Journal of Contemporary Educational Research, 2026, Volume 10, Issue 3
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

Evidence-Based Insights from a Bilingual Big Data Analytics Course: A Comparative Study of Domestic and International Students in Transportation Engineering

Rongsheng Chen, Shasha Liu*, Sida Luo

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

**Author to whom correspondence should be addressed.*

Abstract: As transportation engineering education increasingly relies on data-driven methods, students are expected to develop competencies in programming, analytical reasoning, and interpretation of heterogeneous traffic datasets. This study examines learning outcomes and student experiences in a dual-language, dual-cohort course on big data analytics at Beijing Jiaotong University, delivered as a Chinese-taught class for domestic undergraduates (CN) and an English-medium instruction class for international students (EN) with largely identical curricular content. The analysis integrates three years of CN performance records (2023–2025) and two years of EN performance records (2024–2025), supplemented by a pre-course survey of the EN cohort on demographics, prior preparation, and learning motivation. Results show that the CN cohort achieved a slightly higher median score but with substantially greater variance, whereas EN scores were more tightly clustered with no failures among active students. Survey evidence indicates that most EN students reported limited programming and analytics foundations but strong career-oriented motivation. Correlation patterns further suggest that project-based assessment and collaborative tasks helped translate assignment effort into stable final outcomes in the EN class. Based on these findings, the study discusses practical implications for bilingual technical courses, including early-stage code scaffolding, motivation-sensitive project design, and process-oriented assessment to support learners with diverse backgrounds.

Keywords: Bilingual education; Big Data Analytics; Transportation engineering; Learning outcomes; Pedagogical strategies

Online publication: April 8, 2026

1. Introduction

With the rapid digitalization and intelligence-driven transformation of the transportation sector, traffic data sources have expanded beyond traditional loop detectors and camera-based monitoring to include richer datasets such as vehicle trajectories and metro smart-card records. The diversity, volume, and heterogeneity of these data introduce new challenges for data processing and analytical reasoning, while also requiring transportation engineering students to develop stronger competencies in data analytics, modeling, and interpretation.

At Beijing Jiaotong University, the course Big Data Analytics and Applications in Comprehensive Traffic Systems is designed to address these needs. It systematically introduces data acquisition technologies, the characteristics of multimodal traffic data, analytical workflows, and hands-on applications using real-world transportation datasets. The course aims to strengthen foundational analytical skills while cultivating both engineering-oriented problem-solving capabilities and research readiness.

This course is offered in two parallel versions each academic year: a Chinese-taught class for undergraduate students in the fall semester, and an English-taught class for international students in the spring semester. While the instructional content remains largely identical, the two cohorts differ substantially in academic background, prior technical preparation, and cultural learning norms. This creates what we call a "dual-language, dual-cohort" teaching context. For international students, this course also functions as an English-Medium Instruction (EMI) course, introducing additional linguistic and cross-cultural learning challenges.

This study conducts a comparative analysis of the learning experiences and performance between these two groups. The contributions are three-fold: first, it constructs a quantifiable cognitive profile of international students based on systematic survey data; second, it explores performance disparities under identical curriculum content but different instructional languages; and finally, it proposes targeted pedagogical reforms for bilingual engineering courses at both undergraduate and postgraduate levels.

2. Literature review

Traffic big data courses have increasingly gained traction across universities, driven by the growing availability of multimodal data and the need for data-driven decision-making in transportation engineering, with the main objectives of improving students' data processing capabilities, understanding of transportation systems, and interdisciplinary analysis capabilities. Although few studies focus specifically on traffic data analysis courses, relevant insights can be obtained from three related domains: the evolution of data science education, the pedagogical dynamics of EMI courses, and factors influencing learning outcomes such as prerequisite knowledge, self-efficacy, and motivation.

