

# Assessing Key Performance Factors in Final-Year Civil Engineering Students at Mbeya University of Science and Technology, Tanzania by Using Principal Component Analysis

Zacharia Katambara<sup>1\*</sup>, Matungwa William<sup>2</sup>

<sup>1</sup>Department of Civil Engineering, College of Engineering and Technology, Mbeya University of Science and Technology, Mbeya, Tanzania

<sup>2</sup>Department of Earth Sciences, College of Science and Technical Education, Mbeya University of Science and Technology, Mbeya, Tanzania

Email: \*zacharia.katambara@must.ac.tz, matungwa.william@must.ac.tz

**How to cite this paper:** 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>

**Received:** February 8, 2025

**Accepted:** March 7, 2025

**Published:** March 10, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

This study applies Principal Component Analysis (PCA) to evaluate and understand academic performance among final-year Civil Engineering students at Mbeya University of Science and Technology (MUST) in the academic year 2023/2024. Responding to Tanzania's growing demand for skilled engineers, this research provides data-driven insights into student achievement patterns, revealing the core factors impacting performance. PCA was utilised to reduce dimensionality, transforming course grades into uncorrelated components that capture underlying performance structures. The results identify three primary components: Core Academic Knowledge, explaining 41.38% of the variance; Specialized Applied Skills, contributing 10.07%; and Advanced Independent Skills, accounting for 6.71%. Together, these components explain 58.16% of total performance variance, indicating a robust framework for understanding student success drivers. Additional analysis, including correlation matrices and descriptive statistics, highlights patterns across courses, revealing strong relationships within core competencies and independent distinctions in advanced courses. Findings suggest curriculum enhancements and targeted interventions that could better prepare students for industry needs, focusing on core academic support, practical skills enhancement, and resources for advanced technical areas.

## Keywords

Student Academic Performance, Civil Engineering Education, Principal Component Analysis, Final-Year Civil Engineering Students, Mbeya University of Science and Technology

## 1. Introduction

In academic institutions such as Mbeya University of Science and Technology (MUST), engineering programs have consistently ranked among the most preferred fields of study. Since the University's inception, and with notable growth, this trend reflects a broader societal and economic demand for engineering expertise in developing economies (World Bank, 2020). Engineering programs attract students for various reasons, including high employability, competitive salaries, and the opportunity to contribute directly to national development goals (Kim & Maloney, 2021; Mlambo, 2019). The trend is further illustrated by the Engineers Registration Board's records, which show that over 40% of registered engineers in Tanzania specialise in civil engineering, underscoring the critical need for expertise in infrastructure development, urban planning, and essential construction projects (Engineers Registration Board [ERB], 2023). The prominence of civil engineering aligns with national priorities for expanding infrastructure and supporting sustainable development within the country (Kim & Maloney, 2021).

Given this demand, the performance of civil engineering students must meet stakeholder expectations regarding skills, competencies, and innovation. Assessing student performance, therefore, is essential to ensure graduates are equipped to meet industry requirements. This study focuses on analysing the academic performance of civil engineering students at MUST to identify key success factors aligned with the skills and competencies needed in the field.

Analysing student performance is essential for identifying learning gaps, optimizing teaching methods, and improving academic outcomes, particularly in technical disciplines in engineering where interdependencies between subjects can be complex. Traditional analyses of student performance often rely on single-variable measures, such as individual subject grades or aggregate scores, which may overlook patterns in learning achievements and relationships between subjects (Kim & Kim, 2018). Multivariate methods, such as Principal Component Analysis (PCA), provide a more comprehensive approach by uncovering latent factors influencing academic performance across multiple variables (Jolliffe & Cadima, 2016). PCA can reveal hidden structures in data, allowing for clustering of high and low performers and pinpointing areas where targeted interventions could be most beneficial.

This paper examines the application of PCA in assessing the performance of final-year Civil Engineering students at MUST in the academic year 2023-2024 to identify key success factors. This, in turn, informs targeted academic support and aligns student outcomes with industry needs.

## 2. Previous Application of PCA in Student Performance Assessment

PCA is a dimensionality reduction technique that transforms correlated variables such as grades across various subjects into uncorrelated principal components, capturing the maximum variance in the dataset. By simplifying complex data,

PCA reveals key factors influencing student performance and provides insights into underlying trends and patterns (Abdi & Williams, 2010; Shlens, 2014). In educational research, PCA has become an invaluable tool for handling high-dimensional datasets, including subject grades, engagement metrics, and assessment results (Cai et al., 2020).

