IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol.004, LIR305
DOI: https://doi.org/10.52731/lir.v004.305

Designing Data Science Courses to Support Non-STEM Undergraduate Students—Insights Based on Expectancy-Value Theory

Mio Tsubakimoto *

Abstract

This study employed the Expectancy-Value Theory framework to organize the educational practices of a data science minor course attended by a mixed cohort of humanities and science students. This approach aims to identify the elements that support humanities students in learning data science. The findings indicate that integrating course content with the students' major fields of study, simplifying the process of setting up technical environments, and offering prompt feedback through information and communication technology are essential.

Keywords: Learning Support, Educational Practice, Expectancy Value Theory, Motivation

1 Introduction

1.1 Data Science Education in Japanese Universities

The significance of data science education in Japanese universities is rapidly escalating due to the increasing digitization of society and the industrial sector. The Ministry of Education, Culture, Sports, Science, and Technology (MEXT) has introduced the "Mathematics, Data Science, and AI Education Program Accreditation Systems (MDASH)" to promote these disciplines [1]. This accreditation system is designed for regular university courses (excluding graduate schools but including junior colleges and colleges of technology) to deepen student interest in mathematics, data science, and AI. It systematically accredits programs that promote a basic understanding and utilization of these fields, thereby improving foundational competencies in these areas. As of August 2024, the accreditation includes 382 entries at the literacy level and 147 entries at the foundational application level, indicating that systematic and extensive data science education initiatives are being advanced primarily at the undergraduate level.

Traditionally, data science has been considered a scientific discipline, with education primarily offered in university science departments. However, the initiatives of MEXT have broadened its appeal, leading to its increased recognition among humanities students. This shift has prompted several universities to establish tailored programs for these students. Consequently, humanities students now have opportunities to learn the fundamentals of statistics, data visualization, and basic machine learning, applying these skills to analyze social science and humanities data. Upon text mining the syllabi of ten designated national universities, Tsubakimoto [2] reported that data science courses in humanities departments include computational tasks involving programming. Furthermore, in 2023, Tsubakimoto et al. [3] demonstrated that programming languages such as Python and R are being incorporated into humanities data science classes at

* Tokyo Metropolitan University, Tokyo, Japan; mio@tmu.ac.jp:

institutions accredited at the MDASH literacy level [3]. These educational initiatives equip humanities students with the necessary skills for data-driven decision-making and problem-solving in various fields, including marketing, strategic management, and policy formulation post-graduation.

Thus, the accreditation system of MEXT ensures that data science education in Japanese universities not only maintains the quality of scientific disciplines but also extends its benefits to the humanities. This is particularly advantageous for humanities students, as acquiring data science skills is essential for their significant involvement in various professions. Consequently, this initiative enhances data literacy across society and promotes the creation of new value through data.

1.2 Data Science Education at Tokyo Metropolitan University

Within Tokyo, the capital city, there are 14 national and public universities [4], of which only two are comprehensive. Tokyo Metropolitan University is one of these comprehensive institutions. It offers a "Minor in Mathematical and Data Sciences" program, designed to provide undergraduate students from all disciplines, including both the humanities and sciences, with foundational knowledge and skills in mathematical sciences and data science [5]. This program is distinguished for its inclusive approach, providing robust support and offering mathematics courses specifically tailored for humanities students to cultivate experts who will utilize data across various fields.

The minor program includes a range of data science-related techniques such as statistics, machine learning, programming, databases, and text mining. The curriculum starts with basic courses focused on data-handling techniques and advances to more specialized courses that address key data science technologies (Table 1). It culminates in practical courses aimed at developing skills for real-world problem-solving. Through this program, students gain practical experience by analyzing actual data and using programming languages such as Python and R, as well as tools such as Excel and Tableau.

The minor program offers an excellent opportunity for humanities students to enhance their data literacy, promoting data-driven decision-making and strategic thinking. Data science skills are increasingly vital across various fields, including social sciences, humanities, business, and journalism. Understanding and applying advancements in science and technology within their respective disciplines is essential for humanities students aiming for successful careers in diverse fields.

The "Minor in Mathematical and Data Sciences" program at Tokyo Metropolitan University represents a significant step for humanities students to understand the basics of data science and progress to more complex applications. This program not only makes the curriculum accessible to humanities students but also strengthens their foundation for pursuing data-driven professions. Ultimately, it prepares them to become proficient in utilizing data across various domains.

Table 1: Curriculum in DS minor program in TMU

Course Classifications	Title of Course
Foundational	Information Literacy Practice I
	Differential and Integral Calculus I or Basic Differential and Integral Calculus
	Linear Algebra I or Basic Linear Algebra
	Introduction to Discrete Mathematics
	Statistics II or Information Literacy Practice II A or Probability and Statistics
	Data Literacy as a Liberal Art
Applied	Software Engineering (Minor)
	Data Analysis (Minor)
	Databases (Minor)
	Text Mining (Minor)
	Information Literacy Practice II C
	Programming (Minor) or Information Literacy Practice II B
	Machine Learning I (Minor)
	Machine Learning II (Minor)
Practical	Data Science Project-Based Learning (PBL)

Note: All courses carry two credits each. Participation in Project-Based Learning (PBL) is necessary to complete this minor task. Enrollment in PBL requires the completion of at least 12 credits from foundational and applied subjects combined, with a recommended minimum of eight credits from applied subjects.

