Investigating Data Science Student Confidence and Competency in Kuwait

Dr. Zahraa Marafie, Kuwait University
Safia Malallah, Kansas State University

Dr. Safia Malallah is a teaching assistant professor at Kansas State University, where she completed her Ph.D. in Computer Science and early childhood. Her research is dedicated to advancing computer science and data science education across the PreK-12 and undergraduate levels. Dr. Malallah is particularly passionate about designing innovative and accessible learning experiences that cultivate essential computational skills in students

Dr. Ejiro U Osiobe

.

Lior Shamir, Kansas State University

Associate professor of computer science at Kansas State University.

Jorge L Valenzuela, Kansas State University

Dr. Jorge L Valenzuela earned a bachelor's degree in computer engineering from Monterrey Tech, Mexico, in 1990. He worked in the industry in Mexico developing embedded systems before coming to Kansas State University (K-State) in 1998 to pursue a master's degree in software engineering. After completing his master's, he worked for Motorola as a software engineer from 2000 to 2003. Dr. Valenzuela returned to K-State in 2004, earning a doctorate in computer science in 2014. He later worked as a systems programmer in the College of Education and served as an adjunct instructor in the Department of Computer Science Department, where he currently holds a Teaching Assistant Professor position.

Friday Emmanuel James, Kansas State University

Friday James is a Teaching Assistant Professor at Kansas State University. He has a double-majored Bachelor's degree in Statistics/Computer Science from University of Agriculture, Makurdi - Nigeria. He got a Master's degree in Statistics and a Master's degree in Computer Science from University of Ilorin - Nigeria and Kansas State University - Kansas USA in 2015 and 2021 respectively. Following this, he bagged a PhD in Computer Science from Kansas State University in 2025 and joined as a teaching track Assistant Professor. His research interest cuts across the use of machine learning and data science in Computing Science Education to improve teaching and learning.

Investigating Data Science Student Confidence and Competency in Kuwait

Zahraa Marafie, Kuwait University, Zahraa.Marafie@ku.edu.kw
Safia A. Malallah, Kansas State University, safia@ksu.edu
Ejiro Osiobe, dr.o@aneosiobe.org
Lior Shamir, Kansas State University, lshamir@ksu.edu
Jorge Valenzuela, Kansas State University, jvalenzu@k-state.edu
Friday E. James, Kansas State University, fejames@k-state.edu

Abstract:

Data science has rapidly become a critical global field, yet little research has examined students' data science awareness and self-efficacy in Kuwait—especially before the introduction of formal data science and AI programs in 2025. This study investigates students' motivation, background knowledge, and confidence in performing data science tasks across high schools, community colleges, and universities in Kuwait. Using a bilingual, extended version of the validated Data Science Self-Efficacy and Awareness Instrument, the study examined social–cultural influences, prior coursework exposure, familiarity with tools and concepts, and self-efficacy across the data science lifecycle. Findings reveal a gap between students' strong motivation toward STEM and their limited structured background knowledge in data science. Significant differences were observed across educational levels and institutions, with self-efficacy increasing as students progress academically. The results provide evidence-based guidance for stakeholders to design clearer data science learning pathways that better prepare students for emerging data science and AI careers in Kuwait.

Introduction

Data science and artificial intelligence have rapidly become foundational disciplines that shape modern economies, research, and workforce demands [1]. Globally, educational systems are adapting to prepare students with the knowledge and skills required to participate in data-driven decision making and AI-enabled innovation [2]. However, in many regions, including Kuwait, formal educational pathways in data science have only recently begun to emerge. Until 2025, Kuwait did not offer dedicated undergraduate degree programs in data science or artificial intelligence [3]. Instead, data science concepts were scattered across courses in statistics, programming, computer science, and business analytics without a cohesive curricular structure. As a result, students' exposure to data science has historically depended on institutional initiatives rather than a unified national curriculum. Understanding how students perceive their awareness, background knowledge, and confidence in data science is therefore critical at this transitional stage of educational development. In addition, social and cultural influences—such as family expectations, school environment, societal perceptions of STEM, and gender views—may shape how students engage with data science and STEM pathways [4]. Examining these factors provides valuable insight into how educational systems can better support students as Kuwait expands its data science and AI offerings.

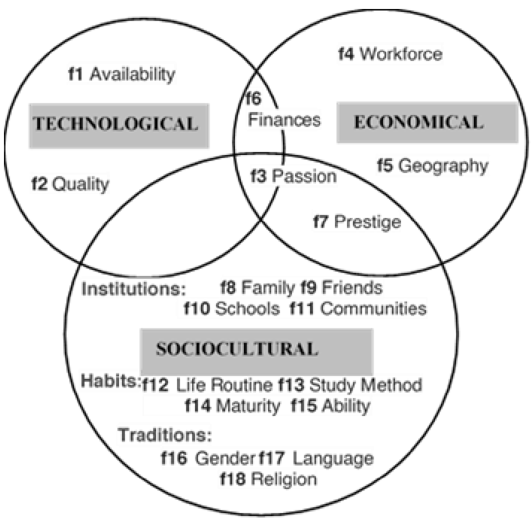
This study investigates data science awareness, background knowledge, and self-efficacy among students across multiple educational levels and institutions in Kuwait, including high schools, community colleges, public and private universities, and specialized institutes. Using a bilingual, validated instrument expanded to include social–cultural dimensions, the study captures students' prior exposure to data science topics, their familiarity with tools and concepts, and their confidence in performing data science tasks across the full data science lifecycle. The findings reveal several important patterns. First, students across all educational levels demonstrate high motivation toward STEM, primarily driven by passion and future career prospects rather than external pressures such as social media or perceptions of ease. Second, while motivation is consistently high, students' background knowledge in data science remains generally low, particularly among high school and bachelor's students. Significant differences appear across institutions, where universities that have integrated data science content earlier report stronger student background knowledge and familiarity with machine learning concepts. Third, data science self-efficacy increases significantly with educational progression, with master's and doctoral students reporting the highest confidence and high school students reporting the lowest. Institutional differences also influence self-efficacy, with students from institutions that emphasize data science reporting higher confidence levels.

These patterns highlight a critical gap between students' strong interest in STEM and their limited structured exposure to data science. This gap reflects the historical absence of formal data science programs in Kuwait prior to 2025 [3] and underscores the need for coordinated educational planning as these programs begin to expand. Based on these findings, this study not only describes current student perceptions but also proposes evidence-based guidelines for stakeholders in Kuwait to support the systematic development of data science education across educational stages. By understanding where students currently stand in terms of awareness, knowledge, and confidence, stakeholders can design targeted interventions that build coherent pathways for data science learning from high school through higher education.

