CHAPTER III
RESEARCH METHODS

3.1 Introduction

This chapter discusses the research methods used during the author's research. Discussion of research methods in this chapter includes research approaches and methods, research subjects, research instruments, data collection techniques, data analysis techniques and research procedures. Before determining the approach and research method to be used, the researcher will first describe the main purpose of this study. The purpose of the study is to determine the levels of EEE students' failure effects of mathematics on the engineering courses they offer, and therefore how best their curriculum mathematics supports the engineering courses they study.

3.2 Object of Research

The focus of this study is on the different roles played by the research variables. The variables in this research are categorized as follows:

1. Free Variable (Independent)

The independent variables that became the cause in this study are students' failure scores in mathematics achievement test (MAT). These were scored in five areas of mathematics: Algebra, Functions, Trigonometry and Complex Numbers, Calculus and Differential Equations, and Probability. Also, real scores were used in some aspects of the investigations.

2. Dependent Variable

The dependent variable in this study is *achievement in electrical and electronic engineering education (ENG)*. This was measured by students' scores in the courses they took in the third and fourth semesters of the HND program.

3. Mediating Variable (mediator)

The mediating variables in this study is students' cognitive failure scores in five areas of mathematics: Algebra, Functions, Trigonometry and Complex Numbers, Calculus and Differential Equations, and Probability, measured at the levels of the Bloom's taxonomy.

4. Variable moderator

The moderator variables that support this study are students' foundation in mathematics, measured by their WASSCE grades in elective mathematics. This was categorized into three: (1) high grade; (2) average grade; and (3) low grade.

3.3 Research Subjects

In research, the term "research subjects," also referred to as "participants" or "subjects," denotes the individuals or entities engaged in a research investigation (Ketefian, 2015). These subjects are the individuals or groups on which researchers collect data and conduct investigations in order to answer research questions or test hypotheses (Cooper et al., 2019).

Following the focus of this research, the participants were the HND electrical and electronic engineering students in their final year in 2021/2022 academic year at the ten technical universities in Ghana who had already completed their engineering mathematics courses in the area of the selected domains during the first and second semesters. These were made up of students who previously studied in the SHS, and the pure Technical Schools for their pre-tertiary education. Four TUs were purposively selected with two from the southern part, one from almost the middle belt and one from the northern part of Ghana. Second year HND EEE students in Cape Coast Technical University who had just completed their fourth semester were selected for testing the MAT instruments. Within the TUs that were purposefully sampled, a cluster sampling technique was used. Those who turned up to respond to the achievement test constituted the total sample. The sample size was 488 students, and this included 273 students from Accra Technical University (ATU), 55 students from Cape Coast Technical University (CCTU), 129 students from Ho Technical University (HTU), and 31 students from Wa Technical University (WTU).

3.3.1 Population

The target group about which a researcher wants to gain information and draw conclusions is known as the population (Mweshi & Sakyi, 2020). This is a group of individuals with common characteristics that are of interest to the researcher. The population from which the researcher can actually select subjects for a study is termed as the accessible population (Pandey & Pandey, 2021). In this study, the target population was the set of 2021/2022 final year HND EEE regular students in the ten TUs in Ghana. They were a total of 1,034 students from ten public TUs (obtained from Ghanaian TUs' matriculation brochures 2019). These are distributed as in Tables 3.1

Table 3.1 Target Population: 2021/2022 Final Year HND Students in Electrical and Electronic Engineering Program in TUs of Ghana

University	Number of Students
Accra Technical University	373
Bolgatanga Technical University	35
Cape Coast Technical University	55
Ho Technical University	147
Koforidua Technical University	74
Kumasi Technical University	144
Sunyani Technical University	57
Takoradi Technical University	68
Tamale Technical University	49
Wa Technical University	32
TOTAL	1,034

Source: Processed by researcher

Only regular students were chosen for the study because unlike their weekend and evening school counterparts, they go through a well-planned academic structured programs which support the learning of mathematics for proper understanding of their engineering courses. The institutions chosen are similar TUs which offer a common HND program, thus the students offer the same engineering courses and the similarity in characteristics helped in assessing the psychometric properties of the individual student (Anwar & Menekse, 2021). This study was of more a cohort than individual. It involved a study of the 2019/2020 group of students who completed their program in 2022/2023. The data taking captured their MAT scores and the scores in various core courses they have studied in order to find the current state of the reliability of the measures.

3.3.2 Sample and Sampling Procedure

Table 3.2 Sample for the Study

No.	University	Sample
1	ATU	273
2	CCTU	55
3	HTU	129
4	WTU	31
Total	TUs	488

Source: Field work

A sample is a collection of individuals, items, or events that accurately reflects the characteristics of the larger group it is taken from (Mweshi & Sakyi, 2020). Having a larger sample size increases the likelihood that the scores on the measured variables will accurately reflect the scores of the entire population. This phenomenon is commonly referred to as the law of large numbers (Descours et al., 2022). However, with the case of this research, the researchers anticipated that the data from the TUs would be homogeneous, with the possible exception of environmental conditional variables, that might have a varied impact on teaching and learning (Gilavand & Hosseinpour, 2016). The homogeneity in the data could be possible as a result of the common entry requirement and syllabus for the HND EEE program for the TUs. This background therefore encouraged the researchers to employ the purposive sampling for the TUs with geographical (or environmental) consideration, that led to a purposive cluster sample of 488 students from ATU, CCTU, HTU, and WTU, as presented in Table 3.2.

3.4 Research Design

A research design is a comprehensive strategy or plan for conducting a research study (Dannels, 2018). It is important to consider the research design based on the type of research being conducted. It is important to select a design that effectively communicates the fundamental structure and objectives of the study(Kish, 2017). Various factors, such as the nature of the hypothesis, the variables at play, and the constraints of the environment, all play a role in determining the appropriate design (Mehboob et al., 2021). In this research, multiple research designs will be employed to answer all the research questions. Three main types of research designs employed are (1) the longitudinal design, (2) the descriptive design, and (3) the causal-comparative type of study design.

Longitudinal design involves collecting data from the same individuals or entities at multiple points in time (Diefendorff et al., 2021). Longitudinal data allows researchers to examine changes in variables over time and to test causal relationships between them. In the case of this research, the stability of relationships between mathematics and EEE courses could be examined over time. The descriptive exploratory design was employed for this study because the study sought to explore the levels of failure effects of mathematics among EEE students with regards to their curriculum mathematics contents of their programs without any manipulation. This descriptive design mainly described, and documented aspects of a situation as it naturally occurred (Miller & Salkind, 2002).

The causal-comparative type of study design was also chosen for this study. By identifying groups of variables where the independent variable occur at various levels, the design helps to establish cause-and-effect relationships. Next, one can check to see if the groups in the study differ on the dependent variables (Asenahabi, 2019). In our case, it could help us to identify different levels of the cognitive domain in which the EEE students could only respond to the HOT problems given to them, in order to ascertain their level and pattern of thinking. As the main characteristic of causal-comparative study, the cognitive level as the independent variable could measure in their categories (high or low).

3.5 Research Instruments

Research instruments are tools used to observe and measure natural and social phenomena (Khaldi, 2017). In other words, instruments are tools to help researchers to collect data (Monday, 2020; Wa-Mbaleka, 2020).

