Data Literacy Skills and Active Teaching Methods:
Consistency Between School Education and Labor Market Demands

**Abstract**
As data becomes part of modern society, schools are incorporating data literacy in the classroom and businesses are also requiring a data literate workforce. This paper goes through the current scientific literature written about data literacy education and business to compare the different data literacy skills at school and businesses. In addition, this research aims to find the best method for teaching data literacy. The literature reveals that schools are using active learning to teach data literacy and what they're teaching is consistent with the demand of businesses. The findings of this research can be used to get an idea of skills and tools that a data-literate person should have, for schools to revise their data literacy curriculum, or people in the labor market to get a sense of what employers are expecting in-terms of data-literacy.

## Introduction

Data Science, Big data, AI, and Machine Learning are terms that we're familiar with nowadays but they would've sounded like words from an extraterrestrial language in the past.

Within the last two decades, our society has become more reliant on technology, and we've approached the age of big data. This means that consumers of technology are generating endless streams of data and to stay competitive businesses need data-literate employees to be able to extract valuable information from the massive amounts of data that's being generated. In particular, skills such as data collection cleaning, and story-telling are seeing a surge in demand, while skills such as traditional research methods and inferential statistics are becoming less valuable (Overton & Kleinschmit, 2023). On the other hand, there is a rise in fake news and misinformation, so a need for citizens to be data literate is also on the rise (Deja et al. 2021). To meet the rising demands, schools are looking for ways to incorporate data literacy into their curriculum across all levels of education from grade school to higher education.

There have been numerous attempts at defining what exactly is data literacy. One of the earliest definitions of data literacy was proposed by scholar Schield as "how to obtain and manipulate data" (2004). A more modern definition of data literacy was proposed in 2020 by scholar Forrester/Belissent et al. as "the ability to recognise, evaluate, work within, communicate, and apply data in the context of business priorities and outcomes."(2020).

Currently, the research front for data literacy and business agree that there is a lack of data literacy skills in the workforce. Hence, to remain competitive, businesses are wanting skilled people who are able to extract invaluable information from data to make decisions. Businesses

want to get in on this action so much, they're even willing to train their employees to be data literate. According to Disseldorp in 2020, 80% of businesses will try to overcome the data literacy gap by creating and developing programs. However, about 50% will not have access to proper AI and data literacy skills. Hence, businesses are requiring people who are able to perform certain data tasks rather than having knowledge to use specific tools. To fulfill this demand, schools are also integrating data literacy into their curriculum. This naturally prompts the following concerns that have not yet been reviewed. It is simply whether data literacy skills taught in schools are consistent with the demands of the workforce and what is the most effective way to teach data literacy for the best results.

Hence, the research aims to unravel the most effective teaching methods for teaching data literacy. In addition, the research examines if what the schools are teaching is consistent with data skill demands in the workforce. This research will help to provide an idea of the skill set of a data literacy individual, for schools to gain insight into the most effective teaching practices and work with school-specific conditions to improve the quality of data literacy education. That way, the school can recalibrate its curriculum to meet the needs of its students to maximize data literacy in the future workplace. Finally, this research will benefit job-seekers and current employees to get an idea of what data literacy skills can be expected by their employers.

**Methodology**

This systematic literature review was conducted as a writing project for Writing 109ST – Writing for Science and Technology at the University of California, Santa Barbara. It is a 6-week summer course and students had under 6 weeks to learn, research, and compile the materials of a systematic literature review paper. Some of the constraints for this research are also to be consistent with the course syllabus. The topic of data literacy was suggested by the course instructor and was accepted by the researcher based on personal interest in statistics, data science, and education.

Since data literacy is a rather broad topic and is dependent on having a specific domain, the research topic was refined and narrowed down by considering data literacy education and valuable data skills that are needed in the labor market. There is no strict restriction on the level of education, but data literacy at the higher education level was preferred to make the search process easier. To get the most credible and latest information, criteria for peer-reviewed articles from 2018 to 2024 were implemented.
The following key terms were used to find more literature for this new research topic with the use of quotation marks and operators:

| | | | |
|---|---|---|---|
| data literacy teaching | data literacy demand | data skill business | data literacy employment |

| | | | |
|---|---|---|---|
| data literacy education | data teaching school | data skill teaching | data literacy teaching study |

After looking through the UC Santa Barbara Library database, the literature search was also conducted on Google Scholar. The same key terms were used and the search was under the same criteria as before to maximize credibility and recency. The "cited by" feature on Google Scholar was also used to find more literature. Since not all search results were available to the public, permission to access certain articles was possible via UC Santa Barbara's VPN supported by Ivanti. In addition, as the number of literature found during the research increased, Zotero was used to keep track of articles that met the criteria.

