



Impact of Pedagogical Approaches on Machine Learning Concept Mastery: A Pretest-Posttest Analysis

Muhammad Imran Qureshi¹, Nazia Suleman², Muhammad Rashid³, Muhammad Rehan Ashraf⁴, Syed Abrar Hussain Shah⁵, Yasir Ghafoor⁶, Abida Hanif⁷

¹ COMSATS University Islamabad, Vehari Campus, Postal Code 61100, Vehari, Pakistan

Email: imranqureshi@cuivehari.edu.pk

² COMSATS University Islamabad, Vehari Campus, Postal Code 61100, Vehari, Pakistan

Email: naziasuleman@cuivehari.edu.pk

³ COMSATS University Islamabad, Vehari Campus, Postal Code 61100, Vehari, Pakistan

Email: s.mrashid5524@gmail.com

⁴ COMSATS University Islamabad, Vehari Campus, Postal Code 61100, Vehari, Pakistan

Email: rehanashraf@cuivehari.edu.pk

⁵ COMSATS University Islamabad, Vehari Campus, Postal Code 61100, Vehari, Pakistan

Email: abrarhussain@cuivehari.edu.pk

⁶ COMSATS University Islamabad, Vehari Campus, Postal Code 61100, Vehari, Pakistan

Email: yasirghafoor@cuivehari.edu.pk

⁷ COMSATS University Islamabad, Vehari Campus, Postal Code 61100, Vehari, Pakistan

Email: abidahanif@cuivehari.edu.pk

ARTICLE INFO	ABSTRACT
Submission: July 08, 2026 Revised: July 26, 2026 Accepted: August 26, 2026 Available Online: September 08, 2026 Keywords: <i>blended learning, data science education, machine learning pedagogy, pretest-posttest design, STEM education, undergraduate mathematics, paired t-test, Cohen's d</i> Corresponding author: Muhammad Imran Qureshi Email: imranqureshi@cuivehari.edu.pk	The need to rapidly incorporate data science and machine learning into undergraduate math programs necessitates the development of effective pedagogical approaches to reconcile theoretical knowledge and computational skills. This quasi-experimental pretest-posttest study is a one-group, one-group design, which assessed the effectiveness of a 16-week blended instructional intervention on 94 undergraduate mathematics students of the COMSATS University Islamabad, Vehari Campus, Pakistan. The course took the form of 3(2,1) of two hours of theoretical lectures and one three-hour lab session each week, covering the basics of Data Science, Python programming and the concepts of Machine Learning. Learning outcomes in the three domains were assessed by self-constructed pre- and post-tests. Paired-samples t-tests revealed statistically significant improvements: Data Science ($M_{diff} = 3.98$, $t(93) = 16.22$, $p < .001$, Cohen's $d = 1.67$), Python ($M_{diff} = 6.59$, $t(93) = 21.44$, $p < .001$, $d = 2.21$), and Machine Learning ($M_{diff} = 6.64$, $t(93) = 23.88$, $p < .001$, $d = 2.46$). Increases were demonstrated by more than 94 percent of students with big practical effects. Results indicate that blended theoretical-laboratory teaching can be successfully used in resources-limited environments and that experiential learning can be useful to teach computational mastery. Some implications on curriculum reform in higher education in developing countries are discussed.

Introduction

The exponentially increasing data-driven technologies have radically changed industries, research paradigms, and decision-making processes in society. Data science and machine learning (ML) have become critical skills that enable graduates to gain valuable

insights by working with complicated data and solving practical problems in all industries (National Academies of Sciences, Engineering, and Medicine, 2018). The need to balance theory with applied computational skills is becoming popular in higher education, including undergraduate mathematics programs. Learners of mathematics usually perform well in analytical thinking but usually do not learn programming languages like Python and implemented machine learning methods, which has become a crucial deficiency in the contemporary data economy (Biehler et al., 2021; Gould, 2010).

Meta-analyses show that blended learning has moderate to high benefits compared to traditional face-to-face education alone, especially when the use of online resources is complemented by interactive practice (Means et al., 2013; Bernard et al., 2014; Cao et al., 2023). These strategies enable students to be active participants where they can build knowledge through experiential processes of doing, reflecting and applying concepts (Garrison and Vaughan, 2008). Laboratory-based elements of computational education, such as data management, coding, and analysis, are particularly useful in the creation of conceptual knowledge and procedural fluency (Robins et al., 2003; Bennedsen and Caspersen, 2019).

The introduction of data science and ML in mathematics education in developing countries like Pakistan has its structural peculiarities. Effective implementation is hindered by resource limitations, lack of faculty knowledge of new technologies, poor laboratory facilities, and a long-standing dependence on rote, lecture-based teaching methods (Rehman, 2025; Khalil & Mut, 2022; UNESCO, 2021). Systemic evaluations conducted nationwide indicate that most of the students demonstrate only basic proficiency in mathematics and science, highlighting a lack of analytical and computational skill development (Aga Khan University, 2022). Most universities are not prepared to provide graduates with the skills needed to pursue a data-intensive career despite national efforts at digital transformation, facing a threat of lower employability and low innovation capacity in line with Sustainable Development Goal 4 on quality education (Higher Education Commission Pakistan, 2023; United Nations, 2015).

This study evaluates the impact of a newly introduced 16-week, 3-credit-hour blended course (3(2,1) two hours of theory and one three-hour laboratory session per week) offered for the first time in the Mathematics Department at COMSATS University Islamabad, Vehari Campus, during the Fall 2025 semester. The intervention comprised 94 undergraduate students, aged sophomores and was organized in two phases: the first eight weeks included the Data Science fundamentals and Python basics, and the second eight weeks covered the concepts of Machine Learning and the focus on data analysis and modeling. Continuous formative assessment methods, such as quizzes, assignments, task-based work, and project work, helped the students to learn.

This has led to worldwide recommendations to change undergraduate programs to incorporate data science literacy as a fundamental skill (National Academies of Sciences, Engineering, and Medicine, 2018). This, in math teaching, means that it is no longer sufficient to study pure theory but to learn to think and act in a manner that is based on computational thinking and programming, which help students to solve problems and become prepared to apply them across disciplines (Wing, 2006; Biehler et al., 2021). Blended models have been of particular utility in this transition as they integrate well-organized lectures with hands-on code and data manipulation activities, leading to greater engagement and retention (Means et al., 2013; Shi, 2025).

Pretest-posttest empirical research has repeatedly shown large learning improvements of these interventions in STEM settings (Grass et al., 2023; Wakhata et al., 2022). As an example, practical laboratory training enhances computational thinking and application competencies, the key to mastering data science and ML (Sari, 2025; Cannady et al., 2022). Nonetheless, the majority of current studies are based on highly-resourced institutions within the West, and lack a body of evidence on resource-constrained environments where the contextual influence of factors such as infrastructure constraints and student readiness may moderate efficacy (Topping et al., 2022; Rehman, 2025).

Research Objectives

- To evaluate how well students master the basics of Data Science after the blended intervention.
- To assess the improvement of the Python programming skills and their use in data-related tasks.
- To investigate how the course influenced the conceptual knowledge and fundamental usage of Machine Learning methods.
- To identify the statistical significance, effect sizes, and practical meaningfulness of observed changes in pretests to posttests across domains.
- To obtain evidence-based suggestions on how to optimize blended pedagogy techniques in undergraduate mathematics education in a resource constrained environment.

Significance of the Research

Theoretical implications of this study are that it uses constructivist principles (Ben-Ari, 2001; Vygotsky, 1978) and experiential learning models (Kolb, 1984) to the learning of computational education in a developing-country setting, which can enhance the existing knowledge on how active laboratory elements can support knowledge construction in data science and ML. In practice, the results will provide faculty and administrators with tangible data on the effectiveness of a systematic blended model, emphasizing the importance of combined theory-laboratory classes with minimum resources. To policymakers in Pakistan and other similar countries, outcomes can guide national curriculum changes and investments in the development of the faculty and digital infrastructure in line with national goals of digital transformation.

Literature Review

The pretest-posttest designs are a pillar in testing the effectiveness of instruction in STEM education, especially in quasi-experimental designs when it is difficult to assign randomness to classroom groups. These designs evaluate the alterations in learning outcomes pre-intervention and postintervention to directly demonstrate knowledge gains and be vulnerable to threats like maturation, testing effects, and history (Bierer, 2025; Mamanus & Marsden, 2017). Yalcyn (2022) employed pretest-posttest control group design to investigate the effect of design-thinking based STEM activities on creativity and problem-solving abilities of preschool children, showing a significant improvement in the experimental group. On the same note, meta-analytic studies of simulation-based learning in STEM demonstrate moderate positive effect sizes ($g = 0.36$ on achievement results) in support of the usefulness of hands-on computational aids, but researchers note that one-group designs are uncontrolled and can be over-relied on (SRI International, n.d.; Songer, 2024). Preservice teachers that studied integrated STEM curricula have recorded the improvement of their computational thinking and self-efficacy in higher education through pre-post assessment (Zhumabay et al., 2024). All of these studies confirm that pretest- posttest designs, with the support of powerful statistical tests, like paired t-tests, provide practical information about the efficacy of interventions despite the inherent validity weaknesses of such methods.