2.1. Evolution of data science education

With the digital and intelligent transformation of engineering disciplines, university courses are increasingly emphasizing data-driven thinking and reproducible data analysis workflows. Educational research in data science suggests that curriculum design should cover the entire workflow from data acquisition, management, and cleaning, to modeling, evaluation, visualization, and even result reproducibility, rather than merely focusing on statistical models or single technical methods [1]. This has also promoted the evolution of data science-related courses from traditional statistical software operations (such as SPSS) to scripted and reproducible programming ecosystems (such as Python/R). In this context, data science is regarded as an interdisciplinary integration of statistics, computer science, and other fields. It not only requires mastering data analysis models and methods but also the ability to model, analyze, evaluate, and interpret based on real engineering problems. However, the difficulty lies in the fact that students' mathematical and programming foundations significantly affect their depth of understanding of models, thereby leading to differences in learning experience and learning outcomes [2].

2.2. Challenges of English-Medium Instruction (EMI)/bilingual courses

English-Medium Instruction has developed rapidly in universities in non-English-speaking countries. It not only serves international student training but also is used to improve students' academic English and cross-cultural collaboration skills. Existing studies indicate that the impact of EMI on learning outcomes is not a one-way gain but is more dependent on conditions such as curriculum design, language support, evaluation methods, and students' prerequisite backgrounds. Especially in professional courses, language and discipline concept learning often compete for cognitive resources, which may introduce additional learning burdens and thus amplify differences among students [3]. Based on this rule, students without a statistical background or with weak programming skills are more likely to encounter both difficulties of learning content and language in the EMI environment.

In the context of most Chinese universities, the teaching effectiveness of EMI courses is also affected by factors such as academic culture and classroom interaction habits [4,5]. Some studies point out that in EMI classrooms, teachers often need more explicit, structured explanations, more visual expressions, and more detailed curriculum frameworks to reduce the threshold for concept learning and maintain classroom participation. At the same time, students' preferences for classroom activities and course assessment methods vary significantly with cultural and educational experiences (e.g., differences in adaptability to discussions, questioning, and collaboration) [6]. Therefore, for non-native English speakers, classrooms taught in their mother tongue will have different teaching effects from those taught in English.

2.3. Impact of prerequisite background, self-efficacy, and learning motivation on learning outcomes

In learning science and higher education research, self-efficacy has been repeatedly proven to be significantly correlated with learning engagement, perseverance, and academic performance. Relevant studies show that there is a stable moderate correlation between self-efficacy and academic performance, which may work through mechanisms such as learning strategies, persistence, and emotional regulation [7]. Furthermore, educational research related to programming clearly points out that programming self-efficacy and learners' mental models will affect their learning paths and problem-solving behaviors. In the initial learning stage, in-sufficient self-efficacy may lead to avoiding debugging and challenging tasks, thereby reducing opportunities to improve skills [8].

At the same time, learning motivation also significantly affects students' level of engagement in the course [9]. Engineering education and higher education research generally distinguish between motivation types such as career-driven, academic-driven, and interest-driven. Different motivation structures will affect students' preferences for homework types (practical vs. theoretical), project themes (engineering application vs. academic research), and evaluation methods (process-oriented evaluation vs. result-oriented evaluation) [10]. For international student groups, motivation is often more closely linked to their goals of the enrolled program, career plans, and original professional backgrounds. Therefore, the result related to international students can show more diverse learning motivation structures compared to domestic students, which means that a unified teaching organization method with consistent difficulty and rhythm may be difficult to balance fairness and effectiveness [6].

Existing research generally emphasizes the significant impact of language factors, differences in disciplinary background, and learning motivation on the learning outcomes of bilingual (or EMI) courses. However, there are relatively few studies that systematically compare the learning backgrounds and academic performance of two types of students when the same course is offered in different language classes. Based on the actual data of the Chinese and English classes of this course, including pre-course questionnaires, classroom observations, and academic performance, this teaching research will conduct a comparative analysis to reveal the learning characteristics and difference trends of the two types of students.

