

Refining final-year engineering outcomes at Mbeya University, Tanzania

Received: 31 December 2025

Accepted: 13 May 2026

Published online: 20 May 2026

Cite this article as: Katambara Z. Refining final-year engineering outcomes at Mbeya University, Tanzania. *Discov Educ* (2026). <https://doi.org/10.1007/s44217-026-01650-3>

Zacharia Katambara

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Refining Final-Year Engineering Outcomes at Mbeya University, Tanzania

Zacharia Katambara

Department of Civil Engineering,
College of Engineering and Technology
Mbeya University of Science and Technology,
Mbeya City,
Tanzania

ORCID: <https://orcid.org/0000-0001-9983-1207>

Email: zacharia.katambara@must.ac.tz

Abstract

Tanzania's Vision 2025 agenda emphasise rapid industrialisation, necessitating a technically proficient workforce, particularly in mechanical engineering, to sustain growth in manufacturing, mining, and infrastructure. This study evaluates the academic performance of 117 final-year mechanical engineering students at Mbeya University of Science and Technology (MUST), employing Exploratory Analysis (EA) and Principal Component Analysis (PCA) to uncover performance patterns and assess curriculum alignment with industrial competencies. Descriptive statistics revealed notable variability in course performance, while correlation analysis identified strong positive associations between GPA and key technical modules, notably ME 8401 (Fluid Power and Control) and ME 8411 (Automation and Robotics). PCA extracted five principal components, explaining 58.2% of the total variance, with Component 1 alone accounting for 30.4%, suggesting a dominant academic structure driven by technical proficiency. High factor loadings for ME 8411, ME 8401, and ME 8415 underscored their central role in student success. At the same time, low-loading courses such as ME 8408 (Industrial Practical Training III) and ME 8413 highlighted potential curricular misalignments. Based on these insights, the study concludes that academic success in mechanical engineering at MUST is tightly linked to performance in core technical subjects. To align educational outcomes with Tanzania's industrial goals, the study recommends targeted curriculum refinement of underperforming modules, enhancing high-impact courses with structured academic support, routine application of multivariate analytics for curriculum monitoring, faculty training in educational data analysis, and the establishment of formal industry-academic partnerships. These data-driven reforms aim to foster a technically adept graduate pool that meets the evolving demands of Tanzania's industrial landscape.

Keywords: Mechanical engineering education; principal component analysis; skills gap; industrialisation; workforce readiness.

Introduction

The Tanzanian economy is transforming, prioritising industrial production as a catalyst for growth and development. Anchored by the Tanzania Development Vision 2025, the country is transitioning from an agrarian to an industrialised economy, focusing on manufacturing, mining and infrastructure development (United Republic of Tanzania, 2021). This transition highlights the crucial need for a highly skilled workforce, particularly in mechanical engineering, to design, maintain, and innovate machinery essential for industrial processes (Bishop et al., 2021). Global examples from China, Malaysia and Germany offer valuable lessons in aligning engineering education with industrialisation goals. China's emergence as a worldwide industrial leader is attributed to its emphasis on engineering education and technological innovation (Wang & Zhou, 2020). Similarly, Malaysia's collaborative approach between academia and industry ensures that skills development meets market demands (Yusoff et al., 2018). Germany's dual education system integrates vocational training with theoretical learning and has proven effective in developing a workforce capable of sustaining industrial innovation (Roth et al., 2019).

Addressing the mechanical engineering skills gap is crucial for Tanzania's industrial ambitions, especially in key sectors such as manufacturing, energy, mining, and transportation. As the nation strives to become a semi-industrialised, middle-income economy by 2025, its Vision 2025 goal of closing this gap is essential for enhancing local technical capacity, increasing productivity, and reducing reliance on foreign expertise. Mechanical engineering, established at Mbeya University of Science and Technology (MUST) in 1985, remains a cornerstone of the country's industrial development strategy. A multivariate analysis of the performance of 2023/2024 final-year mechanical engineering students at the University offers critical insights into their readiness to meet industry demands, focusing on technical competencies, problem-solving abilities and adaptability (Zarate et al., 2017). Aligning engineering education with industrialisation goals requires evidence-based strategies, as recent studies at MUST demonstrate. Katambara (2025) employed the Principal Component Analysis (PCA) approach to evaluate Electrical and Electronics Engineering students, uncovering disparities across theoretical, project-based and specialised training dimensions. Similarly, Katambara and William (2025) applied PCA to Civil Engineering students, identifying three principal performance domains: core academic knowledge, specialised applied skills and advanced independent competencies. These findings underscore the potential of multivariate analysis in uncovering latent performance structures and curriculum misalignments. Mechanical engineering education, therefore, stands to benefit significantly from such analytical frameworks, enabling curriculum refinement that bridges the skills gap, fosters industry-ready graduates, and propels Tanzania's industrial transformation.

Multivariate analysis techniques, particularly PCA and EA, have become indispensable in educational research for analysing and interpreting complex datasets on students' performance. PCA, a robust dimensionality reduction technique, transforms original variables into orthogonal components that maximise variance, enabling the identification of hidden patterns and structures within the data. For instance, Borges et al. (2018) utilised PCA to predict students' performance and uncover implicit patterns in datasets that incorporated social and personal characteristics. The findings demonstrated that PCA effectively retained critical information, even after dimensionality reduction, thereby enhancing the interpretability of the educational data. Similarly, Giovannella et al. (2013) applied PCA within Web 2.0-based learning environments to examine the relationship between student activities and final academic performance, revealing that specific online activities predicted academic success and highlighting PCA's relevance in aligning performance indicators with modern educational technologies. Additionally, Twenefour et al. (2015) employed PCA to address inconsistencies in grading systems, proposing a novel performance metric derived from principal components, which offered a more nuanced evaluation of individual student differences and improved the fairness of assessments.

Complementing PCA, EA offers a flexible framework for identifying latent trends and relationships in data without requiring predefined hypotheses. Hu et al. (2015) demonstrated the effectiveness of integrating PCA into an EA framework for evaluating teaching quality in higher education. They validated their assessment tool using Cronbach's alpha, reinforcing the value of multivariate approaches in robust, statistically sound evaluations of educational programs. Collectively, these studies highlight the pivotal role of PCA and EA techniques in enhancing the understanding of student performance across diverse educational settings, thereby forming a strong foundation for the present study.