For example, Tarabasz et al. (2021) applied PCA to analyse university students' performance across various disciplines and found that quantitative subjects like mathematics and physics were major predictors of overall academic success. Their findings emphasized the foundational role of quantitative competency in STEM achievement, providing useful direction for developing remedial programs in these areas. In a related study, Martínez-Ruiz and García (2019) used PCA to examine the link between student engagement and academic performance in online courses. Their analysis showed that metrics such as time spent on coursework and forum participation were significant predictors of success, supporting PCA's utility in uncovering multidimensional relationships that might be missed by traditional single-variable analyses.

PCA has also been applied to differentiate between technical and non-technical subjects within engineering programs. Praveen and Mahapatra (2017) used PCA to examine engineering students' performance and found that technical subjects, such as core engineering courses, had a more substantial impact on overall performance compared to non-technical subjects, though non-technical subjects still contributed meaningfully. This insight has enabled educators to prioritize targeted interventions and tailored support for students based on specific subject needs. Similarly, Singh and Upadhyia (2019) and Zain et al. (2020) demonstrated that PCA could be used to group students into clusters based on similar performance characteristics, facilitating group-specific support.

In educational settings, PCA is frequently applied to identify clusters of students with shared performance characteristics, enabling targeted interventions and support tailored to each group (Singh & Upadhyia, 2019). Zain et al. (2020) used PCA followed by clustering techniques on high school student performance data, identifying distinct clusters, such as students excelling in scientific disciplines versus those performing better in arts and humanities. Such clustering facilitates efficient resource allocation and allows educators to deliver specialized support based on the unique needs of each cluster.

Recent developments in educational data mining have also seen PCA integrated with predictive models to forecast academic outcomes, identify students at risk of dropping out, and guide proactive interventions (Wang et al., 2021). For instance, Zhang et al. (2021) combined PCA with support vector machines (SVMs) to predict student retention based on initial semester performance, with PCA-driven dimensionality reduction improving SVM model accuracy. In another example, Sharma et al. (2019) incorporated PCA with artificial neural networks (ANNs) to predict final exam outcomes. By selecting the most influential predictors such as early test scores and classroom engagement PCA significantly enhanced

model efficiency, making it a valuable tool for developing precise academic support strategies.

### 3. Methodology

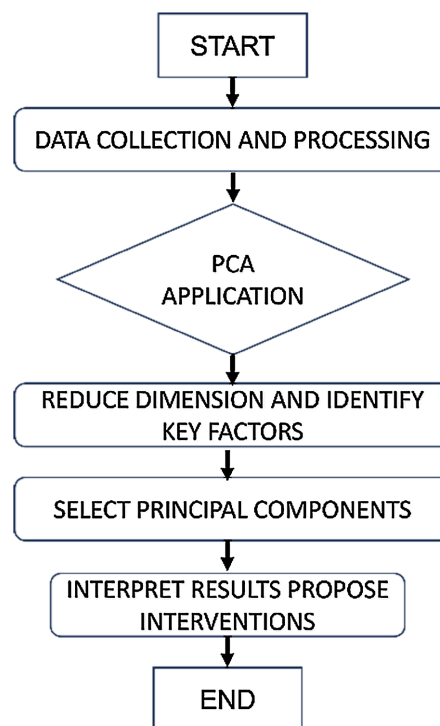
This study employs Principal Component Analysis (PCA) to examine annual academic performance data for final-year Civil Engineering students at Mbeya University of Science and Technology (MUST) in the academic year 2023/2024. The performance data, sourced from the Student Information Management System (SIMS) through the Department of Examinations, focuses on students who completed all required examinations, ensuring data consistency and reliability. The methodological framework comprises of three structured stages: data preprocessing, PCA application, and result interpretation. This provides a robust approach for identifying key performance trends and academic factors influencing student outcomes. The description follows next and is summarised in **Figure 1**.

#### 3.1. Data Collection and Preprocessing

Academic records were sourced from the SIMS. Grades were standardized to ensure consistency across different grading scales, as recommended by [Shlens \(2014\)](#). Outliers were addressed to avoid distorted principal component results ([Nguyen et al., 2020](#)).

#### 3.2. Application of PCA

PCA was applied to the standardized data to reduce dimensionality and identify



**Figure 1.** Flow diagram of PCA.

key factors influencing student performance. Principal components were selected based on their contribution to total variance, and subject loadings were analysed to interpret the academic areas most affecting overall performance (Abdi & Williams, 2010).

### 3.3. Cluster Analysis and Interpretation

Principal component scores were used to perform cluster analysis, grouping students into performance clusters to identify those with similar academic strengths and weaknesses. This approach aligns with methods used by Singh and Upadhyya (2019) and Zain et al. (2020) for categorizing students based on shared characteristics and tailoring academic interventions accordingly.