Background and Related Work

Factors Influencing STEM Programs in GCC Countries

The educational landscape of the Gulf Cooperation Council (GCC) is shaped by a unique interplay of shared demographics, cultural traditions, and rapid technological growth. To systematically analyze these influences, Malallah's study utilizes an adapted version of Veen's three-factor framework, which examines the intersection of Technological, Economical, and Sociocultural domains (Figure 1). While technological availability and economic factors such as workforce demands are critical, this research places particular emphasis on the Sociocultural dimension, which is further deconstructed into specific institutions (family, schools), habits (study routines, maturity), and traditions (gender dynamics, religious regulations). This granular mapping of regional social barriers served as the structural foundation for designing our assessment, allowing us to rigorously evaluate and address the intricate social aspects that uniquely impact STEM learning and student engagement in the region [5] [6].



F1	**TECHNOLOGICAL:** the ways new practices and equipment can affect education with respect to technological capabilities	
f1	Availability	"The opportunity, at a micro or a macro level, to materially access technology at reasonable prices, whether at home, at work, at school or in public places" (such as public institutions or commercial outlets) [284]
f2	Quality	How good and convenient technology is in terms of efficiency and durability with 72 technology qualities, including aesthetically pleasing, functional shape (form), accessible, culture fit [285]
f3	Passion	How much individuals are emotionally connected to technology to facilitate their learning processes for STEM subjects
F2	**ECONOMICAL**: the effect of a country's financial status on students' education and technology	
f4	Workforce	How available jobs influence individuals to pursue or reject STEM education
f5	Geography	How the environment and physical location affect individuals' perceptions of STEM
f6	Finances	How the economic status of a country and individuals affect STEM and individuals
f7	Prestige	Having the latest trend and showing off had affected STEM and individuals
F3	**SOCIOCULTURAL**: the ways habits, traditions, and beliefs consciously or unconsciously reflect a majority of society groups	
F3.1	**Institutions**	Group of people who come together for a common purpose
f8	Family	Certain life situations between common ancestors that help shape preferences
f9	Friends	Relationship of mutual affection between people that helps shape preferences
f10	Schools	Physical, human, materials, and resources used by schools to promote academic performance
f11	Communities	Social gatherings with mutual interests or goals, including institutional and official/no-official clubs or gatherings
F3.2	**Habits**	Recurring actions of individuals
f12	Life Routine	Lifestyles and daily routines that affect individuals' preferences
f13	Study Method	Learning and processing information styles that affect individuals' preferences
f14	Maturity	Quality or state of being fully mentally and physically developed, reflecting judgment and wisdom
f15	Ability	Current possession of a skill
F3.3	**Traditions**	"A belief or behavior (folk custom) passed down within a group or society with symbolic meaning or special significance with origins in the past" [286]
f16	Gender	Gender role, or the outward manifestations of personality that reflect gender identity [287]
f17	Language	System of communication used by a particular country or community [288]
f18	Religion	Belief in and worship of a supernatural power and its influence on individuals' preferences

Figure 1. Factors that Influence the Educational System [5]

Kuwaiti Education System Regulations
Kuwait's education system combines substantial economic resources with a centralized, gender-segregated structure managed by the Ministry of Education (MOE). The curriculum progresses through kindergarten, primary, middle, and secondary levels. The high school stage serves as a critical gatekeeper: students must select either a "Science" or "Literature" track, a decision that rigidly dictates future university eligibility. While Science majors retain broad access, Literature majors are effectively barred from STEM programs unless they repeat high school entirely. Despite these structural constraints, the state heavily subsidizes higher education, offering full scholarships for both university degrees and vocational training [4].

The Data Science Self-Efficacy Awareness Scale
The instrument is a comprehensive assessment designed by Malallah to evaluate learner confidence across the multidisciplinary spectrum of Computer Science, Business, Statistics, and Mathematics. Serving as the structural foundation for this study, the instrument comprises 48 items that assess 13 distinct aspects of proficiency by explicitly mapping six core competency domains against the seven sequential stages of the Data Science Life Cycle (Figure X). These stages include: (1) Domain Knowledge and Research Design, (2) Data Planning and Collection, (3) Data Cleaning, Wrangling, and Feature Engineering, (4) Feature Selection, (5) Model Design, (6) Model Evaluation, and (7) Communication and Action Proposals. For the specific purposes of this research, the original scale was adapted to ensure linguistic and cultural relevance for the Kuwaiti student population. Responses were measured using a 5-point Likert scale, interpreted through a specialized rubric that categorizes mean scores into Low (1.0–2.9), Moderate (3.0–3.6), and High (3.7–5.0) confidence levels, thereby allowing for a granular diagnosis of specific gaps in student readiness [7].

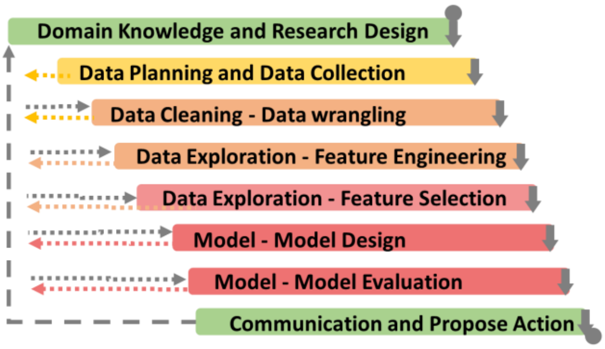


Figure 2. Factors [7]

Method
Design
This study employed a cross-sectional quantitative survey design to investigate data science awareness, competency, and self-efficacy across diverse educational populations in Kuwait.

Instrument
The study utilized an existing validated data science self-efficacy and awareness instrument design by Malallah [7], which was extended to include a social dimension examining peer, community, and environmental influences on data science perceptions. The survey was translated from English into Arabic and underwent a back-translation process to ensure accuracy and consistency in meaning. Language and subject-matter experts reviewed both versions for clarity, cultural appropriateness, and conceptual equivalence. The final

instrument was presented in a dual-language format (English and Arabic). The instrument consisted of four sections: (1) Demographic Information, (2) Social–Cultural Motivation and Attitudes toward STEM, (3) Data Science Background Knowledge and Prior Exposure. And (4) Data Science Self-Efficacy (48 items across seven subscales).