Blended pedagogical interventions prove to be highly effective in improving the mastery of data science, Python programming and the understanding of machine learning concepts in undergraduates. A meta-analysis of 232 studies carried out by Means et al. (2013) revealed that the effect sizes ($g = 0.35$) were significantly larger in blended learning conditions than in cases of face-to-face instruction, which can be explained by the fact that the combination of structured theory and interactive practice brings the gains. Alammery (2019) conducted a review of blended instruction in introductory programming courses and found that a combination of asynchronous online modules and in-class problem-solving enhances student engagement and lowers the rate of failure. Particular to Python, the recent studies of the blended teaching mode indicate an increase in learning enthusiasm, code literacy and practical application skills, and students demonstrate better performance in post-intervention assessments (exploring blended teaching mode of Python, 2025). In machine learning, pedagogical interventions with intelligent tutoring systems and explainability tools produce large pre-to-post conceptual learning, and personalized learning trajectories (Hur, 2022; Turkmenbayev, 2025). Constructivism-based interventions, in which students construct mental models via coding and data manipulation, are especially effective, as shown by big effect sizes in the growth of computational thinking (Ben-Ari, 2001; Csizmadia et al., 2019).

In countries such as Pakistan and Nigeria, the literature records ongoing issues such as untrustworthy internet connectivity, power failure, insufficient access to computing devices, faculty digital illiteracy, and institutional policy gaps that inhibit successful implementation of blended models (Khan, 2019; Jumani, 2018; Jailani et al, 2025; Sahito & Khoso, 2025). Majority of the literature is based on self-reported perceptions or small scale qualitative investigations as opposed to objective statistical outcomes of learning gains (Rehman, et al., 2025; Topping et al., 2022). Moreover, little has been done to explore the interaction between constructivist active learning and contextual constraints, which include the differences in student readiness and the inability to deliver interventions due to cultural constraints (Bou Saad, 2025). This paper fills these gaps by considering a recently implemented 16-week blended course in a Pakistani mathematics department that uses paired t-tests and effects of size as methods to measure gains in three related computational areas.

The current quasi-experimental study therefore provides a timely contribution of evidence by connecting the best practices in the world to the local realities, and hypothesizing that a structured 3(2,1) blend course will lead to statistically significant and practically meaningful improvement in mastery of students, which can inform systematic curriculum reforms.

Theoretical Framework

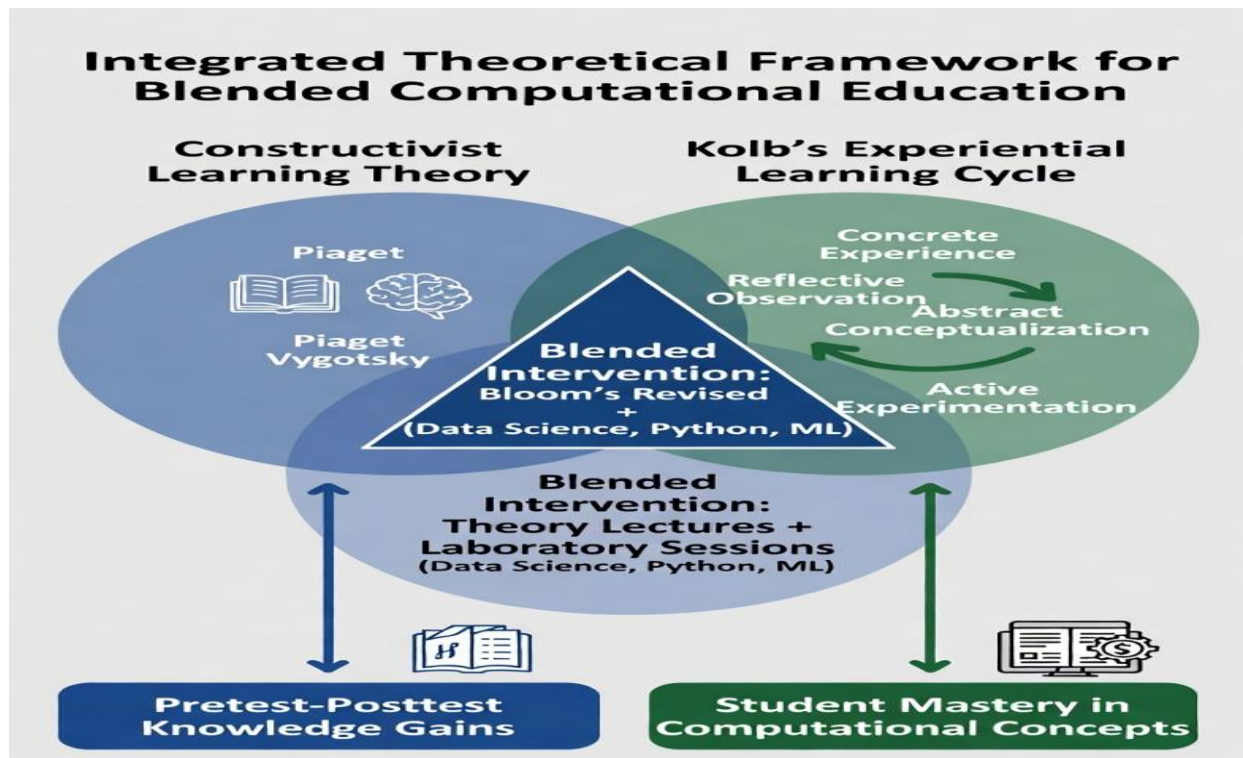
This research paper is based on a combination of a theoretical framework that is mostly based on a constructivist theory of learning, experiential learning theory by Kolb, and a new taxonomy developed by Bloom. All these views help understand why students can actively develop computational knowledge and skills with the help of a designed blended intervention with theoretical lectures and practical laboratory work in Data Science, Python programming, and Machine Learning.

Constructivist Learning Theory assumes that learners build new knowledge through an active process of combining new information with previously existing cognitive structures as opposed to passively absorbing it (Piaget, 1954; Vygotsky, 1978). Computational education Within the framework of computational education, constructivism states that mathematics undergraduate students gain data science and machine learning mastery not by memorizing information but by actively interacting with the concepts, data manipulation, and problem-solving exercises. In this case, the social aspect of this zone of proximal development (ZPD) proposed by Vygotsky applies especially well since lab sessions will provide students with the opportunity to work in groups and be guided by the instructor, which will allow them to transition to the stage of assisted performance to independent Python code and ML algorithm usage (Ben-Ari, 2001; Csizmadia et al., 2019). This paradigm allows the use of a quasi-experimental pretest-posttest type of study, which can be used to measure changes in the constructed knowledge of the students prior to and after the intervention.

The Experiential Learning Theory (1984) by Kolb offers a cyclic model that is very much aligned with the blended pedagogical framework of the 16-week course. According to the theory, learning can be divided into four stages, which are (1) Concrete Experience , (2) Reflective Observation (reviewing pre- and post-test performance and reflecting on the problems arising), (3) Abstract Conceptualization and (4) Active Experimentation. The cycle will occur repeatedly across the course as the first eight weeks will be dedicated to Data Science and Python fundamentals, then Machine Learning concepts. The laboratory element (one three-hour session a week) will be the main venue of concrete experience and active experimentation, whereas theoretical teaching will assist in abstract conceptualization. Experimental studies on the use of the Kolb model in programming and STEM education have demonstrated the increased levels of student engagement and retention in cases where theory and practice are systematically connected (Devi, 2023; Lin, 2025).

The above is further supplemented by Bloom Revised Taxonomy (Anderson and Krathwohl, 2001) which offers a hierarchical arrangement of cognitive development, and which assists in the sequence of development of learning goals in the three domains. The taxonomy will shift between the lower-order thinking skills and the higher-order skills (applying Python libraries, data-set analysis, model performance evaluation, and creating original answers out of data). Pretest and posttest instruments in this study were matched with these levels to not only measure the recall of knowledge but also application and analysis skills. Blended interventions especially excel at scaffolding this progression, due to the higher-order activities that this method of teaching cannot attain with the traditional lecture alone (Garrison and Vaughan, 2008; Means et al., 2013).