3. Course statistics and comparative analysis

This section introduces the basic student information and comparative learning outcomes of the bilingual big data analytics course. The data are from two distinct cohorts taking the same core curriculum: the Chinese-taught cohort (CN) consisting of domestic undergraduate students, and the English-taught cohort (EN) consisting of international students. Both cohorts utilized the same syllabus, covering Python programming basics, data visualization, and statistical or machine learning applications in transportation. However, the assessment structure varied slightly to accommodate the different semester lengths and student backgrounds. The final course score is a composite of attendance, individual programming assignments, a group course project, and a final written exam (for CN students) or a comprehensive final assessment (for EN students).

For the Chinese cohort, the data encompasses three academic years (2023–2025), with a total of 98 students. For the English cohort, data from two academic years (2024–2025) is available, with 50 students in total. Additionally, a pre-course survey was administered to the EN cohort to capture their demographic information, prior programming experience, data analysis knowledge, and learning motivations.

3.1. Participants' profile of the EN cohort

Table 1 summarizes the basic information of the English-taught cohort, and **Figure 1** visualizes the major distributions. The survey reveals a highly diverse cohort in terms of nationality, but a consistent maturity in age (95% are over 26). Most students come from developing countries in Africa (for example, Ethiopia, Rwanda, etc.) and Southeast Asia (Bangladesh, Laos, etc.), reflecting the university's international recruitment focus. 78% of international students rated their programming skills at level 1 or 2, where level 1 represents no experience in programming, and none considered themselves proficient (level 5). Similar trends are observed in data analysis knowledge, with 65% rating themselves at level 1 or 2. The most commonly used tools prior to the course were Excel (57%) and Python (52%), indicating a basic familiarity with data manipulation but limited exposure to advanced programming environments. Learning motivations are primarily career-oriented, with 91% aiming to understand transportation data systems and 74% seeking to improve technical skills. Only 22% cited personal interest as a motivation. This suggests a significant gap between their initial capabilities and the technical demands of a big data analytics course, necessitating a scaffolded instructional approach.

In **Figure 1**, the demographic and background characteristics of the EN cohort are visualized through bar charts. The nationality distribution highlights the diverse origins of the students, with Ethiopia, Rwanda, and Tanzania being the most represented countries. The gender distribution shows a predominance of male students (74%). For Chinese cohorts, similar survey data were not systematically collected; however, classroom observations suggest that domestic students generally possess stronger programming foundations and are more accustomed to exam-oriented assessments. The age distribution is also younger, with most students being senior students (ages 18–22), with 67% students being male.

Table 1. Background characteristics of international students in the English-taught class

Category	Group / Range	Percentage (%)
Nationality	Ethiopia	17%
	Rwanda	13%
	Tanzania	13%
	Myanmar	9%
	Uganda	9%
	Other (e.g., Brazil, Kenya, Fiji, Laos, Liberia, Bangladesh, Namibia)	39%
Gender	Male	74%
	Female	26%
Age group	26–30 years	43%
	31–35 years	48%
	36–40 years	4%
	≥41 years	4%
Programming skills	1 (No experience)	35%
	2	43%
	3	13%
	4	9%
	5 (Very proficient)	0%
Data analysis knowledge	1 (Low)	17%
	2	48%
	3	35%
	4–5	0%
Tools used	Excel	57%
	Python	52%
	SPSS	22%
	MATLAB	4%
	R	4%
Motivation	Understand transportation data systems	91%
	Improve technical/analytical skills	74%
	Prepare for thesis/research	61%
	Enhance job competitiveness	39%
	Personal interest	22%
	Other	4%