Across recent research, several core EA techniques have emerged as essential tools for analysing educational datasets. PCA remains a dominant method for dimensionality reduction, preserving critical variance while simplifying complex data structures (Borges et al., 2018; Giovannella et al., 2013). Complementing this, clustering techniques, particularly K-means clustering combined with the Elbow method and Silhouette scores, have been widely applied to uncover latent groupings and behavioural patterns among students (Sofia & Hema, 2024). Furthermore, correlation analysis and advanced visualisation methods have proven instrumental for pattern detection and mapping relationships between variables in large-scale educational studies (Maphosa et al., 2023). Finally,

exploratory factor analysis (EFA) has been extensively used for validating psychometric instruments, exemplified by the development and validation of the Student Online Learning Readiness (SOLR) model (Yu, 2014), thus highlighting the EA's dual contribution to exploratory insights and the strengthening of measurement precision within educational research. The purpose of this study is to apply multivariate analysis techniques, specifically PCA and complementary EA, to assess the academic performance of final-year mechanical engineering students at Mbeya University of Science and Technology (MUST). By identifying latent performance structures and potential curriculum gaps, the study aims to generate evidence-based insights that can inform curriculum refinement, enhance graduate readiness, and align mechanical engineering education with Tanzania's Vision 2025 industrialisation objectives.

Conceptual Framework: Outcome-based education and Data-Driven Evaluation of Learning Outcomes

Outcome-Based Education (OBE) organises curriculum design, teaching strategies, and assessment around clearly defined learning outcomes that students are expected to demonstrate upon completing a program. Spady (1994) argues that this approach shifts the emphasis from the mere delivery of course content to the achievement of measurable competencies. Within this framework, learning outcomes guide instructional planning and assessment practices to ensure that students acquire knowledge, technical abilities, and professional skills relevant to real-world contexts.

A central principle supporting OBE is constructive alignment, which ensures that learning outcomes, teaching activities, and assessment methods are coherently linked (Walker, 2005). Through this alignment, educational programs can more effectively evaluate whether students have developed the competencies required for professional practice. In engineering education, such alignment is particularly important because programs must demonstrate that graduates possess the analytical, technical, and problem-solving capabilities expected by industry and professional bodies (Harden, 1994).

In this context, quantitative educational analytics provide an important mechanism for evaluating how well academic outcomes reflect actual student performance patterns. Multivariate statistical techniques allow researchers to examine relationships among course results, uncover hidden performance structures, and identify areas where curriculum delivery may require improvement. By integrating such analytical approaches with the principles of OBE, educational institutions can move beyond descriptive evaluation toward evidence-based curriculum development. Consequently, applying advanced exploratory and dimensionality-reduction techniques to student performance data offers a systematic approach for identifying latent learning patterns, diagnosing curriculum strengths and weaknesses, and aligning engineering education outcomes with national industrial development goals.

Methodology

This study employed a quantitative, exploratory approach to analyse students' performance data using EA and PCA (Figure 1). The methodology was designed to uncover hidden patterns, reduce data dimensionality and enhance the interpretation of academic outcomes. A structured dataset, rigorous statistical procedures and ethical safeguards underpinned the research process.

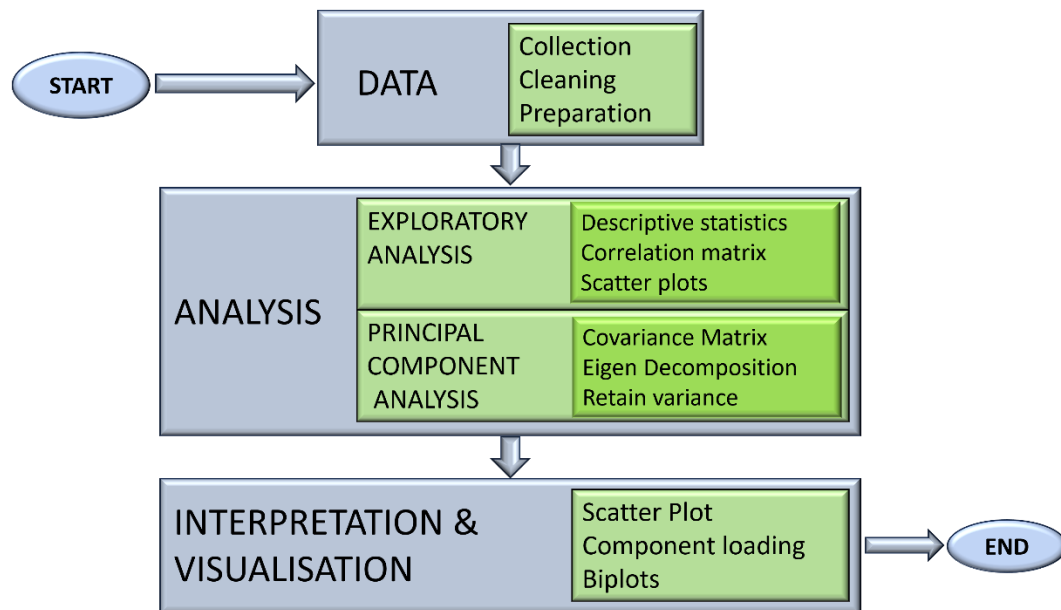


Figure 1: A flow diagram on the application of Exploratory Analysis (EA) and Principal Component Analysis (PCA) in the context of assessing student performance

Research Design

This study employed a quantitative and exploratory research design, utilising EA and PCA to investigate academic performance patterns among mechanical engineering students. The primary objective was to identify key factors influencing student outcomes, reduce data dimensionality, enhance interpretability and inform curriculum development (Abdi & Williams, 2010; Jolliffe & Cadima, 2016).

Population and Sampling

The target population consisted of 117 final-year mechanical engineering students enrolled at Mbeya University of Science and Technology during the 2023/2024 academic year. A purposive sampling approach was applied to select final-year students, as their accumulated academic records provided a comprehensive basis for multivariate analysis. This strategy was chosen to maximise variability in the performance measure and ensure a rich dataset suitable for PCA (Etikan et al., 2016).

Instruments

The primary research instrument was a structured academic dataset comprising final examination results, sourced from the University's Student Information Management System (SIMS) under the institutional authorisation. Data cleaning, transformation and statistical analysis were conducted using Jamovi and Microsoft Excel (Field, 2024). Data preparation included ensuring uniform scaling and handling missing values to meet the prerequisites for PCA.

Validity and Reliability

To enhance data validity, variables were standardised, accounting for PCA's sensitivity to the measurement scale (Hair et al., 2019). Reliability analysis was conducted using Cronbach's alpha and McDonald's omega to evaluate the internal consistency of the performance indicators, ensuring that the items collectively measured a coherent construct of student academic performance. Additionally, the dataset's suitability for PCA was validated using the Kaiser-Meyer-Olkin (KMO) test and Bartlett's Test of Sphericity (Tabachnick & Fidell, 2013), ensuring sampling adequacy and verifying that the correlations among variables were sufficiently strong for PCA.