## 4. Results and Discussion

### 4.1. Descriptive Statistics

**Table 1** provides descriptive statistics for final-year Bachelor of Civil Engineering

**Table 1.** Descriptive statistics of the students' performance in each course.

Description	CE 8316	CE 8401	CE 8402	CE 8403	CE 8404	CE 8405	CE 8406 CE 8407	CE 8408	CE 8409	CE 8410	CE 8411	CE 8412	CE 8413	CE 8414 CE 8415	CE 8416
N	156	155	156	156	156	156	156	154	154	154	154	154	154	156	154
Mean	80.99	56.15	61.34	50.73	68.08	61.6	63.75	70.2	55.88	61.8	59.31	61.97	69.66	55.22	71.92
Median	81	56	61.5	50	68	62	65	70	55.5	62	59.5	62	70.5	56	73
Standard deviation	6.9	8.09	6	8.45	7.68	5.93	13.87	4.88	6.02	8.83	6.08	6.57	10.29	10.29	6.77
Variance	47.62	65.46	36.05	71.37	58.91	35.17	192.38	23.77	36.24	78.03	36.97	43.16	105.85	105.8	45.78
Minimum	54	31	45	31	49	46	0	55	42	40	41	43	29	0	56
Maximum	93	75	75	71	86	78	93	83	71	85	72	78	88	77	86
Skewness	-0.59	-0.37	-0.17	-0.06	-0.15	-0.15	-0.51	-0.02	0.24	-0.11	-0.17	-0.23	-0.83	-2.09	-0.28
Kurtosis	0.75	1	-0.37	-0.69	-0.23	-0.11	1.73	-0.17	-0.45	-0.19	-0.33	-0.1	1.18	9.43	-0.46
Shapiro-Wilk W	0.97	0.98	0.99	0.98	0.99	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.96	0.98
Shapiro-Wilk p	0.001	0.01	0.193	0.087	0.22	0.454	0.001	0.484	0.151	0.761	0.292	0.264	<0.001	<0.001	0.058
25 <sup>th</sup> percentile	77	51	57	44	63	57.75	54	66.25	51	56	55	58	64.25	51	67
50 <sup>th</sup> percentile	81	56	61.5	50	68	62	65	70	55.5	62	59.5	62	70.5	56	73
75 <sup>th</sup> percentile	86	61.5	66	58	74	66	72.25	74	60	67	64	66	77.75	62	77

Where, CE 8316 is industrial Practical Training III; CE 8401 is Engineering Economics; CE 8402 is Structural Steel Design; CE 8403 is Waste Water Management; CE 8404 is Pavement Maintenance; CE 8405 is Bridge Design and Construction; CE 8406 is Pre-Stressed Concrete Design; CE 8407 is Irrigation Engineering; CE 8416 is Project II; CE 8408 is Project I; CE 8409 is Design of Masonry and Retaining Structures; CE 8410 is Structural Timber Design; CE 8411 is Solid Waste Management; CE 8412 is Industrial Building Construction; CE 8413 is Hydraulic Structures; CE 8414 is Transportation Engineering and CE 8415 is Water Resources Management.

courses by providing a detailed analysis of student performance, central tendency, distribution, and variability. The key statistics include the sample size (N), measures of central tendency (mean and median), dispersion (standard deviation and variance), range (minimum and maximum), skewness, kurtosis, and normality tests (Shapiro-Wilk test with  $p$ -values). This analysis provides insights into the overall performance trends and distribution of scores across these courses.

1) *Sample Size (N)*: Most courses have a sample size (N) close to 154 - 156, with only minor missing values in a few courses. Complete records across most courses enhance the reliability of the descriptive statistics (Field, 2018).

2) *Central Tendency (Mean, Median and Mode)*: Mean scores vary significantly across courses, ranging from 50.73 (CE 8403) to 80.99 (CE 8316), indicating variability in overall performance. CE 8316 has the highest mean, suggesting that students performed particularly well, while CE 8403's lower mean indicates potential challenges in that course. Median values are close to the mean for most courses, suggesting generally symmetrical distributions, which typically signify a balanced performance. This is further corroborated by the minimal difference between mean and median in most courses, indicating few extreme outliers (Gravetter & Wallnau, 2017). Mode highlights the most frequent score, which can provide additional insights into common student performance levels. For instance, the mode in CE 8415 is 70.00, close to the mean and median, indicating a concentration of scores around this value (Pallant, 2020).