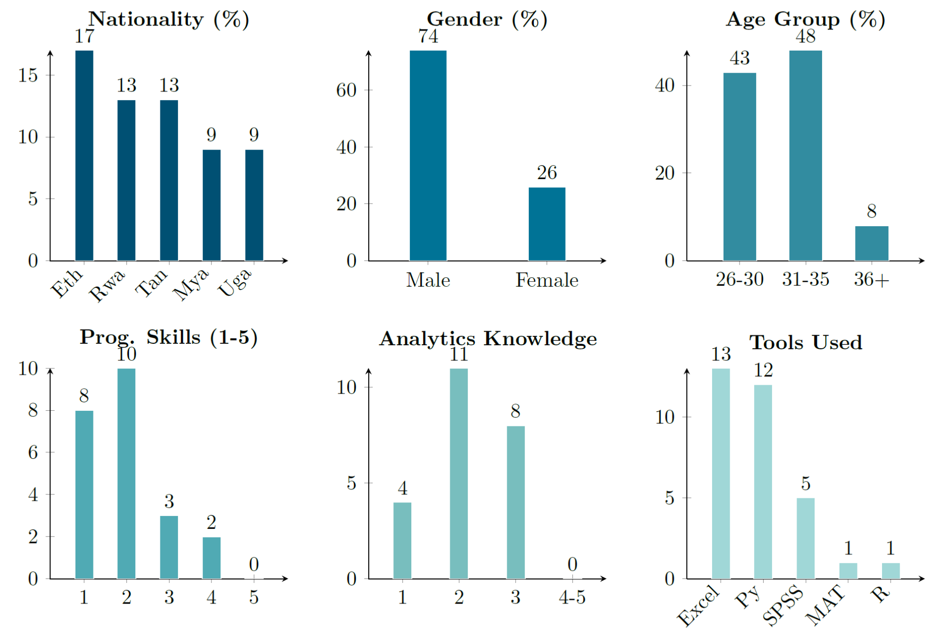

Figure 1. Profile of the English-taught cohort (EN) based on pre-course survey responses

3.2. Comparative analysis of learning outcomes

This section analyzes the academic performance of students from the Chinese-taught (CN, 2023–2025 cohorts) and English-taught (EN, 2024–2025 cohort) classes. The refined dataset comprises 98 CN students and 50 EN students.

Figure 2 presents a boxplot comparison of the final course scores between the CN and EN cohorts. As previously mentioned, the final score of CN students includes a written exam component, while the EN students' final assessment is project-based. To ensure a fair comparison, an alternative final score is computed that effectively approximates a project-only assessment similar to the EN cohort (labeled as "CN (Alt)" in the figure). The "CN (Alt)" scores were calculated by excluding the written exam component and re-weighting the remaining components (attendance, individual assignments, group project) to sum to 100%. As the final exam usually lowers the overall score for weaker students, removing it results in a higher minimum score and median for the "CN (Alt)" group.