3.5.1 The Mathematics Achievement Test (MAT 1 and II)

The MAT used for this research were both objective (MAT 1) and subjective (MAT II) tests. The test items in MAT 1 and MAT II were constructed using the De-Lange's assessment model (see Sub-section 3.5.2) (Verhage & de Lange, 1997). The MAT I was made up of five sections, A to E. Each section consisted of twenty (20) objective test items, covering the course content for the study in the six levels of cognitive domain of knowledge, comprehension, applications and Higher-order thinking (HOT), where HOT consists of analysis, synthesis and evaluation (Rasyidi & Winarso, 2020). MAT I was in the domains of Algebra, Functions, Trigonometry and Complex Numbers, Calculus and Differential Equations, and Probability, making the five main sections (A to E) of a total of 100 mathematics test items. Because of the volume of the test, MAT 1 was administered in three different occasions under standard examinations condition. The many items in MAT 1 were to ensure repeated measurements in the cognitive domain and thus reduce the effect of using multiple choices in the measurements. We were also motivated by the effectiveness in assessment, and positive impact multiple-choice question authoring and regular participation has on students' learning (Butler, 2018; Xu et al., 2016). Table 3.3 shows the test item specifications, and for MAT 1.

The second instrument, MAT II, used for this study was made up of five subjective mathematics problems, 1 to 5, covering Algebra, Functions, Trigonometry, Calculus and Differential Equations, and Probability respectively. This achievement test was carefully prepared such that each of the test items assessed students' abilities in knowledge, comprehension, application, and HOT. All the five questions were prepared such that the peak of the cognitive skills required (cognitive peak) were in the HOT domain. Each of the questions was made to have sub-questions which was somehow generated with order of difficulty, according to the De-Lange's assessment model (Verhage & de Lange, 1997).

Table 3.3 Achievement Test Item Specification in Mathematics (MAT 1)

Domain of Knowledge	Cognitive Level				Total
	Knowledge	Comprehension	Applications	HOT	
Algebra	5	5	5	5	20
Functions	5	5	5	5	20
Trigonometry and Complex Numbers	5	5	5	5	20
Calculus and Differential Equations	5	5	5	5	20
Probability	5	5	5	5	20
Total	25	25	25	25	100

Source: Researcher's construct

3.5.2 Using the De-Lange's Assessment Model

Mathematics ability is considered in three parts: curriculum mathematics; mathematical thinking; and engaging ability. Because the population will be made of students who are engaged with the same curriculum for their programs, it will be convenient to employ the curriculum mathematical ability in our methodology. The methodology employed to assess curriculum mathematics ability in this study was derived from De Lange's mathematics assessment pyramid. De Lange's assessment model organizes mathematics questions in a pyramid based on their mathematical content, difficulty level, and cognitive demands. An effective test encompasses a wide range of questions, covering different subject areas, with varying levels of difficulty and requiring different levels of thinking. As the complexity of the subject matter increases, it becomes more challenging to differentiate between different mathematical content areas. Additionally, the distinction between easy and difficult questions becomes less pronounced, resulting in a pyramidal model (Goold, 2012; Verhage & de Lange, 1997). Here is the pyramid shown in Figure 3.1.

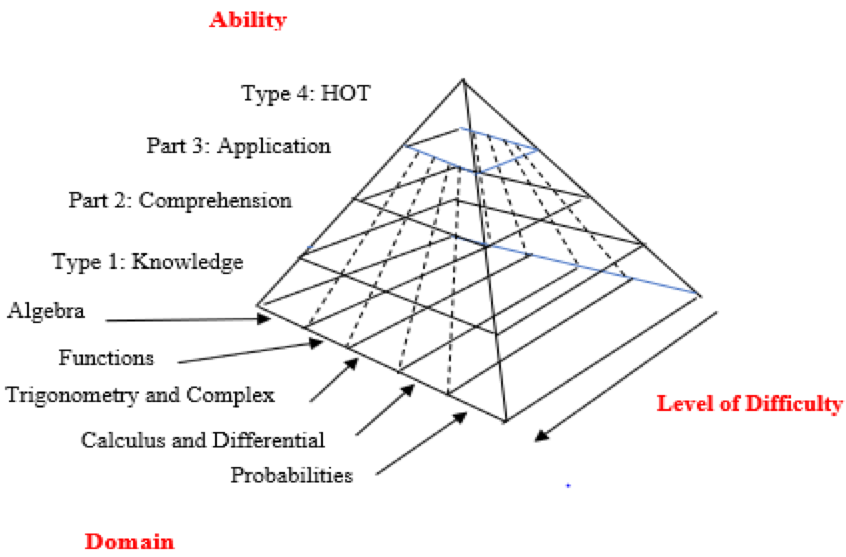

Figure 3.1 Assessment Pyramid for Curriculum Mathematics

3.5.3 Validity Test

Validity of instruments in the context of quantitative research is proposed as "the degree to which it measures what it is supposed to measure" (Dave et al., 2021). This means that the validity of a study is related to the extent to which a researcher measures what should be measured (Heale & Twycross, 2015; Mohajan, 2017). Furthermore, in conducting the validity, three types of validity used to measure research validity are *content validity, criterion related validity,* and *construct validity* (Almanasreh et al., 2019; Cohen et al., 2017). Content validity refers to whether the items or questions within a questionnaire or test adequately cover all the material that needs to be measured (Terwee et al., 2018). Criterion validity is about establishing whether a newly developed measurement tool corresponds effectively with other established instruments or recognized standards in particular scientific fields (Mohajan, 2017). Construct validity is concerned with verifying if the research tools used are built upon suitable and pertinent theoretical foundations (Almanasreh et al., 2019; Mohajan, 2017).

For the MAT, testing the validity of the content was done by comparing the contents of the MAT with the subject matter that has been taught. Technically, testing the construct validity and the content validity can be done with the aid of grid of test tools (Oermann et al., 2024). To assess the construct validity of the test, factor analysis was considered (Park & Kim, 2021; Vakili, 2018). By this, the underlying structure of the MAT items were examined to see if they align with the constructs (domain of mathematics and domain of knowledge). *Principal Component Analysis (PCA)* is a widely used technique in factor analysis that may be applicable in this situation (Greenacre et al., 2022).