Table 1. Revised and Expanded Data Science Self-Efficacy and Awareness Instrument

Section 0: Demographic Information | المعلومات الديموغرافية

Gender | الجنس: Male ذكر — Female أنثى — Prefer not to say أفضل عدم الإفصاح — Other أخرى
Year of Birth | سنة الميلاد: (Short answer)
Governorate | المحافظة: (Dropdown)
High School Track | التخصص في الثانوية: Science علمي — Arts أدبي
High School GPA | معدل التخصص: >90% أكثر من ٩٠٪ — >80% أكثر من ٨٠٪ — >70% أكثر من ٧٠٪ — <70% أقل من ٧٠٪
Highest Education Completed | آخر شهادة: Middle أقل من الثانوية — High School الثانوية — Diploma دبلوم — Bachelor بكالوريوس — Master ماجستير — Doctorate دكتوراه — Other أخرى
Currently Studying | تدرس حالياً: High School الثانوية — Diploma دبلوم — Bachelor بكالوريوس — Master ماجستير — Doctorate دكتوراه — Not studying لا أدرس
Student Level | المستوى الدراسي: Freshman سنة أولى — Sophomore سنة ثانية — Junior سنة ثالثة — Senior سنة رابعة
Current Major | التخصص الحالي: (Short answer)
University Name | اسم الجامعة: (Short answer)
Employment Status | الحالة الوظيفية: Employed موظف — Unemployed غير موظف — Student طالب — Other أخرى
Anticipated Industry | القطاع المتوقع: (Short answer)

Section 1: Motivation Social–Cultural Factors | العوامل الاجتماعية والثقافية

Why did you choose your major? | لماذا اخترت تخصصك؟
Family العائلة — Friends الأصدقاء — School/teachers المدرسة — Passion الشغف — Social media التواصل — Salary الراتب — Prestige المكانة — Future career المستقبل — Job market سوق العمل — GPA المعدل — Language barrier اللغة — Easy major السهولة

Attitudes (Likert Grid) | العوامل الاجتماعية والثقافية

- I am comfortable using technology tools | لدي سهولة في استخدام التكنولوجيا
- My tools are sufficient and good quality | الأدوات التقنية لدي كافية
- CS/Engineering only for strong math students | الحاسب والهندسة للمتفوقين فقط
- I follow science/tech on social media | أتابع محتوى علمي تقني
- My family encourages high-status careers | عائلتي تشجع الوظائف الرفيعة
- Non-STEM careers less prestigious | غير STEM أقل مكانة
- My school promoted STEM | مدرستي ركزت على STEM
- I prefer financially stable careers | أفضل الاستقرار المالي
- I prefer studying in Arabic | أفضل الدراسة بالعربية
- Comfortable in Arabic & English | مرتاح بالعربية والإنجليزية
- Prefer comfortable lifestyle careers | أفضل المهن المريحة
- Islamic values encourage STEM | القيم الإسلامية تشجع STEM
- STEM serves Muslim community | STEM يخدم المجتمع المسلم
- Support women in leadership | أدعم النساء في القيادة
- Women add value to leadership | المرأة تضيف قيمة
- Females succeed in STEM like males | الطالبات ينجحن مثل الطلاب
- Need more support for females | نحتاج دعم أكبر للطالبات
- Many friends like science/tech | أصدقائي مهتمون بالعلوم

If you go back in time | لو عاد بك الزمن:
Different major تخصص مختلف — More STEM دورات STEM — No change لا تغيير

Section 2: Data Science Background | الخلفية المعرفية لعلم البيانات

Course and background | (Likert Grid) (: Never لم يسبق — One مرة — More أكثر — Part جزء)
Statistics الإحصاء — Research methods البحوث — Programming البرمجة — ML/AI الذكاء الاصطناعي — Business Analytics تحليل الأعمال — Data Science علم البيانات

Tools & Concepts: (keep your checkbox list)
Excel | Matplotlib | Pandas | NumPy | Scikit-Learn | Tableau | Power BI | D3.js | QlikView | Apache Flink | MongoDB | SQL | Logistic Regression | Naive Bayes | Decision Tree | Linear Regression | Neural Networks | Random Forest | SVM | KNN | Reinforcement Learning | BigML | RapidMiner | KNIME | Weka | DataRobot | R | Matlab | Julia | NLTK | Python | Jupyter | TensorFlow | Scrapy | SAS | ANOVA | Hypothesis Testing | Google Analytics | Minitab | Clustering | K-Means | PCA | Unsupervised Learning | Probability| Imputation | Coefficients | Central Tendency | Class Imbalance | SOM | Sampling | AIC | SIC | High Cardinality | Partitioning

Section 3: Data Science Self-Efficacy | الكفاءة الذاتية

(Use Likert Grid) I am confident in | أنا واثق من قدرتي على

1. Plan and strategy for DS project | وضع خطة لمشروع علم بيانات
2. Set timelines using DS lifecycle | جدول زمني بدورة الحياة
3. Explore domain knowledge | استكشاف المعرفة المطلوبة
4. Review literature and trends | مراجعة الدراسات والاتجاهات
5. Formulate investigative questions | صياغة أسئلة بحثية
6. Consider ethics (privacy, bias) | مراعاة الجوانب الأخلاقية
7. Document code and results | توثيق الشيفرات والنتائج
8. Identify reliable data sources | تحديد مصادر بيانات موثوقة
9. Design data collection method | تصميم طريقة جمع البيانات
10. Iteratively improve collection/cleaning | تعديل خطوات الجمع والتنظيف
11. Choose data collection tools | اختيار أدوات جمع البيانات
12. Handle different data types | التعامل مع أنواع البيانات
13. Understand dataset structure | فهم بنية البيانات
14. Merge datasets | دمج البيانات
15. Visualize missing/outliers | تصور القيم المفقودة والشاذة
16. Normalize/standardize values | توحيد القيم
17. Handle invalid data | معالجة البيانات غير الصالحة
18. Validate cleaned data | التحقق من جودة البيانات
19. Create subsets | إنشاء مجموعات فرعية
20. Understand features & relations | فهم الخصائص والعلاقات
21. Use EDA/PCA/SOM | استخدام التحليل الاستكشافي
22. Rank features statistically | ترتيب الخصائص إحصائياً
23. Apply feature selection techniques | تطبيق تقنيات اختيار الخصائص
24. Experiment with techniques | تجربة عدة تقنيات
25. Create new features | إنشاء خصائص جديدة
26. Transform variables | تحويل المتغيرات
27. Remove redundancies | إزالة التكرار
28. Detect patterns/anomalies | اكتشاف الأنماط
29. Model building plan | خطة بناء النموذج
30. Choose modeling tools | اختيار أدوات النمذجة
31. Balance complexity/performance | الموازنة بين التعقيد والأداء
32. Choose method (regression/classification/…) | اختيار الأسلوب المناسب
33. Prepare data for learning type | تهيئة البيانات للتعلم
34. Select suitable models | اختيار النماذج المناسبة
35. Determine need for sampling | تحديد الحاجة للعينات
36. Scale model for large data | توسيع النموذج لبيانات أكبر
37. Tune hyperparameters | ضبط المعاملات
38. Apply cross-validation | تطبيق التحقق المتقاطع
39. Define evaluation metrics | تحديد مقاييس التقييم
40. Perform hypothesis testing | اختبار الفرضيات
41. Create visualizations | إنشاء تصورات بيانية
42. Extract insights | استخلاص الرؤى
43. Explain model results | تفسير النتائج
44. Explain to non-technical audience | شرح لغير المتخصصين
45. Connect results to literature | ربط النتائج بالدراسات
46. Use data storytelling | السرد القصصي بالبيانات
47. Tailor visuals to audience | تكييف التصورات للجمهور
48. Follow best visualization practices | اتباع أفضل ممارسات التصور