Any work not written in English and books were excluded. By the end of the initial search process, a total of 29 articles were found but only a small number contained an empirical study. The search terms were expanded to include new search terms such as "data science", "data analysis", and "business". Once the search process concluded, approximately 10 papers were able to be included in the research. The inclusion meant that the articles met the previous criterias and they were able to answer the research questions.

- **RQ 1**: What are the most effective teaching methods for teaching data literacy?
- **RQ 2**: Are data literacy skills taught in schools consistent with the demands of the workforce?

Each article was read through in the following order:

Abstract → Result → Discussion → Introduction → Methodology

Any interesting findings, statistics, and ideas in each article were carefully annotated for later collection. Data collection was conducted after reading through all 10 articles. Annotations from each article were collected and stored for synthesis in the results section.

**Results**

With our updated search terms to include "data science", "data analysis", "statistics", and "business", we were able to find some more empirical studies relating to data literacy than before. A typical paper that includes an empirical paper looks like the paper written by Dzuranin in 2018. The papers adhere to the IMRaD format. The introduction mentions how data is becoming relevant to our lives and data literacy is becoming necessary. The authors also define data literacy in the introduction. If there is a particular teaching method that the author uses in their research, they'll define it in the introduction as well. The definition usually includes the benefits of the teach method and theories on how it improves student's learning. Because of this, the introduction does take up the majority of the paper itself. The methodology section involves the teacher incorporating the method into their course. In this case, the methods are active learning strategies through projects throughout the course. Before starting the projects, some teachers note students' attributes, such as sex, major, and which college they belong to at the

university. However, Tsai's work revealed that the college that the students are studying at is not a no significant indicator of performance in data literacy (2024). A typical project involves students working a data analysis project where they're exposed to data concepts such as data collection through APIs, data analysis through programming languages such as R, and data visualization through Tableau. A more extensive list of tools for data literacy are included in Table 1.

According to Moon et al. in 2024, a recent survey examining the current state of data science in 69 colleges and universities found that there was no uniformity regarding the technologies used in data science courses. However, they were able to uncover the following trend amongst students. It was found that students prefer to use tools that offer the ability to drag and drop menus for data visualization over programming language. Interestingly, programming tool use decreased as datasets got larger while analysis tools were used at a low rate, but it was a consistent rate regardless of dataset size.

After the project is done, the teacher measures students' learning progress through a number of assessments about the course content and a survey for feedback. The teacher then does data analysis of students' responses which includes hypothesis testing and visualization of the results section. There is little mention of a control group to compare results to. The discussion is for the impact, implications, and limitations of the active learning methods.

According to the literature, the most popular method for teaching data literacy is with project-based learning, where students learn data literacy by doing a project. Under this method of teaching, teachers teach enough theory for the students to start a data science project. During the project, teachers help the students by facilitating and supervising their projects instead of teaching. From doing projects, students in the research were able to learn data skills better and became equipped with a multitude of data science tools. The improvement in learning is evident in their assessment at the end and the changes were statistically significant across the board (Tsai, 2024). In addition, while working on their projects, students are given the opportunity to work with their peers. The effectiveness of learning from peers is heavily supported by several authors in this research (Deja et al., 2021; Disseldorp, 2020; Tsai 2024; Chen, 2023; Moon et al., 2023). In terms of student's reactions to project-based learning, the literature reveals that students are adapting rather well to project-based learning. Studies that incorporate a data science project into class received positive reactions in their end surveys. The students felt that they learned about data literacy, interpreting data to others made them try harder to understand, and the skills they learned were applicable in their future careers (Hesford et al., 2024; Chen, 2023; Dzuranin et al., 2018)

Lastly, Disseldorp reveals that businesses want people who can collect data, clean data, and derive meanings from data. However, businesses are not certain of the minimum required skills

for their new applications. Businesses instead, are using “the hiring process to identify what skills someone at least needs for a specific job” (2020).

A visual summary of the research is shown in Table 1.