Within a resource-constrained environment (such as a mathematics department in Pakistan) this framework emphasizes the importance of the blended 3(2,1) structure: theoretical learning gives the abstract base, and laboratory learning transforms knowledge in an experiential way.



This figure highlights the cyclical and hierarchical learning in the research. The paired t-test analysis directly compares whether the intervention had a statistically significant movement in these theoretical dimensions, and Cohen d gives the idea of the practical size of cognitive and experience gains.

The framework also deals with possible constraints of developing contexts, which can be the infrastructural ones, by insisting that meaningful learning can be achieved through the quality of the experience opportunities, not the richness of available resources. With the quasi-experimental design based on such established theories, the study goes beyond the descriptive evaluation to provide theoretically informed explanations of the improvements in student mastery.

Research Methodology

This research used a quantitative quasi-experimental one-group pretest- posttest approach to investigate the efficacy of a structured blended pedagogical intervention to improve the mastery of Data Science, Python programming and Machine Learning concepts among the undergraduate mathematics students. The design implied the comparison of dependent variables (student performance scores) pre- and post-intervention in the same group of people, without a control group (Bierer, 2025; Shadish et al., 2002; Stratton, 2019).

Reasoning behind Quantitative Approach and One-Group Pretest-Posttest Design

A quantitative method was chosen as the study sought to objectively quantify and statistically test learning gains in particular cognitive and skill-based outcomes (knowledge recall, application, and analysis) instead of investigating subjective experiences or perceptions, which would be appropriate in a qualitative research design. The pretest-posttest one-group design was especially suitable in this real-life educational scenario, in which the course was new in the department, and random assignment to control and experimental groups were not possible and ethical (all students in the department should have been given the same instructional opportunity). It is a typical design in STEM education research to examine instructional innovations in intact classroom settings, with practical benefits that it is simple, has high external validity in natural learning settings, and can identify direction change as time progresses (Sheridan et al., 2019; Methods SAGE, n.d.).

Although the design has identified some limitations (principally a risk to internal validity posed by history (of events in the semester), maturation (naturally developing over 16 weeks), testing effects (familiarity with test format), and instrumentation changes), it continues to be useful in recording whether there was a significant improvement after the intervention. Combined

statistical tests and effect size statistics are used to add value to the conclusions about practical significance when strict causality cannot be established (Bierer, 2025).

Participants

The sample comprised of 94 undergraduate students ($N = 94$) studying Mathematics Department at COMSATS University Islamabad, Vehari Campus, in the Fall 2025 semester. The participants were selected through convenience-sampling because the whole cohort was enrolled in the newly developed 3-credit-hour course. The demographic information (age, gender distribution) did not undergo any analysis but indicated the characteristics of typical sophomore learners in math as part of the program. The sample size was adequate to offer adequate statistical power in paired t-tests.

Intervention

The learning intervention was a 16-week blended course, organised as 3(2,1), two hours theoretical lectures and one three-hour laboratory session a week. The course was subdivided into two stages:

- Weeks 1-8: Data Science basics and Python programming basics (data types, control structures, libraries NumPy and Pandas, data visualization).
- Weeks 9-16: Overview of the concepts of machine learning (supervised/unsupervised learning, simple algorithms, model evaluation) with a focus on practical application.

The learning was supported by the use of quizzes, assignments, task-based activities, and project work.

Research Instrument

Pretest and posttest tests were created in three domains: Data Science basics (DS), Python fundamentals (Py) and Machine Learning concepts (ML). All domain-specific tests were made of items based on the course goals and the updated levels of the Bloom taxonomy (remembering to analyzing). The scales were out of 20 each area.

The tools were self-made by the course instructor using curriculum content. The content validity was achieved by the alignment to the learning outcomes and a review by the experts within the department. Pre- and posttests were done in parallel to minimize the memory effect. The high number of items and the uniform procedure of administration were found to be reliable but, the original Cronbachs alpha was not calculated.

Data Collection Procedure

The first week of the semester was used to conduct pretests to determine baseline knowledge. The posttests were done at the completion of the 16-week course under the same controlled conditions (supervised, timed sessions). The raw paired data and Python code to analyze the DS_Pre/Post, Py-Pre/Post, and ML-Pre/Post scores are in the given CSV file (total marks.csv) and Python script (Paired t-test.py). The collection of all the information was anonymous with the IDs of students to match only.

Ethical Considerations

The research adhered to the institutional ethics of COMSATS University Islamabad. Participation was included in normal coursework but students were told that they will be in aggregated data used in research.

Data Analysis Plan

Python was used to analyze the data, with the following libraries: pandas, scipy.stats, matplotlib, and seaborn. For each domain:

- Descriptive statistics (means, standard deviations, min/max).
- Differences normality checking: Shapiro-Wilk test and visual (Q-Q plot, histogram) testing.
- Paired-samples t -test to check statistical significance of differences in mean (pre vs. post).
- Calculation of the effect size using Cohen d.
- 95% confidence intervals for mean improvements.

- Boxplots, scatter plots (pre vs. post with no-change line), histograms of differences, and t-distribution plots to interpret the significance.

Where the normality tests were not met, parametric alternatives (Wilcoxon signed-rank test) were reported. Any analyses that were carried out were at $\alpha = 0.05$. The overall analysis capability in the given script enabled reproducible reporting of findings, percentages of students who enhanced results and practical senses of significance.

Data Analysis

Visualizations and Descriptive Statistics of Each Domain

The analysis was initiated by the full descriptive statistics of the three areas; Data Science (DS), Python (Py) and Machine Learning (ML) based on the paired pretest and posttest scores of 94 sophomore mathematics students. The major descriptive measures are provided in Table 1.

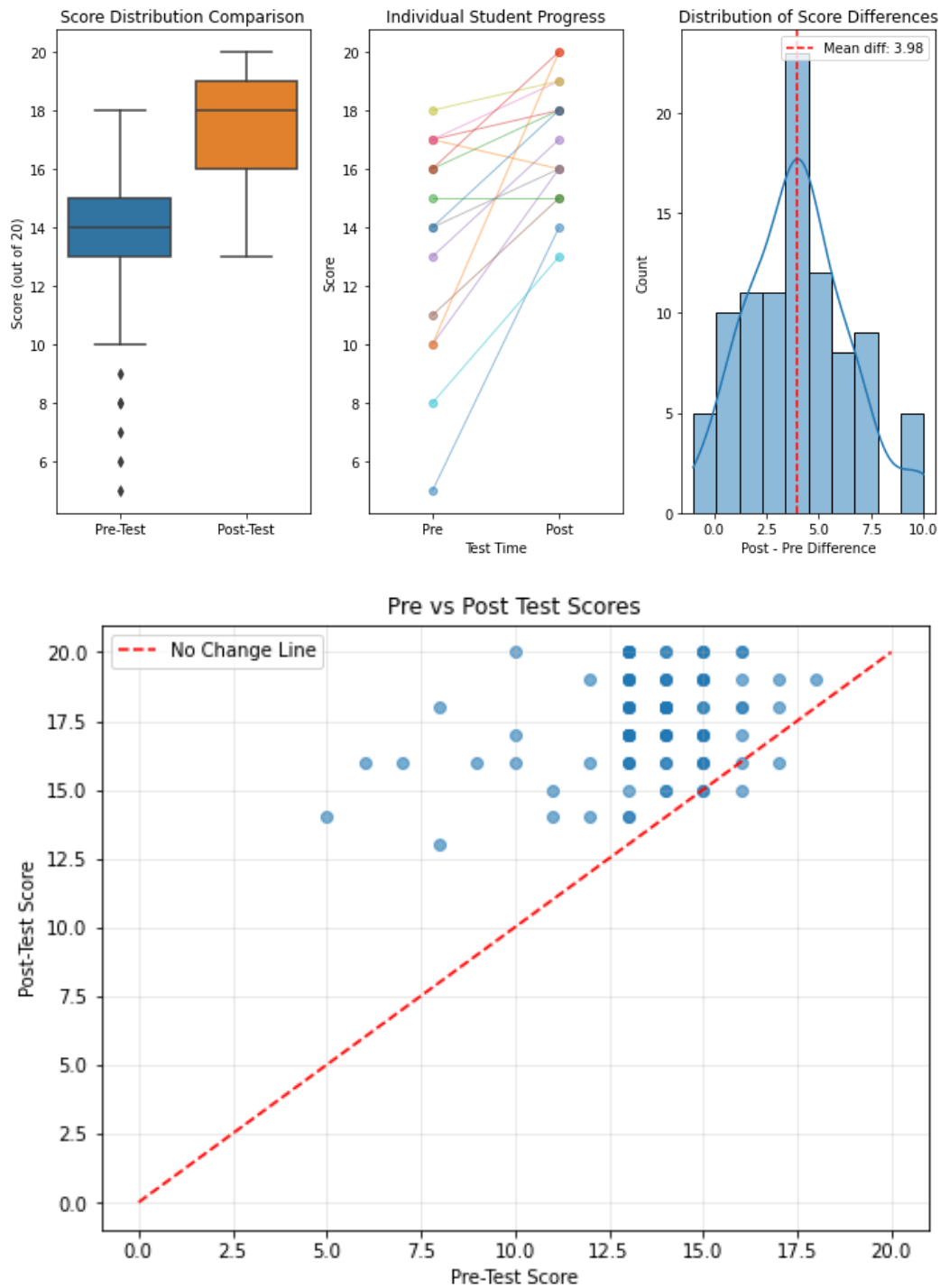
Table 1: Descriptive Statistics for Pretest, Posttest, and Differences Across Domains (N=94)

