

Research Article

Mathematical Model Construction of Network Computer Test Automation System for College English Teaching

Wenjuan Sun ¹, Ping Han ², and Shaohua Pei ³

¹Department of Flight Training and Management, Luoyang Flight College, Civil Aviation Flight University of China, Luoyang, Henan 471001, China

²College of Mathematics and Computer Science, Hanjiang Normal University, Shiyan, Hubei 442000, China

³Department of Digital Construction, Shanghai Urban Construction Vocational College, Shanghai 200433, China

Correspondence should be addressed to Wenjuan Sun; wjsun@cafuc.edu.cn

Received 10 March 2022; Revised 16 April 2022; Accepted 20 April 2022; Published 9 May 2022

Academic Editor: Sheng Bin

Copyright © 2022 Wenjuan Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Networked education represents a development direction of educational reform, has become a feature of modern education, and has formed a new impetus to the development of education. The change from paper-and-pencil examinations to computer network-based machine examinations is not only a reform of the content and form of the University English Level 4 and 6 examinations, but also a reflection of the reform of the comprehensive assessment system and teaching philosophy of University English teaching. Promoting the development and construction of online college English examination systems in universities, applying machine examinations to various types of tests and formal examinations for college English teaching, and updating the college English question bank through continuous accumulation of professional teachers will have very high research value and practical value in improving the quality of examinations and promoting the standardization of college English examinations. In order to show the individualized chemistry report simultaneously, it is necessary to further ask relevant subject experts, educational measurement experts, and front-line teachers to expand the question attributes; that is, the question attributes extend to module specific knowledge points, cognitive levels, and information literacy ability levels, while most of the pages are combined with Ajax technology to reduce hardware resource dependency and optimize software performance. The paper analyzes the requirements analysis and overall design of the online college English examination system, including the design objectives, design principles, functional design, core algorithm design, architecture design, database design, and security design. The main functions of the online college English examination system, such as personal information management, question bank information management, test paper grouping management, online examination management, paper marking management, and system setting management, are discussed in detail. The main functions of the system, such as personal information management, question bank information management, test paper grouping management, online test management, mark and approve management, and system setting management, are discussed in detail, and the performance test and some function test of the system are also briefly discussed. The whole research process involves the identification of test questions and their attributes in the construction of the question bank, the construction of the adaptive test system, the comprehensive analysis of the mock test results, and the thinking about the application of the high school adaptive academic level test.

1. Introduction

Modern information technology, with computer networks at its core, is beginning to be integrated into college English classroom teaching, bringing unprecedented and profound changes to English teaching [1]. More and more students have access to a large number of linguistic learning resources

with rich graphics, sound and visuals, and vivid reality through networked computers. Teachers' teaching concepts, teaching processes, teaching activities, teaching methods and students' learning strategies, and even teaching evaluation criteria have all undergone great changes [2]. College English teaching is gradually showing the characteristics of information education, such as three-dimensional teaching

materials, networked teaching resources, virtualized teaching environment, and personalized teaching and learning assessment process. Computers are gradually becoming a part of college English classroom teaching [3]. Only when teachers and students have truly adapted to modern emerging technologies can technology be most effective in education. The key lies in how to integrate it into daily English classroom teaching activities, making it a tool and a purpose for active thinking and action together with language, so as to develop the ability to use English and the Internet for various kinds of communication and cooperation, and to use information to build new knowledge, which, in my opinion, is the basic content and goal of the training of contemporary innovative foreign language talents [4].