3) *Dispersion (Standard Deviation and Variance)*: Standard deviation and variance are key indicators of score variability. CE 8406 shows the highest standard deviation (13.87), indicating a wide range of scores and possibly greater variability in student performance or course difficulty. In contrast, CE 8407 has a lower standard deviation (4.88), suggesting more consistent performance among students (Tabachnick & Fidell, 2019). High variance values, such as in CE 8406 (192.38), indicate that scores are widely spread around the mean, which could be due to diverse skill levels or other factors affecting student performance.

4) *Range (Minimum and Maximum)*: The range of scores across the courses reflects a broad spectrum of student performance. The minimum scores show considerable variation, with some courses having a minimum of 0 (e.g., CE 8406 and CE 8413), indicating potential outliers or failing grades. The maximum scores are generally high across all courses, with CE 8316 and CE 8406 both reaching the maximum score of 93, indicating that top-performing students excelled consistently (Allison, 2010).

5) *Skewness*: Skewness measures the symmetry of the distribution. Most courses show skewness values close to zero, indicating fairly symmetrical distributions. However, CE 8413 shows a skewness of  $-2.09$ , indicating a strong negative skew with a concentration of higher scores and fewer low scores (Field, 2018). Courses with near-zero skewness, like CE 8408 (0.24), show balanced score distributions around the mean.

6) *Kurtosis*: Kurtosis measures the "peakedness" of the distribution. While most

courses have kurtosis close to zero, suggesting a normal distribution, CE 8413 has a kurtosis of 9.43, indicating a leptokurtic distribution with heavy tails and more extreme values. Such values are often associated with high variability and the presence of outliers (Pallant, 2020).

7) *Normality Testing (Shapiro-Wilk  $W$  and  $p$ -values)*: The Shapiro-Wilk test assesses whether the data follow a normal distribution. Most courses have  $p$ -values above 0.05, indicating a normal distribution assumption cannot be rejected, suggesting that scores for these courses are normally distributed. However, courses like CE 8316 ( $p = 0.001$ ) and CE 8413 ( $p < 0.001$ ) have significant results, indicating that scores for these courses deviate from normality (Ghasemi & Zahediasl, 2012).

8) *Percentiles (25th, 50th, and 75th)*: Percentile values help in understanding the spread of scores. For instance, the 25th percentile for CE 8316 is 77.00, indicating that 25% of students scored below this value, while the 75th percentile is 86.00, showing that 75% scored below this level. Percentiles provide a nuanced view of score distribution, complementing other central tendency and variability metrics (Pallant, 2020).

Overall, these descriptive statistics provide essential insights into performance trends across courses. Understanding these patterns enables more informed decision-making regarding curriculum development, resource allocation, and potential academic support initiatives to enhance student success (Tabachnick & Fidell, 2019).

## 4.2. Correlation Matrix of the Course Performances

**Table 2** shows the correlation matrix provides insights into the relationships between different course scores, and reveals patterns in student performance across these courses. Correlation coefficients range from  $-1$  to  $1$ , with positive values indicating that higher scores in one course tend to be associated with higher scores in another and negative values suggesting an inverse relationship as discussed below.

1) *Strong Positive Correlations*: Certain courses exhibit strong positive correlations, meaning that students who perform well in one course are likely to perform well in another. For example: CE 8404 and CE 8412 have a high correlation ( $r = 0.71$ ), indicating a significant association in performance between these two courses. CE 8410 and CE 8411 also show a strong correlation ( $r = 0.62$ ), suggesting that these courses may share similar content, skills, or assessment criteria (Gravetter & Wallnau, 2017). Such relationships often suggest thematic links, as high correlation values imply overlapping competencies or teaching approaches (Pallant, 2020).

2) *Moderate Correlations*: Several pairs of courses demonstrate moderate correlations, typically ranging from  $r = 0.3$  to  $r = 0.5$ . For instance: CE 8401 and CE 8403 have a correlation of  $r = 0.46$ . CE 8405 and CE 8412 exhibit a correlation of  $r = 0.56$ , suggesting that performance in these courses is related but not strongly dependent on one another. These moderate correlations may arise from shared