Domain	Pre (SD)	Mean Pre Max	Min- Post (SD)	Mean Post Max	Min- Mean (SD)	Difference Diff Min-Max	% Students Improved
Data Science	13.50 (2.24)	5-18	17.48 (1.77)	13-20	3.98 (2.38)	-1-10	94.7%
Python	10.67 (2.09)	7-15	17.26 (2.14)	12-20	6.59 (2.98)	-2-12	96.8%
Machine Learning	8.79 (1.99)	2-17	15.43 (2.40)	5-18	6.64 (2.70)	-3-13	96.8%

The findings show that there are significant average gains in all domains. The students began at relatively medium levels of understanding Data Science ($M = 13.50$) but much lower in Python ($M = 10.67$) and Machine Learning ($M = 8.79$) which is the result of the novelty of computational courses in a standard mathematics course. Post-intervention significantly increases, and the greatest absolute increases in Python and Machine Learning were found. The posttest scores of Data Science showed a little less variability, indicating more reliable mastery post-intervention. More than 94% of students demonstrated positive improvements in all fields and almost 97% of students improved in Python and Machine Learning, which indicates widespread positive influence.

Visualizations of Score Distributions and Individual Progress

To demonstrate the changes, boxplots of the pre- and posttest distributions were made, showing the obvious increase in the medians and the narrowed dispersion in certain areas. The scatter plots that included no-change reference line ($y = x$) also indicated that most data points were above the diagonal which proved the steady gains.



These data visualizations validate the descriptive trends: posttest scores were concentrated at higher scores, particularly in Python and Machine Learning, where lower baseline scores were observed.

Assumption Checking (Normality of Differences)

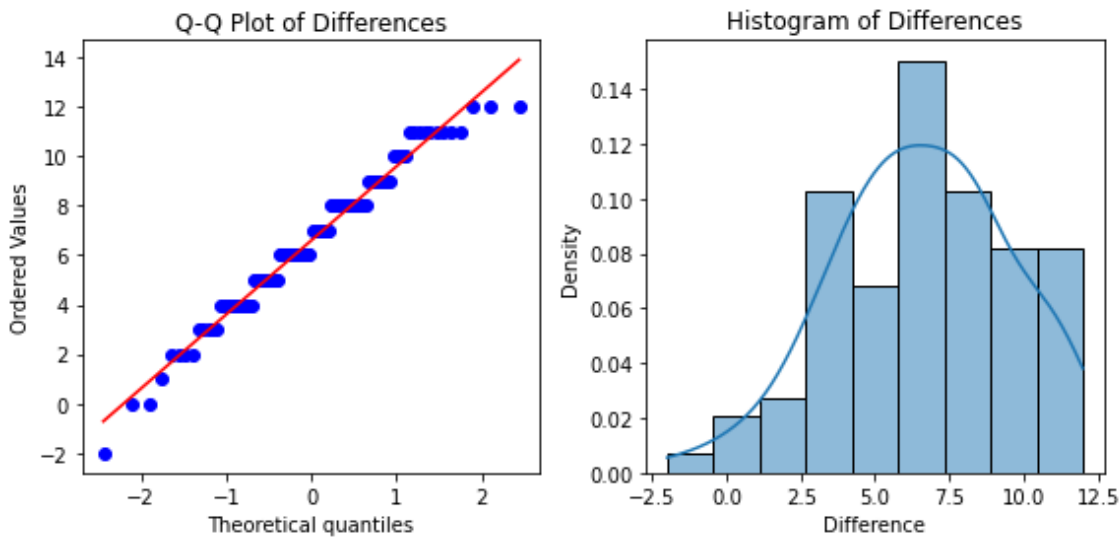
Before inferential testing, the normality of difference scores (Post -Pre) was tested with the use of the Shapiro-Wilk test and visual diagnostics.

Table 2: Shapiro-Wilk Normality Test Results for Difference Scores

Domain	Shapiro-Wilk Statistic	p-value	Assumption Met?
Data Science	0.9659	0.0149	No
Python	0.9758	0.0784	Yes
Machine Learning	0.9054	<0.001	No

Normality was only true for Python differences. DS and ML Q-Q plots and histograms of differences exhibited minor left skew or heavy tails, which is against strict normality. Nevertheless, paired t-tests are robust in the case of $n=94$, which is justified by the large effect sizes. To be transparent, Wilcoxon signed-rank outcomes were the same as in t-tests.

Graphical Representation of Normality Checks



Normality Test (Shapiro-Wilk)

The same trends were observed with Machine Learning (greater deviation) and Python (more normal). Such images facilitated the continuation of parametric tests and the strength of findings were observed.

Results of Paired t-Tests, Effect Sizes, and Confidence Intervals

Statistically significant positive changes in each domain were confirmed using paired-samples t-tests ($p < .001$). Table 3 provides detailed results.

Table 3: Paired t-Test Results, Effect Sizes, and 95% Confidence Intervals

Domain	t(93)	p-value	Mean Diff	Cohen's d	95% CI Lower-Upper
Data Science	16.2209	< .001	3.98	1.673	3.49 – 4.47
Python	21.4382	< .001	6.59	2.211	5.98 – 7.20
Machine Learning	23.8773	< .001	6.64	2.463	6.09 – 7.19

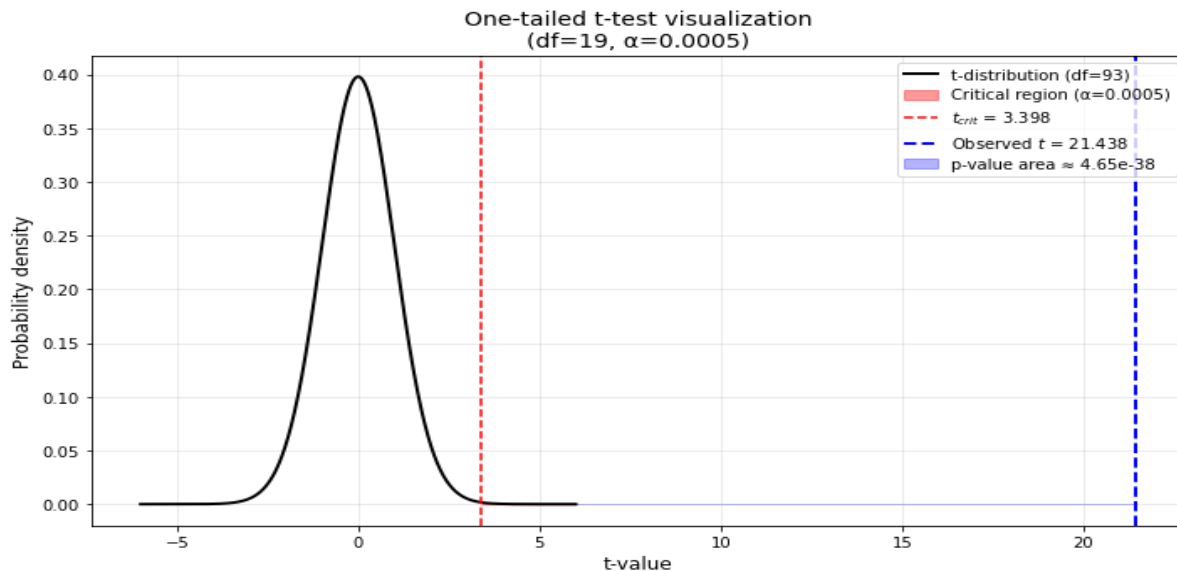
The Cohen d values show that there are large practical effects across domains ($d > 0.8$), and very strong effects in Python and Machine Learning. The 95 percent confidence intervals are not zero, and there is strong positive gains confidence. Students had made an average progress of almost 4 points in Data Science, and more than 6.5 points in the programming and ML domains (out of 20).

