A Comparative Study on the Comprehensive Difficulty of the Mathematics Examination of Malaysia's Unified Examination and China's National Unified Examination

Corresponding Author: Kun An
quinn.an@ywies.com
Academic Coordinator
Yaw Wah International Education School of Shanghai Gubei Shanghai, China

Gopinathan A/L K Raman Kutty
Professor
Dean of the Faculty of Education
University College Fairview
Selangor, Malaysia

Abstract

This research conducts a comparative analysis of the comprehensive difficulty of mathematics examinations in Malaysia and China by examining their examination structure, content, and difficulty level. The study aims to identify similarities and differences in the education systems of the two countries and provide valuable insights for policymakers, educators, and students. A new improved Comprehensive Difficulty Model for Mathematics Examinations (CDMME) is applied to compare the comprehensive difficulty of the mathematics examinations of the Senior Middle Level Unified Examination Certificate (SUEC) and the National Unified Examination for Enrollment of Ordinary Colleges and Universities in China (Gaokao), taking into account subjective and objective comprehensive difficulty factors. The results suggest that the comprehensive difficulty of Gaokao Mathematics Examination is between that of the SUEC Further Mathematics Examination (Level I) and the SUEC Further Mathematics Examination (Level II). The research provides important implications for improving the accuracy and fairness of exams and contributes to the broader field of mathematics education and assessment. Limitations of the study and recommendations for further research are also discussed.

Key Words:

International Curriculum Comparison, Model on Exam Difficulty, Analytic Hierarchy Process, Senior Middle Level Unified Examination Certificate, Gaokao, Mathematics Examination.

Statements and Declarations

the authors of the paper submitted for publication have non-financial interests that are directly or indirectly related to the research. The authors have no affiliations or financial involvement with any organization or entity that has a direct or indirect interest in the research. The authors also have followed all ethical guidelines and protocols in conducting this research and have obtained all necessary approvals and permissions.

1. Introduction

Mathematics is considered a fundamental subject in both Malaysia and China, with rigorous examinations designed to assess the understanding and mastery of students. In recent years, there has been a growing interest in comparing the education systems and examination standards of different countries, particularly those that are globally recognized for their academic excellence. This study aims to conduct a comparative analysis of the comprehensive difficulty of the mathematics examinations of Malaysia's Unified Examination and China's National Unified Examination. By examining the examination structure, content, and difficulty level, this study seeks to identify similarities and differences in the education systems of the two countries, providing valuable insights for policymakers, educators, and students alike, thus contributing to the relevant body of knowledge.

When people talk about the difficulty of mathematics exams, they are often based on their own feelings and understanding of the exams but not a quantified result. As researchers working in international schools, we have observed students and parents struggling to decide which standardized mathematics examination to take each year due to the lack of a generally agreed-upon quantitative assessment of the difficulty of these exams. This has led to confusion and uncertainty surrounding the selection of exams and the preparation required, which can impact students' academic performance and future opportunities. The Comprehensive Difficulty Model for Mathematics Examinations (CDMME) is designed to quantitatively compare the comprehensive difficulty of mathematics exams. The comprehensive difficulty of a mathematics examination paper is an index that can not only reflect the comprehensive characteristics of the examination, but also objectively describe the level of difficulty of the examination (Bao, 2002). However, the CDMME model only considers the difficulty of the questions on mathematics exam papers, but does not take into account the study duration, assessment objectives (AOs) of the syllabus and studying language, which all of them will also affect the comprehensive difficulty of mathematics examination. Additionally, the CDMME model uses the number of questions for each difficulty level to calculate the comprehensive difficulty, whereas using the total marks of the questions of each difficulty level is more accurate. Therefore, to assess the comprehensive difficulty of the mathematics examinations more comprehensively, the researchers implemented but also improved the CDMME model in this research to compare the comprehensive difficulty of the Senior Middle Level Unified Examination Certificate (SUEC) and Mainland China’s National Unified Examination for Enrollment of Ordinary Colleges and Universities in China (Gaokao)’s Mathematics Examination.

According to the latest version of the General Syllabus of Gaokao released by the Chinese National Education Examination Authority (CNEEA), Gaokao is a selective examination taken by qualified Mainland China high school graduates and Chinese nationals with equivalent qualifications for the enrollment of ordinary colleges and universities (CNEEA, 2019). A news article on the official website of China's Ministry of Education stated that in 2021, 10.78 million candidates have taken Gaokao (Fan, 2021). However, according to the statistics from the Ministry of Education of China, in 2020, the admission rate of the 985 project universities in mainland China is only 1.9%, and the admission rate of the 211 project universities in mainland China is 5.2% (Li, 2021). Gaokao has two types of mathematics syllabuses, one for liberal arts students only and one for science students only. Mathematics subject examination is a compulsory examination for Gaokao. Different from Gaokao, the Senior Middle Level Unified Examination Certificate (SUEC) is a standardized examination designed by the United Chinese School Committees’ Association of Malaysia (Dong Zong) for the graduates from the Chinese Independ High Schools in Malaysia (CIHSM) only to take to “create favorable conditions for their further studies and employment” (Dong Zong, 2014). According to the SUEC Booklet Version 2014, the other three purposes of holding the SUEC are to use the exam results as: a unified measurement of the academic level of all participated CIHSM, a reliable academic basis for the admissions of local and foreign academic institutions and a trustworthy basis for companies to recruit talents (Dong Zong, 2014). Since 1975, SUEC has been successfully held for 48 seasons. It is currently the most influential subject-based unified examination with Putonghua (mandarin Chinese) as its examination language outside China. From 2018 to 2022, data released by Dong Zong (2022) showed that more than 11,000 CIHSM graduates participated in SUCE every year. Until 2014, although the Malaysian federal government had not yet allowed the use of SUEC grades to apply to public universities in Malaysia, SUEC grades were recognized by universities in 17 other countries (Dong Zong, 2014). For example, the United States, the United Kingdom, and France. Using the SUEC grades, students can apply to more than 820 universities in China (Dong Zong, 2014). In addition, mathematics subject examination is also a compulsory examination for the SUEC students. Dong Zong (2022) has 4 different syllabuses for its mathematics subject: ① mathematics, ② Further Mathematics, ③ Further Mathematics (Level I) and ④ Further Mathematics (Level II). Science major students must take ③ ④; liberal arts, business and economics major students must take ① ②; engineering students can choose to take either the combination of ① ② or ③ ④ ; art and the other major students must take ① ② (Dongzong, 2022).