**Table 2.** Correlation matrix of the courses.

	CE 8316	CE 8401	CE 8402	CE 8403	CE 8404	CE 8405	CE 8406 CE 8407	CE 8408	CE 8409	CE 8410	CE 8411	CE 8412	CE 8413	CE 8414 CE 8415	CE 8416
CE 8316	1														
CE 8401	0.09	1													
CE 8402	0.11	0.35	1												
CE 8403	0.03	0.46	0.36	1											
CE 8404	0.14	0.52	0.46	0.59	1										
CE 8405	0.2	0.51	0.39	0.46	0.47	1									
CE 8406 CE 8407	0.09	0.27	0.39	0.4	0.39	0.23	1								
CE 8408	0.13	0.25	0.22	0.24	0.33	0.4	0.16	1							
CE 8409	0.06	0.44	0.45	0.56	0.46	0.44	0.48	0.3	1						
CE 8410	0.09	0.43	0.35	0.49	0.42	0.39	0.52	0.18	0.56	1					
CE 8411	0.17	0.49	0.41	0.49	0.52	0.41	0.49	0.23	0.5	0.62	1				
CE 8412	0.11	0.53	0.5	0.57	0.71	0.56	0.44	0.31	0.53	0.48	0.52	1			
CE 8413	-0.11	0.44	0.25	0.36	0.32	0.22	0.29	0.09	0.41	0.44	0.54	0.33	1		
CE 8414 CE 8415	0.12	0.33	0.29	0.45	0.37	0.37	0.19	0.19	0.4	0.42	0.4	0.43	0.33	1	
CE 8416	0.15	0.19	0.24	0.21	0.26	0.39	0.21	0.6	0.27	0.22	0.27	0.32	0.08	0.21	1

foundational knowledge, partial content overlap, or complementary skills, which can create a moderate level of association in student performance (Field, 2018).

3) *Low and Near-Zero Correlations:* Some course pairs show very low or near-zero correlations, indicating minimal relationships in student performance across these courses. For example: CE 8316 and CE 8401 have a low correlation of  $r = 0.09$ . CE 8408 and CE 8413 have a correlation of  $r = 0.09$ , showing minimal association in student performance. Low correlations suggest that success in one course does not predict success in the other, possibly due to differences in course content, skills required, or instructional styles (Gravetter & Wallnau, 2017; Tabachnick & Fidell, 2019).

4) *Negative Correlations:* A single instance of a weak negative correlation exists in the matrix: CE 8316 and CE 8413 show a slight negative correlation ( $r = -0.11$ ), suggesting that high scores in one course might correspond with lower scores in the other. This could point to contrasting skill sets or differing assessment criteria, highlighting a potential divergence in focus or difficulty level (Pallant, 2020).

5) *Course Groupings with Higher Correlations:* The matrix suggests possible clusters of courses with higher inter-correlations, potentially indicating thematic or skill-based connections: CE 8404, CE 8410, CE 8411, and CE 8412 show moderate to high correlations with one another, with values often exceeding  $r = 0.5$ . This may indicate that these courses build on related skills or content areas, al-

lowing students to perform consistently across them (Field, 2018). Another cluster appears with CE 8405, CE 8409, and CE 8412, suggesting a grouping of courses where students' performance is relatively consistent, possibly due to shared topics or aligned assessments (Gravetter & Wallnau, 2017).

6) *High Correlation Between CE 8408 and CE 8416*: The strong correlation ( $r = 0.60$ ) between CE 8408 and CE 8416 suggests a significant performance relationship, potentially due to similarities in content, prerequisite skills, or assessment methods. Such high correlations can indicate that these courses may require related competencies or reinforce similar concepts, benefiting students who excel in these areas (Pallant, 2020).

### 4.3. Eigenvalues, Percentage Variance and Component Numbers for the Courses

Table 3 shows principal components of eigenvalues and variance contributions for each principal component extracted in a factor analysis. The eigenvalues represent the amount of variance explained by each component, with higher eigenvalues indicating greater explanatory power (Field, 2018). Generally, components with eigenvalues greater than 1 are considered significant, as they explain more variance than a single observed variable (Kaiser, 1960). The cumulative percentage column provides insight into the total variance explained as components are added, indicating how much of the total variability in the data is

**Table 3.** The eigenvalues, percentage variance and component numbers derived from the courses.

Component	Eigenvalue	% of Variance	Cumulative %
1	6.21	41.38	41.38
2	1.51	10.07	51.45
3	1.01	6.71	58.16
4	0.93	6.19	64.35
5	0.84	5.61	69.96
6	0.7	4.7	74.66
7	0.65	4.33	78.99
8	0.55	3.66	82.65
9	0.49	3.25	85.9
10	0.43	2.89	88.79
11	0.39	2.62	91.41
12	0.37	2.46	93.87
13	0.37	2.43	96.31
14	0.31	2.06	98.36
15	0.25	1.64	100

accounted for by the components.