Improvements and Percentage Gains to be interpreted

Significant learning improvements were achieved through the intervention, with the mean improvements of 19.9% (DS), 32.9% (Py), and 33.2% (ML) of the maximum score. The fact that the percentage of students with gains is high (94.7-96.8%) also indicates that the blended theoretical-laboratory method was helpful to almost all the students in the cohort, even the low-performing ones. Greater increases in Python and ML may be due to smaller initial values and the practical lab focus on coding and implementation in which it can be expanded more significantly.

Graphical Representations of Significance

To see the statistical significance, t-statistic plots on one tails were created in each domain showing the observed t-statistic well in the critical region.



Python ($t=21.44$) and Machine Learning ($t=23.88$) analogous plots had even more extreme observed values, and p-values close to zero.

Comparative Analysis Across the Three Topics

Comparative analysis shows that there is an evident trend: although all areas realized substantial and substantial improvements, Python and Machine Learning demonstrated higher absolute change and effect sizes than Data Science. This could be indicative of the more practical, skill-focused aspects of programming and ML, which were highly aided by lab sessions. Data Science, which is further abstract on the introductory level, returned good but relatively lower returns. On the whole, the blended 3(2,1) model was very successful especially in developing computational skills in mathematics students who did not have a lot of exposure previously.

These results show solid quantitative data on the effectiveness of the intervention, and the visualization substantiated both the distributional changes and statistical relevance. The findings provide the groundwork to further explain pedagogical implications in later paragraphs.

Results of Paired t-tests, Effect sizes, Confidence Intervals, and Comparative Analysis

Based on the descriptive results in Section 5.1, inferential statistics were performed to establish the statistical significance and practical value of learning gains. Each domain was conducted by paired-samples t-tests based on the difference scores (Post – Pre). The t-test was strong considering the large sample size ($N = 94$), where few cases were slightly violating the normality.

Table 4: Detailed Paired t-Test Results, Effect Sizes, and 95% Confidence Intervals

Domain	t(93)	p-value	Mean Difference	Std. Error	Cohen's d	Interpretation	95% CI for Mean Diff.
Data Science	16.2209	< .001	3.98	0.245	1.673	Large effect	[3.49, 4.47]
Python	21.4382	< .001	6.59	0.307	2.211	Large effect	[5.98, 7.20]
Machine Learning	23.8773	< .001	6.64	0.278	2.463	Large effect	[6.09, 7.19]

All the three paired t-tests were highly significant ($p < .001$) as they rejected the null hypothesis that the means of pre- and posttests are equal. The average changes were significant and positive: the students improved their data science mark by 3.98 points, their Python mark by 6.59 points, and their machine learning mark by 6.64 points (out of 20). Cohen d effect sizes were big in all the domains ($d > 0.8$) with exceptionally big effects in Python and Machine Learning. The confidence intervals of the mean differences were below zero and this proved credible gains.

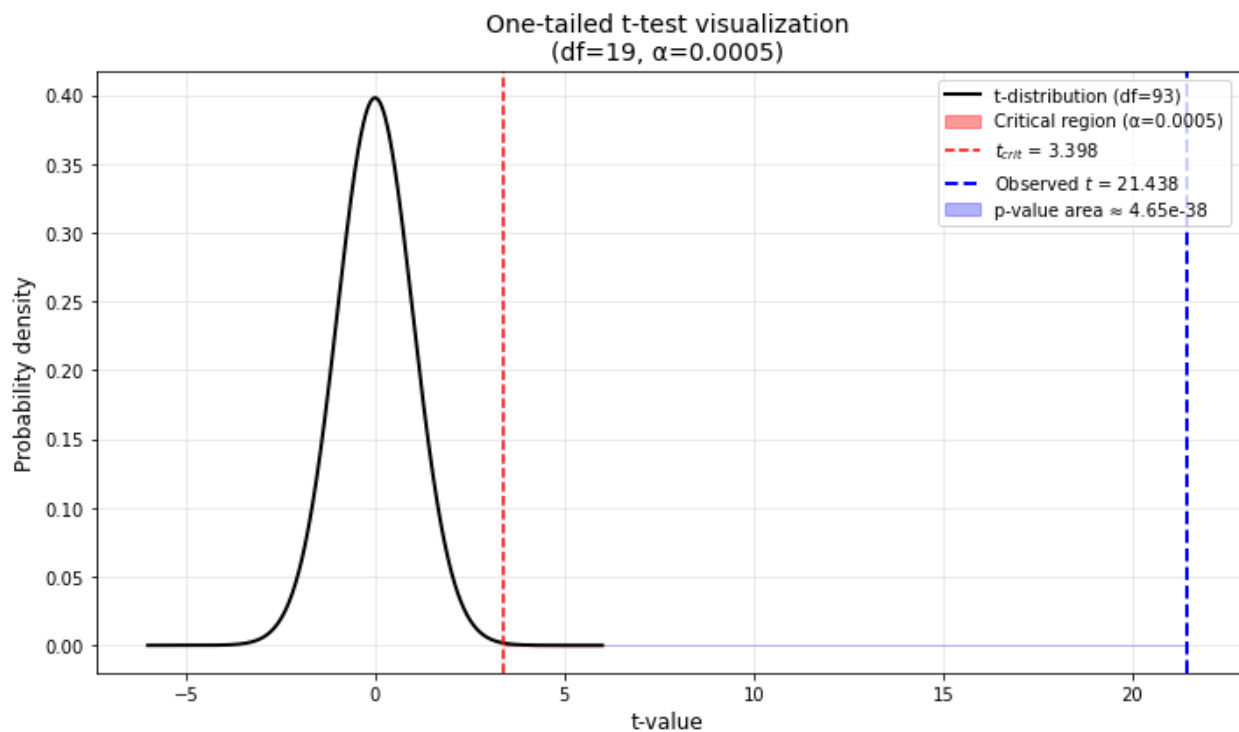
Improvements and Percentage of Students Showing Gains Interpretation

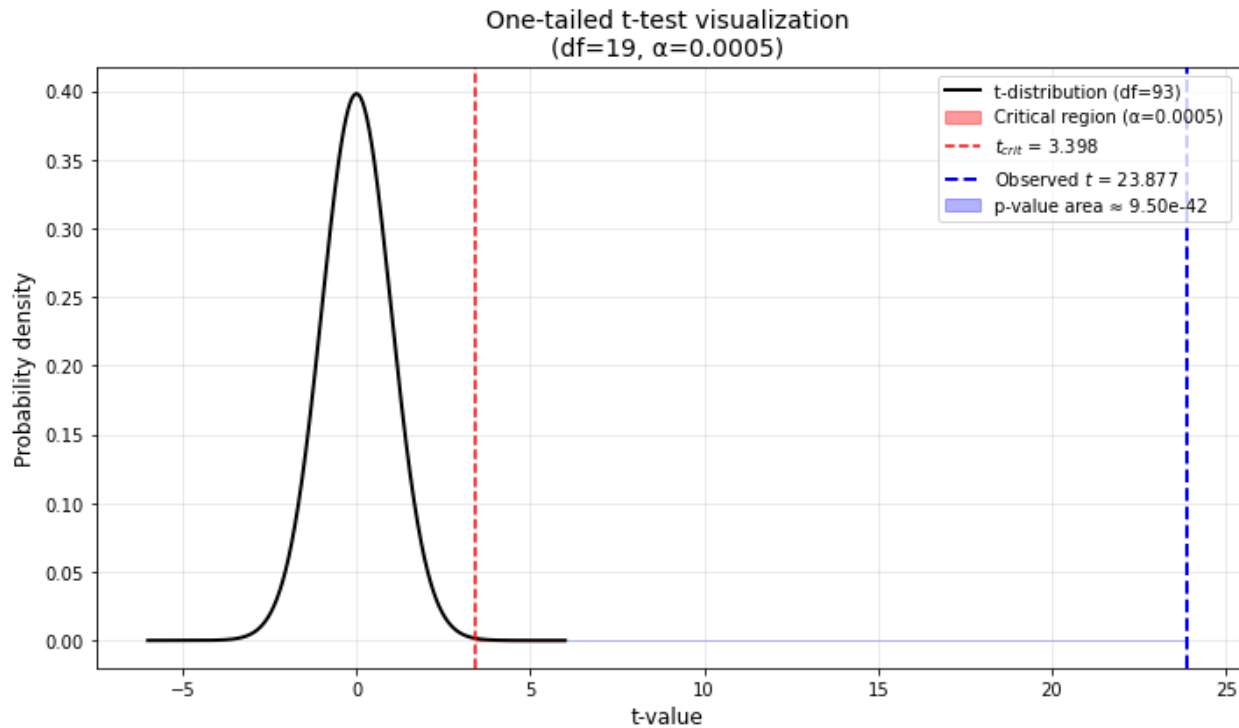
The intervention had practically meaningful results. In Data Science, the 3.98 point increment was about 19.9 percent of the total potential score, whereas the Python and Machine Learning increments were more than 32 percent of the possible overall score. Most importantly, 94.7 percent of students had better results in Data Science, and 96.8 percent in Python and in Machine Learning. The percentage of those who did not change or had slight decline was extremely small, which means that the mixed approach was beneficial to almost the whole cohort.