Participants and Sampling
Participants were recruited from a broad range of educational settings across Kuwait, including high schools, public universities, private universities, community colleges, and specialized institutes. A convenience sampling strategy was used through educator and institutional networks to achieve broad institutional representation. A total of 560 responses were collected, with 430 complete responses used for demographic and background analyses. For analyses specifically related to current university enrollment, 89 responses from participants currently studying in higher education were examined. The sample showed strong female representation (70.6% female, 29.4% male). Most participants (71.2%) reported a science track in high school. Academic performance was high, with 39.9% reporting GPAs above 90% and 33.9% above 80%. Participants represented varied educational attainment levels: 39% had completed high school as their highest qualification, 35.8% held a bachelor's degree, and others reported diploma, master's, or doctoral degrees. Regarding current enrollment, 57.9% were in bachelor's programs and 27.9% were in high school. Participants were drawn from multiple Kuwaiti institutions, with the largest representation from Kuwait University (55.8%), followed by AUK, GUST, PAAET, Arab Open University, KTECH, ACK, KILAW, and other specialized or international institutions.

Data Collection Procedure
The survey was administered online over a one-month period during the academic term. The survey link was distributed through Ministry of Education educator networks and shared by faculty members through classroom announcements and course communication channels. The survey was also promoted during educational events where participants could complete it on-site. Participation was voluntary, and informed consent was obtained prior to participation.

Reliability and Validity Testing
Although the original self-efficacy instrument had been previously validated, modifications were introduced in this study, including the addition of a social–cultural section and translation into Arabic. Therefore, reliability testing was conducted to ensure internal consistency within this new context. Cronbach's alpha was calculated for each construct: Data Science Self-Efficacy ($\alpha = .93$), Data Science Awareness ($\alpha = .88$) and Social–Cultural Factors ($\alpha = .90$). These values indicate strong internal consistency and support the reliability of the adapted bilingual instrument for measuring data science awareness and competency across diverse educational populations in Kuwait.

Results
This study examined students' data science awareness, background knowledge, and self-efficacy across multiple educational levels in Kuwait. The analysis aimed to understand not only the overall status of students' preparedness in data science, but also how educational level, prior learning experiences, social–cultural influences, and gender relate to their confidence and understanding in this field. The following research questions guided the analysis:

- **RQ1.** What are the levels of data science awareness, background knowledge, and self-efficacy among students across educational levels in Kuwait?
- **RQ2.** Are there significant differences in data science awareness and self-efficacy between high school students, community college students, and university students?
- **RQ3.** How does prior exposure to statistics, programming, and data science courses influence students' data science self-efficacy?

- **RQ4.** How do social–cultural factors (family, peers, school environment, societal perceptions, and gender views) relate to students' data science awareness and self-efficacy?
- **RQ5.** Which factors (background knowledge, social–cultural influences, gender, and educational level) best predict data science self-efficacy?

Data analysis was conducted using IBM SPSS to address these research questions through appropriate quantitative statistical methods. Descriptive statistics were used to examine the overall levels of data science awareness, background knowledge, and self-efficacy (RQ1). A one-way ANOVA with Tukey post-hoc tests was performed to identify differences across educational levels (RQ2). To examine the influence of prior exposure to statistics, programming, and data science courses, independent t-tests, ANOVA, and correlation analyses were applied (RQ3). Pearson correlation analysis was used to explore the relationships between social–cultural factors and both awareness and self-efficacy (RQ4). Multiple linear regression analysis was conducted to determine which factors best predict data science self-efficacy (RQ5). Finally, independent samples t-tests were used to investigate differences between male and female participants in awareness and self-efficacy (RQ6).

Social–Cultural Motivation for Choosing STEM

Table 2. Social–Cultural Reasons for Choosing STEM Across Educational Levels (Mean, SD)

Educational Level	*Family*	*Friends*	*School/Teachers*	*Passion*	*Social Media*	*Salary*	*Prestige*	*Future Career*	*Job Market*	*GPA*	*Language Barrier*	*Easy Major*
Master's	0.62 (0.49)	0.30 (0.47)	0.55 (0.51)	0.85 (0.37)	0.25 (0.44)	0.68 (0.47)	0.60 (0.50)	0.88 (0.33)	0.82 (0.39)	0.40 (0.50)	0.18 (0.39)	0.12 (0.33)
Diploma	0.58 (0.50)	0.35 (0.49)	0.50 (0.51)	0.78 (0.42)	0.32 (0.48)	0.64 (0.49)	0.55 (0.51)	0.82 (0.39)	0.78 (0.42)	0.45 (0.51)	0.22 (0.42)	0.20 (0.41)
Bachelor's	0.55 (0.50)	0.28 (0.45)	0.48 (0.50)	0.80 (0.40)	0.30 (0.46)	0.66 (0.47)	0.52 (0.50)	0.84 (0.37)	0.79 (0.41)	0.42 (0.49)	0.25 (0.43)	0.18 (0.38)
Doctorate	0.65 (0.49)	0.22 (0.42)	0.58 (0.51)	0.90 (0.30)	0.20 (0.41)	0.72 (0.46)	0.63 (0.49)	0.90 (0.30)	0.85 (0.37)	0.38 (0.50)	0.15 (0.36)	0.10 (0.31)
High School	0.70 (0.46)	0.40 (0.49)	0.62 (0.49)	0.72 (0.45)	0.38 (0.49)	0.58 (0.50)	0.48 (0.50)	0.75 (0.44)	0.72 (0.45)	0.55 (0.50)	0.35 (0.48)	0.30 (0.46)
P-Value	*0.098*	*0.181*	*0.156*	*0.283*	*0.417*	*0.576*	*0.710*	*0.224*	*0.540*	*0.181*	*0.148*	*0.049*
Eta Squared (η^2)	*0.02*	*0.01*	*0.02*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.02*	*0.02*
Effect Size	*Small*	*Small*	*Small*	*Small*	*Negligible*	*Negligible*	*Negligible*	*Small*	*Negligible*	*Small*	*Small*	*Small*

