such as attitude and values toward data (e.g., act data driven, data ethics).

Research has also emphasized the role of situated learning as pivotal for employees participating in data and analytics activities (Dubey & Gunasekaran, 2015; Lefebvre & Legner, 2022). Hence, data literacy cannot be characterized as a passively ingested skills set, which is detached from actual work applications (Zhu et al., 2019). Thus, we distinguish between skills as static uncontextualized properties and competences, i.e., abilities to apply a job's requisite skills (Bartram, 2005). Competences, then, refer to the ability to put the developed generic (i.e., cross-discipline) and situated (i.e., specific to workplace) skills into practice. Applied to the enterprise context, the goal is to develop employees' competences so that they are able to use and make sense of given data in a way that supports their daily work (Aaltonen et al., 2021). Finally, applying the behavioral competence approach (McClelland, 1973) to data literacy suggests that competences are not innate, and can be taught through programs that combine generic upskilling and workplace-relatable content. Data literacy then becomes a personal trait or set of habits that can lead to better job performance.

Table 1. Generic data literacy skills in the literature

Research field and sources	Examples of data literacy skills
Library Studies / Education (Carlson et al., 2011) (Calzada Prado & Marzal, 2013) (Ridsdale et al., 2015)	• Data discovery and acquisition • Data management • Data visualization • Data curation • Data processing • Data analysis • Data ethics and security • Data culture
Work and society (Wolff et al., 2016) (Sternkopf & Mueller, 2018) (Schüller, 2020)	In addition to the above: • Act data driven • Solve a problem with data • Identify data use cases • Coordinate data use cases • Evaluate impact of data • Trace back data transformations

2.2 Competence development and curriculum

Firms have acknowledged that developing talent and learning is vital for sustaining their business. Thus, they seek to equip their employees with the necessary competences to sustain such a new and competitive environment (Ho & Frampton, 2010; Merchel et al., 2021). Specifically, competence development has become a critical factor in preparing employees for a more technology-driven future (Li, 2022). Such development is often associated with a set of learning outcomes based on expected job qualifications. These learning outcomes support the mapping of learning content into a curriculum (Walker, 2003). A curriculum is defined as a collection of documents and learning activities aiming to deliver a structured series of learning experiences. It includes theoretical and practical content to equip learners with predefined competencies (Prifti, 2019). Clarifying learning outcomes ensures dedication to advanced proficiency levels and defined learning paths (von Konsky et al., 2016). A larger group of employees can share a subset of competences; yet, individuals' competences that are associated with personalized learning outcomes indicate that most competences are expected to address situated practices

(Le Deist & Winterton, 2005). Despite the evidence of recent training success, many training programs still neglect the role of workplace experience, disregarding different learning formats such as secondment or projects (Zhu et al., 2019). Further, learning materials are key elements in training and its success. Companies should, therefore, select and organize learning materials to meet both generic and situated learning expectations (Wang et al., 2014).

As a general framework, Bennet et al.'s (1999) view on learning suits the development of data literacy well, as it is a highly situated competence developed through collective understanding and workplace-like experience. Their curriculum model helps to bridge the gap between classroom training practices and workplace expectations. The model displays five blocks representing components to be enabled for learning success, identified as generic skills, disciplinary content, disciplinary skills, workplace awareness, and workplace experience (see **Figure 1**).

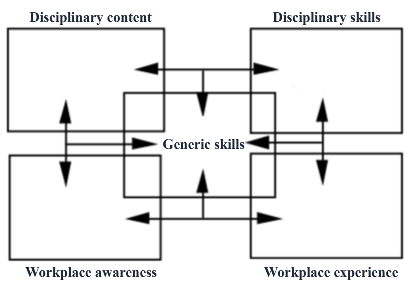

Figure 1. Curriculum model by Bennett, Dunne and Carré (1999)

Despite variations across disciplines, the generic skills in the middle of the model support all the other blocks by providing the necessary skills to engage in situated learning. Disciplinary content refers to conceptual knowledge corresponding to a trainee's own discipline. Trainees can develop a wider set of disciplinary skills relevant to disciplinary content and generic skills, which they can leverage and apply in a simulated (workplace awareness) or real (workplace experience) environment. The connections (arrows) between the blocks indicate the directionality of learning i.e., the options companies have in sequencing the learning blocks.

3. Methodology

Considering our research goal, we chose a qualitative research design using multiple case studies to investigate how companies develop their data literacy learning journeys (Paré, 2004). Case studies, as "*well-suited to capturing the knowledge of practitioners and developing theories from it*" (Benbasat et al., 1987, p. 370), are commonly used for answering "how" questions and multiple cases support better analytical generalization (Miles et al., 2014).