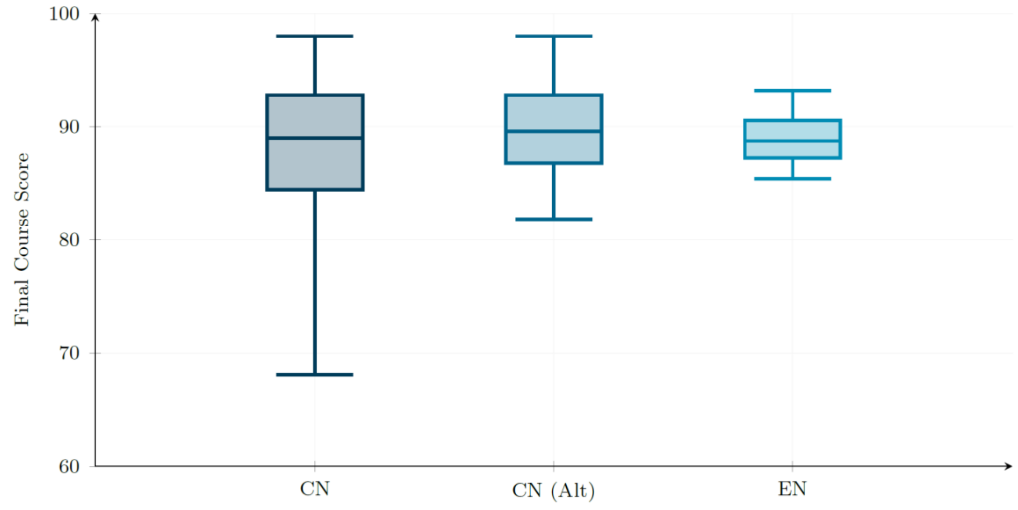

Figure 2. Comparison of final score distributions

It can be observed that the CN cohort displays a higher median score (≈ 91) compared to the EN cohort (≈ 88), but also significantly higher variance. The CN scores span from 70 to 99, whereas the EN scores are tightly clustered between 85 and 93. Even after adjusting for assessment structure (CN Alt), the CN cohort still shows a wider spread. This indicates that while some CN students excel, a substantial portion struggle. Conversely, EN students, despite their self-reported lower programming confidence and the language barrier, achieved a consistent baseline of competence, with no failures recorded in the active cohort. This may suggest the effectiveness of the project-based curriculum design, which emphasizes practical application and peer collaboration over traditional exam performance. It also indicates that language challenges and the background gap in programming or data analysis may have a modest impact on performance, but not to the extent of causing widespread failure. The good outcome of the EN cohort may also be attributed to their strong learning motivation, as 91% of them aimed to understand transportation data systems, which likely drove consistent engagement with the course material, as well as the well-prepared teaching strategies that scaffolded programming skills development.

3.3. Score component analysis: Assignments vs. projects

To further understand the learning dynamics of CN and EN students and the relationship between individual effort and final outcomes, the correlation between individual assignment scores and final course scores for both cohorts is analyzed. **Figure 3** presents scatter plots for each cohort.

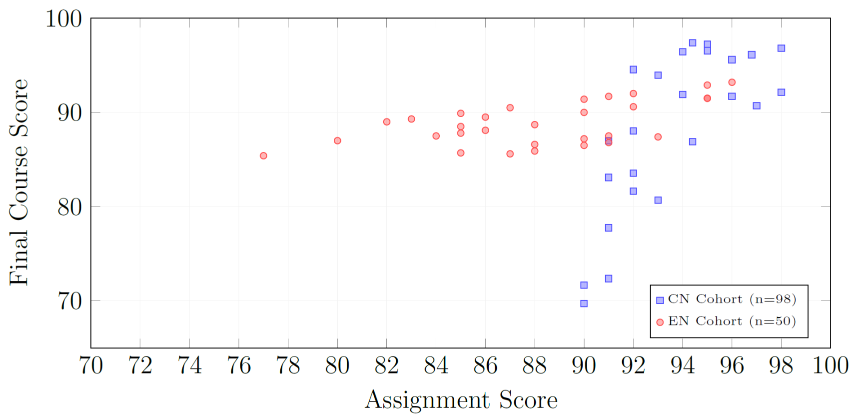

Figure 3. Correlation between assignment scores and final course score

As shown in **Figure 3**, the blue dots represent CN students, while the red dots represent EN students. Both cohorts exhibit a positive correlation between individual assignment scores and final course scores, indicating that consistent effort in assignments generally leads to better overall performance. However, the CN cohort shows a wider dispersion, with several students achieving high assignment scores but relatively lower final scores, likely due to the written exam component. In contrast, the EN cohort's scores are more tightly clustered, suggesting that their project-based assessment allowed for a more consistent translation of assignment effort into final outcomes. For EN students, it can be observed that several students with lower individual assignment scores (e.g., 77–83 range) still managed to achieve respectable final scores (≥ 85). This indicates that the collaborative learning component (group projects) acted as a buffer, allowing students with weaker coding skills to contribute in other ways (e.g., data interpretation, reporting) or learn from peers, thus elevating their overall performance, which can be applied to classes with diverse backgrounds.

4. Pedagogical implications

Based on the comparative analysis of student information and learning outcomes, the following strategies are proposed for bilingual engineering education.

4.1. Bridging the technical gap with scaffolding

The survey indicated that nearly 80% of international students lacked proficiency in programming (Python) and advanced statistics. However, their consistent final grades suggest they can overcome this barrier with appropriate support. Instructors should implement "Code Scaffolding," providing template scripts (e.g., Jupyter Notebooks with "fill-in-the-blank" code blocks) in the early stages. This reduces the cognitive load of syntax errors and allows students to focus on data logic, which is crucial for the 57% of students who were previously only familiar with Excel.