Even though, outside of China, SUEC is the most influential nationwide college entrance examination with using mandarin Chinese as the examination language, only Hou Xianneng (2011) has ever conducted a comparative study on the Chinese subject examination of Gaokao and SUEC. There has been no previous research that has ever compared or discussed the comprehensive difficulty of SUEC or Gaokao’s Mathematics Examinations. As for both the Gaokao and SUEC, mathematics is a compulsory subject. There should have a comparative study on the comprehensive difficulty of the two examinations to make up for the insufficiency of the comparative study between Gaokao and the SUEC. On the other hand, since its establishment, CDMME model has been used for the comparative study of the comprehensive difficulty of the mathematics examination of Gaokao and other influential examinations, including the College Scholastic Ability Test (CSAT) mathematics examination in South Korea, the Cambridge Assessment International Education (CAIE) Advanced Level (A-Level) mathematics examination in the U.K., Taiwan Province’s Advanced Subjects Test (AST) mathematics examination, the Baccalaureat’s mathematics examination in France and the mathematics examination of the Hong Kong Diploma of Secondary Education Examination (HKDSE) (Wu & Zhang, 2018; Xie *et al.*, 2019; Li & Shi, 2020; Zhang & Zhou, 2020; Su, 2020). Therefore, to make up for the insufficiency of the comparative study between Gaokao and the SUEC, it is of certain significance to use the CDMME model to conduct a comparative study on the comprehensive difficulty of the two examinations’ mathematics paper. The results of this comparative study can also be used as a reference for further research on the similarities and differences between these two examinations and serve as a basis for universities to recruit students who have studied the curricula. Additionally, since there is yet no generally agreed quantitative assessment of the difficulty of mathematics examination, the research can also contribute to the quantitative description of the difficulty of mathematics examination. Furthermore, using a more quantitative approach to compare the difficulty of different mathematics examinations can better motivate students’ learning by exciting them to choose the mathematics examination that best suits their learning level.

2. LITERATURE REVIEW

2.1 The Concept of the CDMME Model

The Comprehensive Difficulty Model for Mathematics Examinations (CDMME) model was designed to quantitatively compare the comprehensive difficulty of mathematics exams. It is originated from the Comprehensive Difficulty Coefficient Model (CDCM) which was first developed by Bao Jiansheng using for comparing the comprehensive difficulty of China and the U.K.'s middle school math intended curriculum based on Nohara's overall difficulty analysis of mathematical problem (Bao, 2002; Nohara, 2001). In 2018, Wu Xiaopeng and Zhang Yi extended Bao's CDCM model to calculate the comprehensive difficulty of mathematics examination papers to compare the difficulty of the mathematics examination papers of Gaokao and CSAT. Wu Xiaopeng and Zhang Yi (2018) stated that Bao's CDCM model is mainly aimed at quantifying the difficulty of exercises in the textbook. However, textbook questions put more emphasis on the evaluation of students' literacy, but questions of the national college entrance exams put more emphasis on their measurability and differentiation. Therefore, to make Bao's CDCM model more appropriate, the factors of the model were revised into seven aspects: background, whether the problem contains parameter, mathematical operation level, level of reasoning, amount of knowledge, thinking direction, and level of computing. In addition, each of the factors was coded to different levels and scores. Detailed information of the factors and levels of Wu & Zhang's CDMME is listed in Table 1 below.

Table 1

Factors, Levels and Scores of Factors of Wu & Zhang's CDMME

Factors	Levels	Scores (d_{ij})
Background	No Background	1
	Life Background	2
	Scientific Background	3
Whether the Problem Contains Parameter	No Parameters	1
	With Parameters	3
Computing Level	Simple Numerical Operation	1
	Complex Numerical Operation	2
	Simple Symbolic Operation	2
	Complex symbolic operation	3

Resoning Level	Simple Reasoning	1
	Complex Reasoning	3
Amount of Knowledge	1 Knowledge Point	1
	2 Knowledge Points	2
	More Than 2 Knowledge Points	3
Thinking Direction	Forward Thinking	1
	Reverse Thinking	3
Cognitive Level	Understanding	1
	Using	2
	Analysis	3

Based on Table 1 and Bao's CDCM model, Wu & Zhang constructed the CDMME model to calculate the comprehensive difficulty in order to compare the difficulty of mathematics papers more appropriately.

$$D = \sum_{i=1}^{7} d_i k_i = \frac{\sum_{i=1}^{7}(\sum_j n_{ij} d_{ij}) k_i}{n} \ (\sum_j n_{ij} = n, i = 1,2,3,4,5,6,7),$$

where D is the comprehensive difficulty, d_i the comprehensive difficulty of the *i*-th factor, k_i is the weight coefficient of each factor in the examination, d_{ij} is the weight of the *i*-th factor at its *j*-th level of difficulty, n_{ij} is the number of questions of the *i*-th factor at its *j*-th level of difficulty, n is the total number of the math exam questions.

To calculate the comprehensive difficulty of a mathematics paper, according to Wu & Zhang's CDMME model, the weight coefficient of each factor k_i needs to be determined. In this research, Wu & Zhang selected 20 experts from different places in China to weight these 7 CDMME factors. The results are based on the "background factor" with a weight coefficient of 1.00 as the reference standard. The statistical results of k_i and d_i are shown in Table 2 below.

Table 2
The statistical results of k_i, d_i and D

	Background	Whether the Problem Contains Parameter	Computing Level	Resoning Level
k_i	1.00	1.21	1.19	1.53
Gaokao	1.20	2.35	2.24	2.06
CSAT	1.44	2.33	2.02	1.93
	Amount of Knowledge	Thinking Direction	Cognitive Level	D
k_i	0.91	1.17	1.35	/
Gaokao	1.68	2.50	1.99	17.00
CSAT	1.32	2.18	1.80	15.80

In 2020, Wu Xiaopeng and Kong Qiping applied the Analytic Hierarchy Process (AHP) to further improve the accuracy of the weight coefficient of each factor. The researchers invited 16 math teachers and experts to level and pairwise compare each of the difficulty factors to build the judgement matrix to calculate the weight coefficient more quantitatively than in the previous research, treating all the difficulty factors equally important or weighting the factors by human judgement.

In addition to Wu Xiaopeng's papers on the CDMME model, the model has also been used in comprehensive difficulty comparison studies of the mathematics examination of Gaokao and other influential examinations since the model was established, including the Cambridge Assessment International Education (CAIE) Advanced Level (A-Level) math exam in the U.K., Taiwan Province's Advanced Subjects Test (AST) math exam, the Baccalaureat's math exam in France and the math exam of the Hong Kong Diploma of Secondary Education Examination (HKDSE) (Xie *et al.*, 2019; Li & Shi, 2020; Zhang & Zhou, 2020; Su, 2020).