Data collection happened in two phases. The first entailed a focus group in June 2021 to exchange knowledge of best practices for developing data literacy competences. The participants were 12 experts from 9 companies representing different industries, with differences in scope (e.g., data analytics, data management) and maturity for data literacy. Since combining focus groups and surveys is generally recognized as suitable for sampling cases (Morgan, 1993), we used a survey to capture examples of data literacy initiatives, their target audiences, and their development phase. Following the survey results, we identified a subset of five mature data literacy training programs at five different companies. They differ in scope, audience, and industry (see **Table 2**). The second phase entailed semi-structured one-hour interviews with each of the five companies between July 2021 and November 2022. Preparing for the interview, we asked key informants (e.g., project manager, director analytics) to provide an overview of their data literacy curriculum. Our interview questions covered the theoretical framework's five areas to ensure results would be compatible, and we sent interview notes to the interviewees for validation. To enrich the case database and triangulate primary data, we searched for secondary data (e.g., press reports, presentations, company documentation). In this way, we also ensured reliability of the evidence. The five cases allowed us to reach theoretical saturation as we noticed redundance in incremental learning, for instance in patterns against our theoretical framework.

Table 2. Cases overview

Data literacy program	Industry	Audience (~# employees)
*R&D Academy – AI & Data Analytics (**A**)*	Manufacturing, automotive	R&D community (20,000)
*Enterprise Data Literacy (**B**)*	Packaging, food processing	All employees (20,000)
*Roadmap for data handling & understanding (**C**)*	Manufacturing, automotive	IT & Digitalization (15,000)
*Digital Analytics Academy (**D**)*	Fashion and retail	Digital unit in Sales (400)
Data Literacy Journey (**E**)	FMCG	Operations & Sales (5,000)

First, we ensured a thorough understanding of the context for each case (e.g., target groups, scope). For the analyses, we leveraged the theoretical insights on workforce development (see section 2.2), using Bennet et al.'s (1999) model as framework for individual analysis of the cases (within-case analysis). One researcher coded the case base against the framework dimensions (generic skills, disciplinary content, disciplinary skills, workplace awareness, workplace experience) and a second researcher reviewed the codes. The two researchers cleared the coding during a meeting in June 2023. **Table 3** illustrates the coding process for case B.

Table 3. Within-case coding examples for Case B

Case description	Coding	Explanation
Data ethics class in the form of an e-learning for all based on a LinkedIn playlist.	Generic skills	Data ethics is currently a typical skill in all data and analytics roles.
"Data Playground" as a new data experimentation platform where trainees are assigned data experts as mentors.	Workplace awareness	Application of the skill in the form of simulation fosters situated learning.
70% of learning journey should happen in the workplace (e.g., projects and job rotations).	Workplace experience	Skills development primarily happens on-the-job.

The comparative analysis is particularly relevant to this study as it supports the aggregation, simplification, and generalization of complex cases (Miles et al., 2014). Moreover, natural variation between cases generally strengthens theory building (Dubé & Paré, 2003). For the cross-case analysis, we performed "pattern-matching," thereby identifying differences and commonalities on the learning blocks level to determine similar ways of developing both generic and situated learning. We iteratively searched for similarities between codes (e.g., all "workplace awareness" codes) and then created and grouped types of codes to examine cases for shared configurations. These we summarized in a curriculum based on the identified five blocks (see results in section 5).

4. Cases

Below, we describe each case in details, also showing how every case maps onto Bennet et al.'s (1999) five framework enablers for workplace upskilling. While directionality of learning is briefly addressed in each case narrative, our analysis focuses on each curriculum's learning blocks rather than their sequencing. We provide a figure that summarizes the case analysis and highlights the key blocks enabled by the data literacy curriculum in dark grey (*Major)*, and the blocks enabled but at a lower intensity in light grey (*Minor)*, with unactivated blocks in white.

