Development and Reflection on the Data Literacy General Education Course for Undergraduates at Wuhan University: A Postprint

Authors: Liu Xia, Fang Xiaoli, Zheng Yiping

Date: 2023-04-01T00:00:00+00:00

Abstract

[Objective/Significance] This study aims to provide a reference for peers undertaking similar initiatives by sharing the development process of a data literacy general education course at Wuhan University. [Method/Process] Through case analysis, statistical methods, and data mining visualization techniques, this paper analyzes the teaching objectives, content, distinctive features, and effectiveness of the "Data Literacy and Data Utilization" course within the "Wuhan University General Education 3.0" framework, thereby exploring the rationality of its curriculum system. [Results/Conclusion] The design of teaching methods and practical activities both inside and outside the classroom effectively stimulates student motivation and improves learning outcomes; course assessment results, pre- and post-course questionnaires, and post-class feedback demonstrate that the course significantly enhances students' data literacy.

Full Text

Construction and Reflection on Data Literacy General Education Courses for Undergraduates at Wuhan University

Liu Xia, Fang Xiaoli, Zheng Yiping Wuhan University Library, Wuhan 430072

Abstract: [Purpose/Significance] This paper aims to share the construction process of Wuhan University's data literacy general education course to provide reference for peers undertaking similar initiatives. [Method/Process] Through case analysis, statistical analysis, and data mining visualization methods, this study examines the teaching objectives, content, characteristics, and effectiveness of the "Data Literacy and Data Utilization" course under the "Wuhan General Education 3.0" framework, exploring the rationality of the curriculum system. [Result/Conclusion] The instructional methods and practical components effectively stimulate student engagement and improve learning outcomes. Course assessments, pre- and post-course questionnaires, and student feedback demonstrate significant enhancement of students' data literacy.

Keywords: data literacy; general education course; undergraduate; curriculum construction; Wuhan University

---

Preamble

In the era of big data, data profoundly impacts daily life, scientific research, and social development. In everyday life, people are both data users and intentional or unintentional data producers, making data literacy a crucial skill affecting the pace and quality of life. In scientific research, data serves as both the cornerstone and source of scientific discoveries and as the touchstone for verifying research value [?]. In terms of social demand, data analysis and application capabilities have become driving forces across all fields. Employers now need not only data analysts but also professionals with data literacy in various domains [?], making data literacy a key competency for college graduates' employment [?]. Against this backdrop, strengthening data literacy education for college students has become an industry consensus.

To deepen educational reform, Wuhan University launched the "Wuhan General Education 3.0" initiative in 2017 [?], restructuring its general education curriculum system and requiring all general education courses (except public foundation courses) to reapply for inclusion in the training program. The library applied for a brand-new course—"Data Literacy and Data Utilization"—which was approved and designated as an innovation and entrepreneurship course. Based on two semesters of implementation during the 2019-2020 academic year, this paper introduces the course's teaching objectives, content, characteristics, and effectiveness, aiming to provide a reference case for data literacy general education in Chinese university libraries.

1. Domestic and International Research and Practice Progress

1.1 International Overview Since 2006, national-level funding agencies in the United States and United Kingdom have successively sponsored numerous data literacy education research projects [?], targeting different groups including secondary school students, undergraduates, graduate students, and researchers. These projects conducted needs assessments, identified core competencies, designed curriculum systems, and carried out teaching explorations, accumulating rich theoretical achievements and best practices [?]. Subsequently, data literacy education rapidly proliferated, with university libraries becoming the primary implementing body. Various forms were adopted, including credit-bearing courses, embedded instruction, online education, training workshops, and seminars, targeting specific disciplines [?] and/or specific audiences [?]. Although syllabi and content systems varied, they generally revolved around the research data lifecycle, covering multiple or all of the following knowledge and skills: data management fundamentals, data management policies and planning, data types and formats, data naming and file organization, data description and metadata, data quality control, data preservation, security and backup, data selection and evaluation, data analysis techniques and tools, data sharing and long-term preservation, data discovery and retrieval, data citation, and data ethics and legal issues [?].