Statistical Treatment of Data

The statistical analysis proceeded in two main phases. The first phase involved Exploratory Analysis (EA), which was employed to investigate variable distributions, identify outliers, and generate descriptive statistics, as well as correlation matrices and scatterplots, for initial pattern detection. This

aligns with Tukey's (1977) foundational view of EA as a strategy for revealing structure, detecting anomalies, and guiding subsequent formal analyses without imposing rigid assumptions. Principal Component Analysis (PCA) was subsequently applied, involving the computation of a covariance matrix and eigen decomposition to extract principal components. Retention of components was based on explaining at least 85%–90% of the cumulative variance. Component loadings were analysed to interpret variable contributions, while scree plots and biplots facilitated visualisation of the data structure (Abdi & Williams, 2010).

Ethical Considerations

This study was conducted in strict adherence to established ethical principles to safeguard the rights, privacy, and confidentiality of all participants. Research clearance was obtained from Mbeya University of Science and Technology, confirming compliance with the University's ethical guidelines and institutional research requirements. In addition, authorisation was sought from relevant government authorities to ensure alignment with national research regulations and lawful data collection procedures. To protect participants' rights, anonymity and confidentiality were rigorously maintained throughout the study. No personally identifiable information was collected or disclosed, and all datasets were fully anonymised to prevent the identification of individual students. Confidentiality was further ensured through secure data storage and restricted access limited to authorised personnel only. Data were handled exclusively for research and institutional development purposes. Overall, the study complied with recognised ethical standards for research involving human participants. The ethical approach adopted aligns with Resnik's (2018) framework, which emphasises informed consent, confidentiality, and responsible data stewardship as fundamental principles in ethical research practice.

Results and Discussion

This section presents the findings and discussion, utilising EA and PCA, in alignment with the study's objectives. EA examined data distributions, variability and correlations to identify performance patterns and areas needing intervention. PCA reduced data dimensionality, revealing key academic constructs, such as technical skills and project-based competencies that significantly influenced outcomes. EA and PCA provided a comprehensive view of students' performance, uncovering hidden trends and interdependencies across courses. This dual-method approach supports data-driven insights for curriculum refinement, ensuring that mechanical engineering education aligns with Tanzania's industrialisation goals. This approach aligns with findings from Tanzanian studies, which emphasise the importance of aligning technical education curricula with current industrial skill demands (Toroka & Kafanabo, 2024). Similar dual-method strategies, such as integrating both quantitative course performance data and qualitative employer feedback, have been shown to enhance curriculum relevance in engineering programs, thereby bridging the gaps between academic content and workplace expectations (Wismansyah et al., 2024).

Dataset Suitable for Principal Component Analysis: KMO and Bartlett's Tests

The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was used to assess the suitability of the dataset for Principal Component Analysis (PCA), evaluating the proportion of variance among variables that might represent common variance. Table 1 shows that the overall KMO value was 0.68, which falls within the acceptable range (0.60–0.70), indicating a mediocre yet adequate level of sampling adequacy for factor analysis. However, individual Measure of Sampling Adequacy (MSA) values revealed variation across variables, with some, such as ME 8401 (0.86), ME 8409 (0.89) and ME 8415 (0.88), showing strong adequacy, suggesting these items share substantial common variance with others and are well-suited for inclusion in PCA. In contrast, variables like ME 8407 (0.26) and ME 8417 (0.20) fell below the critical threshold of 0.50, indicating weak shared variance and potentially a poor fit within the factor structure. These low-MSA variables may introduce noise or distortions in the PCA and should be reviewed for possible exclusion or reconsideration in model refinement (Field, 2024).

This mix of adequate and poor MSA values highlights the need for careful variable selection to enhance the interpretability and validity of PCA results.

Table 1: Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy

Description	MSA	Description	MSA
Overall	0.68	ME 8408	0.81
GPA	0.62	ME 8409	0.89
CS 8406	0.59	ME 8410	0.77
ME 8401	0.86	ME 8411	0.76
ME 8402	0.78	ME 8412	0.64
ME 8403	0.61	ME 8413	0.71
ME 8404	0.56	ME 8414	0.74
ME 8405	0.49	ME 8415	0.88
ME 8406	0.77	ME 8416	0.7
ME 8407	0.26	ME 8417	0.2

Bartlett's Test of Sphericity was conducted to evaluate whether the correlation matrix significantly deviates from an identity matrix, thus determining the dataset's suitability for factor analysis. Table 2 shows that the test yielded a chi-square value of $\chi^2(171) = 843.58$, with a p-value of less than 0.001, indicating strong statistical significance. This result confirms the presence of sufficient intercorrelations among the variables, validating the use of Principal Component Analysis (PCA) for dimensionality reduction and pattern recognition (Tabachnick & Fidell, 2013). The test outcome aligns with established guidelines, which recommend proceeding with PCA when Bartlett's test is significant, ensuring that the data structure is appropriate for extracting meaningful components.

Table 2: Bartlett's Test of Sphericity

χ^2	df	p
843.58	171	<.001

Item Reliability Analysis of Academic Performance Indicators

Table 3 presents the item reliability analysis, which demonstrates strong internal consistency among the performance indicators, with Cronbach's alpha of .79 and McDonald's omega approximating .84–.85, indicating that the composite scale cohesively measures student performance (Tavakol & Dennick, 2011; McDonald, 1999). GPA emerged as the most internally consistent item, with a remarkably high item–rest correlation of .96, confirming its value as a robust summary measure of performance. Courses such as ME 8401 and ME 8415 also made notable contributions, with item–rest correlations of 0.66 and 0.57, respectively, indicating strong alignment with the underlying competency construct. In contrast, ME 8407 and ME 8408 exhibited weak or negative item–rest correlations (.12 and –.21), which may reflect misalignment with the composite construct or inconsistent assessment practices. Indeed, the slight increase in Cronbach's alpha upon removal of ME 8408 further identifies it as a potential detractor in scale reliability (Streiner, 2003). Additionally, items like ME 8417 and ME 8412, which have low item–rest correlations and a minimal effect on alpha if removed, offer limited discriminative utility and may warrant further review of their inclusion in the curriculum or assessment. Overall, these findings underscore the importance of reliability analysis—not only to

validate measurement tools (Nunnally & Bernstein, 1994), but also to pinpoint specific course elements that may benefit from targeted pedagogical or evaluative enhancements.