4.1. Case A: *R&D Academy – AI & Data Analytics Landscape*

Company A is a large automotive supplier ($1B–$50B revenue/~150,000 employees) that invests considerably in next generation mobility ($2.5B in 2022), e.g., in automated driving. The firm released a "data enablement strategy" in 2020 in planning for data-driven innovation. Accordingly, they set up a data enablement team to break down data silos and to stimulate collaboration on various data use cases between data and business experts. As a first step toward their data-driven business model, the firm decided to focus on upskilling more than 15 000 employees in the R&D department. The company started developing a data literacy program, the *R&D Academy*, dedicated to the entire R&D community. Before this, only a few data experts had benefited from comprehensive data literacy training programs. It chose to personalize the program centered on three role families: *employees/managers*, *domain developers/subject matter experts*, and *AI experts*. Training is optional and the content is structured in one of the following proficiency levels: I-*Create Awareness* aims to raise awareness of the company's business and data strategies and their impact on R&D, and introduces selected foundational topics to employees/managers and domain developers/subject matter experts. At proficiency level I, role families can benefit from an introduction session on AI, Data Science, and Machine Learning leveraging LinkedIn Leaning Playlists. Level II - *Gain Deeper Understanding* focuses on R&D role families' specific technological and technical competences by offering one-day to three-day qualification courses. Level III-*Achieve Enablement* enables selected R&D engineers to fulfil the requirements of their specific technology domains through longer qualification programs (>10 days). At level III, they offer an expert program that enrolls 40 engineers per semester. So far, the program relies largely on virtual content such as sourced e-learning (e.g., LinkedIn, Udacity) and knowledge sharing via the analytics communities; however, the firm anticipates bringing in other learning formats such as conferences (more than 1000 participants from eight divisions during the 2022 edition). Further, an "AI adventure" program is being planned to raise non-experts' awareness of AI through a collaborative game presenting mini problems to solve with data. **Figure 2** maps case A onto Bennet et al.'s (1999)'s building blocks.

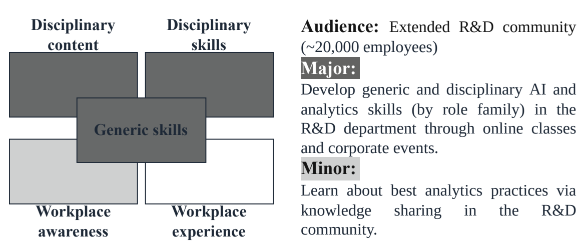

Figure 2. Case A mapped onto Bennet et al. (1999)'s framework

4.2. Case B: *Enterprise Data Literacy*

Company B is a large multinational ($1B–$50B revenue/~25,000 employees) operating in packaging and food processing. It has identified operational excellence as a key enabler in its business strategy, "Company 2030." They described the required data capabilities in their 2019 data and analytics strategy. The firm has been developing a corporate-wide data literacy initiative called *Enterprise Data Literacy* (EDL), a recommended but not obligatory program, by which they aim to upskill 20 000 employees. The company decided to design EDL for three participant groups: *Data citizens* (all employees) who should understand why data is important and how it is used for the firm's business; *business analysts* (e.g., a marketing analyst) who should have strong domain knowledge and be able to talk comfortably with the third group called *citizen data scientists*. The learning outcomes for each participant type are divided into to three proficiency levels, i.e., the *Conceptual*, *Core* and *Advanced* levels. EDL is implemented on an EdCast platform, mostly offering classes sourced from LinkedIn and addressing all proficiency levels. The classes are bundled into introductory "learning journeys" to inspire all role players. For more advanced players it deep dives into defined areas allowing them to "pick-and-choose" what is most relevant for them. However, this "structured learning" represents only 10% of EDL's learning design framework. The next 20% is about "learning from others," which aims to sustain the learning momentum through social and collaborative knowledge sharing. Thereby, employees can benefit from coaching and mentoring opportunities with experts or join communities of practice (e.g., a Business Intelligence (BI) community). Currently under construction, is a "Data Playground" that will offer a safe space for employees to practice data analytics skills. The last 70% is about "learning from experience" and integrating learning with work. This longer-term part of the program expects freshly trained employees to develop sustained autonomy in taking action and solving problems with data.

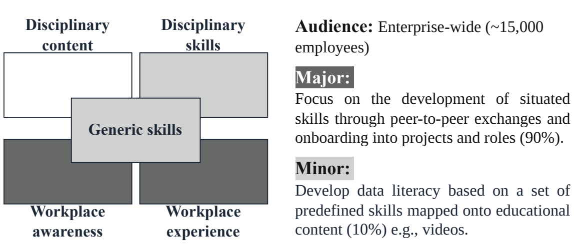

Figure 3. Case B mapped onto Bennet et al. (1999)'s framework

Depending on the specific project or assignment, placements, secondments, and job rotations can be involved. **Figure 3** summarizes our analysis based on the building blocks suggested by Bennet et al. (1999).