2.2 International Literature and Theories Related to the CDMME Model

The CDMME model was developed from the CDCM model, which was first created by Bao Jiansheng (2002) based on the overall difficulty theory of mathematics assessment proposed by David Nohara in his report (2001), *A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)*, prepared for the National Center for Education Statistics (NCES) of U.S. Department of Education. In David's report (2001), he identified four factors that could contribute to the overall difficulty of a mathematics assessment: (1) extended response, which represents the percentage of the scalability problems, (2) context, which represents the percentage of problems with real-life background, (3) Multi-step reasoning, which represents the percentage of problems that requires intermediate steps to solve, (4) computation, which represents the percentage of problems that require operations above an additional degree of difficulty. Bao (2002) developed the four factors into five factors with different levels.

Table 3
Factors and Levels of the CDCM Model for Mathematics Questions

FACTORS	LEVELS			
Cognition	Memory	Understanding	Investigation	
Background	Mathmetical Background	Individual-life Background	Common-life Background	Scientific Background
Computing	No Operation	Numerical Operation	Simple Symbolic Operation	Complex symbolic operation
Resoning	No Reasoning	Simple Reasoning	Complex Reasoning	
Amount of Knowledge	1 Knowledge Point	2 Knowledge Points	More Than 2 Knowledge Points	

Based on these new developments, Bao (2002) created the CDCM model to compare the comprehensive difficulty of the National Mathematics Curriculum of China (2001) and the Framework for Teaching Mathematics: Years 7, 8 and 9 of the United Kingdom (2001). In his model, the comprehensive difficulty of the *i*-th factor for a set of mathematics questions equals to

$$d_i = \frac{\sum_j n_{ij} d_{ij}}{n} \left(\sum_j n_{ij} = n, i = 1, 2, 3, 4, 5; j = 1, 2 \cdots\right),$$

where $d_{ij}$is the weight of the *i*-th factor at its *j*-th level of difficulty, $n_{ij}$is the number of questions of the *i*-th factor at its *j*-th level of difficulty, n is the total number of the math questions. However, Shi Ningzhong et al. (2005) structured a new model, the Course Difficulty Coefficient Model, to indicate the difficulty degree of a course. They argued that the difficulty degree of a course is influenced by three factors: the depth of course, the scopes of course and the course time. In the other words, the following functional relationship can be established: $N = f(S, G, T)$, where *N* is the difficulty degree of course, *S* is the depth of course, *G* is the scopes of course, *T* is the course time. They believe that, if there is enough time, nearly all students can understand the course contents. Therefore, the difficulty degree of course *N* is proportional to the depth of course *S* and the scopes of course *G*, and inversely proportional to the course time *T*: $N = \alpha \frac{S}{T} + (1 - \alpha) \frac{G}{T}$, where $0 < \alpha < 1$ is the weighting coefficient, $\frac{S}{T}$ is defined as the comparable depth and $\frac{G}{T}$ is defined as the comparable scopes. To calculate the depth of course, the authors used the abstraction method of analysis. Also, they take the weighting coefficient $\alpha = 0.5$, which means the comparable depth and comparable scopes are equally important in contributing to the difficulty degree of course.

In 2010, Li Gaofeng argued that there were deviations in comparing the difficulty of courses using the Course Difficulty Coefficient Model of Shi *et al.* (2005). The deviations come from (1) using the number of "knowledge points" to quantify the scopes of course; (2) using the maximum degree of abstraction to describe the depth of course; (3) using the average number of the sum of degrees of abstraction to calculate the depth of course. Correspondingly, he (2010) believed that the following modifications should be made to Shi's model: (1) use the number of all curriculum objectives to quantify the scopes of course; (2) use the sum of the degrees of abstraction to describe the depth of course; (3) use the sum of the curriculum objectives to represent the depth of course.

With reference to the models created by Bao (2002) and Shi *et al*. (2005), Kuang Kongxiu *et al*. (2013) constructed a new model, the three factors model, to calculate the degree of difficulty of primary school mathematics textbooks. After interviewing a group of experts and analyzing questionnaire responses from 1236 primary school mathematics teachers, Kuang *et al*. (2013) identified three main factors that can affect the difficulty of primary school mathematics textbooks: the width of content (the amount of knowledge, the depth of content and the degree of difficulty of the textbook exercises. In addition, Kuang et al. (2013) established a set of functions to quantitively relate the degree of difficulty to the three factors:

$$N = f(C_1, C_2, E);$$

$$f = \alpha_1 C_1 + \alpha_2 C_2 + \alpha_3 E, 0 < \alpha_1 = 0.2, \alpha_2 = 0.5, \alpha_3 = 0.3 < 1, \alpha_1 + \alpha_2 + \alpha_3 = 1;$$

$$C_2 = \alpha_{21} C_{21} + \alpha_{22} C_{22}, 0 < \alpha_{21} = 0.5, \alpha_{22} = 0.5 < 1, \alpha_{21} + \alpha_{22} = 1;$$

$$E = \alpha_{31} E_1 + \alpha_{32} E_2, 0 < \alpha_{31} = 0.6, \alpha_{32} = 0.4 < 1, \alpha_{31} + \alpha_{32} = 1.$$

In this model, N refers to the difficulty of primary school mathematics textbooks, C_1 is the width of content, C_2 is the depth of content, E is the degree of difficulty of the textbook exercises, C_{21} is the style of presenting knowledge, C_{22} is the cognitive demand of knowledge, E_1 is the cognitive level of exercise, E_2 represents the background of exercises. Also, from easy to hard, they divided C_{21}, C_{22}, E_1 and E_2 into three different levels:

Table 4

Stages and Levels of the three factors model for the differentiation of knowledge and exercise

The first stage	The second stage	Level and score		
		1	2	3
The depth of content	The presenting style of knowledge	Intuiting	Inducing	Abstracting
	The cognitive demand of knowledge	Knowing	Understanding	Applicating
The level of exercise	The cognitive level of exercises	Imitating	Transferring	Probing
	The background of exercises	No background	Life background	Scientific background

Moreover, the authors standardized C_1 , C_2 and E in order to more reliably compare the width of content and the depth of content:

$$C_1 = \frac{D}{B};$$

$$C_2 = \frac{n_1 \times 1 + n_2 \times 2 + n_3 \times 3}{n_1 + n_2 + n_3};$$

$$E = \frac{n_1 \times 1 + n_2 \times 2 + n_3 \times 3}{n_1 + n_2 + n_3},$$

where D is the amount of knowledge in a country's primary school mathematics textbook, B is the total amount of knowledge in the selected countries' primary school mathematics textbooks, n_1, n_2 and n_3 are the amount of knowledge about different levels. Using this model, Kuang *et al*. (2013) compared the difficulty of 12 sets of the fourth-grade textbooks from 10 countries in Asia, Europe, America, and Australia.