The main idea behind the PCA is to reduce the dimensionality of a data set which is made up of a large number of variables, which are possibly interrelated, while retaining as much as possible the information in the data set. The new fewer variables are uncorrelated, while there is linear correlation among the PC's and the original variables. In PCA, we can construct a latent common structure of factors and determine their structural meanings. Principal components consist of various components, each with its own unique impact on the overall performance of the PC. The factors with significant coefficients provide strong accounts of the corresponding component. Through an analysis of the variations in factors that impact PCs, we can infer the overall implications of each PC. These implications are typically subjective and not directly observable (Li et al., 2018).

a) Principal Component Analysis Result for Validity of Research Instrument

Table 3.4 Data Layout for Principal Component Analysis

Student	ALG, KNOW	ALG, COMP	ALG, APPL	ALG, HOT	. . .	PROB, APPL	PROB, HOT
1	4	3	3	2		4	2
2	3	3	4	1		3	3
3	3	4	2	2		3	2
.	.	.	.	.		.	..
.	.	.	.	.		.	
.	.	.	.	.		.	.
M	5	4	5	4		4	3

Researcher's construct

To make PCA easier for measuring construct validity based on the results of the MAT data (MAT I and II combined), we needed to arrange the data in such a way that each test item is represented as a separate variable, with columns indicating the participant's responses to those items. The items were categorized according to the cognitive domain and levels they were supposed to measure. Here in Table 3.4 is how the layout could be structured:

By organizing the data in this manner, each test item category serves as a separate variable, allowing PCA to analyze the underlying structure and relationships among these variables to assess construct validity effectively. This format facilitates straightforward input into IBM SPSS for analysis. The results are as presented as follows: First, it is presented as the result of constructing the variance-covariance matrices based on the observational data as seen in Table 3.5.

Table 3.5 Communalities obtained from PCA

Variable	Initial	Extraction
ALG-KNOW	1.000	0.860
ALG-COMP	1.000	0.845
ALG-APPL	1.000	0.810
ALG-HOT	1.000	0.928
FUNC-KNOW	1.000	0.863
FUNC-COMP	1.000	0.841
FUNC-APPL	1.000	0.826
FUNC-HOT	1.000	0.909
TRIG-KNOW	1.000	0.882
TRIG-COMP	1.000	0.837
TRIG-APPL	1.000	0.807
TRIG-HOT	1.000	0.913
CALC-KNOW	1.000	0.854
CALC-COMP	1.000	0.812
CALC-APPL	1.000	0.796
CALC-HOT	1.000	0.885
PROB-KNOW	1.000	0.833
PROB-COMP	1.000	0.828
PROB-APPL	1.000	0.805
PRON-HOT	1.000	0.911

Extraction method: Principal Component Analysis

The extraction value needs to exceed 0.5 to qualify as a factor for all variables (Thien, 2021). As indicated in Table 3.4, all potential variables have extraction values that are above the 0.5 threshold, and so they can all be advanced for further factor analysis. The total variance analysis is then shown in Table 3.6. From Table 3.6, as also supported by Figure 3.2, there are two PCs in the initial solution having eigenvalues greater than 1. Together, they account for 85.229% of the variation in the original variables. Table 3.6 shows the PCs extracted for further analysis. These were confirmed by the Scree plot in Figure 3.2. Thirteen PCs were dropped from the PCA as they had eigenvalues less than 1. The two retained PCs were included in the varimax rotation to precisely see which instrument items each of them represents. In interpreting Table 3.7, what is needed to be understood is that the variable with the highest values of loading correlates strongly with the corresponding PC, and becomes the element of that PC.

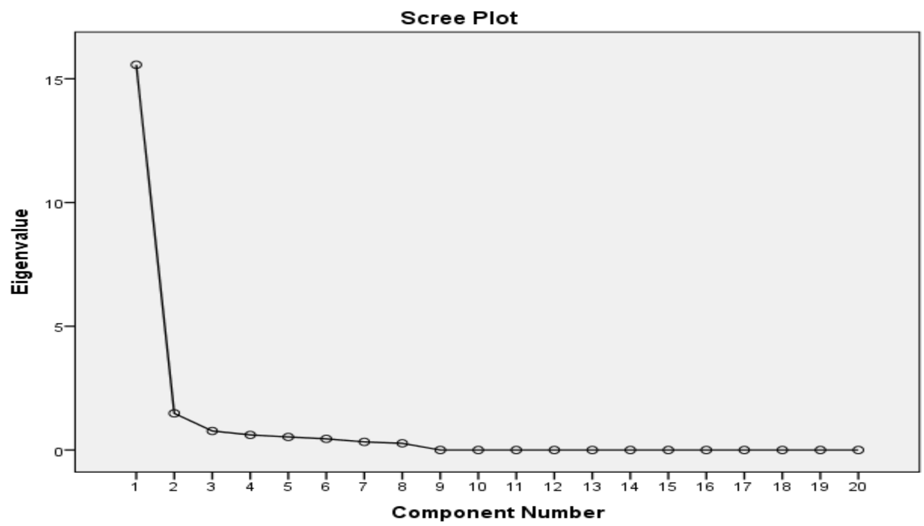

Figure 3.2 Scree Plot for Selection of PCs

Table 3.6 Total Variance Explained by Principal Components

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Comp	Total	% of Var	Cum. %	Total	% of Var	Cum. %	Total	% of Var	Cum %
1	15.565	77.823	77.823	15.565	77.823	77.823	9.315	46.577	46.577
2	1.481	7.406	85.229	1.481	7.406	85.229	7.730	38.652	85.229
3	0.769	3.843	89.071						
4	0.611	3.054	92.125						
5	0.527	2.637	94.761						
6	0.450	2.250	97.012						
7	0.329	1.643	98.654						
8	0.269	1.346	100.000						
9	3.421E-15	1.711E-14	100.000						
10	2.325E-15	1.162E-14	100.000						
11	1.918E-15	9.592E-15	100.000						
12	2.965E-16	1.483E-15	100.000						
13	-3.512E-17	-1.756E-16	100.000						
14	-9.809E-16	-4.905E-15	100.000						
15	-1.401E-15	-7.004E-15	100.000						
16	-1.786E-15	-8.929E-15	100.000						
17	-2.287E-15	-1.143E-14	100.000						
18	-3.361E-15	-1.681E-14	100.000						
19	-3.996E-15	-1.998E-14	100.000						
20	-5.075E-15	-2.537E-14	100.000						

Extraction Method: Principal Component Analysis

Based on the Rotated Component Matrix obtained from the PCA, we can draw conclusions regarding the validity of the MAT that measures students' abilities in Algebra, Functions, Trigonometry and Complex Number, Calculus and Differential Equations, and Probability across the domains of Knowledge, Comprehension, Applications, and HOT. Variables related to knowledge in each domain of mathematics exhibit high loadings on the Knowledge component. Similarly, variables related to comprehension, applications, and HOT show high loadings on their respective components. This indicates that the instrument effectively captured the intended constructs across these domains. Secondly, the loadings of variables on their corresponding components exhibit a consistent pattern. For instance, variables related to Algebra have high loadings on the Knowledge, Comprehension, Applications, and HOT components, suggesting that the instrument consistently measured Algebra knowledge across different cognitive levels. Furthermore, There might be some cross-domain relationships observed in the loadings. For example, certain variables might have moderate loadings on multiple components, indicating overlap or shared variance between different cognitive domains. This could suggest interrelatedness between certain concepts or skills across domains, which is not uncommon in educational assessments.

Concluding on the construct validity of the MAT instrument, the components extracted through PCA align well with the theoretical framework of the instrument. The Knowledge component captured factual understanding, the Comprehension component reflected the ability to interpret and explain concepts, the Applications component assessed the practical use of knowledge, and the HOT component measured critical thinking and problem-solving skills(Chandio et al., 2016). This indicates that the instrument assessed a broad range of cognitive abilities as intended.