4.3. Case C: *Roadmap for data handling and understanding*

Company C is a large multinational ($1B–$50B revenue/~90,000 employees) involved in the automotive and manufacturing business. The firm seeks to develop a data culture to accompany its recently released data and analytics strategy (2021) focusing on industry 4.0 and AI in business processes. After releasing a new data organization, the firm needs to provide improved data access and to develop data and analytics skills for new roles. The company is developing a project, *Roadmap for data handling and understanding,* to increase awareness and to upskill various roles in several digitalization areas through a structured learning program. The program is designed for three proficiency levels, i.e., *Basic Knowledge – Interested and Affected by Digitalization*, *Advanced – Participate in Digitalization*, and *Experienced – Realization of Digitalization Projects*. While all employees in the IT and digitalization department are expected to know the foundations for data handling, most of the training is role-specific. For instance, the basic level includes generic and role-specific courses: generic courses intended for all participants offer e-learning content (e.g., What is BI? What is a digital twin? Introduction to data management), while role-specific courses (e.g., Data management basics, Data-driven decision making; Self-service BI basics) address different kinds of data expertise. The advanced and experienced levels are fully role-specific. They cover different specialized topics depending on the trainee's role: data analysis, data science, digital twinning, semantic models, and data management. To illustrate, the advanced level includes classes such as Consuming Analysis for Office, Consuming SAP Analytics Cloud, Digital Twin API hands-on, Semantic Modelling Fundamentals, or Data Modelling & Data Catalogue. Eventually, the experienced level aims to upskill "data-citizen" roles and IT roles by offering classes focused on creation and innovation, such as Design Principles for Self-Service BI, Data Science Workbench, Semantic Modelling – Projects, Data Management Processes, How to think like a Data Scientist. Also, the company is exploring alternative formats such as mini projects, while investigating the synergies with several existing communities of practice. Overall, **Figure 4** maps case C onto Bennet et al. (1999)'s framework.

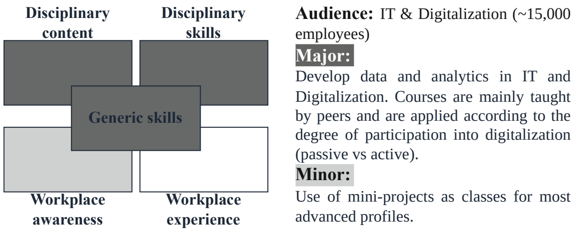

Figure 4. Case C mapped onto Bennet et al. (1999)'s framework

4.4. Case D: *Digital Analytics Academy*

Company D is a large fashion company ($1B–$50B revenues /~60,000 employees) experiencing a digital transformation of its sales channels, notably triggered by a surge in digital sales during the Covid-19 period. In this context, data literacy is mentioned as a strategic enabler for their large digital sales unit which is responsible for e-commerce and digital activities, including sales growth and advertising. The department expects the employees to be able to generate and leverage data-driven insights that help digital sales growth (e.g., using metrics to track net sales or to monitor product demand within and across e-commerce channels). This includes business roles, such as product category managers or digital marketing specialists, as well as data experts who currently lack integrating the business context when developing analytical products. For instance, product category managers currently struggle to find the information they need for decision-making or do not act on the analytical insights provided by the analytics team. To design its data literacy program (focused on analytics and data-driven insights), the company first reviewed the organization and the different analytical roles and value drivers. Then, they created a list of 13 job families (e.g., digital activation, digital marketing, product ownership, decision science). Through a comprehensive analysis of skills and job descriptions (internal and external), the firm derived 25 analytical skills groups (and more than 400 skills) to map onto the 13 job families. Company D then created a skill finder tool which is fed by raw data from the skill mapping project to support skills discovery for each employee. This enabled the development of individual learning paths for the different job families across three core areas: tools/dashboards, KPIs and data, technique and skills. Each of these areas is linked to six learning outcomes representing the different cognitive steps of learning: awareness, meaning, adoption, interpretation, communication, creativity. Accordingly, by Q1 2023, the company expects 100% of the employees in the digital sales unit (e.g., campaign managers, data scientists) to be data aware (i.e., achieved the *awareness* learning outcome) and by Q2 2023 they expect 60% of the digital sales unit to use a set of key dashboards in self-service at least on a quarterly basis. As a first step, the company offered several awareness sessions on MS Teams with 100+ digital sales employees to emphasize the importance of data (e.g., how metrics can provide business insights). The program has already shown progress: within a year, the number of consumers on the academy's SharePoint has tripled, as have the visits on the key dashboards. **Figure 5** maps Case D onto Bennet et al. (1999)'s framework.