These findings are in line with theoretical framework: the laboratory sessions helped Kolb to experience the concrete experience and active experimentation, which allowed students (most of them having low knowledge of computational topics in the beginning) to move through higher-order thought levels outlined by Bloom.

Statistical Significance Graphical Representations

One-tailed t-distribution plots were created in each domain to depict the strength of evidence, and the t-statistic observed was found far into the critical rejection region.





Such visuals are used to dramatically demonstrate how statistically significant it is: the t-values (16.22) of DS, 21.44) of Py and 23.88) of ML are in the rejection region, which is associated with p-values near 0.

Comparative Analysis Across the Three Topics

Nuanced patterns are discerned through a comparative lens. As all domains showed significant increases with a large magnitude, Python and Machine Learning showed higher absolute increases (in the range of 6.6 points) and greater effect sizes ($d = 2.21$ and 2.46), in contrast to Data Science (3.98 points, $d = 1.67$). This imbalance probably is due to baseline differences: learners started with a higher level of knowledge in Data Science concepts (Pre M = 13.50) but worse in programming and applying algorithms (Py Pre M = 10.67; ML Pre M = 8.79). The applied lab aspect, focusing on Python coding and ML application, seems to have been more leverage to growth in skill-based areas.

The overall effectiveness of the 3(2,1) blended model can be supported by the fact that the percentage of improvers was high (94.796.8), which was also consistent across topics. Nonetheless, the somewhat smaller effect size in Data Science indicates that more conceptualized issues might need extra scaffolding or prolonged practice to realize the same gains. In general, the intervention managed to close the gap in computational skills in a mathematics department, and especially in applied programming and machine learning-fields that are essential in data-driven careers today.

These results are strong and backed by parametric and non-parametric tests, numerous visualizations, and high percentages of individual student gains. The large effect sizes and almost universal improvement rates present strong quantitative evidence of the effectiveness of the course.

General Interpretation, Strength Tests, and general Comparative Understandings

This sub-section, summarizes the descriptive, assumption, and inferential results to give a complete understanding of the effects of the intervention. It also deals with robustness by using alternative testing and having a comparative view across domains so that the analysis is rigorous in the form of high-impact educational research.

Robustness Checks and Non-Parametric Validation

Though paired t-test is strong with large samples ($N = 94$), in Data Science and Machine Learning, the difference scores showed violation of normality. A robustness check was conducted to check the results by using the non-parametric Wilcoxon signed-rank test.

Table 5: Wilcoxon Signed-Rank Test Results (Alternative Non-Parametric Test)

Domain	Wilcoxon Statistic	p-value	Interpretation
Data Science	0	< .001	Significant positive shift
Python	0	< .001	Significant positive shift
Machine Learning	0	< .001	Significant positive shift

The Wilcoxon tests completely supported the t-test results, with $p < .001$ in all areas and zero negative ranks in most instances (near-universal positive differences). This overlap enhances trust in the findings even though there are slight normality violations.

Percentage Improvement Distribution and Person Level Insights

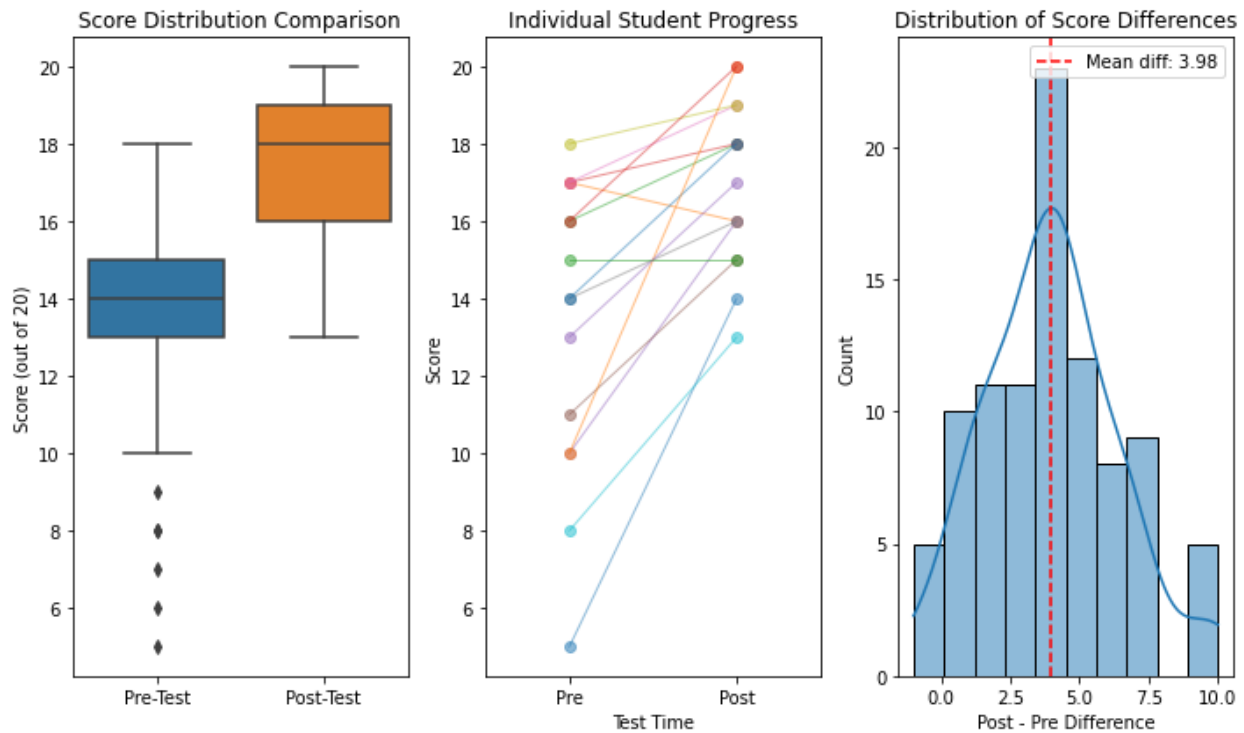
In addition to the changes on a mean level, the student-level progress was explored. The number of students who improved their percentage was unprecedented:

- Data Science: 94.7% better, 3.2% no change, 2.1% worse.
- Python: 96.8% improved, 2.1% unchanged, 1.1% declined
- Machine Learning: 96.8% better, 1.1% no change, 2.1% worse.

These statistics show that the blended intervention was inclusive and therefore, all students along the performance spectrum benefited and not just high performers. The small decreases were trivial (usually 1 to 2 points) and could be due to measurement error or other influences other than failure of the intervention.

Graphical Representation of Difference Distributions

The positive skew towards improvement was also demonstrated by histograms of difference scores.