Table 3.7 Rotated Component Matrix*

Variables Measured	Component	
	PC1	PC2
TRIG-KNOW	0.881	0.326
CALC-KNOW	0.870	0.310
FUNC-KNOW	0.863	0.344
ALG-KNOW	0.833	0.408
PROB-KNOW	0.821	0.399
TRIG-COMP	0.804	0.437
CALC-COMP	0.797	0.420
FUNC-COMP	0.796	0.456
ALG-COMP	0.761	0.515
PROB-COMP	0.754	0.510
CALC-APPL	0.640	0.622
TRIG-APPL	0.636	0.635
ALG-HOT	0.365	0.892
PROB-HOT	0.352	0.886
TRIG-HOT	0.382	0.876
FUNC-HOT	0.383	0.873
CALC-HOT	0.830	0.859
PROB-APPL	0.356	0.679
ALG-APPL	0.592	0.678
FUNC-APPL	0.636	0.649

Extraction Method: Principal Component Analysis.
Rotation Method: Verimax with Kaiser Normalization.

3.5.4 Reliability Test

The most appropriate internal consistency test to use was the *Cronbach* Alpha or also called *alpha coefficient* (Oktavia et al., 2018). Alpha coefficient values ranges from 0 (no reliability) to 1 (perfect reliability)(Arifin, 2018). Furthermore, to determine the level of reliability, data is classified as reliable if the value of the alpha coefficient ≥ 0.7(Taber, 2018). *Cronbach's Alpha* method using SPSS version 26 was used to measure the reliability of the test items in the MAT instruments. The results of the ability test for the MAT test items are presented in Tabel 3.8.

Table 3.8 Reliability Statistics for MAT Test Failure Scores in Domain of Mathematics Knowledge

Cronbach's Alpha	Number of Items
0.815	100

SPSS Output

From Table 3.8, it is stated that the reliability coefficient of the MAT Test Failure scores is 0.815. It can therefore be concluded that the MAT test items are reliable since the Cronbach Alpha is greater than 0.70. The results of reliability tests using *Cronbach's Alpha* method for the MAT scores in cognitive domain according Benjamin Bloom is also presented in Table 3.9.

Table 3. 9 Reliability Statistics for MAT Failure Scores in Cognitive Domain

Cronbach's Alpha	Number of Items
0.891	100

SPSS output

Table 3.9 shows that the reliability coefficient of the MAT scores in the cognitive domain is 0.89, and since this score is greater than 0.70 it can be concluded that the test items used to assessed students' cognitive failure (ability) are reliable.

3.5.5 Definition of Variables and Data for the Research

The data for the study constituted both the real scores and failure scores from the MAT, described in Subsections 3.5.1 and 3.5.2. Also, a secondary data which constituted ten core EEE courses offered in Semesters 3 and 4 of the HND EEE program was used for the investigations. Tables 3.10a and 3.10b provides the definition of the basic variables used for the research.

Table 3.10 Research Variables from the MAT Data

No.	Variable	Interpretation
Domain of Knowledge		
1	ALG	Knowledge or competence in Algebra
	FALG	Failure or incompetence in Algebra
2	FUNC	knowledge or competence in Functions
	FFUNC	Failure or incompetence in Functions
3	TRIG	Knowledge or competence in Trigonometry and Complex Numbers
	FTRIG	Failure or incompetence in Trigonometry and Complex Numbers
4	CALC	Knowledge or competence in Calculus and Differential Equations
	FCALC	Failure or incompetence in Calculus and Differential Equations
5	PROB	Knowledge or competence in Probability
	FPROB	Failure or incompetence in Probability
Domain Cognition		
6	KNOW	Knowledge. Ability in Knowledge (or recall). Ability to recall or recognize information and understand the basic facts and concepts within a particular domain.
	FKNOW	Failure in Knowledge. Failure in Knowledge (or recall). Inability to recall or recognize information and understand the basic facts and concepts within a particular domain.
7	COMP	Comprehension. Ability to grasp the meaning of information, explain concepts in one's own words, and make connections between different pieces of information.
	FCOMP	Failure in Comprehension. Failure or inability to grasp the meaning of information, explain concepts in one's own words, and make connections between different pieces of information
8	APPL	Application. Ability to transfer knowledge and concepts to new contexts, applying principles to solve problems, and using strategies or procedures to address specific tasks or challenges.
	FAPPL	Failure in Application. Failure or inability to transfer knowledge and concepts to new contexts, applying principles

		to solve problems, and using strategies or procedures to address specific tasks or challenges
9	HOT	Higher-order thinking. Ability to apply skills such as analysis, evaluation, synthesis, critical thinking, and creativity. Higher-order thinking tasks require individuals to analyze information, evaluate arguments, generate solutions to novel problems, and engage in reflective and creative thinking.
	FHOT	Failure in Higher-order thinking. Failure or inability to apply skills such as analysis, evaluation, synthesis, critical thinking, and creativity. Higher-order thinking tasks require individuals to analyze information, evaluate arguments, generate solutions to novel problems, and engage in reflective and creative thinking.

Table 3.11 Research Variables from Pre-tertiary and EEE Examination Records

No.	Variable	Interpretation
	Pre-tertiary	
1.	PMA	Pre-tertiary mathematics. Foundation in mathematics
		Third Semester
2.	EEE207	Digital Electronics
3.	EEE211	Telecommunication I
4.	MCE211	Thermodynamics
5.	EEE231	Electrical Machines II
6.	EEE241	Power Systems I
		Fourth Semester
7.	EEE212	Telecommunications II
8.	EEE222	Control Systems
9.	EEE225	Instruments and measurements
10.	EEE232	Electrical machines III
11.	EEE242	Power Systems II

Source: EEE Curriculum

3.6 Research Procedure

The procedure of this research outlines the different stages of the research, specifically focusing on the research design. This research is divided into multiple stages that can be grouped into three phases: research preparation and research data processing.

Phase one: In order to confirm the existence of the problem, the initial stage of this research involved conducting a thorough literature analysis. The researcher obtained an overview of the problem related to the research topic by analyzing previous findings in a systematic manner (Snyder, 2019). Now we moved on to the next stage in our research: identifying the problem at hand.

To enhance the researcher's focus on the research problem, it is important to establish delimitations that help organize the research problems and yield clear and concrete results (Hancock et al., 2021). In the third stage, the researcher first wrote the theoretical background to demonstrate the purpose and the research focus. The theoretical backgrounds showed in this research is to bring to light, the relationship between mathematics and EEE education and also, clear meaning of the *failure effects* of mathematics, and *ripple failure effects* of mathematics. The transformative theory, cognitive load theory, and attribution theory of success and failure were considered as theoretical basis for this research

After gathering the references, the next stage was the designing of the MAT instruments used to assess the competence level and thus failure effects of the EEE students' mathematics. After the preparation, the MAT instruments were tested for reliability and validity before using it to test the participants under uniform and standard examination conditions. Validation is crucial because accurate variable measurement is vital for research. It is during this stage that reliable and correct data can be obtained.

Phase two: Once the data was gathered, the subsequent step involved analyzing the data using descriptive and inferential statistics. Descriptive statistics serve the purpose of offering a structured overview of data, providing a description of the relationship between the variables that constitute the sample or population (Baffoe-Djan & Smith, 2019; Mooi et al., 2018). The data processing produced findings that can serve as informative material for the reader. The results of the inferential statistics revealed the significant impact of mathematics on the HND EEE curriculum. Additionally, it shed light on how cognitive and/or mathematics failure can hinder one's achievement in EEE education.