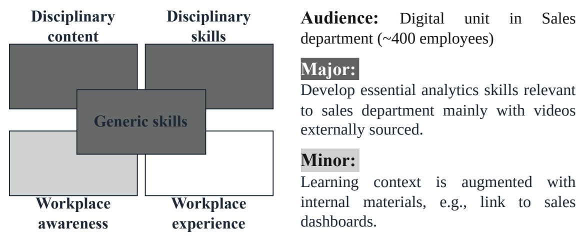

Figure 5. Case D mapped onto Bennet et al. (1999)'s framework

4.5. Case E: *Data Literacy Journey*

Company E is a large FMCG company ($1B–50B revenue/~60,000 employees) halfway through a large business transformation started in 2017. The radical shift toward an electronic device product line required that the firm invest considerably in its digital capabilities. In this context, the firms embarked on a large data and analytics journey which started with a data and analytics organization of eight people tasked with setting up the data foundation (e.g., data governance, management, and quality control, and a business glossary) while drafting the initial analytical roadmap and engagement. Within six years, the firm managed to roll-out a 100+ FTE data organization with strong data management and analytics capabilities. For instance, in 2020, they had started developing 28 data science use cases (400+ million USD). In developing data roadmaps for various business functions, the firm realized soon that data literacy is a central skill in decentralized data enablement. Starting in the department with the highest needs, in this case the Operations and Sales department, company E identified four role families to be trained, i.e., *senior executives* (C-Suite roles and their direct reports (e.g., CEO, SVPs, VPs), *business leaders*, (e.g., directors and managers, data owners and stewards, subject matter experts, analytics product owner), *specialists in data-and-analytics functions* (e.g., business analysts, visualization experts), and *technical experts* (e.g., data architects, data engineers, data scientists, source systems specialists). After performing a skill gap analysis, the firm developed a pilot data literacy program called *Data Literacy Journey* focusing on a cohort of 400+ business leaders. They were considered the primary consumers of data-driven insights (e.g., in defining and using metrics, identifying opportunities for data use cases, taking responsibility for local data collection and quality). The training program offers 1) a three-hour pre-work self-paced awareness course); 2) 20 hours of virtual instructor lead training (VILT) fostering engagement and interaction in two modules, 9 hours of introduction to data and data products, articulating a business problem, metrics, data management basics, visualization, and storytelling, and 11 hours of a course introducing ML, data governance, and digital platforms.; 3) ongoing engagement after training through self-paced assignments (creating an individual data product plan with expert coaching). Upskilling materials are sourced from externally available programs and platforms (e.g., IMD business school, Coursera) and augmented with relevant company-specific content such as real-life use case examples. After being piloted, the program was scaled up to train 5000+ business leaders within 18 months, including 3000+ in the consumer and commercial department. **Figure 6** maps case E onto the Bennet et al. (1999) framework.

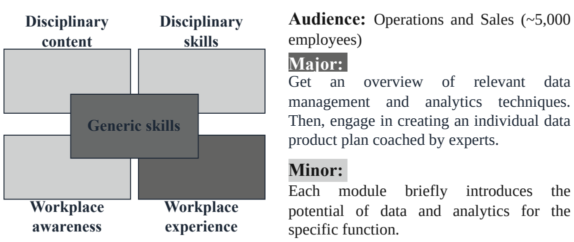

Figure 6. Case E mapped onto Bennet et al. (1999)'s framework

5. A data literacy curriculum built on five blocks

We generalize our findings in the form of five building blocks for a data literacy curriculum (see **Figure 7**). For each, we highlight the key findings on audience, scope, and delivery mode. Our model is generic enough to offer flexibility in customizing the learning outcomes and direction of learning.

A key motivation for developing data literacy is to enable employees with different data backgrounds to work, collaborate, and communicate with others about data or in projects. We observed that data literacy programs have a common baseline or include foundational skills for the entire audience, which we characterize as **generic skills** since they should be transferrable to any work environment. In all cases, we found that common learning outcomes encompass motivational topics on the value of data in the context of the company's strategy. As highlighted in case D, these topics could also support the development of certain essential skills such as foundations of statistics, data tool landscape overview, or high-level impact of data on enterprise processes. Trainees are mostly expected to ingest basic concepts and be able to apply them in a meaningful way. It is also important for all role players to understand the impact of data literacy on their career progression, also in using success stories. Curiosity to engage in upskilling should be fostered at this stage.

Beyond generic skills, employees further need to engage in additional modules that show how data can be used in their specific (business) context, i.e., they need **disciplinary content**. For instance, after having understood what tools (e.g., BI tools) and techniques (e.g., checking duplicates) are available to analyze data, one needs to understand how these relate to their disciplines i.e., their specific working environment. Here, both data specialists and experts should know about data's impact on specific business processes and other possibilities of value creation from data. As highlighted in case C, disciplinary content can be taught by peers, for instance data experts or subject matter experts.