There are few specialized courses for undergraduates abroad, most adopting an embedded teaching approach to enhance undergraduate data utilization capabilities. For example, Purdue University Library embedded data literacy instruction into two required STEM undergraduate courses, teaching partial courses or assigning coursework to develop students' abilities to discover, evaluate, and acquire data [?]. UCLA integrated data literacy content into sociology information literacy courses, using two sessions to develop students' abilities to read and evaluate data tables, properly cite them, use tools to create tables, and understand charts in context [?].

1.2 Domestic Progress Domestic research can be divided into three categories: (1) Comprehensive investigations of data literacy education practices in the United States, United Kingdom, Canada, and other countries, examining teaching objects, objectives, forms, content, and evaluation [?, ?], or case studies summarizing theoretical foundations and practical experiences [?]. (2) Theoretical discussions and strategic research on data literacy education, involving goal positioning and system construction [?], educational models [?], content frameworks and curriculum systems [?], and implementation strategies [?]. (3) Investigations of domestic data literacy education practices. Zhang Xinhong [?] surveyed and compared 14 domestic and foreign universities, concluding that foreign countries have established complete data literacy education systems while domestic efforts mostly rely on information literacy education. Song Jiali et al. [?] surveyed 42 "Double First-Class" universities, finding that while various teaching and training forms were offered covering scientific data fundamentals, database usage, data processing, analysis, and application, none addressed the entire data lifecycle process, and no university offered specialized scientific data literacy courses.

In summary, data literacy education is already well-established abroad with diverse teaching systems for different levels and disciplines. Domestically, efforts remain largely at the theoretical discussion stage with few reported practical cases. Therefore, this paper aims to timely summarize the construction of the "Data Literacy and Data Utilization" course to provide a reference case for data literacy general education in Chinese university libraries.

2. Course Objectives and Content System Construction

2.1 Teaching Objectives Teaching objectives and content should fully consider the actual needs of enrolled students while maintaining forward-looking vision. To understand these needs, the course conducted a real-name questionnaire survey before class, designing over 20 questions from perspectives of data awareness, knowledge, ability, and ethics. The survey found that most undergraduates had data utilization experience across various scenarios including coursework, student activities, daily life, research projects, thesis writing, competitions, and social practice (see Figure 1). Their primary concern was how to acquire and utilize data, necessitating a focus on developing data discovery, collection, processing, and analysis capabilities. Meanwhile, undergraduates are also data producers, especially at research universities where most graduates pursue further study or overseas education [?] and will engage in scientific research to varying degrees. Therefore, while enhancing individual foundational data literacy, research-based data literacy education is also necessary [?]. In fact, the pre-course questionnaire also indicated that students generally lacked data application skills and related knowledge, such as using tools for heterogeneous dataset processing, understanding disciplinary dataset description methods, judging and cleaning data, and synchronously recording data background information.

Based on this analysis, the course objectives are positioned to help students spontaneously establish data value awareness, master approaches and methods for acquiring, analyzing, evaluating, managing, and utilizing data, develop their ability to address data issues in learning and practical activities, and lay foundations for research innovation.

2.2 Teaching Content and Schedule According to the teaching objectives, the course is structured from a data user's perspective around the data lifecycle, divided into four stages: (1) Data acquisition: involving primary and/or secondary data, including library collection data, government data, research data, and other internet data; (2) Data processing: including data evaluation and data cleaning; (3) Data analysis: including method selection and result presentation; (4) Data publication: including formal and informal release methods. The first three stages may be repeated as needed until satisfactory results are obtained before entering the data publication stage. Additionally, process management is required throughout the entire data lifecycle, including data file naming, format selection, storage and backup, real-time recording of background information or metadata description, and full-cycle quality control (see Figure 2).

Based on this lifecycle and referencing the core data literacy competencies and related skills developed by the U.S. Data Information Literacy (DIL) project [?], the course developed an 11-week, 32-credit-hour teaching schedule (see Table 1). The data fundamentals section occupies 5 credit hours, with student learning exchanges and seminars accounting for 3 credit hours. These two components are delivered across three sessions, while remaining content follows the four stages across eight teaching sessions.

3. Teaching Implementation and Characteristics

The course first opened in September 2019 and has completed two semesters, enrolling 96 students total. Student grade and discipline distributions are shown in Figure 3. 97% were first- through third-year students, with sophomores comprising the largest group (37%). Discipline distribution favored social sciences (45%), with engineering, information science, natural sciences, and humanities having similar proportions. The medical school had fewer enrollees due to campus distance. Overall, the grade and discipline distribution reflects the characteristics of a general education course.