In fact, machine tests are better than traditional paper-and-pencil tests and computer-based tests in many aspects, which are summarized as follows: (1) regarding test duration, on the one hand, machine tests can provide convenience to test takers by booking test slots online, which can truly realize “test-on-demand” and effectively reduce the human and material resources loss caused by large-scale administration [5]. On the other hand, the machine test can adopt the fixed-length method in the design of the termination strategy, which can effectively measure students’ real ability in a shorter period of time and can not only accurately assess students’ ability level but also reduce the examination burden for students [6]; (2) regarding the test content, the machine test can implement the concept of “individualized testing” and calculate the candidates’ ability level in real time according to each candidate’s answers [7]. The machine test calculates the test taker’s ability level in real time based on each test taker’s answers, so that each test taker answers questions near his or her own ability level, so that high-ability test takers can avoid answering easy questions, and low-ability test takers can avoid answering difficult questions, so that every question the test taker takes serves to accurately assess his or her ability level [8]; (3) regarding the organization and management of the test, the “personalized” presentation of the test questions can avoid fraud during the test and reduce the burden of the invigilators; (4) regarding the test results, the machine test can not only provide timely feedback on the relative performance of the test takers at the end of the test, but also provide the test takers with the ability level values of the test, and the computerized adaptive test with cognitive diagnostic function (referred to as CD-computerized test), on the one hand, allows the test takers to understand their own knowledge in a timely manner after the test [9]. On the other hand, it allows teachers to pinpoint students’ knowledge misconceptions and to teach with precision, so that they can better implement remedial measures. For the traditional Web application model, the user triggers an action on a Web page and sends an Http request to connect to a Web server [10]. The server processes it (such as receiving data, processing calculations, and accessing databases) and finally returns a new HTML page to the client.

Object-oriented methods have been widely used in programming languages, formal definitions, design

methodologies, operating systems, distributed systems, artificial intelligence, real-time systems, databases, human-machine interfaces, computer architectures, and concurrent engineering, comprehensive integration engineering, and so on, and their applications have been greatly developed in many fields. The main content of this project is to design and implement an online university English examination system based on B/S structure, which can be applied to various types of tests and formal examinations of university English teaching, and can effectively improve the quality of examinations, greatly simplify the examination process, significantly reduce the examination operation and management costs, and make the examinations more notarized, objective, scientific, formal, and standardized through the continuous accumulation of updated question banks by English teachers every year. At the same time, the application of the system also allows students to adapt to the pace of reform of the university English IV and VI machine examinations earlier. The paper first analyzes the background of the topic and the current research situation at home and abroad and summarizes the previous development experience, studies the theoretical basis of software development, comes up with the requirement analysis and overall design of the online college English examination system, and carries out the implementation and case test analysis.

2. Related Work

Online examination is the extension and expansion of the traditional paper-based examination and is the product of the combination of information technology and examination method theory [11]. The online examination system has a more profound impact on teaching management, teachers, and students. Teaching management departments can use the computer network online examination system and will constantly reuse relevant teaching materials and information; this system will make the school teach the standardization of learning examinations, thus greatly improving the quality of teaching and teaching effectiveness [12]. The computer network online examination system greatly improves the flexibility of education students teaching and learning, can make students examinations are not restricted by geographical time, and improve the flexibility of various examinations in schools and mobility, the maximum realization of the process of paperless examination, with economic and efficient characteristics. The online examination system facilitates teachers to quickly prepare, distribute, and modify examination papers through information technology, reducing teachers’ workload, improving teaching effectiveness and efficiency, and providing effective assurance of fairness and seriousness of examinations [13]. The traditional method of examinations is that the examinations start from the teachers’ artificial questioning, to the arrangement of the examination room and the examination time, the arrangement of the marking of the papers, etc., until the test scores and the analysis of the students’ reactions to the test papers. Each step of the gathering consumes a considerable amount of time [14]. Modern Internet computer online examination system can use the relevant

conditions set by the system to automatically assemble papers from the question bank and can automatically review the test papers, reducing the workload of teachers.

In the traditional C/S mode, the examination paper information is stored on the remote server, and the examination program is installed in the examination machine client, which will cause more installation and maintenance configuration work on the examination side [15]; if we put the examination installation that is stored on the client computer, which will reduce the inconsistent diversity of the volume test questions of the machine examination system, the repetition rate of the examination paper will be higher, and the ability to group the paper will be limited. If implemented by mode B/S, that is, using Web technology to achieve a three-tier architecture, user interface layer, business logic layer, and database layer [16], there is no need to install any exam program on the examiner client, and the exam questions are stored on the database server.