3. Research Design

3.1 Overall Design of the Study

This research intends to use a new improved CDMME model to compare the comprehensive difficulty of the mathematics examination of the SUEC and Gaokao. To assess the comprehensive difficulty of the mathematics examinations more comprehensively, the researchers categorized and quantified the difficulty factors into subjective comprehensive difficulty factors (SCDFs) and objective comprehensive difficulty factors (OCDFs). In this research, the SCDFs are defined as the factors that can affect the comprehensive difficulty of a mathematics examination paper and can be quantified by marks on the examination paper. And the OCDFs are defined as the factors that can affect the comprehensive difficulty of a mathematics examination paper but cannot be quantified by marks on the examination paper. In the study, the researchers applied the AHP method to develop the accuracy of the weight coefficient of each SCDFs and the weight coefficient of each SCDF levels. To calculate the comprehensive difficulty of OCDFs, the researchers refer to the method of using comparable difficulty factors developed by Shi Ningzhong et al. (2005). Furthermore, to get the judgement matrix, 6 mathematics researchers and 14 high school mathematics teachers are equally divided into 2 groups to compare the two kinds of weight coefficients respectively and discuss to obtain the final pairwise comparison results.

3.2 Data Collection & Analysis Method

Our study employs qualitative research approach to collect its data. In the research, 20 selected researchers and teachers will be equally divided into two groups. According to the 7 factors and 19 levels classification of the SCDFs in Table 5, the two groups of researchers and teachers will level and pairwise compare 9 sets of mathematics examinations for Gaokao and SUEC. The SUEC questions are selected from the 2020, 2021 and 2022 SUEC mathematics examinations for science major students, including ③ Further Mathematics (I) and ④ Further Mathematics (II), and the Gaokao questions are selected from the 2020, 2021, and 2022 Gaokao Mathematics National Volume I for Science Students. Considering the impact of different marks of questions, in the classification and pairwise comparison coding process, the structured questions are coded by parts to balance the marks of each question. Also, to ensure the consistency of the classification and pairwise comparison coding results, the two groups of researchers and teachers will level and pairwise compare the mathematics examinations respectively and discuss to obtain the final classification and pairwise comparison coding results. In addition, to assess the comprehensive difficulty of the mathematics examinations more comprehensively, the data of study duration, number of assessment objectives (AOs) of the syllabus and the studying language will be also quantified and considered as the OCDFs in this research. Detailed information of the SCDFs of CDMME used in the research is listed in Table 5 below.

Table 5

SCDFs and Levels of SCDFs of the CDMME Model

SCDFs	Levels
Background	No Background
	Life Background
	Scientific Background

Whether the Problem Contains Parameter	No Parameters
	With Parameters
Computing Level	Simple Numerical Operation
	Complex Numerical Operation
	Simple Symbolic Operation
	Complex symbolic operation
Reasoning Level	Simple Reasoning
	Complex Reasoning
Amount of Knowledge	1 Knowledge Point
	2 Knowledge Points
	More Than 2 Knowledge Points
Thinking Direction	Forward Thinking
	Reverse Thinking
Cognitive Level	Understanding
	Using
	Analysis

To improve the accuracy of the weight coefficient of each subjective comprehensive difficulty factors (SCDFs) and the weight coefficient of the SCDF levels, the researchers used the AHP method. Based on the pairwise comparison results, the researchers set up a judgment matrix A that consists of 7 by 7 entries since there are total 7 factors for SCDFs in the research. The numbers along the diagonal of the 7 by 7 matrix are all number 1 because same factors have an equal rating.

$$A = \begin{bmatrix} 1 & \cdots & a_{17} \\ \vdots & \ddots & \vdots \\ a_{71} & \cdots & 1 \end{bmatrix}$$

The other elements are from the pairwise comparison coding results of each factors using the following ranking:

- 1 means that the two factors are equally important.
- 3 means that one factor is moderately more important than the other factor.
- 5 means that one factor is strongly more important than the other factor.
- 7 means that one factor is very strongly more important than the other factor.
- 9 means that one factor is extremely more important than the other factor.

To calculate the weight coefficient of each subjective difficulty factors (SCDFs) and the weight coefficient of the SCDF levels:

1. Calculate the product of the elements in each row of the criteria comparison matrix, $A_i = \prod_{j=1}^{n} a_{ij}$ (i = 1, 2, ⋯ n), where n is the number of SCDFs or SCDF levels.
2. Calculate the nth root of A_i, $\overline{a_i} = \sqrt[n]{A_i}$
3. Normalize $\overline{a_i}$ can get $k_i = \frac{\overline{a_i}}{\sum_{i=1}^{n} \overline{a_j}}$

The last step of using the AHP method to calculate the weight coefficient of each SCDFs and the weight coefficient of the SCDF levels is to check to see if the rankings are consistent. In the research, researchers used the Consistency Ratio (C.R.) to check for consistency. It is defined as the Consistency Index (C.I.) divided by the Random Index (R.I.): $C.R. = \frac{C.I.}{R.I.}$. If $C.R. \leq 0.10$, the rankings are consistent. If $C.R. > 0.10$, the comparisons should be recalculated.