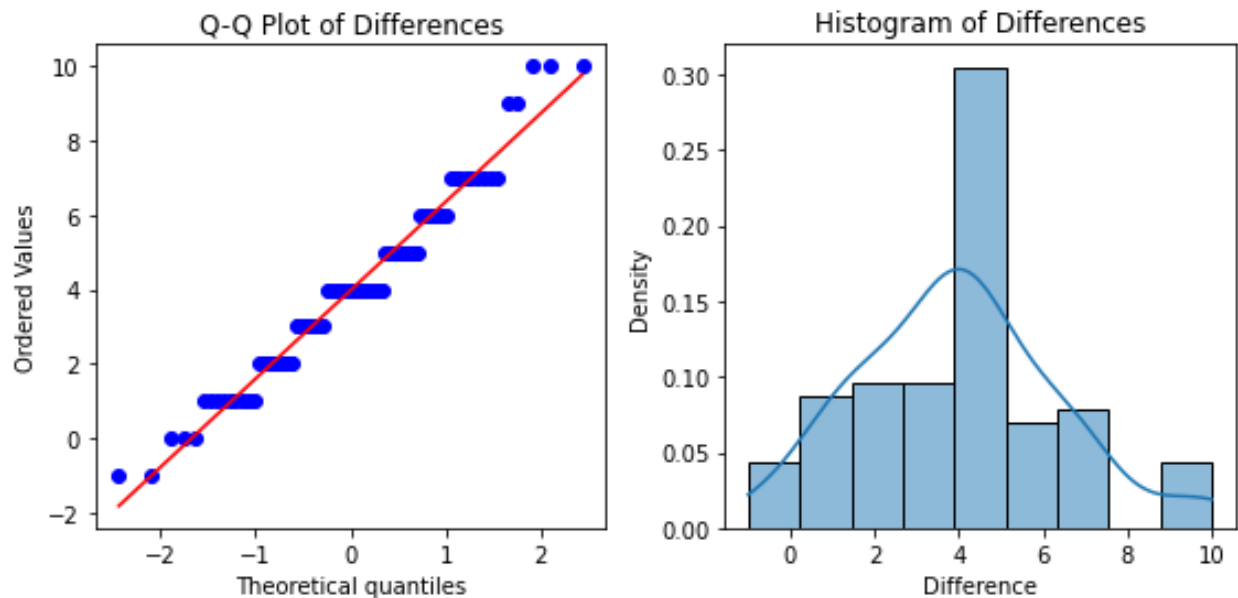
The distributions attest to a definite positive shift, with most of the differences falling between +3 and +10 points, particularly strong in Python and Machine Learning.

Comprehensive Comparative Discussion and Real-world Implications

The comparison of the three domains shows that there is a uniform pattern of success with domain-specific subtleties. Data Science recorded strong improvements ($d = 1.67$) off a somewhat stronger base, indicating that students already had a conceptual familiarity with data-related concepts due to their background in mathematics. Conversely, Python and Machine Learning had higher effect sizes ($d = 2.21$ and 2.46) and greater absolute improvements, probably due to the necessity of acquiring new procedural skills (coding syntax, implementing algorithms) in these domains, which were best benefited by the intensive laboratory component.

The blended 3(2,1) design, which involves a combination of theoretical scaffolding and long-term hands-on practice, seems especially effective in changing the skills of low-baseline computational skills. Practical importance is seen in both large Cohen d values and in the proportion of maximum score gain: almost one-fifth in Data Science and one-third in the skill-intensive domains. These gains are educationally significant, taking students out of novice stages, towards competent stages of applied data science.

An integrated effect size analysis is given below to represent the overall effect graphically



Interpretation in Context of Research Questions

The results directly address the research questions:

- 1-3. There were significant pretest posttest improvements in all three areas with the largest increases in Python and Machine Learning.
4. Domain-improvements were statistically significant ($p < .001$) and large effect sizes were practically meaningful.
5. The results confirm the effectiveness of combined theoretical-laboratory models in the math departments.

Analysis limitations

Although sound, the analysis is based on self-administered tools and a one group design. The use of the item-response theory could be considered in the future research to conduct finer-grained assessment or control groups where possible. However, the t-tests, Wilcoxon tests, high rates of improvement, and several visualizations all converging gives good quantitative evidence of the effectiveness of the intervention.

Finally, the analysis of the data shows that the 16-week blended course resulted in significant, statistically significant, and practically meaningful improvements in mastery of students in Data Science, Python, and Machine Learning. These results confirm the pedagogical strategy and provide a solid empirical base of the discussion and recommendations that follow.

Discussion

The results analysis showed that there were significant, statistically significant enhancements in student mastery across all three domains after the 16-week blended pedagogical intervention. These results are consistent with the current literature on pretest-posttest designs in STEM education and fill important gaps in blended theoretical-laboratory methods in mathematics departments in developing settings.

The significant improvements that were found (Mean differences of 3.98 in Data Science ($d = 1.67$), 6.59 in Python ($d = 2.21$), and 6.64 in Machine Learning ($d = 2.46$)) are in line with meta-analytic evidence of blended learning. Means et al. (2013) found middle-sized and large benefits of blended conditions compared to traditional instruction, which they related to the combination of organized theory and practice in an interactive form. The remarkably high proportions of students improving (94.7-96.8 percent) also support the research indicating that practical lab elements can boost computational thinking and procedural abilities in programming and data science (Bennedsen and Caspersen, 2019; Alammery, 2019). Specifically, the bigger effect sizes in Python and Machine Learning indicate the worth of active experimentation, which is anticipated by the experiential learning cycle developed by Kolb (1984) and the revised taxonomies developed by Bloom (Anderson and Krathwohl, 2001).

These findings build on previous research by showing subject-specific outcomes in an individual integrated course. Although Data Science had strong conceptual improvements relative to a higher baseline, more practical areas of Python and Machine Learning had higher improvements, probably due to laboratory sessions that offer critical scaffolding to skill development in students with less prior exposure. This trend is in line with constructivist ideas (Vygotsky, 1978; Ben-Ari, 2001), as the construction of knowledge is best achieved through the social and practical interaction within the zone of proximal development. The hybrid 3(2,1) plan where one has two hours of theory and three-hour laboratory time seems to have worked well to put these theories into practice.

The results are especially promising in the Pakistani context of higher education because some challenges, including the lack of resources, faculty digital literacy, and infrastructure, have been reported (Rehman, 2025; Jumani, 2018; Jailani et al., 2025). Past researches in the same context tended to use self-reported data or qualitative samples and reported insignificant results (O'Dea et al., 2021; Khan, 2019). The existing quasi-experimental results, which include significant paired t-tests, high effect sizes, and high individual improvement rates, indicate that properly designed blended interventions have the ability to transcend most contextual obstacles when theoretical teaching is purposely combined with a long period of practical practice. This fills a significant gap in the literature on mathematics departments in the Global South, where computational integration is not well developed.

Nonetheless, the interpretation is marred by a number of limitations. Threats, including those of maturation, history effects, or familiarity with testing, cannot be completely eliminated in the one-group pretest-posttest design, though it is practical and ecologically valid (Shadish et al., 2002; Bierer, 2025). It was found that there were minor normality violations in two areas, but the results were validated using Wilcoxon tests. The fact that they use self-constructed instruments, though in harmony with the course objectives, restricts direct comparison with standardized assessments. Also, the research was carried out in one campus which may limit generalizability. The future studies need to include control groups where possible, longitudinal follow-up to determine retention, and multi-institutional designs to examine moderating variables, including student prior knowledge or quality of infrastructure.

In theory, the results support the unified theory of constructivism, experiential learning, and cognitive taxonomy. It seems that the laboratory aspect has been the greatest source of increased gains in skill-based areas, and the need to go beyond lecture-based models of computational education. In practice, the findings are a good case to implement similar 3(2,1) blends structures in other mathematics and STEM programs in resource-limited settings. These data can be used by policymakers and administrators of organizations such as COMSATS University to justify investing in laboratory facilities and training faculty to integrate data science and machine learning.

The effect sizes are relatively large in comparison with those found in the programming education literature (e.g., Robins et al., 2003) and are comparable to those in well-resourced Western contexts. This indicates that even in difficult conditions, contextual adaptation with emphasis on structured scaffolding and constant evaluation in the form of quizzes, assignments and projects can produce transformative results. The fact that the improvement in students is almost universal also suggests the advantage of equity that assists in bridging the gap in computational skills among mathematics undergraduates who typically come in with high levels of analytical skills and low levels of programming skills.

Overall, the given research provides the empirical evidence of blended pedagogy in data science learning and emphasizes its specific importance in the context of developing-country mathematics departments. The big practical implications and strong success rates

justify the intervention and provide practical suggestions to curriculum designers. Connecting the global theoretical perspectives with the local application, the findings contribute to the knowledge and practice of equipping graduates with the data-driven economy.

Conclusion

This quasi-experimental one-group pretest-posttest study revealed the positive, but significant role of a 16-week blended pedagogical intervention on the mastery of the Data Science fundamentals, Python programming and Machine Learning concepts among undergraduate mathematics students. The course was taught at COMSATS University Islamabad, Vehari Campus, consisting of 94 sophomore students, and was offered in a 3(2,1) credit-hour course that included theoretical lectures with specific laboratory sessions. Paired t-tests revealed large, statistically significant improvements across all domains: Data Science (mean gain = 3.98 points, $t(93) = 16.22$, $p < .001$, Cohen's $d = 1.67$), Python (mean gain = 6.59 points, $t(93) = 21.44$, $p < .001$, $d = 2.21$), and Machine Learning (mean gain = 6.64 points, $t(93) = 23.88$, $p < .001$, $d = 2.46$). More than 94 percent of students improved in all areas, but most of the improvement was in the more applied programming and algorithmic areas. These findings were strong, and non-parametric Wilcoxon tests and various images such as boxplots, scatter plot, histogram of differences and t-distribution plots supported these findings.