1.3 Purpose & Aims

This study aims to explore the specific elements of educational practice necessary for the data science education of non-STEM undergraduate students. It reflects on the practices implemented in the author's educational practice, using the framework of "Expectancy Value Theory" (Figure 1) [6]. Expectancy Value Theory explains how individuals choose actions and allocate effort based on two primary components: expectancy and value. This framework is particularly useful in the domains of teacher education and the improvement of educational methods.

- **Expectancy:** This is a learner's belief or confidence in their ability to complete a task successfully. Influenced by prior experience, perceived self-efficacy, and perceived difficulty of the task, expectancy dictates the likelihood that a learner will engage in and persist in a learning activity.

- **Value**: This reflects the subjective attractiveness or worth of the activity to the learner. This value can be subdivided into several categories, including
  - **Intrinsic value:** This refers to the enjoyment, interest, or satisfaction derived directly from an activity.
  - **Instrumental value:** The refers to utility of the activity in achieving future goals or rewards.
  - **Attainment value:** The contribution of task success to one's sense of identity.
  - **Cost:** The effort, time, or other opportunity costs associated with undertaking an activity.

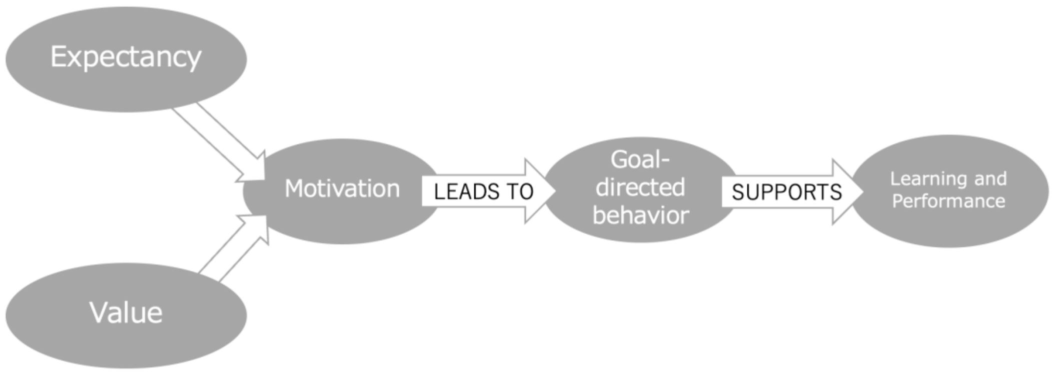

Figure 1: Expectancy Value Theory

Note: This figure was created by the author for the purpose of referencing and citing Figure 3.1. on page 70 of reference [6].

Reference [6] posits that an increase in motivation results from the interplay of three factors: moderate expectancy, high value, and a supportive, cooperative environment. This interaction is illustrated in Fig. 3.2 on page 80 of the reference.

From the perspective of learner analysis, the application of Expectancy Value Theory enables educators and trainers to gain a deeper understanding of the reasons and methods by which learners engage in educational activities. By leveraging this theory can enhance learner motivation and allow for the customization of instruction to meet individual needs. For example, if a learner has low confidence in a subject, increasing their expectations through supportive feedback and setting appropriate challenge levels can be effective. Similarly, clarifying the relevance and practicality of learning activities can enhance their perceived value to learners.

The author was appointed at Tokyo Metropolitan University during the inaugural year of the minor courses, assuming responsibilities in managing minor course operations—including course orientations, open campuses, and advising sessions—and teaching "Text Mining." In this study, the author analyzes a data science course that they personally taught and conducted.

2 Method

2.1 Subjects

The study examined educational practices in the "Text Mining" course, which was part of the Minor in Mathematical and Data Sciences at Tokyo Metropolitan University. This course was attended by eleven non-STEM undergraduate students during the academic years 2022 and 2023, who earned credits upon completion. Table 2 illustrates the characteristics of the "Text Mining" courses. The proportion of non-STEM undergraduates who successfully completed the "Text Mining" course was 25% in 2022 (5 out of 20 students, including those from science fields) and 30% in 2023 (6 out of 20 students). These students hailed from various humanities faculties, including the Faculty of Humanities and Social Sciences, the Faculty of Economics, and the Faculty of Law.

Table 2: Characteristics of the Targeted Courses "Text Mining"

	Text Mining (Minor)
Target Academic Year	Undergraduate freshmen to Graduate students
Proportion of non-STEM Students	25–30%
Compulsory or Elective	Elective
Course Content	・Natural Language Processing and Visualization Using R ・Individual Research Utilizing Text Mining Techniques
Course Offering Period	Second Semester

2.2 Procedures

The course was offered twice over two years, with no research-oriented interventions. At the onset of the course, a readiness survey was administered to gauge students' motivation. Furthermore, the course spanned 15 class sessions, during which students were required to write a "minute paper" at the end of each session. These minute papers provided text-based feedback on their understanding of the course material and their learning progress.