The first semester used face-to-face instruction, while the second semester shifted to online video instruction due to COVID-19. Although delivery methods differed, the teaching philosophy and content remained consistent. After two semesters, the course demonstrates several distinctive features:

3.1 Rich Teaching Resources with Diverse Reference Materials Due to the lack of specialized textbooks, the course provides 7 reference books from the library collection covering foundational data literacy, research data management, and introductory statistics, data analysis, and data science. As shown in Table 1, each chapter includes supplementary learning materials including courseware, policy documents, standards, relevant websites, academic papers, and special reports, with key content designated as required reading. Additionally, eight practical micro-videos were produced for after-class self-study, such as data cleaning demonstrations, SPSS operation tutorials, and creating word clouds with Tableau.

3.2 Integration of Theory and Practice in Classroom Teaching In terms of time allocation, the "teaching knowledge points" in Table 1 constitute the lecture content, with teaching videos totaling 802 minutes (approximately 18 credit hours, 56% of the total). These systematically introduce relevant knowledge while combining cases to explain typical methods and tools, laying foundations for future self-directed learning. The remaining 44% of class time is dedicated to in-class reading, exercises, and discussion (see "In-class exercises/discussion assessment knowledge points" in Table 1). In-class exercises aim to consolidate learning and enhance practical skills, divided into test questions and practical tasks. Test questions check knowledge mastery through multiple-choice and short-answer formats, while practical tasks guide hands-on computer operations to complete specific assignments. Class discussions stimulate interest through personal experience sharing to identify problems and exchange experiences.

Attendance, class interaction, and assignments primarily measure student engagement and learning attitudes, with assignment quality serving as teaching reference only. The mid-term disciplinary data resource report guides students to collect professional resources using data discovery and collection methods, building resource foundations for future specialized study. The final data research report and peer evaluation comprehensively assess students' integrated data utilization, teamwork, and presentation skills. Additionally, the course introduces the "Sharing Cup" College Students' Scientific Data Resources Sharing Service Innovation Competition [?] and the "Huiyuan Sharing" University Open Data Innovation Research Competition [?], encouraging students to participate based on course learning or competition topics to enhance data understanding and application through practice.

3.3 Comprehensive Use of Teaching Assistance Tools The first semester used face-to-face instruction, utilizing the university's online teaching platform (Luojia Online for computers, Xuexitong for mobile) to share resources and organize interactions (roll call, tests, assignment submission, discussions, surveys), supplemented by a course QQ group for information dissemination and communication. The second semester used online teaching with recorded videos uploaded to the platform for pre-class self-study and review, plus weekly scheduled QQ live sessions for assignment explanations and key point discussions. Based on both semesters' experiences, using the teaching platform for face-to-face instruction enables easy transition to MOOC format by simply recording videos. Moreover, the platform accurately records student learning for teaching optimization. During online teaching, video access reached 10,655 total views, with students repeatedly watching videos (up to 6 times maximum), helping instructors identify interest points and difficulties.

3.4 Emphasis on Process Management and Comprehensive Application Ability The course uses comprehensive assessment: attendance (5%-10%), class interaction and assignments (25%-30%), mid-term disciplinary data resource report (10%-20%), and final data research report (50%), with the first three items adjusted slightly by delivery format. Attendance, interaction, and assignments measure participation and attitudes. The group-based final report, while considering individual contributions, makes it difficult to differentiate scores significantly. This may cause discrepancies between final grades and regular performance, and primarily assesses data literacy at course completion rather than learning progress. Future improvements should refine group assessment with individual and group indicators, the former evaluated by group leaders [?] and the latter through peer and instructor evaluation [?], while incorporating pre- and post-course survey results to comprehensively assess literacy improvement.