Table 2 presents the descriptive patterns and the statistical testing of social–cultural reasons that influence students' decisions to choose STEM across different educational levels. As shown in Table 2, passion and future career prospects consistently received the highest mean values across all groups. For example, doctoral participants reported very high agreement with passion (M = 0.90, SD = 0.30) and future career (M = 0.90, SD = 0.30), and similar patterns appeared among master's students (passion M = 0.85, future career M = 0.88) and bachelor's students (passion M = 0.80, future career M = 0.84). Even high school students showed relatively strong agreement with these reasons (passion M = 0.72, future career M = 0.75). Job market opportunities followed a similar trend, with means ranging from 0.72 (high school) to 0.85 (doctorate), indicating that career-related thinking is a dominant factor across all stages. In contrast, social media, language barrier, and easy major were among the lowest-rated reasons overall. For instance, social media ranged from 0.20 (doctorate) to 0.38 (high school), while language barrier ranged from 0.15 (doctorate) to 0.35 (high school). The perception of STEM as an easy major was also low for most groups, especially doctorate participants (M = 0.10) and master's students (M = 0.12), but noticeably higher for high school students (M = 0.30). Additionally, high school students reported higher influence from friends (M = 0.40), school/teachers (M = 0.62), and GPA (M = 0.55) compared to higher education groups, suggesting that external contextual influences play a larger role earlier in the academic pathway.

The last three rows of Table 2 confirm these patterns statistically. The ANOVA results show that almost all these factors are not significantly different across educational levels, with small or negligible effect sizes (η^2 between 0.01 and 0.02). For example, passion ($p = 0.283$), future career ($p = 0.224$), and job market ($p = 0.540$) show no statistical difference between groups despite their high means. The only factor approaching statistical significance is easy major ($p = 0.049$, $\eta^2 = 0.02$), where high school students (M = 0.30) were more likely to report this reason than doctoral students (M = 0.10); however, the small effect size indicates that this difference has limited practical impact.

Table 3. Overall STEM Motivation Across Educational Levels

Educational Level	*N*	*Mean motivation*	*Std. Deviation*	*P-Value*	*Eta Squared (η²)*	*Effect Size*
Master's	20	4.66	0.44	**< .001**	**0.16**	**Large**
Diploma	25	4.25	0.62			
Bachelor's	249	4.21	0.53			
Doctorate	16	3.70	0.00			
High School	120	3.67	0.90			

Table 4. Overall STEM Motivation Across Institutions

Institution (University/College)	*N*	*Mean motivation*	*Std. Deviation*
Kuwait University	170	4.20	0.50
AUM – Middle East University	19	4.58	0.41
GUST – Gulf University for Science & Technology	28	4.18	0.76
AUK – American University of Kuwait	18	4.14	0.37
PAAET – Public Authority for Applied Education & Training	36	4.06	0.48
Arab Open University	14	3.82	0.65
Others (International / Specialized Institutes)	56	4.05	0.70

Variable	*P-Value*	*Eta Squared (η²)*	*Effect Size*	*Interpretation*
Motivation Attitudes	**0.003**	**0.06**	**Small/Medium**	**Significant.** Students at AUM reported notably higher motivation (4.58) compared to those at Arab Open University (3.82) or PAAET (4.06).

Tables 3 and 4 present the overall levels of STEM motivation and examine how this motivation varies across both educational levels and institutions. As shown in Table 3, overall STEM motivation differs noticeably across educational stages. Master's students reported the highest motivation (M = 4.66, SD = 0.44), followed by diploma (M = 4.25, SD = 0.62) and bachelor's students (M = 4.21, SD = 0.53). In contrast, doctoral participants (M = 3.70, SD = 0.00) and high school students (M = 3.67, SD = 0.90) reported lower motivation levels. The ANOVA results indicate that these differences are statistically significant ($p < .001$) with a large effect size ($\eta^2 = 0.16$), meaning that educational level strongly influences how motivated students feel toward STEM. This pattern suggests that motivation increases as students' progress into higher education, peaks at the master's level, and is comparatively lower at the early (high school) and terminal (doctorate) stages.

Table 4 examines motivation across institutions and shows more moderate variation. Students at AUM reported the highest motivation (M = 4.58, SD = 0.41), noticeably higher than those at Arab Open University (M = 3.82, SD = 0.65) and PAAET (M = 4.06, SD = 0.48). Kuwait University (M = 4.20), GUST (M = 4.18), and AUK (M = 4.14) displayed very similar motivation levels, while students from other institutions reported M = 4.05 (SD = 0.70). The statistical test confirms that these differences are significant (p = 0.003) with a small-to-medium effect size ($\eta^2 = 0.06$). This indicates that while institution does influence motivation, the effect is weaker than the influence of educational level.

Data Science Background Knowledge

Table 5. Prior Exposure to Data Science–Related Courses by Institution

Institution	*Statistics*	*Research Methods*	*Programming*	*ML / AI*	*Business Analytics*	*Data Science*
Kuwait University (170)	1.82 (0.61)	2.02 (0.73)	2.30 (0.75)	1.72 (0.59)	2.36 (0.53)	1.74 (0.60)
AUM – Middle East University (19)	2.25 (0.70)	2.18 (0.73)	1.95 (0.68)	2.05 (0.69)	2.12 (0.72)	2.08 (0.71)
GUST (28)	1.95 (0.66)	1.88 (0.68)	2.02 (0.70)	1.85 (0.63)	1.80 (0.61)	1.86 (0.64)
AUK (18)	2.48 (0.52)	2.42 (0.55)	2.55 (0.50)	2.30 (0.56)	1.68 (0.57)	2.40 (0.54)
PAAET (36)	2.10 (0.71)	1.76 (0.64)	2.15 (0.75)	1.92 (0.68)	1.95 (0.66)	2.00 (0.69)
Arab Open University (14)	1.88 (0.63)	1.80 (0.65)	1.92 (0.67)	1.70 (0.61)	1.72 (0.60)	1.76 (0.62)
Others (56)	2.05 (0.69)	1.98 (0.72)	2.12 (0.73)	1.90 (0.66)	1.94 (0.65)	1.98 (0.67)
P-Value	**< .001**	**0.033**	**0.029**	**0.003**	**< .001**	**< .001**
Eta Squared (η^2)	**0.07**	**0.04**	**0.04**	**0.06**	**0.16**	**0.07**
Effect Size	**Medium**	**Small**	**Small**	**Small/Medium**	**Large**	**Medium**

Table 6. Overall Background Knowledge Across Educational Levels

Educational Level	*N*	*Mean Background Knowledge*	*Std. Deviation*
Master's	20	2.23	0.92
Diploma	25	2.14	0.72
Bachelor's	249	1.59	0.56
Doctorate	16	1.67	0.00
High School	120	1.62	0.55