Principal Component 1 has a significantly high eigenvalue (6.21), capturing 41.38% of the total variance, which suggests it is a dominant factor in the dataset. This component likely represents a core underlying structure that influences a substantial portion of the observed variables, which is common in datasets with a single strong factor (Gravetter & Wallnau, 2017). High variance explained by Component 1 indicates that the variables share a considerable amount of common variance, potentially reflecting a foundational factor or skill set that is pervasive across the variables. Principal Component 2, with an eigenvalue of 1.51, explains an additional 10.07% of the variance, bringing the cumulative explained variance to 51.45%. This component is still significant as its eigenvalue is above 1, indicating it explains more than an individual variable's worth of variance. It may represent a secondary, distinct factor or skill area that complements the dominant factor identified in Component 1, possibly addressing more specific or specialized aspects within the dataset (Pallant, 2020). Principal Component 3 has an eigenvalue slightly above 1 (1.01), explaining an additional 6.71% of the variance. This component brings the cumulative explained variance to 58.16%, indicating that the three components together capture over half of the total variance in the data. The fact that Component 3's eigenvalue is barely above 1 suggests it may be a less robust but still meaningful factor, potentially representing a narrow or niche area relevant to certain variables but not as influential as the first two components (Tabachnick & Fidell, 2019).

Principal Components 4 through 15 all have eigenvalues below 1, with each explaining less than 7% of the variance. These components cumulatively explain additional variance, but the diminishing eigenvalues suggest that they capture unique, minor variances or residual noise rather than substantial underlying factors (Field, 2018). In line with Kaiser's criterion (Kaiser, 1960), components with eigenvalues below 1 are generally considered insignificant, meaning they explain less variance than a single observed variable. The first three components together explain 58.16% of the total variance, which is substantial for social science data, where explained variance of 50% - 60% is considered adequate (Hair et al., 2010). This cumulative variance suggests that the primary factors provide a reasonable account of the variability across the observed variables, capturing the essential underlying structure without excessive complexity.

#### 4.4. Factor Score Coefficient of the First Three Principal Components

**Table 4** shows the component loadings table presents the results from a factor analysis using "varimax" rotation, which enhances interpretability by producing orthogonal (uncorrelated) components. Each course loads onto one of three identified components, with "Uniqueness" values representing the proportion of variance in each variable that is not explained by the components. Loadings closer to 1 indicate a stronger association with a specific component, while higher uniqueness values suggest that certain courses contain unique variance not captured by

**Table 4.** Component loading for the courses.

	Component			Uniqueness
	1	2	3	
CE 8411	0.77			0.39
CE 8410	0.76			0.42
CE 8412	0.73	0.31		0.36
CE 8403	0.72			0.44
CE 8409	0.72			0.43
CE 8404	0.7			0.42
CE 8401	0.67			0.51
CE 8406 CE 8407	0.64			0.59
CE 8413	0.64		-0.4	0.43
CE 8414 CE 8415	0.61			0.6
CE 8402	0.58			0.6
CE 8405	0.53	0.49		0.44
CE 8408		0.87		0.23
CE 8416		0.84		0.27
CE 8316			0.92	0.15

the components.

*Principal Component 1 (Core Knowledge or General Academic Performance):* Courses Loading Highly on Principal Component 1 include CE 8411 (0.77), CE 8410 (0.76), CE 8412 (0.73), CE 8403 (0.72), CE 8409 (0.72), CE 8404 (0.70), CE 8401 (0.67), CE 8406/CE 8407 (0.64), CE 8413 (0.64), CE 8414/CE 8415 (0.61), CE 8402 (0.58), and CE 8405 (0.53). This principal component likely represents a foundational or core academic performance factor. High loadings on Principal Component 1 suggest that students who perform well in one of these courses also tend to perform well in others, potentially due to shared content or similar assessment methods (Gravetter & Wallnau, 2017). This component may represent courses that are content-heavy, theoretical, or require similar cognitive skills, indicating a general academic performance area within the curriculum (Pallant, 2020).

*Principal Component 2 (Specialized or Applied Skills):* Courses Loading Highly on Principal Component 2: CE 8408 (0.87) and CE 8416 (0.84). Principal Component 2 appears to capture a specialized skill set, particularly in applied knowledge or practical skills, unique to CE 8408 and CE 8416. These courses' high loadings on Principal Component 2 imply that their content or assessment methods are distinct from the foundational courses in Principal Component 1. This could suggest that these courses focus more on hands-on applications or require practical competencies, thereby creating a distinct grouping from the theoretically oriented

courses in Principal Component 1 (Tabachnick & Fidell, 2019).

*Principal Component 3 (Independent Knowledge or Advanced Skills):* Course Loading Highly on Principal Component 3: CE 8316 (0.92). CE 8316's high loading on Principal Component 3 (0.92) and minimal loading on other components suggest it is uniquely distinct from the other courses, possibly covering an advanced or specialized topic. The high loading and low uniqueness value indicate that this course emphasizes a unique set of skills or knowledge not addressed in other components. It may likely require independent thinking, advanced problem-solving skills, or specific technical knowledge, distinguishing it from the broader curriculum (Pallant, 2020; Field, 2018).