Phase three: The last stage of this research was report writing. If a research work or information is not published or disseminated, it would be of little use in the development of science and would not have high practical value (Brownson et al., 2018). Therefore, it is the responsibility of the researcher to accomplish a number of scientific tasks and produce an auditable written scientific report.

3.7 Data Analysis

The data analysis for this research employed both descriptive statistics and inferential statistics. For the inferential statistics, the Covariance-Based Structural Equation Modelling (CB-SEM), Bayesian multivariate regression (BMR), and the Dependency Graph Analysis were all reviewed and applied to either the real scores or failure score data, according to the research questions in Section 1.2 to estimate the failure effects of mathematics on the EEE courses in various dimensions. Further, the Receiver Operating Characteristics (ROC) Curve was reviewed to assess the sensitivity of the models generated by these failure effects, and used to classify the EEE courses according to the sensitivities to mathematics failure. Review of these statistical tools was deemed relevant to the study after the review of relevant literature on the theories underpinning the research was made.

3.7.1 Analysis by Structural Equation Modelling (SEM)

Structural Equation Modeling (SEM) is a statistical method used in the social sciences, economics, psychology, and other disciplines to examine the connections between latent and observable variables (Tarka, 2018; Westland, 2015). It allows researchers to test complex theoretical models that depict relationships among variables. SEM is often used for hypothesis testing, model comparison, and theory development (Kline, 2023). It allows researchers to analyze complex relationships among variables and test theoretical models using empirical data. SEM encompasses several types or variations, each suited to different research contexts and objectives. The Covariance-Based Structural Equation Modelling (CB-SEM) and Partial Least Square Structural Equation Modelling (PL-SEM) appear to be popular among the approaches to SEM, with CB-SEM focusing on covariance structure and hypothesis testing, while PL-SEM emphasizes latent variable estimation and prediction. The choice between the two depends on the research objectives, the complexity of the measurement model, and the nature of the data being analyzed (Kacprzak, 2018; Makatjane & Makatjane, 2017).

The researchers made a number of considerations before employing the CB-SEM approach. These included the following: CB-SEM works better when evaluating complex theoretical models and computing correlations between latent variables, such as in hypothesis testing and model confirmation (Dash & Paul, 2021). Also, When the measurement model (the connections between observable and latent variables) is well-established or when researchers have compelling theoretical arguments to support their measurement model, CB-SEM is appropriate (Hair Jr et al., 2017; Yıldız & Kelleci, 2023). Again, CB-SEM typically requires larger sample sizes compared to PL-SEM, especially when estimating complex models with many parameters. CB-SEM relies on asymptotic theory for parameter estimation, which requires sufficient sample sizes to produce reliable estimates (Hair Jr et al., 2017). This research made use of a sample size of 488 students which is enough for CB-SEM. Further, CB-SEM assumes that the data are normally distributed or approximately normal, especially for smaller sample sizes, as it relies on maximum likelihood estimation (Usakli & Rasoolimanesh, 2023). CB-SEM is suitable for estimating complex models with multiple latent variables and pathways, as it allows for testing of specific hypotheses and model constraints (Hair Jr et al., 2021). The researchers observed that with respect to the data for the study and the focus, it was appropriate to choose CB-SEM over any other SEM method.

a) Analytic Strategy

The study employed the CB-SEM techniques to examine the research hypotheses using Amos 26. CB-SEM is commonly employed to evaluate process models created by a theory (Hayes, 2009; Lei & Wu, 2007). When using CB-SEM, researchers do not search for a model that perfectly matches the data. Instead, they examine a theory by creating a model that illustrates the connections between the concepts described in that theory. These concepts are then measured using reliable observed variables (Hair Jr & Sarstedt, 2019). In doing so, researchers can “evaluate the validity of substantive theories with empirical data" (Lei & Wu, 2007), which in turn helps develop a theory (Anderson & Gerbing, 1988). Hence, the present study employed CB-SEM to reveal how the learning or understanding of EEE courses are affected if students fail to grasp the concepts of certain aspects of engineering mathematics. Also, how unsuccessful they would be if students fail to engage in cognitive training to improve their creative thinking, which can be done by recommended methods of learning including Realistic Mathematics Education (RME) (Van den Heuvel-Panhuizen & Drijvers, 2020; Zakaria & Syamaun, 2017) The maximum likelihood estimation method was chosen due to its reputation for producing more reliable parameter estimates than other estimators, such as generalized least squares. This method has been found to be effective even when the observed variables do not perfectly follow a multivariate normal distribution (Curran et al., 1996; Iacobucci, 2010).

To test the research hypotheses, this study administered three analysis. Firstly, confirmatory factor analysis (CFA) was performed to assess the measurement model. Secondly, mediation analysis was conducted by employing bootstrapping to test research hypotheses 1 (See Section 2.4). The reliability test was followed by a confirmatory factor analysis to look at the factor structure. With the help of standard absolute fit indices, the entire model fit was evaluated. Fit indices such as Chi-square, the Turker-Lewis index (TLI), comparative fit index (CFI), incremental fit index (IFI), root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) were employed to measure the fit of the model. Finally, the subgroup method (-1 SD and +1 SD) and bootstrapping were applied to conduct a moderated mediation analysis to test research hypothesis 2 (See Section 2.4). All the bootstrapping procedures were conducted with 5000 bootstrap samples (Banjanovic & Osborne, 2019; Hayes, 2009).

Effect size was discussed (Gignac & Szodorai, 2016). Hedges' g was also calculated to gauge how failure effects of various mathematics domains varied (Ellis, 2010). Besides, Pearson's Product moment correlation coefficient (r), and coefficient of multiple determination (R^2) were applied to measure the strengths of the relationships between constructs (Ellis, 2010).

b) Mediation Analysis

A mediation model was developed in this study, as shown in Figure 3.3. The path coefficients from MF to CF and CF to ENG are represented by a_1 and b_1 respectively. c_1' is the path coefficient from MF to ENG, representing the direct effect of MF on ENG. This model includes one specific indirect effect (SIE_1). The product of a_1 and b_1 represents the indirect effect of MF on ENG through CF (SIE_1). Thus the total effect of MF on ENG is quantified SIE + c_1'. Assessing such process models involves conducting a mediation analysis, which helps researchers understand how a predicting variable influences an outcome variable (Preacher et al., 2007). The mediation effect or indirect effect deserves proper attention, otherwise, "the relationship between two variables of concern may not be fully considered" (Raykov & Marcoulides, 2006,p.7).