Tables 5–7 present students' prior exposure and overall background knowledge in data science across both institutions and educational levels. As shown in Table 5, students' exposure to data science–related courses varies notably by institution. AUK students consistently reported the highest background knowledge across most course areas, including Statistics (M = 2.48), Research Methods (M = 2.42), Programming (M = 2.55), ML/AI (M = 2.30), and Data Science (M = 2.40). Similarly, AUM students reported relatively strong exposure, particularly in Statistics (M = 2.25) and ML/AI (M = 2.05). In contrast, Kuwait University students reported lower exposure in several areas such as Statistics (M = 1.82) and Data Science (M = 1.74), but notably higher knowledge in Business Analytics (M = 2.36). The ANOVA results indicate statistically significant differences across institutions for all course areas. The strongest effect was observed in Business Analytics ($\eta^2 = 0.16$, large effect), while Statistics and Data Science showed medium effects ($\eta^2 = 0.07$). Programming and Research Methods demonstrated small effects ($\eta^2 = 0.04$), and ML/AI showed a small-to-medium effect ($\eta^2 = 0.06$). These findings indicate that institutional curriculum emphasis plays a clear role in students' background exposure to data science topics.

Table 6 examines overall background knowledge across educational levels. Master's students reported the highest background knowledge (M = 2.23, SD = 0.92), followed closely by diploma students (M = 2.14, SD = 0.72). Bachelor's students (M = 1.59, SD = 0.56), doctoral participants (M = 1.67, SD = 0.00), and high school students (M = 1.62, SD = 0.55) reported notably lower levels. This pattern suggests that background knowledge increases with advanced study but remains relatively similar among high school, bachelor's, and doctoral participants, likely reflecting differences in formal coursework exposure.

Finally, Table 7 shows overall background knowledge by institution. AUK reported the highest overall background knowledge (M = 2.41, SD = 0.17), followed by PAAET (M = 2.01, SD = 0.75) and AUM (M = 1.83, SD = 0.78). Kuwait University (M = 1.58, SD = 0.55) and GUST (M = 1.59, SD = 0.75) reported similar levels, while Arab Open University reported slightly lower levels (M = 1.72, SD = 0.68). These results reinforce the pattern observed in Table 5, indicating that institutional differences in curriculum and course offerings are strongly associated with students' perceived background knowledge in data science.

Table 7. Overall Background Knowledge Across Institutions

Institution (University/College)	*N*	*Mean Background Knowledge*	*Std. Deviation*
Kuwait University	170	1.58	0.55
AUM – Middle East University	19	1.83	0.78
GUST – Gulf University for Science & Technology	28	1.59	0.75
AUK – American University of Kuwait	18	2.41	0.17
PAAET – Public Authority for Applied Education & Training	36	2.01	0.75
Arab Open University	14	1.72	0.68
Others (International / Specialized Institutes)	56	1.88	0.70

Table 8. Familiarity with Key Data Science Tools and Concepts by Institution

Institution	*Excel*	*Python*	*SQL*	*Pandas*	*NumPy*	*Tableau*	*Power BI*	*ML Concepts*	*Statistics Concepts*	*Visualization Concepts*
Kuwait University	92%	48%	36%	32%	30%	41%	38%	35%	72%	66%
AUM	95%	71%	52%	55%	53%	60%	58%	62%	78%	70%
GUST	90%	50%	40%	38%	35%	48%	46%	42%	74%	68%
AUK	96%	74%	60%	58%	55%	65%	63%	68%	82%	75%
PAAET	88%	52%	42%	40%	38%	44%	42%	46%	76%	69%
Arab Open University	84%	36%	28%	26%	25%	30%	29%	32%	68%	61%
Others	90%	49%	37%	35%	33%	45%	43%	40%	73%	67%
P-Value	**0.948**	**0.269**	**0.352**	**0.304**	**0.203**	**0.302**	**0.326**	**0.041**	**0.956**	**0.976**
Eta Squared (η^2)	**0.01**	**0.03**	**0.02**	**0.03**	**0.03**	**0.02**	**0.02**	**0.04**	**0.00**	**0.00**

Table 8 presents students' familiarity with key data science tools and concepts across institutions, reported as percentages of students who indicated prior familiarity. The results show that Excel is universally familiar across all institutions, with very high usage rates ranging from 84% to 96%. The statistical test confirms no significant differences between institutions for Excel ($p = .948$, $\eta^2 = .01$), indicating that spreadsheet use is widespread and consistent across groups. For programming and analysis tools such as Python, SQL, Pandas, NumPy, Tableau, and Power BI, familiarity levels vary slightly across institutions but without statistically significant differences. For example, Python familiarity ranges from 36% at Arab Open University to 74% at AUK, while SQL ranges from 28% to 60%. Despite these visible differences, the small effect sizes (η^2 between .02 and .03) and non-significant p-values indicate that these variations are not strong enough to be considered statistically meaningful. A notable exception appears in ML Concepts, where significant differences were observed ($p = .041$, $\eta^2 = .04$). Students at AUK (68%) and AUM (62%) reported considerably higher familiarity with machine learning concepts compared to Arab Open University (32%) and Kuwait University (35%). This suggests that certain institutions place greater emphasis on introducing machine learning topics within their curriculum. In contrast, familiarity with Statistics Concepts (68–82%) and Visualization Concepts (61–75%) is consistently high across all institutions, with negligible effect sizes ($\eta^2 = .00$) and no significant differences. This indicates that foundational statistical and visualization knowledge is relatively well distributed among students regardless of institution.

Data Science Self-Efficacy

Table 9. Self-Efficacy Subscales Across Educational Levels

Educational Level	*Domain Knowledge & Research Design M(SD)*	*Data Planning & Collection M(SD)*	*Data Cleaning & Feature Eng. M(SD)*	*Feature Selection M(SD)*	*Model Design M(SD)*	*Model Evaluation M(SD)*	*Communication & Action M(SD)*
Master's	4.68 (0.36)	4.61 (0.41)	4.55 (0.39)	4.58 (0.42)	4.66 (0.34)	4.71 (0.31)	4.74 (0.29)
Diploma	4.35 (0.49)	4.28 (0.53)	4.21 (0.51)	4.26 (0.50)	4.33 (0.47)	4.38 (0.44)	4.41 (0.42)
Bachelor's	4.22 (0.55)	4.16 (0.58)	4.09 (0.57)	4.14 (0.56)	4.20 (0.54)	4.24 (0.52)	4.27 (0.51)
Doctorate	4.75 (0.00)	4.69 (0.00)	4.63 (0.00)	4.67 (0.00)	4.72 (0.00)	4.78 (0.00)	4.80 (0.00)
High School	3.62 (0.86)	3.55 (0.88)	3.48 (0.85)	3.51 (0.87)	3.58 (0.83)	3.60 (0.80)	3.66 (0.78)
P-Value	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**
Eta Squared (η^2)	**0.18**	**0.16**	**0.16**	**0.18**	**0.19**	**0.19**	**0.19**