4. Teaching Effectiveness

4.1 Analysis of Learning Outcomes To fully understand knowledge mastery, we analyzed final group data research reports from the second semester (see Table 2). Regarding data sources, each group used an average of 3 different sources across 1.9 resource types, primarily library databases, government open data, and internet resources, indicating students mastered various data sources. For data cleaning and analysis, all groups used Excel, no longer viewing it as simple spreadsheet software but as a common and comprehensive data cleaning and analysis tool. Most groups also used multiple cleaning/analysis tools including OpenRefine, SPSS, and Python, demonstrating students' ability to understand and master professional analysis tools under instructor guidance.

Overall, students could select topics based on their majors or interests, collect data from various sources, choose appropriate cleaning and analysis tools, collaborate to complete data research reports, present findings, and provide objective, insightful peer evaluations, achieving excellent exchange and learning outcomes.

4.2 Student Feedback Analysis Post-course data literacy questionnaires were administered. Overall, all observation indicators improved significantly across both classes, with "disagree" responses dropping from 18% to 2% and "agree" responses rising from 45% to 82%. Improvement varied by indicator, with the fastest gains in: understanding data and lifecycle knowledge, using tools for heterogeneous dataset processing, understanding disciplinary dataset description methods, judging data quality by source, mastering data analysis and visualization tools, understanding disciplinary data sources, synchronously recording background information on data acquisition/processing, mastering basic statistical methods, reasonably interpreting analysis results, saving data files from different stages for traceability, and other areas with improvement exceeding 40%—most of which were knowledge and skills students initially lacked.

Additionally, we collected 43 final reflections written by students (individual or group) from both semesters and conducted word frequency analysis using Python. Figure 4 shows the word cloud after filtering course-related terms like "data," "data literacy," "data utilization," and "course" with a frequency threshold of 3. High-frequency terms included "data analysis," "analysis," "knowledge points," "professional knowledge," "database," "data processing," "cleaning," "excel," "importance," "tools," and "software," indicating students gained significantly in data awareness, data sources, processing, and analysis tools—aligning with positive changes in indicators like "ability to identify and express data needs," "understanding disciplinary data sources," "ability to verify data integrity/consistency and clean data," and "mastering at least one data analysis and visualization tool."

5. Summary and Reflection

The "Data Literacy and Data Practice" general education course achieved positive results in its first year but faces issues requiring continuous improvement.

5.1 Teaching Content: Balancing Methods and Skills Education As noted, enrolled students span all grades and almost all faculties. Disciplinary and grade-level differences create varying learning starting points that directly affect outcomes, particularly in data processing and analysis. Instructors focus on cultivating data thinking, systematically introducing basic principles, general methods, and implementation tools with case explanations. However, limited class time prevents in-depth tool instruction. Feedback indicates some students found it "difficult to follow" and "unfriendly to those with weak computer backgrounds," requesting "more accessible content" and "detailed software instruction," while others suggested "increasing content depth." This conflict requires ongoing exploration of solutions, potentially through tiered instruction by grade [?] or major background for certain content in small classes or groups.

5.2 Teaching Methods: Organic Integration of Face-to-Face and Online Instruction COVID-19 enabled the course to trial both face-to-face and online methods within one year. Pre- and post-course surveys showed similar learning effectiveness. Face-to-face instruction excelled in progress control, hands-on guidance, and group interaction, while online teaching offered richer content, repeatable videos, and observable learning processes beneficial for personalized learning. Future iterations should combine both strengths through MOOC-based flipped classrooms [?]. Fundamentals, data acquisition, and publication could use MOOC-based instruction, freeing instructors to focus on assignment grading and Q&A, while data processing and analysis could adopt flipped classroom formats with theory and methods learned pre-class and hands-on case work led by instructors in class, supplemented by case videos for repeated study.

5.3 Assessment Method: Combining Individual and Group Evaluation Although the course uses comprehensive assessment, university regulations require final grades to account for at least 50% of the total. The group-based final report, while considering individual contributions, makes it difficult to differentiate scores significantly, potentially causing discrepancies between final grades and regular performance. Moreover, this method primarily assesses literacy at course completion rather than learning gains. Future improvements should refine group assessment with individual indicators (evaluated by group leaders [?]) and group indicators (through peer and instructor evaluation [?]), while incorporating pre- and post-course survey results to comprehensively assess literacy improvement.