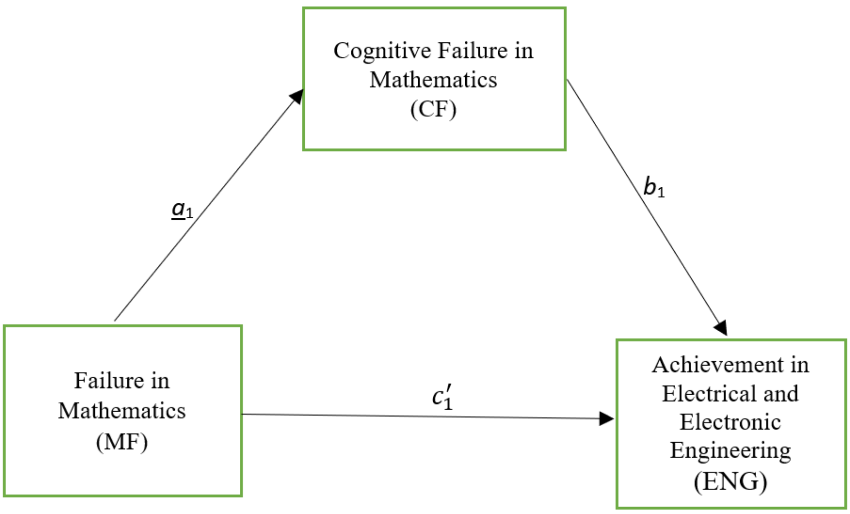

Figure 3.3 Mediation Model (Cognitive Failure Investigation)

While there are many ways to estimate the size of an indirect influence, the causal steps technique from (Baron & Kenny, 1986) has been the most popular (Hayes, 2009; MacKinnon et al., 2004). Nevertheless, the method has been criticized by some researchers for its low statistical power (Fritz & MacKinnon, 2007; Hayes, 2009), and it is comfortably applied to only the simple mediation model (Preacher et al., 2007). We therefore applied the Baron and Kenny's (1986) mediation analysis to our mediation model in Figure 3.3. The approach is based on the grounds that the product of a_1 and b_2 follows the normal distribution, which is not easy for researchers to achieve (Bollen & Stine, 1990; Preacher et al., 2007). Thus, for this study the bootstrapping technique was used (Bollen & Stine, 1990; Hair Jr & Fávero, 2019; A. F. Hayes, 2009; Preacher et al., 2007). The proposed structural model, descriptive statistics, reliability, convergent validity, and discriminant validity tests, as well as the measurement model test, test of the hypotheses, and analysis of the results are presented in the next section.

Two forms of bootstrapping were adopted in this study: the Bollen–Stine bootstrapping (Bollen & Stine, 1992) and the naive bootstrapping (Yung & Bentler, 1996). The Bollen–Stine bootstrapping was applied to modify the enlarged χ^2 due to imperfect normality of some variables and also multivariate nonnormality (Enders, 2005), whereas the naïve bootstrapping was used to conduct a mediation analysis (Hayes, 2009; Preacher et al., 2007),

c) Moderated Mediation

When the effect of one variable on another is influenced by the presence of a third variable, we refer to this third variable as a moderator (Ramayah et al., 2018). As mediation analysis has aroused considerable attention, many researchers show interest in the condition under which an indirect effect occurs, which is thus referred to as conditional indirect effects (Preacher et al., 2007) or moderated mediation (Hayes, 2018). This phenomenon is commonly referred to as conditional indirect effects or moderated mediation.

An effective way to analyze moderated mediation is by examining the mediation effect at each level of the moderator. This approach, known as the subgroup approach, is widely used in research (Edwards & Lambert, 2007; Fabrigar & Wegener, 2014). In line with Preacher et al.'s (2007) recommendation, mediation effects were estimated using the bootstrapping procedure within each subgroup. This involved calculating the effects one standard deviation below the mean (-1 SD) and one standard deviation above the mean (+1 SD) using the PROCESS micro.

Conducting Bayesian Regression analysis after the CB-SEM can offer several advantages, including flexibility in model specification, robustness to violations of assumptions, and handling of uncertainty in parameter estimates. It can be particularly useful when dealing with complex models, uncertainty in parameter estimates, or the need for model comparison and evaluation which is of interest to the researchers.

3.7.2 Analysis with Bayesian Multiple Regression and the Receiver Operating Characteristics Curve

In our case, where there are multiple response variables, it becomes necessary to employ a multivariate multiple linear regression model. The structure of a data table for multivariate multiple regression data was arranged as depicted in Table 3.12.

Table 3.12 Multivariate Multiple Regression Data

Observation i	Response Variable				Predictor			
	y_1	y_2	...	y_m	z_1	z_2	...	z_r
1	y_{11}	y_{12}	...	y_{1m}	z_{11}	z_{12}	...	z_{1r}
2	y_{21}	y_{22}	...	y_{2m}	z_{21}	z_{22}	...	z_{2r}
.	.	.		.	.	.		.
.	.	.		.	.	.		.
.	.	.		.	.	.		.
N	y_{n1}	y_{n2}	...	y_{nm}	z_{n1}	z_{n2}	...	z_{nr}

Source: Researcher's construct

Bayesian multiple regression provides a flexible, robust, and comprehensive framework for estimating the effects of each of mathematics variables on academic performance, accommodating prior knowledge, quantifying uncertainty, and offering intuitive interpretations. To answer Research Question 4, the researcher performed Bayesian multiple regression analysis using the rstanarm package. This involved fitting a regression model where the dependent variables, *EEE courses,* were predicted based on each set of multiple independent variables, *domain of knowledge* and *domain of cognition* defined in Section 3.5.5. The Bayesian approach incorporated prior distributions for the parameters, which were updated with the data to form posterior distributions. The researcher specified the model, defined the priors, and run the model to obtain the posterior estimates and diagnostic statistics. The theory behind the Bayesian Analysis employed by the researcher is found in APPENDIX 3.

The next analysis of this research will be made use of the Receiver Operating Characteristics (ROC) curve. Employing the ROC curve analysis, following the Bayesian multiple regression in this study helped in many ways: It enhanced model evaluation, helped in threshold selection, facilitated model comparison, helped to identify important features, and assess generalization performance, particularly in binary classification tasks or predictive modeling scenarios (Hellmich et al., 1998; Obuchowski & Bullen, 2018). Irwin & Irwin (2011) present a comprehensive overview of these indices and other metrics that have been developed. However, for the sake of simplicity in our model, the researchers choose to use the Youden index. Given that sensitivity and specificity are both probabilities ranging from 0 to 1, the Youden index J also falls within this range. In addition, when J is at either extreme (J = 1 or J = 0), it can be observed that both sensitivity and specificity are either both 1 (indicating 100% accuracy with complete separation of the two groups) or both 0 (indicating 0% accuracy with complete overlap of the two groups). Therefore, the Youden index offers a way to evaluate the performance of the classifier by measuring its closeness to 1.

This research therefore relies on the concepts of the ROC curves for model predictions classifications to interpret as many parameters as possible that will give further meaning to the Bayesian multiple regression analysis. The AUC value ranges from 0 to 1, where a higher AUC indicates better discrimination or predictive ability of the model (Kam et al., 2020; Nahm, 2022) gave the interpretation of the AUC as in Table 3.13.

Table 3.13 Interpretation of the Area Under the Curve

Area Under the Curve (AUC)	Interpretation
$0.9 \leq AUC$	Excellent
$0.8 \leq AUC < 0.9$	Good
$0.7 \leq AUC < 0.8$	Fair
$0.6 \leq AUC < 0.7$	Poor
$0.5 \leq AUC < 0.6$	Fail

Note: For a diagnostic test to be meaningful, the AUC must be greater than 0.5. Generally, an AUC $\geq$ 0.8 is considered acceptable(Nahm, 2022).