Table 9 presents students' data science self-efficacy across seven subscales and compares these perceptions across educational levels. A clear pattern is visible across all subscales: self-efficacy increases with educational progression. Doctoral participants reported the highest confidence across every subscale, with means ranging from 4.63 to 4.80, followed closely by master's students whose scores ranged from 4.55 to 4.74. Diploma and bachelor's students reported slightly lower but still high confidence levels, with means generally between 4.09 and 4.41. In contrast, high school students reported noticeably lower self-efficacy across all areas, with means ranging from 3.48 to 3.66 and larger standard deviations, indicating greater variability in their confidence levels. Across the specific dimensions, the highest confidence for advanced students appeared in Model Evaluation and Communication & Action, where doctoral students reached 4.78 and 4.80, and master's students reached 4.71 and 4.74, respectively. The lowest scores overall were consistently reported by high school students in Data Cleaning & Feature Engineering (3.48) and Feature Selection (3.51), suggesting these more technical aspects of data science present greater challenges at earlier educational stages. The ANOVA results show that differences across educational levels are statistically significant for all subscales ($p < .001$) with large effect sizes (η^2 ranging from 0.16 to 0.19). This indicates that educational level is a strong predictor of students' perceived ability to perform key data science tasks.

Table 10. Self-Efficacy Subscales Across Institutions

Institution	*Domain M(SD)*	*Planning M(SD)*	*Cleaning M(SD)*	*Feature M(SD)*	*Model Design M(SD)*	*Evaluation M(SD)*	*Communication M(SD)*
Kuwait University	4.23 (0.52)	4.17 (0.55)	4.10 (0.54)	4.15 (0.53)	4.21 (0.51)	4.25 (0.49)	4.29 (0.48)
AUM (N=19)	4.70 (0.34)	4.63 (0.36)	4.59 (0.35)	4.62 (0.33)	4.68 (0.32)	4.73 (0.29)	4.76 (0.28)
GUST (N=28)	4.28 (0.61)	4.21 (0.64)	4.14 (0.63)	4.19 (0.62)	4.25 (0.60)	4.30 (0.58)	4.33 (0.57)
AUK (N=18)	4.58 (0.31)	4.51 (0.33)	4.47 (0.32)	4.49 (0.30)	4.55 (0.29)	4.60 (0.27)	4.63 (0.26)
PAAET (N=36)	4.12 (0.48)	4.06 (0.51)	4.00 (0.50)	4.04 (0.49)	4.10 (0.47)	4.15 (0.45)	4.18 (0.44)
Arab Open University	3.84 (0.69)	3.79 (0.72)	3.72 (0.70)	3.76 (0.71)	3.82 (0.68)	3.86 (0.66)	3.90 (0.64)
Others (N=56)	4.18 (0.64)	4.12 (0.66)	4.05 (0.65)	4.10 (0.63)	4.16 (0.61)	4.20 (0.60)	4.24 (0.59)
P-Value	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**
Eta Squared (η^2)	**0.09**	**0.09**	**0.09**	**0.09**	**0.10**	**0.10**	**0.10**

Table 10 compares students' data science self-efficacy across institutions for the seven subscales. A consistent pattern appears across all dimensions: students from AUM and AUK reported the highest levels of confidence in nearly every data science skill. AUM students' means range from 4.59 to 4.76, while AUK students range from 4.47 to 4.63. These are noticeably higher than the scores reported by students from Kuwait University, GUST, PAAET, Arab Open University, and the "Others" group. Students from Arab Open University reported the lowest self-efficacy across all subscales, with means between 3.72 and 3.90, indicating comparatively lower confidence in performing data science tasks. Kuwait University, GUST, PAAET, and the "Others" group fall in the mid-range, with means generally between 4.00 and 4.33, suggesting moderate to high confidence but still below AUM and AUK. Across specific subscales, the highest reported confidence overall appears in Model Evaluation and Communication, particularly for AUM (4.73, 4.76) and AUK (4.60, 4.63). The lowest values consistently appear in Data Cleaning and Feature Selection for Arab Open University (3.72, 3.76). The ANOVA results indicate that differences between institutions are statistically significant across all subscales ($p < .001$). The effect sizes are medium ($\eta^2 \approx 0.09$–0.10), meaning that institutional affiliation has a meaningful, though not large, influence on students' perceived data science ability.

Discussion

RQ1. What are the levels of data science awareness, background knowledge, and self-efficacy among students across educational levels in Kuwait?

The results indicate clear patterns in students' levels of data science awareness, background knowledge, and self-efficacy across educational stages. Overall awareness is highest among master's students, followed by diploma and bachelor's students, while high school students report the lowest awareness and the greatest variability. Background knowledge shows a similar trend, with master's and diploma students reporting higher exposure to data science–related coursework than bachelor's and high school students. Self-efficacy demonstrates the strongest progression across educational levels. Doctoral and master's participants report the highest confidence across all seven data science skill areas, while high school students report consistently lower confidence across each subscale. Statistical testing confirms that differences in self-efficacy across educational levels are significant with large effect sizes, indicating that educational progression is strongly associated with students' perceived competence in data science tasks.

RQ2. Are there significant differences in data science awareness and self-efficacy between high school students, community college students, and university students?

Significant differences are observed between these groups. Awareness levels are noticeably higher among postsecondary students compared to high school students, with diploma and university students reporting greater familiarity with data science concepts. The most pronounced differences appear in self-efficacy. Across all seven self-efficacy subscales, educational level differences are statistically significant with large effect sizes. High school students report the lowest confidence in performing data science tasks, diploma students report moderate to high confidence, and university students—particularly at the master's and doctoral levels—report the highest levels of confidence. These findings demonstrate that both awareness and self-efficacy increase substantially as students' progress through higher levels of education.

RQ3. How does prior exposure to statistics, programming, and data science courses influence students' data science self-efficacy?
Descriptive patterns across institutions suggest a strong alignment between prior coursework exposure and students' perceived data science competence. Institutions where students report higher exposure to statistics, programming, machine learning, and data science courses also report higher self-efficacy across all data science skill areas. For example, students at institutions such as AUK and AUM report both higher prior exposure to data science–related coursework and higher self-efficacy scores across the seven subscales. In contrast, institutions reporting lower coursework exposure also show comparatively lower self-efficacy levels. This consistent pattern suggests that prior academic exposure to data science topics is associated with increased confidence in performing data science tasks.