To calculate C.R., in the research

1. A_k is the product of matrices A and k: $A_k = A \cdot k_i$
2. λ_{max} is the largest principal eigenvalue of matrix A: $\lambda_{max} = \frac{1}{n} \sum_{i=1}^{n} \frac{A_k}{k_i}$
3. $C.I. = \frac{\lambda_{max} - n}{n-1}$

Table 6

Random Consistency Index (R.I.) (Saaty, 1981)

Matrix Size	1	2	3	4	5	6	7
R.I.	0	0	0.58	0.9	1.12	1.24	1.32

Using the data collected from the classification and calculation, the comprehensive difficulty of the SCDFs of a mathematics examination paper will be calculated by

$$D_{SCDFs} = \sum_{i=1}^{u} d_i k_i \ (i = 1, 2, 3 \cdots u),$$

where d_i is the comprehensive difficulty of the *i*-th SCDF, k_i is the weight coefficient of each SCDFs, and *u* is the total number of the SCDFs. The comprehensive difficulty of each SCDF factor will be calculated by

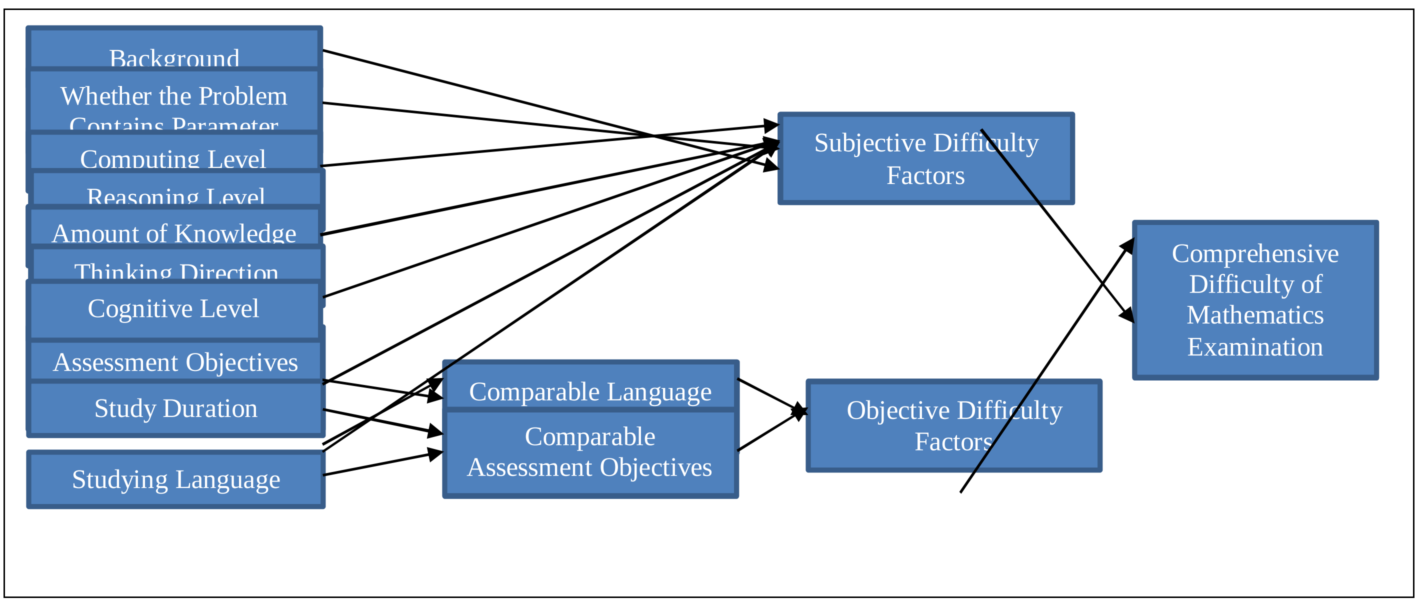
$$d_i = \frac{\sum_{j=1}^{v} m_{ij} d_{ij}}{s} \left(\sum_{j=1}^{v} m_{ij} = m, i = 1, 2, 3 \cdots u; j = 1, 2 \cdots v\right),$$

where d_{ij} is the weight coefficient of the *i*-th SCDF at its *j*-th level of difficulty; m_{ij} is the total marks of the *i*-th SCDF at its *j*-th level of difficulty in a math paper, *s* is the total sets of the selected papers for each mathmetics examination, and v is the total number of the levels of an SCDF. The d_{ij} can be calculated from the same method as the k_i.

On the other hand, in this research, the comprehensive difficulty of the OCDFs of a mathematics examination paper is affected by three factors: the studying language, the assessment objectives (AOs) of the syllabus and the study duration. In the other words, the following functional relationship can be established: $D_{ODFs} = f(L, A, T)$, where D_{ODFs} is the comprehensive difficulty of the OCDFs of a mathematics examination paper, L is the difficulty score of studying language, A is the number of assessment objectives (AOs) of the syllabus, T is the study duration in years. For students taking the math exam in their native language, L=0; for students taking the exam in a second language, L=2. The comprehensive difficulty of the OCDFs of a mathematics examination paper is directly proportional to the difficulty score of studying language and the number of assessment objectives (AOs) of the syllabus, but inversely proportional to the study duration. So, the comprehensive difficulty of the OCDFs of a mathematics examination paper D_{OCDFs} can be defined as $D_{OCDFs} = \alpha \frac{L}{T} + (1-\alpha) \frac{A}{T}$, where $0 < \alpha < 1$ is the weighting coefficient, $\frac{L}{T}$ *is* the comparable language capability and $\frac{A}{T}$ is the comparable assessment objectives. Therefore, the comprehensive difficulty of a mathematics examination paper D in the research will be defined as the sum of the weighted D_{SCDFs} and the weighted D_{OCDFs}, D=D_{SCDFs} + D_{OCDFs}. Figure 1 below illustrates the conceptual framework of this study.

Figure 1

Comprehensive Difficulty of Mathematics Examinations



3.3 The Methodological Limitations

To ensure the consistency of the classification and coding results, two groups of researchers and teachers will level and pairwise compare the mathematics examinations respectively and discuss to obtain the final classification and coding results. The AHP method is also used in this research to develop the accuracy of the weight coefficient of each subjective comprehensive difficulty factor and the weight coefficient of its levels. Even though, in the process of getting the final classification and pairwise comparison coding results, it is impossible to completely eliminate personal bias from teachers and researchers when pairwise comparing the difficulty factors of examinations to decide the judgement matrix. However, by using a standardized and objective method, such as the improved CDMME model, the impact of personal bias can be minimized, and the comparison process can be made more reliable and consistent. Starting from 2021, Gaokao changed from having 9 different sets of examination papers to 8 sets of different examination papers. Even though they are all based on the same Gaokao Curriculum issued by the Ministry of Education, the comprehensive difficulty of each mathematics paper of the Gaokao may vary with different examination sets. Another limitation of this research is that, although the effect of the studying language on the comprehensive difficulty of mathematics examinations was considered in this research, students took the SUEC and the Gaokao mathematics examinations all in their native language. Moreover, the generation of the weight coefficient of each difficulty factor and level in the CDMME model is only based on theoretical analysis but lacks previous exam data support. In the later stage, the difficulty factors can be corrected through real exam data.