Using this as threshold for the predictive power of the model, determined by the ROC curves, the relations between each set of dependent and independent variables in accordance with the area under the ROC curve will be classified on a classification table for discussions on dependencies in relation to the EEE courses, with the aid of Dependency Graph Analysis.

3.7.3 Analysis with Dependency Graph

Our method for predicting the failure effects (paths) of ripple mathematics involves two distinct steps. The initial step focuses on establishing the knowledge graph that depicts potential routes through the items. A combination of different approaches is used to build the knowledge graph, including teacher-based, constraint-based, and user-interaction-based methods. The direct transitions (edges) from one item (node) to another indicate a connection between the past and present courses. By analyzing the historical paths of learners, a knowledge graph is created along with the corresponding correlations. Put simply, if users consistently follow a specific course order, there is a greater chance that curriculum developers will recommend including that part of the path in future implementations. In the given situation, an individual learning path is predicted for each learner as the second step.

a) Algorithm Design

Collaborative filtering algorithms are a class of techniques employed in recommendation systems to make predictions or recommendations about items such as movies, books, products to predictions, rather than requiring explicit information about the items being recommended to users (Alhijawi & Kilani, 2020; Martins et al., 2020; Nilashi et al., 2018) These algorithms operate under the premise that users are likely to express similar preferences going forward if they have done so in the past (Chartier et al., 2020). Collaborative filtering techniques use past user-item interaction data to infer trends and generate paths for users.

The real algorithm divides the first step into two sections. Initially, the semester core courses are converted into nodes within a graph database. Let us denote a collection of items associated with the course C by I. Prerequisites, which are specified according to the curriculum, are taken into account for constructing the primary dependency graph. The prerequisites associated with the courses are already defined as semester-based grouping and can be viewed as constraints on a pathway that should not be violated. For example, EEE211 (Telecommunication I) which is offered in the third semester is a prerequisite for EEE212 (Telecommunication II) offered in the fourth semester. The outcome is the dependency graph that shows all permitted learning transitions. Figure 1 illustrates the dependency graph containing six items (A,..., F), where B serves as a prerequisite for C (C →B), C is a prerequisite for D (D→ C), and E is a prerequisite for F (F→ E). When there is a loop in the network, there can be circular dependencies, which indicate that one item depends on another, which in turn depends either directly or indirectly on the first item. (Rötzer et al., 2022). Thus, the circular shape of a dependency graph is a result of the relationships between courses. This is illustrated in Figure 3.4.

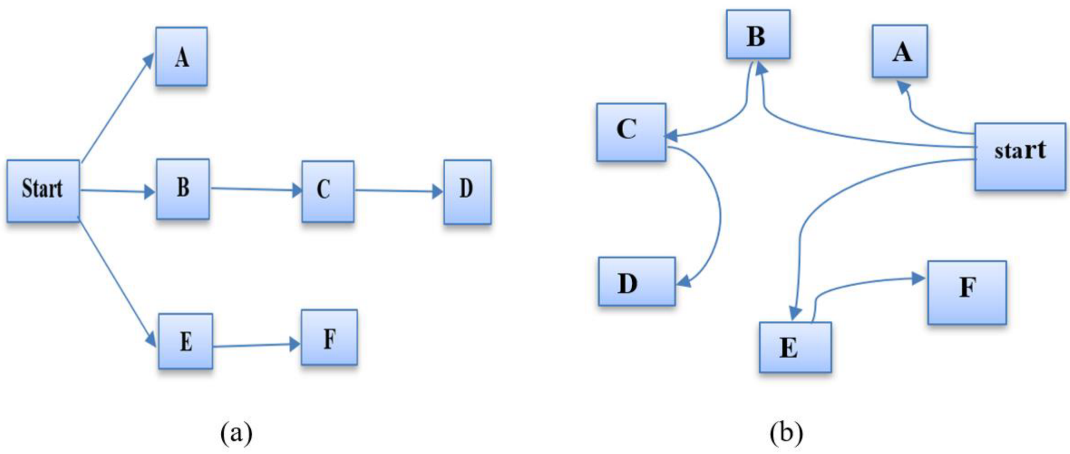

Figure 3.4 Dependency Graph

Dependency graph created with six Learning Objects and three defined prerequisites (D → C, C → B and F→ E): (a) items (except the next set) do not depend on the first (b) items depend directly or indirectly on the first item (start).

b) Knowledge Graph

A knowledge graph is a structured representation of information that shows the connections between concepts, entities, or data points within a domain. Information is arranged in a graph style, with nodes representing concepts or entities and edges representing the connections between them. Knowledge networks are made to connect and encode many kinds of data, enabling context-aware understanding and extensive semantic connections (Lampropoulos et al., 2020). The functioning of a knowledge graph revolves around its core components: nodes, edges, and the relationships they represent. Here's a simplified overview of how a knowledge graph works.

Data integration, which is the process of gathering and organizing information from multiple sources, is usually where knowledge graphs begin. Web pages, unstructured text, structured databases, and other types of data could all fall under this category. Once the data is collected, entity extraction techniques are employed to identify and extract relevant entities. Entities could be anything from people, organizations, locations, events, to abstract concepts. After extracting entities, the next step is to identify and define relationships between them. This involves understanding the semantics of the data and establishing meaningful connections. For example, if one entity is "Theodore" and another is "Cape Coast," a relationship could be "Theodore lives in Cape Coast."

A graph structure is then created using the retrieved entities and their connections as nodes and edges. An entity is represented by each node, and a connection between entities is represented by each edge. Once the knowledge network is built, users can query it to get specific data or run other kinds of analyses. This could entail more complicated queries that require taking into account many relationships and traveling multiple paths, or it could involve simpler queries such as determining the shortest path between two things. Knowledge graphs can also be integrated with machine learning and artificial intelligence algorithms to enhance their capabilities. For example, machine learning models can be trained to predict missing relationships or infer new knowledge based on existing information in the graph. Knowledge graphs are dynamic structures that evolve over time as new data becomes available. Therefore, it is essential to continuously update and maintain the graph to ensure its accuracy and relevance.

Considering the specified prerequisite conditions, we compile a comprehensive list of permissible combinations of items within the set I. These subsets of items within I signify a potential learning stage, encompassing all items consumed thus far. Set S encompasses all feasible states attainable within the course,

$$S = \{s[s \subseteq C\} \qquad (3.1)`$$

Every transition from a particular state to another is symbolized by a directed edge $e_{s1,s2}$ in E, barring instances where this movement conflicts with the prerequisite conditions. An edge establishes a connection between two states, s_1 and s_2, indicating the addition of an item to s_1 reulting in state s_2. Here, E is the set of all edges or all possible connections. The directed transition edges are augmented with an attribute that denotes the number of other learners who have previously employed the same transition. Specifically, this attribute reflects a ratio derived from the historical transitions $e_{transitions}$ between the two states, s1 and s2, and the total number of historical transitions $s_{transitions}$ leaving s_1. The outcome of the probability function p(e) determines the transition probability per edge e based on the historical movements of all learners, expressed as a percentage:

$$p(e) = \frac{e_{transitions}}{s_{transitions}} \qquad (3.2)$$

As an intermediate outcome, a knowledge graph K is produced, illustrating the full range of states and transitions within the dependency graph spanning from an initial state B, devoid of any items, to a final state $T = C$, encompassing all items from the course.

$$K = (B, T, S, E, p) \qquad (3.3)$$

Figure 3.6. illustrates an example of knowledge graph. It showcases the six items in the example of a dependency graph in Figure 3.5. This graph is calculated offline and stored in a database to optimize the future calculation of individual paths.