Table 3: item reliability statistics

	Mean	SD	Item-rest correlation	If item dropped	
				Cronbach's α	McDonald's ω
GPA	3.59	0.31	0.96	0.79	0.81
CS 8406	53.01	12.91	0.5	0.78	0.84
ME 8401	67.81	6.41	0.66	0.77	0.83
ME 8402	65.03	8.85	0.51	0.77	0.84
ME 8403	63.5	9.24	0.39	0.78	0.84
ME 8404	62.51	7.06	0.33	0.79	0.85
ME 8405	58.48	7.06	0.33	0.78	0.85
ME 8406	58.11	5.62	0.49	0.78	0.84
ME 8407	72.4	6.06	0.12	0.8	0.85
ME 8408	77.82	7.72	-0.21	0.82	0.86
ME 8409	66.9	6.85	0.52	0.77	0.84
ME 8410	70.56	6.15	0.39	0.78	0.84
ME 8411	65.6	8.77	0.49	0.77	0.84
ME 8412	53.78	3.64	0.3	0.79	0.85
ME 8413	63.63	6.97	0.47	0.78	0.84
ME 8414	68.93	6.83	0.43	0.78	0.84
ME 8415	58.73	7.01	0.57	0.77	0.83
ME 8416	61.82	6.49	0.39	0.78	0.84
ME 8417	64.16	6.39	0.15	0.79	0.85

Exploratory Analysis of Performance Patterns

Understanding how students perform across various academic indicators is essential for identifying strengths, weaknesses, and hidden trends that may influence overall educational outcomes. This section presents an exploratory analysis of mechanical engineering students' performance, utilising descriptive statistics, correlation matrices, and visual tools, including scatterplots. The goal is to generate preliminary insights that inform further statistical modelling and guide curriculum development efforts in alignment with Tanzania's industrial transformation agenda.

Distributions and central tendencies of students' academic scores

To establish a foundational understanding of student performance, this section presents descriptive statistics and normality assessments across core mechanical engineering courses and overall GPA. These summary metrics not only offer insights into central tendencies and variability but also help identify data distributions suitable for multivariate techniques.

Figure 2 provides graphical descriptive statistics and normality tests for students' performance across various courses and the overall GPA.

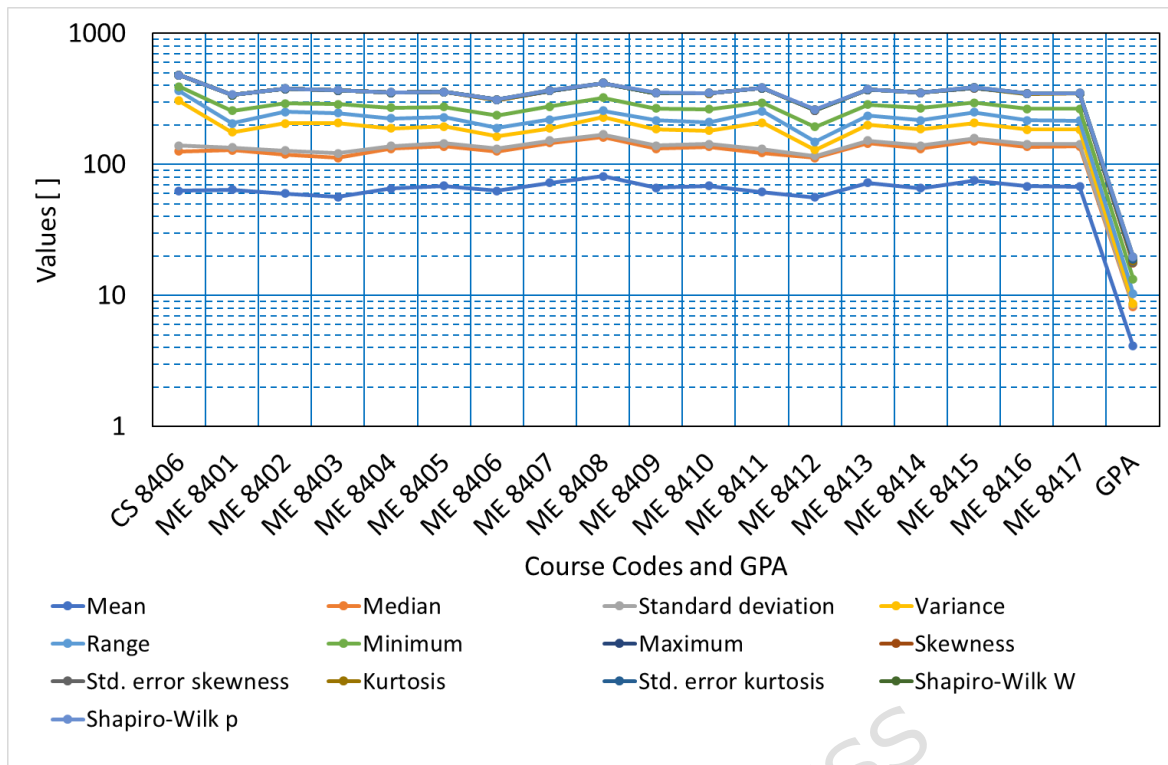


Figure 2: Descriptive statistics of the students' performance

Central Tendency

The mean and median values are used to represent the central tendency of scores for each course and the overall GPA. When these two measures are closely aligned, as is typically the case in most courses, it suggests that the distribution of scores is approximately symmetrical. For instance, ME 8414 shows a mean of 66.07 and a median of 66, indicating minimal skew in the data. This alignment implies that the distribution is not heavily skewed toward higher or lower scores, making it suitable for parametric analysis. According to Hair et al. (2019), when the skewness statistic falls between -1 and +1, the data distribution is considered approximately normal. The skewness for ME 8414 is 0.11, which confirms that the data is only mildly right-skewed and thus approximately normal. This interpretation is further supported by the Shapiro-Wilk test, which yields a p-value of 0.253 for ME 8414, indicating no significant deviation from normality at the 0.05 level. Together, these findings suggest that the data for most courses is reasonably well-behaved statistically, supporting the appropriateness of subsequent multivariate techniques such as PCA.

Spread

The standard deviation and variance provide insights into the degree of variability in student performance across different courses. A higher standard deviation and variance indicate greater dispersion in scores, meaning students performed more unevenly in that subject. For example, CS 8406 exhibited the highest variability (SD = 12.91, variance = 166.55), suggesting that while some students excelled, others struggled significantly, possibly reflecting differences in prior exposure to computing or inconsistencies in instructional delivery. In contrast, ME 8401 exhibited much lower variability (SD = 6.41, variance = 41.12), indicating a more uniform understanding or assessment structure among students. Notably, GPA showed the lowest variability (SD = 0.31, variance = 0.10), suggesting that overall student performance was relatively consistent. This narrow spread could imply standardisation in grading across courses or a generally homogenous student cohort in terms of academic ability. According to Abdi and Williams (2010), understanding variability is essential for interpreting multivariate patterns because variables with high dispersion often exert a disproportionate influence on dimensionality reduction techniques, such as PCA. Therefore, identifying courses with high variance

(e.g., CS 8406) is important not only for pedagogical review but also for ensuring balanced inputs in statistical modelling. Moreover, low-variability measures such as GPA may serve as stable outcome indicators in predictive models, but may contribute less to identifying latent groupings due to their limited spread.