4. Presentation, Analysis and Interpretation of Data

Table 7
The Judgment Matrix

SCDFs	Background	Whether the Problem Contains Parameter	Computing Level	Reasoning Level	Amount of Knowledge	Thinking Direction	Cognitive Level
Background	1	1	1	1/3	1	1	1/3
Whether the Problem Contains Parameter	1	1	3	1	1	1	1
Computing Level	1	1/3	1	1	3	1	1
Reasoning Level	3	1	1	1	3	1	3
Amount of Knowledge	1	1	1/3	1/3	1	1	1
Thinking Direction	1	1	1	1	1	1	1
Cognitive Level	3	1	1	1/3	1	1	1

$$A = \begin{bmatrix} 1 & 1 & 1 & 1/3 & 1 & 1 & 1/3 \\ 1 & 1 & 3 & 1 & 1 & 1 & 1 \\ 1 & 1/3 & 1 & 1 & 3 & 1 & 1 \\ 3 & 1 & 1 & 1 & 3 & 1 & 3 \\ 1 & 1 & 1/3 & 1/3 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 1 & 1 & 1/3 & 1 & 1 & 1 \end{bmatrix}$$

Table 8

CDC of Each Factor

SCDFs	Background	Whether the Problem Contains Parameter	Computing Level	Reasoning Level	Amount of Knowledge	Thinking Direction	Cognitive Level
A_i	1/9	3	1	27	1/9	1	1
$\overline{a_i}$	0.731	1.17	1	1.601	0.731	1	1
k_i	0.101	0.162	0.138	0.221	0.101	0.138	0.138

Table 9

Weight Coefficient of Each Factor Level

SCDFs	Levels	Marks in 100 Marks Scale for Each Paper			Percentage			Judgment Matrices of Each Factor	d_{ij}
		Gaokao	SUEC Further Mathematics (I)	SUEC Further Mathematics (II)	Gaokao	SUEC Further Mathematics (I)	SUEC Further Mathematics (II)		
Background	No Background	20	90	30	6.7%	30%	10%	$\begin{bmatrix} 1 & 1/5 & 1/3 \\ 5 & 1 & 3 \\ 3 & 1/3 & 1 \end{bmatrix}$	0.105
	Life Background	44	57	33	14.7%	19%	11%		0.637
	Scientific Background	236	153	237	78.7%	51%	79%		0.258
Whether the Problem Contains Parameter	No Parameters	40	126	30	13.3%	42%	10%	$\begin{bmatrix} 1 & 1/5 \\ 5 & 1 \end{bmatrix}$	0.167
	With Parameters	260	174	270	86.7%	58%	90%		0.833
Computing Level	Simple Numerical Operation	20	33	24	6.7%	11%	8%	$\begin{bmatrix} 1 & 1/3 & 1/3 & 1/5 \\ 3 & 1 & 1 & 1/3 \\ 3 & 1 & 1 & 1/3 \\ 5 & 3 & 3 & 1 \end{bmatrix}$	0.078
	Complex Numerical Operation	44	69	12	14.7%	23%	4%		0.201
	Simple Symbolic Operation	84	87	57	28.0%	29%	19%		0.201

	Complex Symbolic operation	152	111	207	50.7%	37%	69%		0.520
Reasoning Level	Simple Reasoning	148	168	120	49.3%	56%	40%	$\begin{bmatrix} 1 & 1/5 \\ 5 & 1 \end{bmatrix}$	0.167
	Complex Reasoning	152	132	180	50.7%	44%	60%		0.833
Amount of Knowledge	1 Knowledge Point	148	246	135	49.3%	82%	45%	$\begin{bmatrix} 1 & 1/3 & 1/5 \\ 3 & 1 & 1/3 \\ 5 & 3 & 1 \end{bmatrix}$	0.105
	2 Knowledge Points	74	42	102	24.7%	14%	34%		0.258
	More Than 2 Knowledge Points	78	12	63	26.0%	4%	21%		0.637
Thinking Direction	Forward Thinking	212	288	198	70.7%	96%	66%	$\begin{bmatrix} 1 & 1/5 \\ 5 & 1 \end{bmatrix}$	0.167
	Reverse Thinking	88	12	102	29.3%	4%	34%		0.833
Cognitive Level	Understanding	42	48	30	14.0%	16%	10%	$\begin{bmatrix} 1 & 1/3 & 1/5 \\ 3 & 1 & 1/3 \\ 5 & 3 & 1 \end{bmatrix}$	0.105
	Using	62	117	75	20.7%	39%	25%		0.258
	Analysis	196	135	195	65.3%	45%	65%		0.637

Table 10

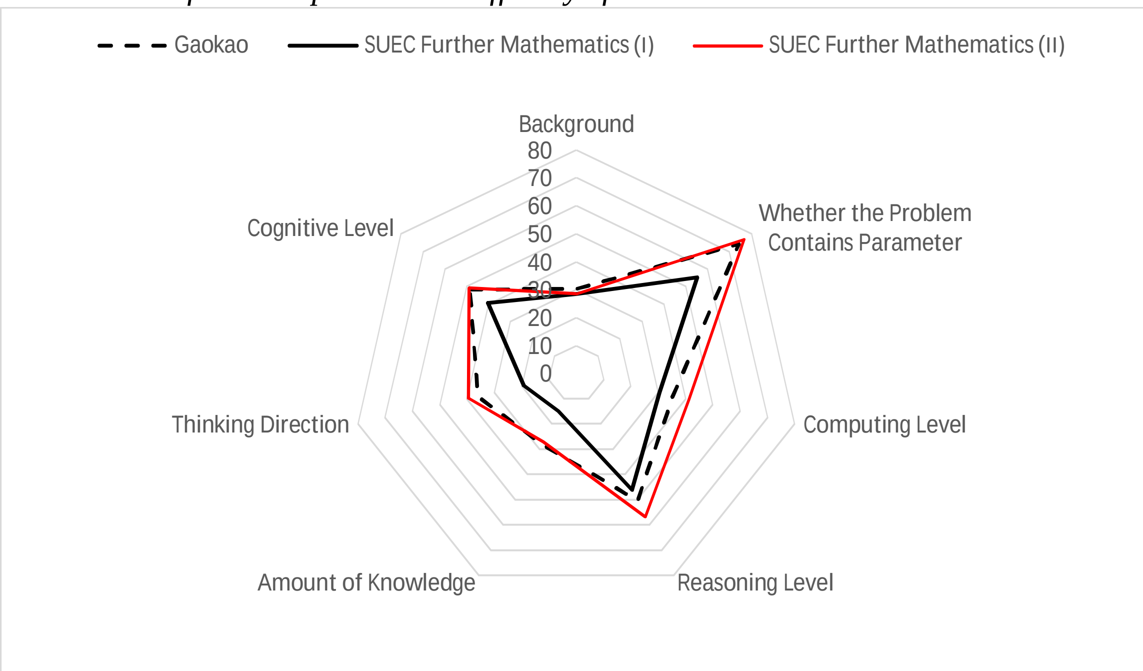
Consistency Check Statistics

SCDFs	Background	Whether the Problem Contains Parameter	Computing Level	Reasoning Level	Amount of Knowledge	Thinking Direction	Cognitive Level
C.I.	0.0193	0	0.0145	0	0.0193	0	0.0193
R.I.	0.58	0	0.9	0	0.58	0	0.58
C.R.	0.0332	0	0.0161	0	0.0332	0	0.0332