As a participant of the focus group mentioned: "*Not everyone needs to be a data scientist*." **Disciplinary skills** aim to transform disciplinary content into situated data activities, i.e., they are the skills necessary to use data in daily work. They are typically aligned with job descriptions. Hence, firms should communicate skills expectations for different roles and job levels, for instance in the form of a skills framework. In all cases, we found the development of disciplinary skills should be stimulated with advanced modules either cultivating basic concepts taught as part of generic skills, or with new learning materials specific to the working context. Firms should also offer trainees the possibility of requesting additional training and certification to sustain engagement.

Courses providing **workplace awareness** aim to support the application of theoretical knowledge in a simulated environment, as authentically as possible. To do so, these courses are organized for specific personas or role families. As in cases A, B, and C, firms can set up playful activities (e.g., gamification, workshops, or mini-projects) to immerse trainees (any

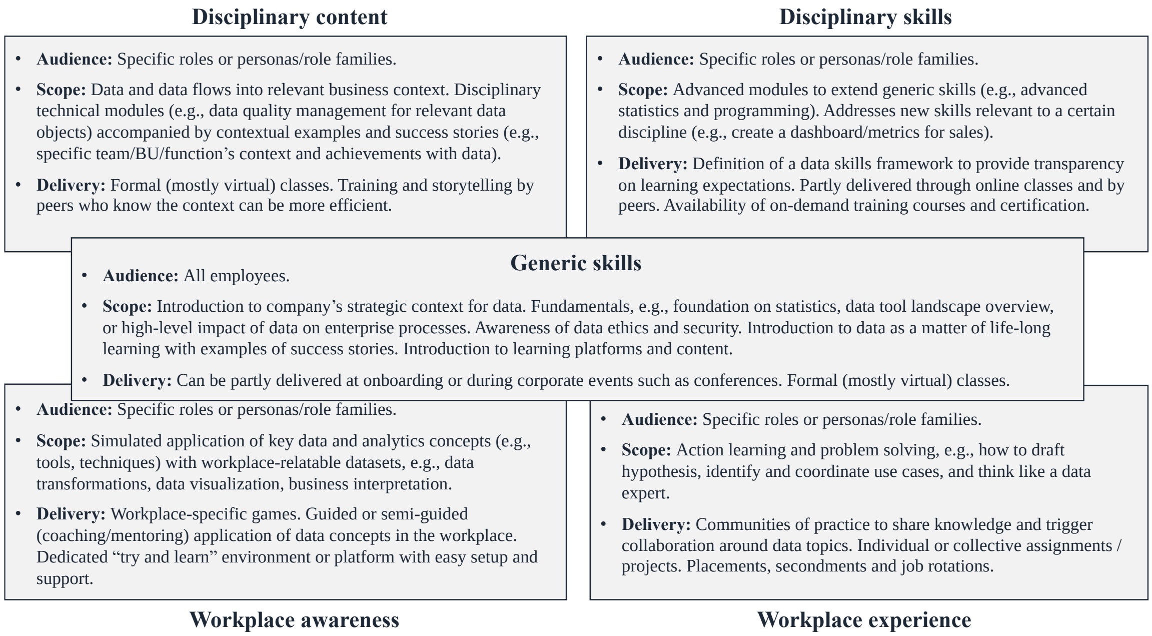

Figure 7. Data literacy curriculum

persona or role families) in workplace-relatable problems and challenges. They can also offer dedicated "data sandbox" environments (e.g., an analytics platform) which approximate the workplace activities. Trainees can then benefit from ongoing support from their peers and especially from data experts to learn about their use of data. Workplace awareness is critical for trainees' sustained engagement and satisfaction since it becomes a first bridge between theory and practice.

Workplace experience is about inviting trainees to take on data responsibilities and commit to a continuous learning journey. Trainees are part of a data users' community from whose experience they can benefit. Employees benefit from knowledgeable community members by deriving mental frameworks to address typical data-related challenges or to work on solutions. As in cases B and E, employees can then be paired with experts on projects so that together they can contribute to visualizing data use cases. To gain expertise trainees can also be seconded, placed in temporary positions, or in a job rotation.

6. Discussion and implications

Overall, our results resonate with the ongoing discourse on data as a matter of practices (Aaltonen et al., 2021). Users interest in training offers and their desire to develop the required workplace competences depend on a proper fit between the curriculum and realistic workplace expectations. This is highlighted in Case B that offers a new and highly situated pattern of curriculum provision adding to the six patterns already identified by Bennet et al. (1999). Our results show that data literacy curricula should offer personalized learning paths that address specific audience needs, including those of existing data roles and of data experts who have often been neglected in existing data literacy literature. We derive and propose three typical persona requiring data literacy training: data amateurs (e.g., casual data consumers with no data responsibility), data specialists (data consumers or creators for whom data is a part of their work routine, e.g., business managers, data owners), and data experts (data professionals who can act as coach e.g., data scientists, data quality manager).