5.4 Learning Outcomes: Constructing Interdisciplinary Mixed-Group Learning Models Although overall effectiveness is positive, disciplinary differences among students create varying starting points, cognition, and final outcomes. Surveys show natural science students had the weakest data literacy foundation, followed by engineering and medical students, with humanities and social science students having the strongest foundation. After the course, literacy levels converged across disciplines, with natural science students achieving the highest level and over 50% improvement in 14 indicators, particularly in data management fundamentals and cleaning/analysis abilities. While surveys reflect self-assessment, group reports demonstrate practical application ability. Table 2 shows humanities and social science students used fewer data sources and tools than STEM students, while multidisciplinary teams generally scored higher. Therefore, against the national backdrop of accelerating interdisciplinary talent cultivation [**?**], the course should leverage its general education characteristic of student disciplinary diversity to consciously guide early formation of interdisciplinary learning teams, promoting vision broadening and learning through disciplinary integration and knowledge sharing, while cultivating interdisciplinary thinking alongside data literacy.

References

[1] Guo Huadong. Where does the clear water come from? The inaugural editorial of *China Scientific Data* [J]. *China Scientific Data*, 2016(1): 2-5.

[2] Pothier WG, Condon PB. Towards data literacy competencies: business students, workforce needs, and the role of the librarian [J/OL]. *Journal of Business & Finance Librarianship*, 2019 [2020-08-28]. https://www.tandfonline.com/doi/full/10.1080/08963568.2019.168

[3] Shorish Y. Data information literacy and undergraduates: a critical competency [J]. *College & Undergraduate Libraries*, 2015, 22(1): 97-106.

[4] Wuhan University Undergraduate School. Notice on applying for "Wuhan General Education 3.0" foundation, core, and general courses [EB/OL]. [2020-05-15]. http://gec.whu.edu.cn/info/1009/1166.htm.

[5] Meng Xiangbao, Chang E, Ye Lan. Data literacy research: origins, current status, and prospects [J]. *Journal of Library Science in China*, 2016, 42(2): 109-126.

[6] Qin J, D'Ignazio J. Lessons learned from a two-year experience in science data literacy education [C/OL] // International Association of Scientific and Technological University Libraries, 31st Annual Conference 2010 [2020-08-28]. https://docs.lib.purdue.edu/iatul2010/conf/day2/5/.

[7] Carlson J, Johnston L, Westra B, et al. Developing an approach for data management education: a report from the data information literacy project [J/OL]. *International Journal of Digital Curation*, 2013, 8(1) [2020-08-29]. http://www.ijdc.net/article/view/8.1.204/306.

[8] Carlson J, Johnston LR. Data information literacy: librarians, data, and the education of a new generation of researchers [J]. *College & Research Libraries News*, 2014, 75(8): 431-434.

[9] Lisa Johnston JJ. Stealth idea: a library instructor's guide to creating the data credibility checklist [C/OL] // Proceedings of the American Society for Information Science and Technology [2020-08-28]. https://docs.lib.purdue.edu/lib_{fsdocs}/89/.

[10] Zilinski LAS. Thinking critically about data consumption: educating students in data management skills [J]. *College & Research Libraries News*, 2014, 75(8): 431-434.

[11] ICPSR. Guide to social science data preparation and archiving [EB/OL]. [2020-08-20]. https://www.icpsr.umich.edu/web/pages/deposit/guide/index.html.

[12] New England collaborative data management curriculum [EB/OL]. [2020-08-28]. https://library.umassmed.edu/resources/necdm/.

[13] Open access post-graduate teaching materials in managing research data [EB/OL]. [2020-08-28]. https://www.icpsr.umich.edu/web/pages/deposit/guide/index.html.

[14] DATA TREE [EB/OL]. [2020-08-28]. https://datatree.org.uk/.

[15] Hu Hui, Wu Ming, Chen Xiujuan. Data literacy education in UK and US university libraries [J]. *Library and Information Service*, 2016, 60(8): 53-58.

[16] Zilinski L, Sapp Nelson MR, Van Epps AS. Developing professional skills in STEM students: data information literacy [EB/OL]. [2020-08-20]. http://dx.doi.org/10.5062/F42V2D2Z.

[17] Elizabeth Stephenson PSC. Incorporating data literacy into undergraduate information literacy programs in the social sciences: a pilot project [J]. *Reference Services Review*, 2007, 35(4): 525-540.