Therefore, the B/S model has obvious advantages in terms of security and ease of maintenance [17]. In the country's institutions of higher learning, primary and secondary schools in recent years have been used to improve education information technology, modern education examination, and other reform of the mass education behavior and are given policy support. Our higher education and primary and secondary school education are compared with the education in developed countries, and after that, various types of computer network examinations on various types of levels have been set up according to the actual situation of education and teaching in China. Since the 1990s, we have been researching and developing ways and means of responding to theoretical examinations and computer-aided adaptive tests in university English examinations [18]. The CEF200 computer network examination system developed by Peking University is a multifunctional computer examination system developed according to the above theory, which is accurate, high-speed, and informative, and the computer network of examination and study is based on the computer Internet technology, thus realizing the automation, programmed, highly scientific, and efficient process of students' study [19], regular practice, and mid-term and final tests. In the existing computer network, computer education examinations are based on standardized test questions; in the examination process, the computer provides subjective test questions, and its approval and scoring are also done manually by the teacher in charge of the class on the computer for various tasks [20].

3. Construction of a Web-Based Machine Examination System for University English

3.1. Question Bank Construction. The question bank is a collection of questions of a certain subject implemented in a computer system according to a certain educational measurement theory, which is an educational measurement tool based on a precise mathematical model and plays a crucial

role in the whole test. Therefore, the overall planning of the question bank is an essential process before the preparation of the test questions.

First, analyze the high school IT curriculum standards, combine the specific course contents, determine the number of test questions attributed to each knowledge point according to a certain ratio, form the content structure table of the question bank system, and finally develop a detailed two-way breakdown table according to Bloom's education goal classification theory and content structure table. In order to be able to initially realize the analysis report of personalized chemistry, enrich the information of the subjects obtained from the test, construct the corresponding information literacy ability map based on the information literacy ability framework, and finally realize the generation of the report sheet of personalized chemistry based on the analysis of different students' response results, the second stage is the determination of parameters in the middle of construction. After the content of the test was determined in the first stage, four sets of parallel test questions were generated, and the traditional paper-and-pencil test was used to conduct a large-scale test to collect the response data of the subjects, and then the response data were analyzed with the statistical analysis software BILOG-MG3.0 based on item response theory. The logistic model is the item response model, and then calculate the difficulty, differentiation, and guessing coefficient of each item (test question). The third stage is the quality check at the later stage of construction.

Although the parameters of each test set have been obtained in the second stage, the test reliability and validity need to be improved by verifying the important hypotheses of item response theory to achieve both theoretical and practical levels to ensure the scientific validity of the question bank test questions, so the following four aspects are tested, respectively.

- (11) First, in the unidimensionality test, in which a certain number of subjects' responses are randomly selected with the help of the statistical analysis software SPSS, the factor analysis method is used to verify whether the test measures are only one potential trait of the subjects.
- (22) Secondly, partial independence tests are conducted, which mainly involve both subjects and items. From the perspective of the subjects, it is essentially a test of whether the subjects' response data can truly reflect their level of psychological traits; from the perspective of the items, it is mainly to verify whether there is a certain correlation between the items, reflecting whether the scores are fair between different test questions.
- (3b) Then, with the help of BILOG-MG3.0, we analyze the amount of information provided in each test question, and the questions that provide less

information are less differentiated, so we can eliminate some of the questions based on the amount of information.

- (42) Finally, the item characteristic curve test is conducted to check whether the item characteristic curve form can fit the actual response behavior and data of the subjects, and the curve form can be used to determine and screen the items again. Finally, according to the quality test results and test requirements, targeted corrections and deletions are made to make the question bank system more measurable, so that the screened question bank system can be used directly for adaptive testing. The questions are added to the database. Therefore, updating the question bank is a tedious process, but the construction of the questions based on item response theory ensures a certain degree of scientific validity and provides the maximum contribution to the test results. This is illustrated in Figure 1.

Each question bank system is a unique container of questions developed for a specific domain to ensure the scientific and effective implementation of the test. Since the new standards for high school IT curriculum have just been introduced, and schools around the world do not widely use the new standards version of high school IT curriculum, this study follows the high school IT textbook as the basis for forming the question bank, which is more representative and reasonable, as shown in Figure 2. At the same time, the question bank system is based on the computer adaptive test as the core and based on the item response theory. The uniqueness of its management functions is mainly reflected in the following three aspects, as shown in Table 1. The system provides the teacher with five parameters for grouping papers, which are the total number of exam questions, the number of exam questions contained in each type of question, the score corresponding to each question type, the difficulty factor of each question type, and the score occupied by each chapter.