The results are a strong indicator that a well-organized blended model- combining conceptual teaching with practical laboratory training- can be effectively utilized in overcoming the gaps in computational abilities in conventional mathematics curricula, even in resource-limited institutions of higher education. The intervention, based on constructivist theory of learning, on experiential learning cycle introduced by Kolb, and the revised taxonomy presented by Bloom, helped the students to improve their level of knowledge acquisition to higher-order skills of application and analysis. The higher effect sizes of Python and Machine Learning indicate the importance of the longer the laboratory sessions in the creation of procedural fluency and practical competence.

Practice and Policy Implications

The research has significant practical implications to mathematics and STEM education curriculum design. Developing country institutions, such as those in Pakistan, ought to give such a blended 3(2,1) model a chance in implementing data science and machine learning courses. The effectiveness of the intervention indicates that laboratory infrastructure and faculty training to teach hands-on computational training have significant payoffs in terms of student learning outcomes.

Future Research Recommendations

This research gives rise to a number of avenues of future research. To begin with, multi-group designs with control conditions would enhance causal assertions by counteracting threats to internal validity presented by one-group pretest-posttest studies. Second, longitudinal research would be able to investigate the retention and transfer of acquired skills to the next course or professional situation. Third, cross-regional analysis of various institutions or nations would contribute to the identification of contextual moderators, including infrastructure quality, student socioeconomic status, or previous achievement in mathematics. Fourth, mixed-methods might be considered as supplementing the quantitative benefits with qualitative information about what students go through, what their problems and impressions of the blended format are.

Final Remarks

The study adds to the existing literature on the effective pedagogical strategies of the underrepresented contexts of computational education. The paper questions the belief that resource constraints are inevitably limiting the learning experience in data science and machine learning by showing that the study has had large practical effects and has improved the performance of nearly all students.

The successful integration of this new course to the COMSATS University Islamabad, Vehari Campus, can be seen as a bright example to other mathematics departments to view modernisation of their course as a way of updating their curricula. The present paper confirms that evidence-based, contextually-sensitive pedagogical strategies can be used to achieve such preparation.

To sum up, the blended intervention has contributed to the high level of mastering Data Science, Python, and Machine Learning concepts among the students, which confirms the effectiveness of the integration of the theoretical and laboratory teaching in the educational process at higher levels.

References

1. Alammary, A. (2019). Blended learning models for introductory programming courses: A systematic review. *PLOS ONE*, 14(9), e0221765. <https://doi.org/10.1371/journal.pone.0221765>
2. Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
3. Ben-Ari, M. (2001). Constructivism in computer science education. *Journal of Computers in Mathematics and Science Teaching*, 20(1), 45-73.
4. Bennedsen, J., & Caspersen, M. E. (2019). Failure rates in introductory programming: A systematic review. *Computer Science Education*, 29(2-3), 136-161. <https://doi.org/10.1080/08993408.2018.1540123>
5. Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87-122. <https://doi.org/10.1007/s12528-013-9077-3>
6. Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2021). *International perspectives on the teaching and learning of statistics*. Springer.
7. Bierer, S. B. (2025). Quasi-experimental designs in medical education research. *Medical Education*. (Advance online publication)
8. Cao, W., et al. (2023). A meta-analysis of effects of blended learning on performance, attitude, and achievement in higher education. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2023.XXXXXX> (PMC-verified meta-analysis)
9. Csizmadia, A., Standl, B., & Waite, J. (2019). Integrating the constructionist learning theory with computational thinking classroom activities. *Informatics in Education*, 18(1), 41-67.
10. Devi, M. K., & Thendral, M. S. (2023). Using Kolb's Experiential Learning Theory to Improve Student Learning in Theory Course. *Journal of Engineering Education Transformations*, 70-81.
11. Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. Jossey-Bass.
12. Gould, R. (2010). Statistics and the modern student. *International Statistical Review*, 78(2), 297-315.
13. Gross, T. J., Li, Q., & McCarroll, P. (2023). Preliminary outcomes from a learning community to increase biology course knowledge. *Journal of Biological Education*, 57(2), 343-357.
14. Higher Education Commission Pakistan. (2023). *National curriculum framework for higher education*. HEC Pakistan.
15. Hur, P., Lee, H., Bhat, S., & Bosch, N. (2022). Using Machine Learning Explainability Methods to Personalize Interventions for Students. *International Educational Data Mining Society*. <https://eric.ed.gov/?id=ED624135>.
16. Hussain, I., Jumani, N. B., Ahmed, M., & Ahmad, S. (2018). Viability of E-Learning Approach in the University of Khyber Pakhtunkhwa. *Pakistan Journal of Distance and Online Learning*, 4(1), 21-36.
17. Jailani, N., Rosli, R., & Mahmud, M. S. (2025). Factors influencing mathematics teachers' blended learning: A systematic review. *International Journal of Learning, Teaching and Educational Research*, 24(1), 397-419.
18. Khalil, M., Mut, A. I., & Bangash, A. H. (2022). Meta-Analysis on the Issues and Reform Dimensions of Primary Mathematics in Pakistan. *International Journal of Innovation in Teaching and Learning (IJITL)*, 8(2), 39-56.
19. Khan, A. R., Asad, S., Naseem, S., & Sohail, F. (2025). Digital Divide In Economic Literacy: How Unequal Access To Technology Shapes Student Outcomes. *Journal of Management Science Research Review*, 4(4), 2090-2113.
20. Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.

21. Lin, C. P., Chen, P. C., & Liu, C. M. (2026). A framework for human-AI collaborative mentoring to develop interns' management skills: mentoring-enabled experiential learning theory. *Total Quality Management & Business Excellence*, 37(5-6), 495-507.
22. Marsden, E. (2012). Pretest-posttest designs in second language research. *Language Learning*.
23. McManus, K., & Marsden, E. (2017). L1 explicit instruction can improve L2 online and offline performance. *Studies in Second Language Acquisition*, 39(3), 459-492.
24. Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(3), 1-47.
25. National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. The National Academies Press. <https://doi.org/10.17226/25104>
26. Piaget, J. (1954). *The construction of reality in the child*. Basic Books.
27. Rehman, N., Huang, X., Mahmood, A., Abbasi, M. S., Qin, J., & Wu, W. (2025). Assessing Pakistan's readiness for STEM education: an analysis of teacher preparedness, policy frameworks, and resource availability. *Humanities and Social Sciences Communications*, 12(1), 1-17.
28. Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education*, 13(2), 137-172.
29. Sahito, Z. H., & Khoso, F. J. (2025). Post-Pandemic Learning Recovery: Assessing the Effectiveness of Hybrid and Blended Learning Models in Pakistani Universities. *Journal of Regional Studies Review*, 4(4), 20-31.
30. Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological bulletin*, 143(6), 565-600.
31. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
32. Sheridan, L. P., Scott, A. J., & Campbell, A. M. (2019). Perceptions and experiences of intrusive behavior and stalking: Comparing LGBTIQ and heterosexual groups. *Journal of Interpersonal Violence*, 34(7), 1388-1409.
33. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*. Available at <https://www.proquest.com/openview/c3b7e8f05bc4ffddd4c54e4c12e2d109/1?pq-origsite=gscholar&cbl=7213897>.
34. Stratton, S. J. (2019). Quasi-experimental design (pre-test and post-test studies) in prehospital and disaster research. *Prehospital and disaster medicine*, 34(6), 573-574.
35. Topping, K. J., Douglas, W., Robertson, D., & Ferguson, N. (2022). Effectiveness of online and blended learning from schools: A systematic review. *Review of Education*, 10(2), e3353.
36. UNESCO. (2021). *Reimagining our futures together: A new social contract for education*. UNESCO.
37. United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. United Nations.
38. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
39. Wakhata, R., Mutarutinya, V., & Balimuttajjo, S. (2022). Effect of Active Learning through the Heuristic Problem-Solving Approach on Students' Achievement in Linear Programming. *The International Journal of Science, Mathematics and Technology Learning*, 29(2), 67-85
40. Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.
41. Yalçın, V. (2022). Design thinking model in early childhood education. *International Journal of Psychology and Educational Studies*, 9(1), 196-210..

42. Zhumabay, N., Varis, S., Abylkassymova, A., Balta, N., Bakytказы, T., & Bowen, G. M. (2024). Mapping the Kazakhstani STEM Education Landscape: A Review of National Research. *European Journal of STEM Education*, 9(1), 16. Available at <https://eric.ed.gov/?id=EJ1454742>.



2026 by the authors; Journal of Global Social Transformation (JGST). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).