Range

The range indicates the spread between the highest and lowest scores, offering a direct view of performance extremes. CS 8406 exhibited the widest range (58), while ME 8412 had the narrowest (20), suggesting substantial differences in score dispersion between courses. A wide range, as seen in CS 8406, may reflect diverse levels of student preparedness, variations in instructional quality, or potentially inconsistent assessment standards. In contrast, a narrow range like that in ME 8412 could point to more uniform performance, either due to standardised assessments or limited differentiation in student abilities. According to Jolliffe and Cadima (2016), variables with wider ranges and higher variability can disproportionately influence the structure of multivariate techniques such as Principal Component Analysis (PCA), as they introduce more dominant variance into the analysis. This means that CS 8406, with its broad range, may heavily influence the principal components unless the data is standardised. Understanding range is thus not only crucial for interpreting student performance disparities but also for preparing data appropriately for robust multivariate modelling.

Skewness and Kurtosis

Skewness values for most courses hover around zero, indicating that their score distributions are approximately symmetric. This symmetry suggests a balanced spread of scores around the mean, which is ideal for applying parametric techniques like PCA. However, notable exceptions exist. For example, ME 8414 exhibits a skewness of -1.2, indicating a strong left-skewed (negatively skewed) distribution, where a majority of students scored relatively high, while a few scored much lower. This pattern may point to an assessment that was generally easy or one where only a few students underperformed. Kurtosis values near zero suggest that many distributions resemble the standard curve in terms of tail weight and peakedness. However, some courses deviate substantially. ME 8408, with a kurtosis of -1.16, is platykurtic, indicating a flatter distribution with lighter tails, suggesting a wider spread of scores and fewer extreme values. In contrast, ME 8414 has a kurtosis of 2.86, classifying it as leptokurtic, which reflects a sharper peak and heavier tails, implying that scores are tightly clustered around the mean with more frequent extreme values on either side. According to Shlens (2014), understanding skewness and kurtosis is essential in multivariate analysis because these distributional features can influence the effectiveness of linear models, such as PCA, which typically assume distributed variables. Highly skewed or kurtotic variables may distort component loadings or require transformation before analysis.

Normality Tests

The Shapiro-Wilk W and corresponding p -values are used to assess the normality of data distributions. In this analysis, most courses yielded W values between 0.98 and 0.99, which are close to 1, the ideal value for a perfectly normal distribution. According to statistical conventions, W values above 0.95 generally suggest that the data is approximately normal, particularly when accompanied by non-significant p -values ($p > 0.05$). Based on this scale, the majority of the course scores can be considered approximately normally distributed. However, ME 8408 is a notable exception. Despite a high W value of 0.93, its p -value was less than 0.001, indicating a statistically significant deviation from normality. This finding suggests that the distribution of scores in ME 8408 does not conform to the assumptions of normality, and therefore may necessitate the use of non-parametric techniques or data transformation in subsequent analyses. As emphasised by Hair et al. (2019), testing for normality is critical when preparing data for multivariate analyses, such as Principal Component Analysis (PCA), which assumes a linear relationship and regular distribution of variables. Violations of these assumptions can distort factor extraction and interpretation, making preliminary diagnostics such as the Shapiro-Wilk test a vital step in data screening.

Correlation Between Course Scores and GPA in Mechanical Engineering Students

Table 4 presents a Pearson correlation matrix assessing linear associations between students' performance in individual mechanical engineering courses and their overall Grade Point Average (GPA). Notably, strong positive correlations were observed for ME 8401 – Fluid Power and Control ($r = 0.72$), CS 8406 – Object-Oriented Programming ($r = 0.64$), ME 8415 – Computer-Aided Manufacturing (CAM) ($r = 0.63$), and ME 8411 – Automation and Robotics ($r = 0.61$), suggesting that achievement in these courses aligns closely with overall academic success. While Pearson's correlation does not imply causality, it indicates that students who excel in these modules tend to achieve higher GPAs. However, the statistical significance of these correlations requires confirmation through accompanying p-values, which are not reported here. In contrast, ME 8408 – Industrial Practical Training III showed a weak negative correlation ($r = -0.21$), potentially indicating misalignment in the assessment or distinct evaluation criteria that do not accurately reflect general academic performance. Similarly, ME 8407 – Senior Project I ($r = 0.22$) and ME 8417 – Senior Project II ($r = 0.33$) showed low correlations with GPA, implying limited influence on cumulative performance. Inter-course correlations, such as those between ME 8401 and ME 8411 ($r = 0.60$) and between ME 8415 and ME 8416 – Industrial Supervisor Skills and Leadership ($r = 0.51$)- suggest potential overlap in competencies or thematic content. According to Cohen (2013), correlation values above 0.50 are considered significant, those between 0.30 and 0.49 are moderate, and those below 0.30 are small, which guides the interpretation of these findings for curriculum evaluation and academic planning.

Table 4: The correlation matrix of the courses

	GPA	CS 8406	ME 8401	ME 8402	ME 8403	ME 8404	ME 8405	ME 8406	ME 8407	ME 8408	ME 8409	ME 8410	ME 8411	ME 8412	ME 8413	ME 8414	ME 8415	ME 8416	ME 8417	
GPA	1																			
CS 8406	0.64	1																		
ME 8401	0.72	0.48	1																	
ME 8402	0.54	0.29	0.45	1																
ME 8403	0.47	0.34	0.25	0.22	1															
ME 8404	0.39	0.2	0.28	0.18	0.2	1														
ME 8405	0.44	0.21	0.25	0.16	0.21	0.12	1													
ME 8406	0.54	0.3	0.44	0.4	0.09	0.17	0.09	1												
ME 8407	0.22	0.03	0.05	0.1	0.04	-	0.09	-	1											
ME 8408	-0.2	-	-	-	0	-	-	0.03	-	1										
ME 8409	0.55	0.29	0.36	0.28	0.33	0.22	0.21	0.39	0.07	-	1									
ME 8410	0.42	0.21	0.19	0.38	0.22	0.17	0.15	0.08	0.18	0.13	-	1								
ME 8411	0.61	0.38	0.6	0.31	0.06	0.36	0.33	0.38	-	0.09	-0.3	0.24	0.21	1						
ME 8412	0.33	0.2	0.26	0.19	0.23	0.11	-	0.29	-	0.01	0.1	0.19	0.15	-	1					
ME 8413	0.56	0.31	0.29	0.21	0.31	0.2	0.33	0.21	0.13	0.04	-	0.26	0.31	0.35	0.07	1				
ME 8414	0.54	0.28	0.36	0.36	0.19	0.12	0.15	0.43	0.07	0.19	0.05	0.39	0.22	0.19	0.37	0.26	1			
ME 8415	0.63	0.32	0.5	0.28	0.22	0.24	0.31	0.34	0.18	-0.3	0.39	0.26	0.41	0.2	0.43	0.18	0.1	1		
ME 8416	0.45	0.13	0.4	0.29	0.19	0.14	0.1	0.24	0.08	-0.2	0.3	0.19	0.23	0.24	0.23	0.17	0.51	0.1	1	
ME 8417	0.33	0.08	0.14	0.01	0.03	0.01	0.03	0.12	0.2	-	0.15	0.06	0.17	0.07	0.02	0.17	0.1	0.04	0.08	1