The first aspect is the management of test parameters. In essence, the management of system parameters is mainly a statistical analysis process of data with static storage and dynamic calculation as the core. The test parameters involved in this study include difficulty, discrimination, and guessing coefficient, and each test question in the system must be strictly calibrated with parameters, and then the system obtains the ability value of the subject according to the great likelihood estimation method in item response theory and then calculates the probability of correct response reflection of items according to the established three-parameter Logistic function model, and then the item characteristic curve of each item can be plotted, in which the model fitting effect can be verified visually.

The second aspect is the management of test question allocation. In this study, instead of presenting all of the test questions in groups like traditional machine examinations, the test questions are dynamically presented; that is, the

subjects are provided with questions that match their ability level in real time according to their responses.

The third aspect is the management of students' knowledge level and ability level results. Since all the prepared test questions correspond to Bloom's educational goal levels according to the high school IT standards and emphasize the cultivation of information literacy in the standards, each question in the question bank system corresponds to its reflected cognitive level and information literacy ability level, so that it can obtain students' ability and also diagnose the students' mastered knowledge and reflect different students' response results. This way, it is possible to capture students' abilities, diagnose their knowledge, and reflect the different levels of information literacy of different students.

3.2. Mathematical Model for Quality Check of Question Bank Test Questions. Usually, the automatic mode of paper formation can greatly meet the conditions set by the user. The user examines the quality of the examination paper through several parameters while forming the paper automatically, such as the total number of questions and difficulty ratio. After logging into the system, the teacher first selects the type of questions to be assessed. When assembling the paper, one question is taken from the question bank with reference to the above five constraints. The selected question should first be identified with five indicators and represented by a five-dimensional vector (question number, difficulty, question type, section, and test score), denoted as $(a_1, a_2, a_3, \dots, a_n)$. If n is used to represent the total number of questions in the paper, then a paper can be represented by a matrix of $n * 5$:

$$n * 3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \quad (1)$$

This is an objective morphological matrix for problem solving, which meets the following constraints.

The total marks of the examination papers are as follows:

$$A_{ij} = \sum_{i=1, j=1}^{n,n} a_{ij}. \quad (2)$$

Marks occupied by each question type:

$$S_j = \sum_{j=1} c_{ij} * a_{ij}. \quad (3)$$

If $a_{13} = j$, then $c_{ij} = 1$; if $a_{13} > 1$, then $c_{ij} = 0$. j represents the question number:

Fractions occupied by each chapter:

$$S_j = \sum_{j=1} c_{ij}. \quad (4)$$

Regarding the difficulty of type i questions,

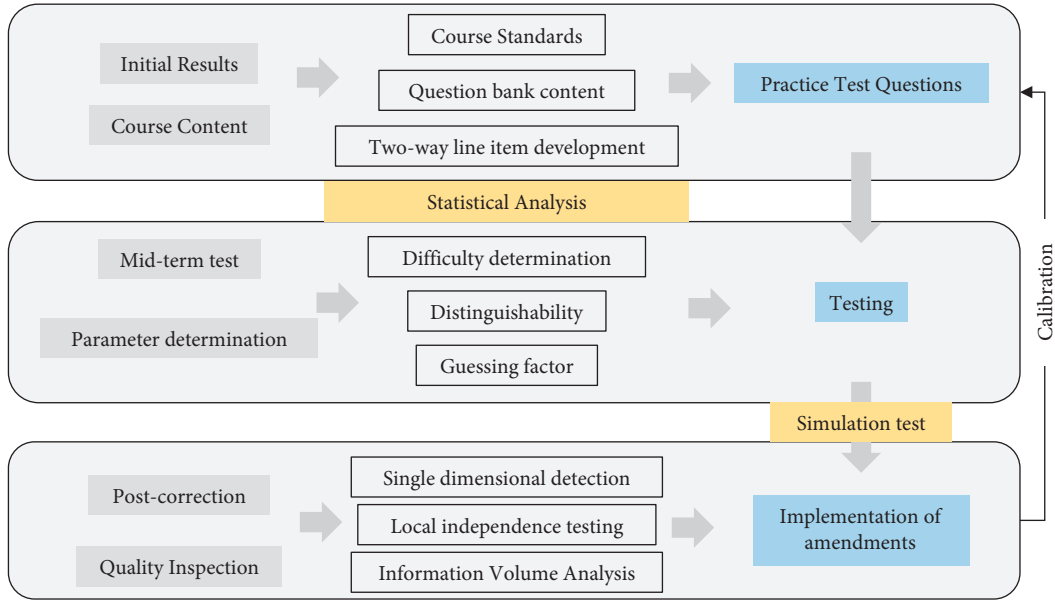


FIGURE 1: The construction of the question bank of the machine examination system.