[18] Hu Hui, Wu Ming, Chen Xiujuan. Data literacy education models in Canadian university libraries [J]. *Library and Information Service*, 2016, 60(8): 53-58.

[19] Yu Junjie. Practice and reflection on data literacy education in foreign university libraries [D]. Taiyuan: Shanxi University, 2018.

[20] Hu Hui, Wu Ming. Research on best practices of data literacy education in foreign libraries and their implications [J]. *Modern Information*, 2016, 36(8): 66-74.

[21] Hao Yuanling, Shen Tingting, Gao Shan. Reflections and suggestions on data literacy education practice in universities—analysis based on Harvard University case and interviews with Chinese library and information professionals [J]. *Library and Information Service*, 2015, 59(12): 44-51.

[22] Zhang Qun, Liu Yumin. Research on constructing a scientific data literacy education system model for university libraries [J]. *Journal of Academic Libraries*, 2016, 34(1): 96-102.

[23] Zhang Minghai, Zhou Yanhong. Goal positioning and system construction of college student data literacy education in the big data era [J]. *Library*, 2016(10): 84-88.

[24] Zhang Qun, Liu Yumin. Research on scientific data literacy education for graduate students in university libraries [J]. *Journal of Academic Libraries*, 2017, 35(3): 93-98.

[25] Chen Mingxing. Research on the content framework of graduate student data literacy education [D]. Nanjing: Southeast University, 2017.

[26] Zhang Bin, Liu Sannüya, Liu Zhi, et al. Research on constructing a data literacy curriculum system for normal university students oriented to big data [J]. *China Distance Education*, 2018(4): 62-68.

[27] Wei Lai, Wang Siming. Research on constructing data literacy course content in Chinese university libraries [J]. *Information and Documentation Services*, 2018(6): 90-95.

[28] Jiao Jiao. Research on data literacy education in university libraries in the big data era [J]. *Library Work and Study*, 2020(3): 81-84.

[29] Guan Jin, Chen Wenkai, Ou Chunmei. Exploration of literature retrieval course teaching models based on data literacy education [J]. *Information Research*, 2017(9): 48-51.

[30] Zhang Xinhong. Comparative analysis of scientific data literacy education in domestic and foreign university libraries [J]. *Library Science Research*, 2018(11): 9-13.

[31] Song Jiali, Cheng Jiejing. Investigation and suggestions on scientific data literacy education in university libraries [J]. *Library Science Research*, 2018(19): 2-9.

[32] Song Jiali. Research on scientific data literacy education in university libraries [D]. Qufu: Qufu Normal University, 2018.

[33] Wuhan University. Wuhan University 2019 graduate employment quality report [EB/OL]. [2020-05-16]. https://info.whu.edu.cn/info/1781/184351.htm.

[34] Zhang Juan, Yang Wenjian. Research on foreign data literacy education and its implications [J]. *Digital Library Forum*, 2020(4): 60-67.

[35] "Sharing Cup" College Students' Scientific Data Resources Sharing Service Innovation Competition [EB/OL]. [2020-05-31]. http://share.escience.net.cn/nav/index.

[36] "Huiyuan Sharing: Data Reading" 2nd University Open Data Innovation Research Competition [EB/OL]. [2020-05-20]. http://www.huiyuan.sh.edu.cn/competition/#/home.

[37] Song Haiyan, Yu Xiaowei, Pan Wei. Practice exploration and reflection on university information literacy general education courses [J]. *Library Science Research*, 2014(7): 11-14.

[38] Wen Na, Lü Haixia, Li Baoming. Exploration of flipped classroom teaching models for university general education courses based on MOOCs [J]. *Education Teaching Forum*, 2019(31): 93-95.

[39] Xu Juxiang. Design and practice of flipped classrooms for information literacy general education courses [J]. *Library and Information Knowledge*, 2018(4): 119-128.

[40] Wang Zhikang. Interdisciplinary studies will become China's 14th discipline category [N]. *China Science Daily*, 2020-08-11(6).

Author Contributions: Liu Xia: Data analysis, paper writing and revision Fang Xiaoli: Data analysis, paper revision Zheng Yiping: Questionnaire design, results processing

Note: Figure translations are in progress. See original paper for figures.