RQ4. How do social–cultural factors (family, peers, school environment, societal perceptions, and gender views) relate to students' data science awareness and self-efficacy?
The results show that social–cultural reasons for choosing STEM are remarkably consistent across educational levels. Passion, future career opportunities, and job market prospects are the most strongly endorsed reasons across all groups, while social media influence, language barriers, and the perception of STEM as an easy major are among the least endorsed. Statistical testing indicates that these social–cultural factors do not significantly differ across educational levels, with only small or negligible effect sizes. Although these factors shape students' motivations for entering STEM, the consistency across groups suggests that they do not vary strongly by educational stage. These findings indicate that while social–cultural motivations are important in shaping students' entry into STEM fields, their relationship to awareness and self-efficacy appears less differentiated across groups.

RQ5. Which factors (background knowledge, social–cultural influences, gender, and educational level) best predict data science self-efficacy?
Among the examined factors, educational level emerges as the strongest predictor of data science self-efficacy. The large effect sizes observed across all self-efficacy subscales demonstrate that students' educational stage has a substantial influence on their perceived data science competence. Institutional affiliation also shows a meaningful, though moderate, effect on self-efficacy, suggesting that curriculum emphasis and learning environment contribute to students' confidence. Background knowledge patterns further support this relationship, as groups and institutions reporting higher prior exposure to data science coursework also report higher self-efficacy. In contrast, social–cultural motivations, while influential in shaping students' reasons for choosing STEM, show limited variation across groups and therefore appear to play a smaller role in directly predicting self-efficacy compared to educational level and background knowledge.

Implications and Proposal for Educational Stakeholders in Kuwait
The findings of this study offer guidance for educational stakeholders in Kuwait currently designing Data Science (DS) programs. Students across all educational levels demonstrated strong motivation toward STEM, driven primarily by passion and future career aspirations. However, the results reveal noticeable gaps in structured exposure to data science concepts, uneven curricular emphasis across institutions, and significantly lower self-efficacy among high school students compared to university students. This suggests that the primary challenge is not a lack of student interest, but rather the absence of a coherent pathway that introduces data science progressively throughout the educational journey. High school students reported the lowest levels of self-efficacy and appeared most influenced by external factors such as peers, GPA, and the school environment, highlighting a critical opportunity

for early intervention. Integrating foundational data science and data literacy concepts into secondary mathematics, science, and computer courses is essential for building early confidence. Students should be exposed to the full data science lifecycle—including data collection, cleaning, visualization, and interpretation—using real-life Kuwaiti datasets to render the learning meaningful and culturally relevant.

Furthermore, the study indicates that students' background knowledge and confidence vary significantly by institution. Students from AUK and AUM consistently reported higher exposure to data science courses and greater familiarity with machine learning concepts, suggesting stronger curricular integration at these campuses. This disparity underscores the need for a National Data Science Learning Progression Framework to define competency standards from high school through postgraduate study. Standardizing foundational exposure would ensure that data science literacy is a universal outcome rather than a privilege dependent on where a student enrolls.

Finally, faculty preparation is paramount. Because student exposure is tied so closely to institutional emphasis, professional development for educators is essential. Training programs aligned with a national framework can empower teachers to effectively integrate data concepts into their courses, moving instruction beyond mere tool usage toward a conceptual understanding of data thinking and modeling. While familiarity with tools like Excel is high, the understanding of machine learning and deeper data concepts remains uneven. By introducing project-based activities, competitions, and community-centric data projects, educators can help students visualize themselves as capable data thinkers, thereby strengthening their sense of belonging in the field.

Conclusion

This study examined data science awareness, background knowledge, and self-efficacy among students across educational levels and institutions in Kuwait. The results show a clear pattern: students' confidence in performing data science tasks increases as they progress through higher levels of education, and this growth is strongly linked to prior exposure to statistics, programming, and data science courses. Educational level was the strongest factor influencing self-efficacy. Students at master's and doctoral levels reported the highest confidence across all data science skill areas, while high school students reported the lowest. Background knowledge followed the same trend, showing that increased academic exposure supports stronger self-efficacy. Institutional differences were also evident, particularly where curricula emphasized statistics, programming, and machine learning, though these effects were moderate compared to educational level. In contrast, social–cultural motivations for choosing STEM were similar across groups. Passion, future career opportunities, and job market prospects were the dominant reasons, while factors such as social media influence and perceived ease of the major were least important. These motivations did not significantly explain differences in awareness or self-efficacy. Overall, the findings highlight that data science self-efficacy is primarily shaped by educational progression and coursework exposure. Introducing structured data science learning earlier in the educational pathway may help strengthen students' confidence and readiness for data-driven fields.

Limitations

The study relied on self-reported survey data, which reflects students' perceptions of their awareness, knowledge, and confidence rather than direct measurement of their actual skills. This may introduce response bias or over/underestimation of ability. Another limitation is that some groups had small sample sizes (e.g., doctoral and certain institutions), and in a few cases zero standard deviation appeared, which limits the strength of statistical comparisons and may affect the reliability of ANOVA results. Relatedly, uneven participation across institutions and educational levels created imbalance in the data. Also, the background knowledge was measured based on prior course exposure rather than verified transcripts or performance, so it reflects perceived exposure rather than confirmed mastery. Similarly, familiarity with tools and concepts was measured as recognition, not demonstrated competence.

Future Work

To extend this study, we propose comparing self-reported confidence with actual skill demonstrations to identify potential discrepancies between competence and confidence. Research should also examine the longitudinal growth of these attributes across the educational pipeline. Ultimately, this instrument offers a foundation for diagnostic assessment, empowering educators to detect specific gaps in data science understanding and deploy precise instructional interventions

References

[1] A. Di Battista *et al.*, "Future of jobs report 2023," in *World Economic Forum*, 2023, pp. 978-2.

[2] F. Miao and W. Holmes, "Artificial intelligence and education. Guidance for policy-makers," 2021.

[3] K. University, "Bachelor of Science in Data Science Program Announcement,," ed, 2025.

[4] S. Malallah, S. Alfailakawi, T. Y. Alkhurafi, and J. L. Weese, "Reversing Gender Stereotypes in STEM Education in a Gender-Segregated Region," in *2021 ASEE Virtual Annual Conference Content Access*, 2021.

[5] M. S. Malallah, "A Replicate Study: Adoption of a STEM Outreach Program in Kuwait," 2020.

[6] S. Malallah, K. N. Alsalmi, and J. L. Weese, "Measuring Awareness of Computational Thinking in Kuwaiti Educational Institutions," in *2021 ASEE Virtual Annual Conference Content Access*, 2021.

[7] S. Malallah *et al.*, "Developing an instrument for assessing self-efficacy confidence in data science," in *2024 ASEE Annual Conference & Exposition*, 2024.