4.2. Differentiation for diverse motivations

The CN cohort showed high variance in performance, while the EN cohort was highly motivated by "understanding systems" (91%) and "job competitiveness" (39%). For CN students, who may be more exam-oriented, introducing open-ended challenges can stimulate the top performers. For international students, linking curriculum content to real cases in their home regions (e.g., traffic data from developing countries) can leverage their maturity and professional experience, transforming the "mixed background" from a liability into a resource for cross-cultural discussion.

4.3. Refining assessment for bilingual contexts

The high completion rate of the EN cohort validates the use of project-based assessment over traditional exams for EMI courses. Language barriers often mask technical understanding in timed, written exams. By shifting weight to Process-Oriented Assessments (weekly coding assignments + semester projects), educators can obtain a more accurate measure of an international student's analytical capability, as evidenced by the correlation between assignment effort and final success.

5. Conclusion

This study compared the learning outcomes of domestic and international students in the course of "Big Data Analytics and Applications in Comprehensive Traffic Systems." By integrating quantitative grade data with qualitative background surveys, we found that despite significant gaps in prior programming knowledge, international students achieved consistent and satisfactory learning outcomes comparable to their domestic peers. The success of the English-taught cohort is attributed to a project-driven curriculum that mitigates language barriers and encourages peer collaboration.

Future improvement can focus on developing modular, tiered programming resources to support novices and designing culturally inclusive data cases. Additionally, assessment strategies should prioritize process-oriented evaluations to better capture analytical skills over rote memorization. This research contributes to the understanding of bilingual education in engineering disciplines and offers practical recommendations for educators aiming to improve student outcomes in diverse learning environments.

Acknowledgments

The authors would like to thank the Graduate School, the College of International Education, and the School of Traffic and Transportation for providing the research environment and administrative support, as well as all students who participated in the course evaluations and data collection.

Funding

This research was supported by the Graduate Education Reform Project of Beijing Jiaotong University: "Construction of the Online Open Course: Big Data Analytics and Applications in Comprehensive Traffic Systems."

Disclosure statement

The authors declare no conflict of interest.

References

[1] Baumer B, 2015, A Data Science Course for Undergraduates: Thinking with Data. The American Statistician, 69(4): 334–342.

[2] Shuntaro I, 2025, Practice and Evaluation of Data Science Education for First-Year Students in Foreign Language Faculties. IIAI Letters on Institutional Research, 5: 1.

[3] Macaro E, Curle S, Pun J, et al., 2018, A Systematic Review of English Medium Instruction in Higher Education. Language Teaching, 51(1): 36–76.

[4] Liao SJ, Zhang LJ, May S, 2025, English-Medium Instruction (EMI) Language Policy and Implementation in China's Higher Education System: Growth, Challenges, Opportunities, Solutions, and Future Directions. Current Issues in Language Planning, 1–20.

[5] Jiang L, Zhang LJ, May S, 2019, Implementing English-Medium Instruction (EMI) in China: Teachers' Practices and Perceptions, and Students' Learning Motivation and Needs. International Journal of Bilingual Education and Bilingualism, 22(2): 107–119.

[6] Wu W, Hammond M, 2011, Challenges of University Adjustment in the UK: A Study of East Asian Master's Degree Students. Journal of Further and Higher Education, 35(3): 423–438.

[7] Honicke T, Broadbent J, 2016, The Influence of Academic Self-Efficacy on Academic Performance: A Systematic Review. Educational Research Review, 17: 63–84.

[8] Ramalingam V, LaBelle D, Wiedenbeck S, 2004, Self-Efficacy and Mental Models in Learning to Program, Proceedings of the 9th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, 171–175.

[9] Pintrich PR, 2003, A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. Journal of Educational Psychology, 95(4): 667.

[10] Matusovich HM, Streveler RA, Miller RL, 2010, Why Do Students Choose Engineering? A Qualitative, Longitudinal Investigation of Students' Motivational Values. Journal of Engineering Education, 99(4): 289–303.