Key:

Course Code	Course Name	Course Code	Course Name
CS 8406	Object Oriented Programming	ME 8409	Quality Assurance and Control Engineering Ethics and Professional Conduct
ME 8401	Fluid Power and Control	ME 8410	Automation and Robotics
ME 8402	Power Plants	ME 8411	Engine Design Technology
ME 8403	Material Handling Design		

ME 8404	Turbo Machinery	ME 8413	Foundry Technology
ME 8405	Production Engineering	ME 8414	Air Conditioning System
ME 8406	Refrigeration System	ME 8415	Computer-Aided Manufacturing (CAM)
ME 8407	Senior Project I	ME 8416	Industrial Supervisor Skills and Leadership
ME 8408	Industrial Practical Training III	ME 8417	Senior Project II

PCA of Performance

PCA was applied to reduce the complexity of student performance data and identify key underlying dimensions. By transforming correlated course scores into a smaller set of uncorrelated components, PCA highlights core performance domains and reveals patterns not evident from raw scores. This aids in curriculum evaluation and supports evidence-based improvements in mechanical engineering education.

Scree Plot for Factor Retention in Exploratory Factor Analysis

To determine the optimal number of components for retention, a scree plot was used as a visual diagnostic tool (Figure 3). According to the Kaiser criterion, components with eigenvalues greater than 1 are considered meaningful, as each explains more variance than a single original variable (Kaiser, 1960). In this analysis, Component 1 had an eigenvalue of approximately 6.72, indicating it accounted for a substantial portion of the variance, while Component 2 had an eigenvalue of approximately 1.89. These are the only two components that met the Kaiser threshold, confirming their relevance for further analysis. All subsequent components had eigenvalues below 1 and were therefore excluded due to their limited explanatory power. The “elbow” point in the scree plot—where the rate of decline in eigenvalues visibly flattens, further supports retaining only the first two components. This visual method, as described by Cattell (1966), identifies the point at which additional components contribute minimal incremental variance, thus ensuring that only the most informative dimensions are retained. This dual-criterion approach (Kaiser + elbow) supports a parsimonious solution that balances dimensionality reduction with the retention of essential information. Simplifying the factor structure in this way enhances interpretability and strengthens the analytical robustness of the subsequent Exploratory Factor Analysis.

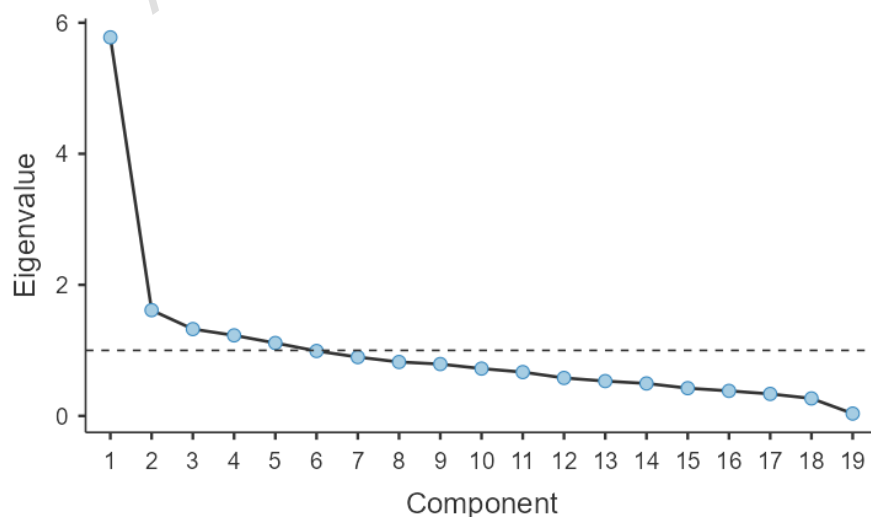


Figure 3: The scree plot for principal component analysis for the course performance

ME 8415		0.79		0.28
ME 8416		0.67		0.54
ME 8417			0.8	0.32
GPA	0.3		0.4	-0.04

Uniqueness Values

High uniqueness values for specific courses, such as ME 8404 (0.78) and ME 8402 (0.55), suggest that a substantial portion of their variance is not accounted for by the common latent factors. This suggests that these courses may be shaped by unique instructional approaches, specialised content, or assessment methods not shared with other courses. As noted by Abdi and Williams (2010), uniqueness reflects variable-specific variance that remains unexplained by the factor model, implying that such variables may not align well with the general performance structure identified through factor analysis.

Latent Structure

Factor 1 appears to represent general academic performance, as it exhibits strong associations with several key variables, including GPA, CS 8406, ME 8401, and ME 8411, suggesting a broad, underlying competency that influences overall achievement. In contrast, Factors 4 and 5 seem to capture more specialised or course-specific trends, with ME 8403 and ME 8410 contributing heavily to each respective factor. The use of oblimin rotation was appropriate for this analysis, as it allows for factors to be correlated, reflecting the real-world nature of educational performance, where skills and knowledge areas often overlap. According to Jolliffe and Cadima (2016), such rotation methods are essential when interpreting data from domains like education, where orthogonality (i.e., complete independence between components) is unrealistic and may mask meaningful relationships between latent traits.