Table 11
The Comprehensive Difficulty of Each SCDF Factor

SCDFs	Background	Whether the Problem Contains Parameter	Computing Level	Reasoning Level	Amount of Knowledge	Thinking Direction	Cognitive Level
Gaokao	30.4	74.4	35.5	50.4	28.1	36.2	48.4
SUEC Further Mathematics (I)	28.4	55.3	30.6	46	14.8	19.3	40.4
SUEC Further Mathematics (II)	28.5	76.7	41.2	56.7	26.9	39.3	48.9

Figure 2
Radar Chart of the Comprehensive Difficulty of Each SCDF Factor



Because the CIHSM students' first language is Chinese, which is the same as their SUEC studying and exam language, L=0 for both the Gaokao and the SUEC mathematics exams. In this research, the comparable language capability and comparable assessment objectives are treated equally important in contributing to the D_{OCDFs}. Therefore, the weighting coefficient $\alpha = 0.5$ in this research.

Table 12
Comprehensive Difficulty of Each Mathematics Examination

	Number of AOs in Syllabus (A)	Study Duration in Years (T)	D_{OCDF}	D_{SCDF}	D
Gaokao	16	3	2.67	45.7	48.4
SUEC Further Mathematics (I)	10	3	1.67	36.0	37.6
SUEC Further Mathematics (II)	9	3	1.5	51.9	53.4

5. Conclusion, Interpretation and Discussion

5.1 Summary of Research

The goal of this research is to improve the CDMME model to conduct a comparative study on the comprehensive difficulty of the SUEC and Gaokao's Mathematics Examination. During the research process, the researchers selected three sets of examination papers from each of the Gaokao Mathematics Examination, the SUEC Further Mathematics Examination (I), and the SUEC Further Mathematics Examination (II) and used the new improved CDMME model to compare their comprehensive difficulty. The SUEC questions are selected from the 2020, 2021 and 2022 SUEC mathematics examinations for science major students, including ③ Further Mathematics (I) and ④ Further Mathematics (II), and the Gaokao questions are selected from the 2020, 2021, and 2022 Gaokao Mathematics National Volume I for Science Students. According to the results in Table 12, the calculation using the new improved CDMME model shows that the comprehensive difficulty of Gaokao Mathematics Examination is between that of the SUEC Further Mathematics Examination (I) and the SUEC Further Mathematics Examination (II)， $D_{SUEC\ Further\ Math\ (I)} < D_{Gaokao} < D_{SUEC\ Further\ Math\ (II)}$.

5.3 Discussion and Interpretation of Findings.

In this study, the researchers divided the factors that influence the comprehensive difficulty of mathematics examinations into two categories: the subjective comprehensive difficulty factors (SCDFs) and the objective comprehensive difficulty factors (OCDFs). The SCDFs consisted of seven different factors with a total of 19 different comprehensive difficulty levels. The OCDFs included language, study duration, and assessment objectives, which are factors can also affect the comprehensive difficulty of mathematics examinations. Moreover, the comparable language capability and the comparable assessment objectives are also defined and used in the research to calculate the results of the OCDFs more accurately to reflect the comprehensive difficulty of the mathematics examinations. From Table 8, it can be seen that using the AHP method for calculation, the reasoning level in the SCDFs has the greatest impact on the comprehensive difficulty of examinations, with k_i=0.221, while the background and amount of knowledge had the smallest impact on the comprehensive difficulty of the mathematics examination, with k_i=0.101. The radar chart can clearly demonstrate that the difficulty of each SCDFs for SUEC Further Mathematics

Examination (I) are lower than those of Gaokao and SUEC Further Mathematics Examination (II). The three examinations had the closest results in the background factor, with a maximum difficulty value difference of only 2.

The biggest difference is in whether parameters were included, with differences of 19.1 and 21.4 between the SUEC Further Mathematics Examination (I) and Gaokao and the SUEC Further Mathematics Examination (II), respectively. According to the final results in Table 12, the comprehensive difficulty of SUEC Further Mathematics Examination (II) is the highest, and the comprehensive difficulty of Gaokao is between that of the SUEC Further Mathematics Examination (I) and the SUEC Further Mathematics Examination (II).

According to Table 10, all the results of C.R. are less than 0.10, which indicate that the results of the study are consistent. The comprehensive difficulty result also corresponds to the general perception of the difficulty of these examinations, which demonstrates that the newly improved CDMME model can effectively quantify the difficulty of mathematics examinations.

5.4 Impact and Implications of the Research

Based on the research results, the research has the potential to make an important contribution to the field of mathematics education and assessment. By improving the understanding and evaluation of exam difficulty, this research can help to ensure that exams are fair, accurate, and effective in promoting student learning and success. The research contributes to a better understanding of the complex factors that influence the overall difficulty of mathematics exams. This can help educators and policymakers to design more effective and fair exams that better align with students' abilities and needs. By improving the CDMME model, the research provides a more accurate and reliable method for evaluating exam difficulty. This is particularly valuable for comparing the difficulty of different exams and identifying areas for improvement. Through this research, the researchers uncovered new insights into the subjective and objective factors that contribute to exam difficulty. This helps to inform the development of new mathematics curricula and assessments. The findings of this research have significant implications for the educational outcomes of students. By improving the accuracy and fairness of exams, the research can help to improve the quality of mathematics education and ultimately contribute to the success and well-being of students. Furthermore, the research can inspire future studies in the field of exam difficulty and evaluation, leading to further refinements and improvements in the CDMME model and other methods for evaluating exam difficulty.