The course under consideration was conducted twice over a period of two years without any interventions desired for research purposes. Upon completion of the course, reflective practices were employed, utilizing various course records such as syllabi, content, and student outputs for evaluation purposes.

3 Results and Discussion

The results of the educational practices are presented in Table 3. The left side of the table lists the components of Expectancy Value Theory alongside the corresponding desirable teaching practices. The right side of the table outlines the fundamental elements of teaching data science at the university level, with elements specifically utilized in this course marked by an "x."

Table.3: Elements of Educational Practices in Expectancy Value Theory

Elements of Expectancy Value Theory	Desirable state	Elements of Educational Practices					
		ICT	TA	Students (peer)	Students (self)	Assignments	Lecture (explanation)
Expectancy	The course content is of an appropriate difficulty level, neither too hard nor too easy, described as "within reach if one jumps."	x					
	Immediate and accurate feedback is provided.	x	x				
	Explanations of effective learning methods are given.		x			x	x
Value (Intrinsic)	Satisfaction is derived from the execution of the task itself.				x		
	Connect the learning content to students' interests.				x	x	x
Value (Instrumental)	The content is useful for achieving other important goals.		x			x	x
	Demonstrate the future importance of the learning content.		x			x	x

Value (Attain-ment)	Satisfaction is derived from task accomplishment and the acquisition of knowledge and skills.				x		
Environ-ment	A cooperative environment interacts with expectancy and value.	x	x	x			

From the educational practices described, the following key aspects were deemed particularly important for reducing learning barriers and enhancing the motivation of non-STEM students in data science:

To address the perception of the subject being "too difficult":

- **Provision of programming environments:** Establish and manage a classroom RStudio Server to circumvent environmental setup issues.
- **Prompt responses to queries:** Ensured swift answers by having teaching assistants and instructors monitor the classroom, utilize chat on teams, and consistently respond to questions raised on reflection sheets in the subsequent week.
- **Encouraged peer and group teaching:** Promoted mutual teaching among peers and groups.
- **Given specific example or data through exercises and lectures:** Informed students regarding the abundance of text data for analysis in their major fields of study, such as law students classifying authors of legal rulings.

To convey the "value" of the subject:

- **Introduced relevance to students' major fields of study:** Demonstrated specific connections between their major studies and data science.
- **Provided a comprehensive experience of the data science research process:** Offered opportunities to learn data science in a broader, meta-contextual sense, extending beyond mere text mining skills and knowledge.

For creating a cooperative environment:

- **Utilized TAs, ICT, and peers continuously:** Encouraged the practice of asking questions and teaching each other, leveraging available resources and collaborative tools.

4 Limitations and Further Tasks

The primary limitation of this study is its narrow focus on a specific course. To enhance the research objectives, the following methodologies are considered essential.

- Comparing groups of liberal arts and science students within the same course.
- Analyzing the learning behaviors of specific students in other courses that are minors, preferably within the same “applied subjects” group as the course targeted in this study.
- Analyzing the relationship between the academic performance of liberal arts students and their learning behaviors based on Expectancy-Value Theory.

Future research will involve conducting interview surveys with liberal arts students enrolled in data science courses. These interviews aim to provide a more detailed description of the "Characteristics of Learning Among Liberal Arts Students in a Text Mining Course" identified in this study. Through the interviews, it was possible to describe the features identified in this study in a more dimensional manner. The findings will help clarify the general characteristics of data science education for liberal arts students and highlight important considerations for educational practices.

References

[1] Ministry of Education, Culture, Sports, Science and Technology, Japan, "Data Science and AI Education," [Online]. Available: https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm.

[2] M. Tsubakimoto, "Content Classification of Data Science Education at 10 National Universities," MJIR, vol. 11, 2022; doi: 10.50956/mjir.11.0_56_9.

[3] M Tsubakimoto, S. Hirokawa, and T. Shimbaru, “Text Mining of Data Science Education Syllabus at MDASH-selected Universities Utilizing the Cross Tabulation System," MJIR, vol. 12, 2023; doi: 10.50956/mjir.12.0_176_1.

[4] Tokyo Metropolitan Government Bureau of Statistics, "Educational Statistics 2023," [Online]. Available: https://www.toukei.metro.tokyo.lg.jp/gakkou/2023/gk23qg10000.htm.

[5] Tokyo Metropolitan University, "Data Science Program Website," 2024; dsprogram.fpark.tmu.ac.jp.

[6] S. A. Ambrose, W. M. Bridges, M. DiPietro, C. M. Lovett, and K. M. Norman, How Learning Works—Seven Research-Based Principles for Smart Teaching, Jossey-Bass, 2010. [Online]. Available: http://firstliteracy.org/wp-content/uploads/2015/07/How-Learning-Works.pdf