Eigenvalue Analysis for Principal Component Extraction and Variance Assessment

Table 6 details the eigenvalues derived from Principal Component Analysis (PCA), highlighting how variance is distributed across components and guiding the dimensionality reduction process. Component 1, with an eigenvalue of 5.78, explains 30.40% of the total variance, capturing the most dominant pattern in student performance. Component 2, with an eigenvalue of 1.61, contributes an additional 8.49%, bringing the cumulative explained variance to 38.90%, following the Kaiser criterion, which recommends retaining factors with eigenvalues greater than 1 (Kaiser, 1960), the first five components are considered significant. Indeed, Component 3 adds 6.97%, Component 4 contributes 6.47%, and Component 5 yields 5.86%, together elevating the cumulative variance explained to 58.20%. Components beyond the fifth have eigenvalues below 1 and thus are deemed less informative, typically capturing noise or minor patterns. The scree (elbow) plot further justifies this selection, showing a clear inflexion point past Component 5 where eigenvalues begin to plateau (Cattell, 1966), indicating diminishing returns. As Jolliffe (2002) emphasises, retaining components that balance explained variance with interpretability enhances the analytical clarity and robustness of PCA, ensuring that essential information is preserved while reducing redundancy.

Table 6: Component number, eigenvalue and variance explained

Component	Eigenvalue	% of Variance	Cumulative %
1	5.78	30.4	30.4
2	1.61	8.49	38.9
3	1.33	6.97	45.87
4	1.23	6.47	52.34

5	1.11	5.86	58.2
6	0.99	5.21	63.41
7	0.9	4.72	68.13
8	0.82	4.33	72.47
9	0.79	4.16	76.63
10	0.72	3.8	80.43
11	0.67	3.52	83.95
12	0.58	3.05	87
13	0.53	2.8	89.79
14	0.5	2.61	92.4
15	0.42	2.23	94.63
16	0.38	2.01	96.64
17	0.34	1.77	98.41
18	0.27	1.4	99.81
19	0.04	0.19	100

Patterns of Course Contributions and Interdependencies Revealed by PCA Visualization

The PCA variables plot (Figure 4) illustrates the contributions and interrelationships among performance indicators with respect to the first two principal components—Dim1, which explains 30.4%, and Dim2, which accounts for 8.5% of the total variance. Each variable is represented by an arrow, whose length indicates the magnitude of its contribution, and direction reflects its correlation with the component axes. Courses such as ME 8414 and ME 8412, with long arrows near the perimeter, are strong contributors to Dim1, whereas ME 8408 and ME 8403, though less clearly positioned, appear more associated with Dim2.

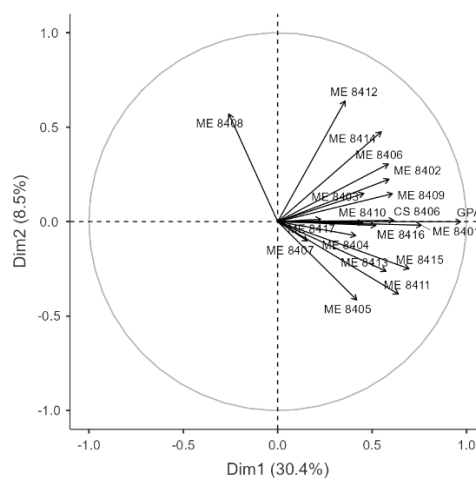


Figure 4: The PCA variables plot

The circle of correlations, representing a theoretical maximum correlation of 1.0, helps interpret variable strength; those nearer the edge (e.g., ME 8414) show strong alignment with one or both dimensions. In contrast, variables near the origin (e.g., ME 8405, ME 8417) contribute minimally. Dim1 likely captures a broad academic competence domain, while Dim2 may reflect a secondary, more specialised performance trait. Arrows pointing in similar directions imply positive correlations, while those in opposite directions indicate negative correlations among variables. According to Jolliffe and Cadima (2016), such biplots are essential for identifying which variables are most influential in defining principal components and how they cluster based on shared variance. Similarly, Abdi and Williams (2010) emphasise that PCA visualisations reveal underlying patterns and dependencies that guide interpretation, supporting the development of more nuanced academic insights and data-driven curriculum improvements.

Limitations of the Study

This study is subject to several limitations. First, the analysis is based on a single cohort from one institution (MUST), which limits the generalisability of the findings to other contexts. Second, the study relies solely on quantitative academic performance data, without incorporating qualitative insights such as student experiences, teaching practices, or industry feedback that could enrich interpretation. Third, PCA, as an exploratory method, identifies patterns in performance but does not establish causality or directly evaluate curriculum effectiveness. Finally, factors such as students' academic preparedness, learning behaviours, and socio-economic influences were not included in the dataset, which may also affect academic outcomes.

Conclusion

This study examined patterns of student performance in final-year mechanical engineering courses at Mbeya University of Science and Technology using EA and PCA. The results indicate that student achievement tends to cluster around technically intensive modules, suggesting that these courses play a central role in developing the core engineering competencies expected of graduates. In particular, ME 8411 (Automation and Robotics) and ME 8401 (Fluid Power and Control) showed strong associations with overall academic performance, while courses such as ME 8408 (Industrial Practical Training III) exhibited weaker relationships with the broader performance structure. The PCA results further revealed five latent components explaining 58.2% of the total variance, with the first component alone accounting for 30.4%, indicating a dominant dimension associated with technical engineering skills. High factor loadings for courses such as ME 8411, ME 8401, and ME 8415 (Computer-Aided Manufacturing) indicate strong contribution to the underlying academic performance structure, whereas courses such as ME 8408 and ME 8413 demonstrate limited alignment with the shared variance structure. From an Outcome-Based Education perspective, these findings suggest that different courses contribute unevenly to the development and assessment of programme learning outcomes, particularly in balancing theoretical knowledge, practical training, and project-based learning. Overall, the results provide an evidence-informed basis for reflecting on curriculum alignment, strengthening competency development, and improving assessment structures to better support industry-relevant engineering outcomes in line with Tanzania's Vision 2025 industrialisation goals.

Recommendations

To align mechanical engineering education at Mbeya University of Science and Technology (MUST) with Tanzania's Vision 2025 industrial goals, several improvements are recommended. Curriculum refinement should target low-impact courses such as ME 8408 and ME 8407 to better align them with practical, industry-relevant competencies, while enhanced academic support (e.g., tutorials, peer mentoring, and skill-based assessments) should be provided for technically intensive modules such as ME 8401 and ME 8411. In addition, greater attention should be given to student-centred strategies, including academic support programmes, flexible learning pathways, and early identification of students with learning difficulties to address variations in preparedness and study habits. Integrating

data-driven monitoring approaches (e.g., PCA and EA), strengthening faculty capacity in educational analytics, and fostering industry–academic partnerships will further support continuous curriculum improvement. Collectively, these measures promote a more balanced, outcome-based engineering education that integrates curriculum quality, teaching effectiveness, and student success.

Acknowledgement

The Author acknowledges the support from the Mbeya University of Science and Technology, particularly the Department of Mechanical and Industrial Engineering and the Examination Office

Declarations

Ethical approval: The Mbeya University of Science and Technology Research Ethical Committee (MUST REC) provided the ethical approval for the study. The author confirms that all experiments were performed in accordance with relevant guidelines and regulations. Also, the study adhered to the Helsinki Declaration.