Ten papers were included in the research analysis, while one paper was a literature review for reference. Eleven papers were used in total. The topic category is a way to classify each paper by which aspect of the research it contributes to. The business and the school categories are collections of data skills that companies are demanding and skills that schools are teaching. Since the data skills are more like concepts, the tools category serves as a large collection of data science tools that were mentioned in the paper for reference. The methods category is the teaching methods that were mentioned in papers about data literacy education.

Table 1: Complete Chart of Research Findings, n = 10

| Author, Date | Topic | Business | School | Tools | Method |
|---|---|---|---|---|---|
| Overton & Kleinschmit, 2022 | Skill in School & Business, Teaching Methods | Collecting, cleaning, managing, deriving meanings from data. | | | Project-Based |
| Tsai, 2024 | Teaching Methods | | | Python, ChatGPT API | Project-Based |
| Donoghue et al., 2021 | Skill in School & Business, Teaching Methods | Experience with common languages and tools: Python, R, SQL, Jupyter Notebook, R Markdown, R-Studio | Data communication, Visualization, Storytelling, Data Ethics | data wrangling, tabular data, EDA, datavix heterogeneous data, web scraping, python standard library, integrated datasets, reidentification data cleaning, regression text data, Term Frequency-Inverse Document Frequency (TF-IDF, version control, , Python, pandas, numpy, | Project-Based |

| | | | | | |
|---|---|---|---|---|---|
| | | | | matplotlib, requests, BeautifulSoup, statsmodels, scipy, nltk, scikit-learn Github, Microsoft Azure, Google Colab, SourceTree, Scratch, API, SQL Database, JSON, CSV, XML files | |
| Kovacs et al., 2021 | Teaching Methods | | | | Project-Based |
| Deja et al., 2021 | Skill in Business, Teaching Methods | Data Collection, Data Calculation, Data Analysis, Data Communication, IT Skills, | | | Project Based |
| Dzurani n et al., 2018 | Teaching Methods | | Data analysis, Communicating data, | | Project-Based |
| Hesford et al., 2024 | Teaching Methods | | data analysis, programming, data wrangling, visualization | R, Excel, SAS, Stat, SPSS, JMP, Minitab | Project-Based |
| Chen et al., 2023 | Skills in Business, Teach Methods | Data computing, analytic, communicatio n, storytelling, | | | Project-Based |

| | | | | | |
|---|---|---|---|---|---|
| Moon et al., 2023 | Skill in School & Business | Data Collection, Data Analysis, Programming, Data Visualization | | Spreadsheet, Visual Interfaces, Scripting Languages, other interfaces: CODAP, Tuva, Tinkerplts, Fanthom, SimSalc, NetLogo, Python, R, SPSS, YouCubed, Google Colab, Tableau, API, EduBlocks, RStudio, IDE, | Project-Based |
| Disseldorp, 2020 | Skills in Business | Information Literacy, Statistical Literacy, Ability to extract meaning from data, predict trends, track customers' behavior, improve performances. | | | |

**Discussion and Limitations**

Although this research suggests that active learning methods are significantly better than traditional methods for teaching data literacy in school, these findings are not safe from criticism. When going through the literature, one notices that there is a lack of a controlled group and most of what's written so far is by people who were successful in incorporating active learning methods into their teaching. While one can argue that if success is abundant, then it must mean that teaching data literacy through active learning must be the best method. While the results seem reasonable, the methodology is still questionable. One possible reason for this could be due to the hype for the success of active learning. Because active learning has shown positive results, it gains popularity and everyone wants to write about it. Hence to improve the objectiveness in

the methodology, future researchers should develop a legitimate control group in the form of students who were taught data literature through traditional methods.

Intriguing findings that were brought up in Moon et al., in 2024 were how students prefer to use data visualization tools that offer the ability to drag and drop menus over programming language. Perhaps this is the case that drag and drop tools are simply more user-friendly. Although programming languages offer lots of flexibility, most of that flexibility is behind the documentation of packages. For example, in order to rotate the labels of an x-axis of a graph in ggplot, one needs to read through a long series of documentation in order to figure it out. Not to mention the syntax. Hence, students prefer data visualization tools that are more user friendly in the form of drag and drop menus. In addition, Moon et al., noted that programming tool use decreased as datasets got larger while analysis tools were used at a low rate, but it was a consistent rate regardless of dataset size. Perhaps this is simply due to the user-friendly experience of analysis tools. The tool is already made for the purpose of data analysis, so students can click on an option rather than typing up codes. However, it can be argued that expert data scientists do generally prefer to use programming languages because it's more flexible.