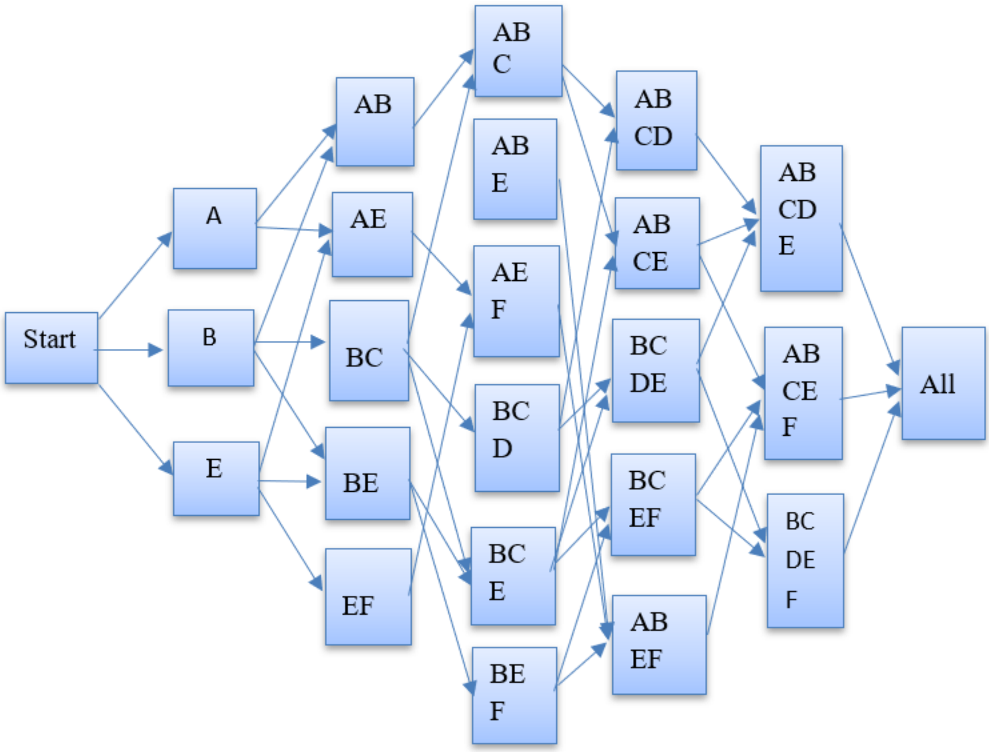

Figure 3.5 Knowledge Graph

c) Path Generation

When considering a recommendation for a new direction, all previously accessed items by the user are taken into consideration. These accessed items signify a state within the knowledge graph K, serving as a new beginning for user u at time t. As an example, when the user accesses Course C, B, and D, the state {B, C, D} is used as the starting point for the routing algorithm. If the user did not meet the necessary requirements and studied a different item beforehand, the starting state is determined by the state that best aligns with the known items.

The routing algorithm calculates the most efficient route from the initial point to the target node within the knowledge graph. The target node T encompasses all the items studied in the course, representing the course objective. R represents the set of all possible routes connecting point B to point T. During this process, the probabilities of traversed edges along each route are combined. Multiplication was not considered in this study to avoid unfairly penalizing smaller edge probabilities.

Indicates the total number of edges in route r. The most favorable option is the pathway with the highest cumulative probability, $p_{total}(r)$, is deemed preferable.

$$p_{total}(r) = \sum_{e=0}^{E_r} p(e), \tag{3.4}$$

The routing plan algorithm was employed in the evaluation (Li et al., 2022; Zografos & Androutsopoulos, 2008). It examines the transitions from T to $B_{u,t}$ in a barkward manner by taking into account the probability p(e) associated with each edge. Furthermore, since the algorithm is tailored for public transportation, it accommodates waiting times during connections. When applied to the learning context, this functionality is used to discourage shifts between higher-level topics, as learners may benefit from concentrating on similar topics consecutively, and to promote shifts only after all lower-level items pertaining to a higher-level topic have been addressed. The result is a compilation of alternative routes, each node having a specified number of branches denoted as b. Only the b branches displaying the highest transition probability p(e) from one edge to another are taken into account. Instructors have the flexibility to modify the total number of branches considered and presented.

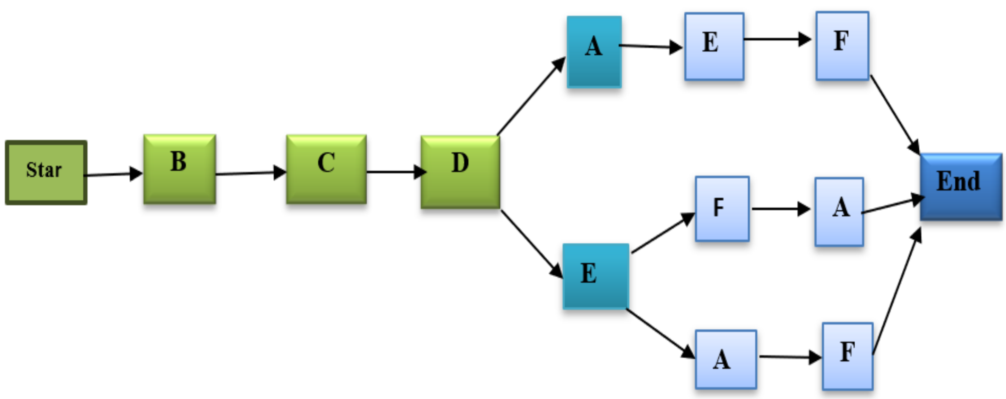

Figure 3.6 Path Presented to the Learner

(Example of a path presented to the learner when B, C and D have been consumed)

In summary, the routing algorithm for dependency graphs aims to find optimal paths through a network of dependencies between variables or entities. Using packages such as "igraph" or "networkD3" in R, the researchers were able to construct and analyze dependency graphs to identify efficient routes of the failure effects of mathematics and cognition through two consecutive semesters of the HND EEE program, which the researchers identified as the “ripple failure effect of mathematics” of the EEE courses. These algorithms are essential for tasks such as determining prerequisite chains in educational courses or optimizing workflows in complex systems (Page Risueño et al., 2020; Wang et al., 2021).

e) Application in R

Creating a dependency graphs in R, the researcher used the igraph package. This process involved defining nodes (representing entities) and edges (representing dependencies or relationships between the entities). Graph objects were created from the edge lists or the adjacency matrices and these graphs were visualized using various plotting functions available in the package. The graphs were customized with different attributes such as vertex size, labels, and edge directions to clearly represent the dependencies.