The "Uniqueness" values quantify the proportion of variance in each course not explained by the identified components. Courses with lower uniqueness scores have most of their variance explained by the components, while higher uniqueness scores indicate that a considerable portion of their variance is specific to that course. For example: CE 8316 has a low uniqueness (0.15), meaning that most of its variance is captured by Component 3, reinforcing its distinct positioning. CE 8408 has a uniqueness of 0.23, suggesting that much of its variance is captured by Component 2. CE 8414/CE 8415 have relatively higher uniqueness values (0.60), indicating that a significant portion of variance is unique to these courses, suggesting additional factors may influence performance in these courses (Field, 2018).

## 5. Conclusion

This study utilised Principal Component Analysis (PCA) to examine the academic performance of final-year Civil Engineering students at Mbeya University of Science and Technology (MUST) in the academic year 2023/2024, revealing three primary factors influencing student outcomes. The first principal component, Core Academic Knowledge, accounts for the highest variance (41.38%) and encompasses foundational courses such as CE 8411 (Solid Waste Management), CE 8410 (Structural Timber Design), and CE 8412 (Industrial Building Construction). These findings highlight the importance of a structured curriculum that progressively reinforces essential engineering concepts to enhance student mastery. The second principal component, Specialized Applied Skills, explains 10.07% of the variance, primarily comprising project-based courses (CE 8408 and CE 8416). This indicates the need for enhanced practical training through field projects, internships, and industry collaborations to ensure students gain hands-on experience aligned with industry demands. The third principal component, Advanced Independent Skills, contributes 6.71% of the variance and is uniquely represented by CE 8316 (Industrial Practical Training III), emphasizing the need for tailored support, such as one-on-one mentoring and specialized software resources, to equip students with high-level technical expertise.

Additionally, the study identified strong correlations between specific courses, suggesting thematic and skill-based groupings that could inform cluster-based ac-

ademic support programs. Regular application of PCA and predictive analytics could further enhance the early identification of at-risk students, facilitating timely interventions and improved retention rates.

## 6. Recommendations

*Enhanced Curriculum for Core Academic Knowledge:* Since Core Academic Knowledge contributes the most to student performance, as evidenced by its 41.38% explained variance, these courses should be systematically structured to build a strong foundational skillset. This could involve aligning instructional goals across courses such as CE 8411 (Solid Waste Management), CE 8410 (Structural Timber Design), and CE 8412 (Industrial Building Construction) to reinforce concepts progressively. Integrating advanced tutorials, collaborative learning, and resource support can improve mastery in these courses, which have consistently shown high loadings on the first component (Tabachnick & Fidell, 2019).

*Targeted Support for Specialized Skills Courses:* Courses such as CE 8408 (Project I) and CE 8416 (Project II) load heavily on the Specialized Applied Skills component, which accounts for 10.07% of the variance. This finding suggests a strong need for additional resources in these applied courses, such as field-based projects, labs, or internships. These practical experiences could enhance students' skill acquisition, aligning their competencies more closely with industry requirements. Partnerships with local engineering firms could provide students with real-world application opportunities, ensuring they are practice-ready upon graduation (Pallant, 2020).

*Dedicated Resources for Advanced Courses:* CE 8316 (Industrial Practical Training III) uniquely loads onto the Advanced Independent Skills component, with an eigenvalue-driven variance contribution of 6.71%. This course represents a specialized focus area, suggesting that students may benefit from unique support such as one-on-one mentoring, specialized software, or advanced lab facilities. Providing these resources could better prepare students for the demands of high-level engineering roles, addressing the specific technical skills required for industry positions.

*Cluster-Based Academic Support Programs:* Using the correlation matrix, student performance clusters identified through PCA and clustering techniques can guide tailored support. Students in the Core Knowledge cluster, for example, could benefit from a structured progression of conceptually linked courses. Conversely, those excelling in applied courses but needing core support could benefit from integrative review sessions or skill workshops, ensuring a well-rounded skill set aligned with program goals (Field, 2018).

*Regular Application of PCA and Predictive Modelling:* Regular application of PCA, combined with predictive analytics, could enable early identification of performance trends, helping to address gaps proactively. For instance, future research could incorporate predictive models such as neural networks or machine learning algorithms to track student progress over time, enabling timely interventions that

improve academic outcomes and support retention efforts (Wang et al., 2021). Integrating PCA-driven dimensionality reduction with predictive models has been shown to improve predictive accuracy, facilitating data-driven decisions for academic support (Sharma et al., 2019).