The role of data literacy at businesses at its core is to capitalize on the information that processing data can provide and use it to get ahead of competitors. One limitation of this research is that it falls victim to the ever-changing demand for business. The results are not definitive, but they can serve as a reference point for future researchers to get an idea of the data skills they at least need. In addition, remember that this research paper only looks at peer-reviewed articles accessible through the UCSB library database and Google Scholar. It doesn't interview all businesses for what they specifically need. As businesses are training their workforce to be more data literate in the next few years, their demands are subject to change. Hence, specific skills required by a company should be researched by the individuals before applying for a job there.

The results on data literacy skills taught at school are easier to generalize. Although there are discrepancies between data skills taught by each school, the research has shown that the ideas of what it means to be data literate is consistent in the literature. In addition, with the development of data science tools, this helps to keep the usage generally consistent from one school to another. The only discrepancy in tool use is in preference by the school. This is probably due to the level of expertise from teachers and the school's budgeting.

One of the issues that this paper doesn't address is worth considering is the balance between teaching other subjects and data literacy in schools. While teaching data literacy through project-based methods has been shown to be effective, it does require students to put in a significantly more amount of effort than traditional methods. Students nowadays are busier than ever with juggling multiple different classes, being involved with extracurricular activities, some

even have jobs. Since projects are meant to be comprehensive and detailed, it can eat up a lot of time that the students need to live life. One potential consequence of having a heavy project is a decrease in effort and project quality due to burn out. Perhaps schools can find more ways to incorporate data literacy into their science and math classes without having to make a separate subject for it. This could lead to maximizing the time students spend in class and maximize their learning, without burning the students out.

## References


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.

Chen, L., Gillan, J., Decker, M., Eteffa, E., Marzan, A., Thai, J., & Jewett, S. (2023). Embedding Digital Data Storytelling in Introductory Data Science Course: An Inter-Institutional Transdisciplinary Pilot Study. *Journal of Problem Based Learning in Higher Education, 11*(2), 126–152. https://doi.org/10.54337/ojs.jpblhe.v11i2.7767

Deja, M., Januszko-Szakiel, A., Korycińska, P., & Deja, P. (2021). The Impact of Basic Data Literacy Skills on Work- Related Empowerment: The Alumni Perspective. College & Research Libraries, 82, 708–729. https://doi.org/10.5860/crl.82.5.708

Disseldorp, E. M. van. (2020, July 1). *Data Literacy: Detecting Data Literacy Gaps within Businesses* [Info:eu-repo/semantics/bachelorThesis]. University of Twente. https://essay.utwente.nl/81934/

Donoghue, T., Voytek, B., & Ellis, S. E. (2021). Teaching Creative and Practical Data Science at Scale. *Journal of Statistics & Data Science Education*, *29*, S27–S39. https://doi.org/10.1080/10691898.2020.1860725

Dzuranin, A. C., Jones, J. R., & Olvera, R. M. (2018). Infusing data analytics into the accounting curriculum: A framework and insights from faculty. Journal of Accounting Education, 43, 24–39. https://doi.org/10.1016/j.jaccedu.2018.03.004

Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G., & Dey, S. (2023). A systematic literature review of data literacy education. *Journal of Business & Finance Librarianship*, *28*(2), 112–127. https://doi.org/10.1080/08963568.2023.2171552

Hesford, J. W., Pizzini, M., & Turner, M. J. (2024). Incorporating Data Analytics in Management Accounting: A Teaching Case on Cost Estimation. *Issues in Accounting Education*, *39*(2), 133–149. https://doi.org/10.2308/ISSUES-2022-064

Kovacs, P., Kuruczleki, E., Kazar, K., Liptak, L., & Racz, T. (2021). Modern teaching methods in action in statistical classes. *Statistical Journal of the IAOS*, *37*(3), 899–919. https://doi.org/10.3233/SJI-210843

Moon, P. F., Israel-Fishelson, R., Tabak, R., & Weintrop, D. (2023). The Tools Being Used to Introduce Youth to Data Science. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 150–159. https://doi.org/10.1145/3585088.3589363

Overton, M., & Kleinschmit, S. (2023). Transforming research methods education through data science literacy. *Teaching Public Administration*, *41*(2), 149–169. https://doi.org/10.1177/01447394221084488

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

Tsai, Y.-C. (2024). Empowering students through active learning in educational big data analytics. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00300-1