Consent to participate: In accordance with the approval conditions stipulated in the Director of Undergraduate Studies' letter, freely given and informed consent was obtained from all participants prior to data collection. Participants were adequately informed about the purpose of the study, the voluntary nature of participation, confidentiality safeguards, and their right to withdraw at any stage without any academic or personal consequences.

Consent to publish: Not Applicable.

Data: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Conflict of interest: The authors declare no conflict of interest.

Funding: no Funding

Clinical trial number: not applicable

References

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- Bishop, D., Akinyemi, A., & Maina, J. (2021). Engineering workforce needs for industrial transformation in Tanzania. *Journal of African Industrial Development*, 12(3), 45–67.
- Borges, V. R. P., Esteves, S., Araújo, P. D. N., de Oliveira, L. C., & Holanda, M. (2018). Using Principal Component Analysis to support students' performance prediction and data analysis. *Proceedings of the Brazilian Symposium on Computers in Education*. <https://typeset.io/papers/using-principal-component-analysis-to-support-students-dv35by90kf>
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioural Research*, 1(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10
- Cohen, J. (2013). *Statistical power analysis for the behavioural sciences*. Routledge.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics (5th ed.)*. Sage Publications Limited.

- Giovannella, C., Scaccia, F., & Popescu, E. (2013). A PCA study of student performance indicators in a Web 2.0-based learning environment. *Proceedings of the International Conference on Advanced Learning Technologies*. <https://typeset.io/papers/a-pca-study-of-student-performance-indicators-in-a-web-2-0-16yiao98ze>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.
- Harden, R. M. (1999). AMEE Guide No. 14: Outcome-based education: Part 1-An introduction to outcome-based education. *Medical teacher*, 21(1), 7-14.
- Hu, S., Gu, Y., & Jiang, H. (2015). Teaching quality assessment based on principal component analysis and Elman neural network. *International Symposium on Computational Intelligence and Design*. <https://doi.org/10.1109/ISCID.2015.270>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <https://doi.org/10.1177/001316446002000116>
- Katambara, Z. (2025). Determinants of Students' Performance in Electrical and Electronics Engineering at Mbeya University of Science and Technology, Tanzania. *East African Journal of Education and Social Sciences* 6(1), 88-101. DOI: <https://doi.org/10.46606/eajess2025v06i01.0427>.
- Katambara, Z., & William, M. (2025). Assessing Key Performance Factors in Final-Year Civil Engineering Students at Mbeya University of Science and Technology, Tanzania by Using Principal Component Analysis. *Creative Education*, 16, 253-267. <https://doi.org/10.4236/ce.2025.163016>
- Maphosa, M., Doorsamy, W., & Paul, B. S. (2023). Student performance patterns in engineering at the University of Johannesburg: An exploratory data analysis. *IEEE Access*, 11, 48977–48988. <https://doi.org/10.1109/ACCESS.2023.3277225>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Resnik, D. B. (2018). *The ethics of research with human subjects: Protecting people, advancing science, promoting trust* (Vol. 74). Springer.
- Roth, F., Schneider, H., & Wendler, C. (2019). Evaluating interdisciplinary competencies in engineering programs: Insights from Europe. *Journal of Vocational Education and Training*, 71(2), 234–250. <https://doi.org/10.1080/13636820.2019.1564320>
- Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.
- Sofia, G., & Hema, D. (2024). Clustering-based recommendations for enhancing students' academic performance by recognizing prevalent assessment method using exploratory data analysis. *Indian Journal of Science and Technology*, 17(6), 503–513. <https://doi.org/10.17485/IJST/v17i6.2175>
- Spady, W. G. (1994). *Outcome-based education: Critical issues and answers*. American Association of School Administrators.
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.

Toroka, M. I., & Kafanabo, E. (2024). Assessing the relevance of technical education curricula to current industrial skill demands in Tanzania: A case study of civil, electrical, and biomedical engineering curricula. *Papers in Education and Development*, 42(1).

Twenefour, D., Osei, S., & Arthur, E. (2015). A principal component analysis approach to students' assessment scores. *International Journal of Education and Development using Information and Communication Technology*, 11(2), 1–11.

United Republic of Tanzania. (2021). *Tanzania Development Vision 2025: Framework for industrial and economic growth*. Dar es Salaam: Ministry of Industry and Trade.

Walker, P. (2005). Teaching for quality learning at university. *Higher Education*, 49(4), 535-538.

Wang, X., & Zhou, L. (2020). Engineering education and innovation as drivers of China's industrial leadership. *International Journal of Engineering Studies*, 15(4), 312–329. <https://doi.org/10.1016/j.engstud.2020.02.005>

Wismansyah, A. R., Kadir, A. R., Baja, S., & Amar, M. Y. (2024). Sustainable vocational education models for industrial revolution role in development of job market absorption in Tangerang City, Indonesia. *Journal of Infrastructure, Policy and Development*, 8(6). <https://doi.org/10.24294/jipd.v8i6.3958>

Yu, T. (2014). An exploratory factor analysis and reliability analysis of the Student Online Learning Readiness (SOLR) instrument (Doctoral dissertation, Purdue University). ProQuest Dissertations & Theses Global.

Yusoff, R., Ismail, H., & Mohamad, F. (2018). Academia-industry collaboration: Strengthening skills development for industrial growth in Malaysia. *Asian Journal of Engineering and Technology*, 10(1), 14–23. <https://doi.org/10.1016/ajet.2018.01.004>

Zarate, L., Perez, R., & Cid, M. (2017). Evaluating engineering graduate performance: A multivariate approach for curriculum improvement. *International Journal of Educational Development*, 52, 105–114. <https://doi.org/10.1016/j.ijedudev.2017.05.001>

Appendix

Course Code	Course Name
CS 8406	Object-Oriented Programming
ME 8401	Fluid Power and Control
ME 8402	Power Plants
ME 8403	Material Handling Design
ME 8404	Turbo Machinery
ME 8405	Production Engineering
ME 8406	Refrigeration System
ME 8407	Senior Project I
ME 8408	Industrial Practical Training III
ME 8409	Quality Assurance and Control
ME 8410	Engineering Ethics and Professional Conduct
ME 8411	Automation and Robotics
ME 8412	Engine Design Technology
ME 8413	Foundry Technology

ME 8414	Air Conditioning System
ME 8415	Computer-Aided Manufacturing (CAM)
ME 8416	Industrial Supervisor Skill and Leadership
ME 8417	Senior Project II

ARTICLE IN PRESS