Our cases also show that learning outcomes vary considerably across persona. Data literacy encompasses more than a simple set of generic skills (such as the ones in **Table 1**). The context-specific nature of data literacy also requires situated enablement by means of disciplinary content, disciplinary skills, workplace awareness, and workplace experience. In other words, apprenticeship will hopefully lead employees from novice levels to mastery (Gherardi, 2000). Further, we find that many data literacy skills (e.g., communicating with data, presenting with data) can be interpreted as generic and transferrable to various work environments. These skills become disciplinary depending on the associated level of proficiency. In fact, a single skill can be observed at various cognitive levels, i.e., ingested rather passively or by enacting it in practice. This marks the distinction between "*knowing that*" and "*knowing how*" (J. R. Anderson, 1983). For instance, the seminal Bloom's taxonomy suggests six progressive levels of cognitive learning identified as remember, understand, apply, analyse, evaluate, and create (L. Anderson et al., 2001). Hence, firms should clarify learning outcomes in terms of the level of cognition applied to the skills and should ask themselves questions such as: *When I conceptualize data analysis as a skill, what do I expect concretely from a given target group?* We believe this crucial point unlocks opportunities for further research on cognitive expectations for different data and analytics roles and on the pre-requisites and skills at the boundaries between roles.

Further, our cases add to existing evidence that shows how a diverse learning toolbox is a success criterion for skill transformation in enterprises (Billing et al., 2021). Companies should complement their own business-specific materials addressing disciplinary content with content from third-party providers or external mainstream sources. Such learning design is essential to trigger behavior change toward establishing a data culture. For instance, several cases in our study by default relied on external learning platforms, such as the prominent LinkedIn learning. A data manager in our focus group said about the latter: "*We are pragmatically using what learning opportunities are already available to us, and ideally they should be free*." Additionally, researchers could do a more detailed study of what makes a successful data literacy learning environment.

To conclude, we contribute to data literacy research on various levels. First, we offer a theory-inspired and situated curriculum concept that relies on successfully enabling learning blocks to develop data literacy in enterprise. Second, we provide detailed descriptions of five data literacy programs with different scopes and target groups, and we highlight data literacy curriculum patterns. Third, by offering a blueprint for developing data literacy curricula, this research will also inform the practitioner community.

Our study comes with certain limitations. Our sample includes only large multinational companies with a certain level of experience in data management and analytics, and with access to human and financial resources. Therefore, our findings may not be generalizable to smaller companies and their specific challenges (e.g., a smaller audience for data literacy, lack of data awareness and organization).

7. References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. *Journal of Management Information Systems*, *38*(2), 401–429.

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. *Organization Science*, *33*(1), 19–37.

Anderson, J. R. (1983). *The Architecture of Cognition* (Cambridge, MA: Harvard University Press).

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., & Wittrock, M. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York : Longman.

Bartram, D. (2005). The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology*, *90*(6), 1185–1203.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369–386.

Bennett, N., Dunne, E., & Carré, C. (1999). *Patterns of core and generic skill provision in higher education*. *Higher Education*(37), 71–93.

Billing, F., De Smet, A., Reich, A., & Schaninger, B. (2021). *Building workforce skills at scale to thrive during—And after—The COVID-19 crisis*. McKinsey. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/building-workforce-skills-at-scale-to-thrive-during-and-after-the-covid-19-crisis

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri*, *63*(2).

Carlson, J., Fosmire, M., Miller, C., & Sapp Nelson, M. R. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Portal: Libraries and the Academy*, *11*(2), 629–657.

Carlson, J., Johnston, L., Westra, B., & Nichols, M. (2013). Developing an Approach for Data Management Education: A Report from the Data Information Literacy Project. *International Journal of Digital Curation*, *8*(1), 204–217.

Cordes, A.-K., & Weber, J. (2021). Design Principles for Digital Upskilling in Organizations. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Innovation Through Information Systems* (pp. 509–525). Springer International Publishing.

Demchenko, Y., Belloum, A., Los, W., Wiktorski, T., Manieri, A., Brocks, H., Becker, J., Heutelbeck, D., Hemmje, M., & Brewer, S. (2016). EDISON Data Science Framework: A Foundation for Building Data Science Profession for Research and Industry. *Proceedings of the 8th IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, 620–626.