5.5 Recommendations for Further Research

The newly improved CDMME model considers a more comprehensive range of factors that influence the difficulty of examinations, which enables a more accurate quantification of examination difficulty. The model provides a more scientific method for future question setting and comparing the difficulty of different examinations. However, although this study used the composite results of 20 mathematics experts and mathematics teachers to level the exam questions and form the judgement matrix for the study, it is not possible to completely avoid subjective factors from personal judgement. Moreover, the generation of the weight coefficient of each difficulty factor and level in the CDMME model is only based on theoretical analysis but still lacks previous exam data support. In the later stage, the difficulty factors can be corrected through real exam data. With the development of artificial intelligence technology, the accuracy of this new improved CDMME model can be further validated by using ChatGPT to simulate mathematics exam-taking and comparing its exam results ranking with the comprehensive difficulty ranking of the model. Another limitation and further research recommendation of this research is that, although the effect of the learning and testing language on the comprehensive difficulty of mathematics examinations is considered in the research, students take the SUEC and the Gaokao mathematics examinations all in their native language. In further research of the comprehensive difficulty of mathematics examinations, this model can be used for studying on comparing the comprehensive difficulty of mathematics examinations that require students to take the exam in their second or third language, such as more internationalized examinations like the AP, A-level, IB, etc. Furthermore, this model can also be applied to other subjects to verify its applicability.

In conclusion, this research has shed light on building a more accurate CDMME model to compare the comprehensive difficulty. By using the new improved CDMME model to compare the comprehensive difficulty of SUEC Mathematics Examination and Gaokao Mathematics Examination, the practicality and accuracy of the model were verified. By deepening the understanding of the CDMME model and the comprehensive difficulty, the researchers hope to pave the way for further research and ultimately contribute to more accurately compare the comprehensive difficulty of examinations in a more quantitative basis.

References

Bao Jiansheng (2002). A comparison on the comprehensive difficulty of Sino-British middle school mathematics curriculum expectations. *Global Education*, 31,48-52.

Dong Zong (United Chinese School Committees' Association of Malaysia). 2023. SUEC Further Mathematics (I) Syllabus. https://uec.dongzong.my/exam-info/examination-outline/senior/33-smathematics-1 (accessed March 28, 2023).

Dong Zong (United Chinese School Committees' Association of Malaysia). 2023. SUEC Further Mathematics (II) Syllabus. https://uec.dongzong.my/exam-info/examination-outline/senior/34-smathematics-2 (accessed March 28, 2023).

Dong Zong (United Chinese School Committees' Association of Malaysia). 2014. UEC Booklet. https://www.dongzong.my/uec/images/pdf/2014/UECbookletv2014.pdf (accessed April 2, 2022).

Dong Zong (United Chinese School Committees' Association of Malaysia). 2022. Number of CIHSM Students Taking the SUEC. https://uec.dongzong.my/ (accessed April 2, 2022).

Dong Zong (United Chinese School Committees' Association of Malaysia). 2022. Registration Guide for the 48th CIHSM SUEC in 2022. https://uec.dongzong.my/exam-info/guidelines/9-senior-application-notes (accessed April 2, 2022).

Fan Weichen. (2021, June 3). 10.78 Million Candidates for Gaokao in 2021. *China Youth Daily*. http://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/2021/2021_zt12/meiti/2021 06/t20 210603_535277.html.

Hou Xianneng. (2011). *Analysis and comparison between the Mandarin Papers of China's NCEE and Malaysia's UEC – Basing on the 2009 papers.* (Master's thesis, Northeast Normal University).

Kuang Kongxiu, Xinrong, Y., Qingyou, C., & Naiqing, S. An International Comparative Study on the Degree of Difficulty of Primary School Mathematics Textbooks. International Statistical Institute (ISI). *Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session CPS111)*, pp. 4997-5002.

Li Baozhen, & Shi Ye. (2020). Comparative Study on the Difficulty of Mathematics Test in the College Entrance Examination between Chinese Mainland and Taiwan——Take the 2016 to 2018 Mainland National Examination Paper I and Taiwan Department Required Test Paper as a Case. *Journal of Mathematics Education*, *29*(1), 58-64.

Li Gaofeng. (2010). Deviations and Corrections of Using the Course Difficulty Coefficient Model. *Journal of Shanghai Educational Research, 2010* (03),46-49.

Li Xiaoya. (2021).”Double Reduction”: A New Opportunity for the Cultivation of Innovative Talents in the Stage of Basic Education. *Journal of Guangxi Normal University (Philosophy and Social Sciences Edition)*, 57(6), 107-115.

Nohara, D., & Goldstein, A. A. (2001). *A comparison of the national assessment of educational progress (NAEP), the third international mathematics and science study repeat (TIMSS-R), and the programme for international student assessment (PISA)*. US Department of education, National center for education statistics (NCES).

National Education Examination Authority. (2019, January 31). 2019 National Unified Exam Syllabus for Enrollment in Ordinary Colleges and Universities (General Syllabus). http://gaokao.neea.edu.cn/html1/report/19012/5989-1.htm.

Saaty, T. L. (1990). *How to make a decision: The analytic hierarchy process. European Journal of Operational Research, 48*(1), 9–26.

Saaty, T. L. (1981). *The Analytic Hierarchy Process*. McGraw- Hill Inc..

Shi Ningzhong, Kong Fanzhe & Li Shuwen (2005). Course Difficulty model: China Compulsory Geometry Course difficulty comparison. *Northeast Normal University (Philosophy and Social Sciences)*, 218,51-55.

Su Juan. (2015). *Comparative study of the mathematics test questions in the University Entrance Exam between the mainland and Hong Kong* (Master's thesis, Gannan Normal University).

Wu Xiaopeng, & Kong Qiping. (2020). The Construction and Application of the Comprehensive Difficulty Model of Mathematical Advanced Examination Questions Based on AHP Theory. *Journal of Mathematics Education*, *29*(2), 29-34.

Wu Xiaopeng, & Zhang, Yi. (2018). Comparative Study on the Comprehensive Difficulty of Mathematics Questions of College Entrance Examination in Chinese and Korean. *Journal of Mathematics Education*, 3.

Xie Xiaochuan, Fang Haoying, Chen Nan, Luo Chaorong, & Ma Zhendi. (2019). *A Comparative Study on the Mathematics Examination Questions of Gaokao and the CAIE A-Level*. *Research on Middle School Mathematics (South China Normal University): First half of the month*, (10), 12-14.

Zhang Yuhuan, & Zhou Xia. (2020). A Comparative Study of Chinese and French College Entrance Examination Mathematics Items (2015—2019): A Focus on Overall Difficulty——Based on the Chinese and French College Entrance Examination Mathematics Papers for 2015—2019. *Journal of Mathematics Education*, *29*(3), 43- 50.