*Expansion of the Study for Comprehensive Insights:* While the current findings provide valuable insights into academic performance factors, expanding the study to include additional variables such as socio-economic background, learning styles, and engagement levels could offer a more comprehensive understanding of student success determinants. Longitudinal studies tracking students over multiple academic years could also help identify trends and intervention impacts more effectively. Furthermore, integrating qualitative data from student feedback and faculty evaluations could enrich the findings, allowing for a more holistic approach to curriculum enhancement and student support strategies (Creswell & Plano Clark, 2018).

## Acknowledgements

The authors sincerely appreciate the University's support, with special recognition to Dr. Muya Somo Mugaza, the Director of Undergraduate Studies, Ms. Anith Shirima, the Examination Officer, the members of the Department of Civil Engineering and the Examination Office.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Abdi, H., & Williams, L. J. (2010). Principal Component Analysis. *WIREs Computational Statistics*, 2, 433-459. <https://doi.org/10.1002/wics.101>
- Allison, P. D. (2010). Missing Data. In R. E. Millsap, & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 632-657). SAGE Publications.
- Cai, Y., Xu, B., & Zou, Y. (2020). Application of Robust PCA in Student Performance Analysis. *Education and Information Technologies*, 25, 4725-4741. <https://doi.org/10.1007/s10639-020-10151-x>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). Sage Publications.
- Engineers Registration Board (ERB) (2023). *Annual Report on Engineering Registrations and Professional Development*. Engineers Registration Board.
- Field, A. (2018). GLM 5: Mixed Designs. In A. Field (Ed.), *Discovering Statistics Using IBM SPSS* (5th ed., pp. 703-734). SAGE Publications.
- Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism*, 10, 486-489. <https://doi.org/10.5812/ijem.3505>
- Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the Behavioral Sciences* (10th ed.). Cengage Learning.

- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Prentice Hall.
- Jolliffe, I. T., & Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A*, *374*, Article 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*, 141-151. <https://doi.org/10.1177/001316446002000116>
- Kim, D., & Maloney, T. (2021). Higher Education, Employment, and Economic Growth: Evidence from Developing Economies. *Journal of Development Studies*, *57*, 820-836.
- Kim, H., & Kim, J. (2018). Analysing Academic Performance: Multivariate Approaches in Educational Contexts. *Educational Assessment*, *23*, 345-364.
- Martínez-Ruiz, J. M., & García, A. (2019). The Role of Engagement Metrics in Online Academic Performance: A PCA Approach. *Interactive Learning Environments*, *27*, 391-406.
- Mlambo, E. (2019). Education and National Development in Sub-Saharan Africa: The Impact of Higher Education on Economic Growth. *Africa Development Review*, *31*, 230-246.
- Nguyen, P., Lee, Y., & Choi, J. (2020). Robust PCA in International Student Performance Analysis. *Asian Journal of Education*, *41*, 15-30.
- Pallant, J. (2020). *SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using IBM SPSS* (7th ed.). Routledge.
- Praveen, K., & Mahapatra, D. (2017). Using PCA to Evaluate Technical and Non-Technical Subject Performance in Engineering Programs. *Journal of Engineering Education*, *106*, 80-94.
- Sharma, P., Bansal, A., & Mehta, S. (2019). ANN-Based Prediction of Academic Performance with PCA for Dimensionality Reduction. *Computers & Education*, *135*, 31-40.
- Shlens, J. (2014). *A Tutorial on Principal Component Analysis*. arXiv: 1404.1100. <https://arxiv.org/abs/1404.1100>
- Singh, V., & Upadhyaya, M. (2019). Clustering Student Performance Using PCA: Educational Applications. *International Journal of Advanced Computer Science and Applications*, *10*, 68-73.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics* (7th ed.). Pearson.
- Tarabasz, A., Kowalik, M., & Jankowski, M. (2021). Quantitative Assessment in Higher Education: PCA in Performance Evaluation. *Studies in Higher Education*, *46*, 706-719.
- Wang, F., Liu, S., & Zhang, Y. (2021). Socio-Economic Influences on Academic Success: A PCA Approach. *Journal of Educational Sociology*, *29*, 443-459.
- World Bank (2020). *The World Bank Annual Report 2020: Supporting Countries in Unprecedented Times*. The World Bank.
- Zain, M., Ahmad, A., & Rafiq, M. (2020). Identifying Student Performance Clusters in Secondary Education: A PCA-Clustering Approach. *Journal of Educational Research and Development*, *10*, 252-267.
- Zhang, T., Sun, Q., & Li, H. (2021). Predicting Student Retention Using PCA and Support Vector Machines. *International Journal of Educational Technology*, *12*, 90-102.