D'Ignazio, C., & Bhargava, R. (2015). Approaches to Building Big Data Literacy. *Proceedings of the Bloomberg Data for Good Exchange Conference*, 6.

Dubé, L., & Paré, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices,

Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597–636.
Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in Big Data and Business Analytics. *Industrial and Commercial Training*, *47*(4), 174–181.
Gherardi, S. (2000). Practice-based Theorizing on Learning and Knowing in Organizations. *Organization*, *7*(2), 211–223.
Goel, K., Sindhgatta, R., Ilavarasan, V., & Kar, A. K. (2021). Building a knowledge society: Learnings from a digital literacy program. *Proceedings of the 29th European Conference on Information Systems (ECIS)*.
Grover, V., Chiang, R. H. L., Liang, T.-P., & Zhang, D. (2018). Creating Strategic Business Value from Big Data Analytics: A Research Framework. *Journal of Management Information Systems*, *35*(2), 388–423.
Ho, S. Y., & Frampton, K. (2010). A Competency Model for the Information Technology Workforce: Implications for Training and Selection. *Communications of the Association for Information Systems*, *27*.
Jones, M. (2019). What We Talk about When We Talk about (Big) Data. *The Journal of Strategic Information Systems*, *28*(1), 3–16.
Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, *49*(1), 3–14.
Le Deist, F. D., & Winterton, J. (2005). What Is Competence? *Human Resource Development International*, *8*(1), 27–46.
Lefebvre, H., & Legner, C. (2022). How Communities of Practice Enable Data Democratization Inside the Enterprise. *Proceedings of the 30th European Conference on Information Systems (ECIS)*, 16.
Lennerholt, C., Van Laere, J., & Söderström, E. (2021). User-Related Challenges of Self-Service Business Intelligence. *Information Systems Management*, *38*(4), 309–323.
Li, L. (2022). Reskilling and Upskilling the Future-ready Workforce for Industry 4.0 and Beyond. *Information Systems Frontiers*.
McClelland, D. C. (1973). Testing for competence rather than for “intelligence.” *American Psychologist*, *28*(1), 1–14.
Merchel, R., Süße, T., Iqbal, T., & Knop, S. (2021). Digital Competencies and IT Skills as Employees’ Resources to Cope with Digitalization Demands. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*.
Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, *7*(2).
Miles, M. B., Michael, H. A., & Johnny, S. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.
Morgan, D. L. (1993). *Successful Focus Groups: Advancing the State of the Art*. SAGE Publications.
Paparova, D. (2023). Exploring the ontological status of data: A process-oriented approach. *Proceedings of the 31st European Conference on Information Systems (ECIS)*.
Paré, G. (2004). Investigating Information Systems with Positivist Case Research. *Communications of the Association for Information Systems*, *13*(1), 18.
Prifti, L. (2019). *Professional Qualification in “Industrie 4.0”: Building a Competency Model and Competency-Based Curriculum*. Universitätsbibliothek der TU München.
Redman, T. C. (2022). Your Data Initiatives Can’t Just Be for Data Scientists. *Harvard Business Review*. https://hbr.org/2022/03/your-data-initiatives-cant-just-be-for-data-scientists
Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B., & Ali-Hassan, H. (2015). *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*.
Saltz, J., Armour, F., & Sharda, R. (2018). Data Science Roles and the Types of Data Science Programs. *Communications of the Association for Information Systems*, *43*, 615–624.
Schüller, K. (2020). *Future Skills: A Framework for Data Literacy* (53). https://hochschulforumdigitalisierung.de
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*.
Vohra, S., & Morrow, J. (2020). *The Human Impact of Data Literacy*. Accenture. https://www.accenture.com/us-en/insights/technology/human-impact-data-literacy
von Konsky, B. R., Miller, C., & Jones, A. (2016). The Skills Framework for the Information Age: Engaging Stakeholders in Curriculum Design. *Journal of Information Systems Education*, *27*, 37–50.
Walker, D. F. (2003). *Fundamentals of curriculum: Passion and professionalism* (2nd ed). L. Erlbaum Associates.
Wang, Y., Gray, P. H., & Meister, D. B. (2014). Task-driven learning: The antecedents and outcomes of internal and external knowledge sourcing. *Information & Management*, *51*(8), 939–951.
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).
Zhu, S., Gupta, A., Paradice, D., & Cegielski, C. (2019). Understanding the Impact of Immersion and Authenticity on Satisfaction Behavior in Learning Analytics Tasks. *Information Systems Frontiers*, *21*(4), 791–814.