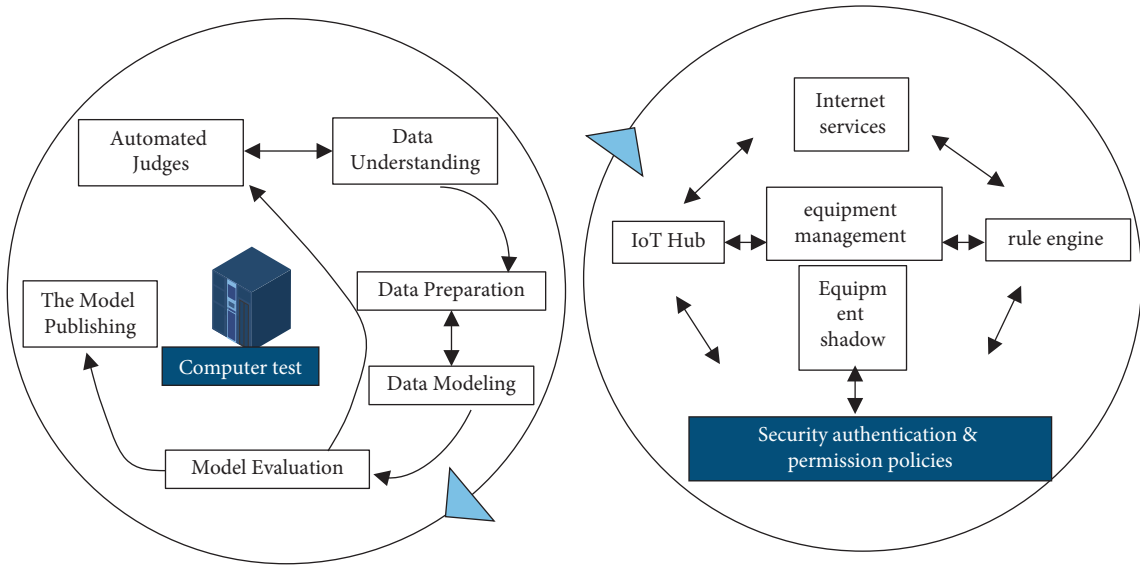


FIGURE 2: Automated judges for setting knowledge competency levels.

TABLE 1: Management functions.

	Grade	Number of individuals	Average value	Standard deviation (SD)	T-value	P-value
Traditional exam	Speaking	37.22	37.71	38.23	2.69**	0.006
	Reading	13.70	13.58	12.61		
Computer exam	Speaking	42.67	42.82	38.14	2.51**	0.012
	Reading	19.73	20.89	15.61		

$$\frac{s_j}{S_j} = \frac{\sum_{j=1} c_{ij}}{\sum_{j=1} c_{ij} * a_{ij}} \quad (5)$$

First of all, from the subject's point of view, the partial independence test is to determine that the individual subject did not receive any assistance from external factors in the process of answering the test and answered the test based on

his or her own true level, including copying others' answers, finding relevant information, and using external communication devices. In fact, the most objective and reliable means of testing the partial independence of the test taker's perspective is to determine whether or not the test taker used external factors to obtain answers that were not at his or her own level during the test, based on the monitoring of the test process by the proctor. The hypothesis of partial independence of the subjects is valid because the supervisors of the test schools were informed of the test before the test, and the supervisors were interviewed after the test to find out that the entire test process was strictly organized and completely free from interference from external factors.

According to the characteristics of the high school academic level examination and the advantages of the adaptive test, the question bank system mainly has the following two functions: the adaptive academic level examination question bank system has the basic management functions of the traditional question bank system, including the functions of adding, deleting, modifying, automatically assembling papers, and the result statistics.

Therefore, the questions in this study were carefully prepared by experienced IT teachers and regional teachers and researchers, so that the knowledge, skills, and relationships between chapters tested in the four sets of parallel papers woven together were more accurately grasped and finally reviewed by IT teachers and teachers and researchers to exclude certain questions as potential prerequisites and foundations for the knowledge tested in other questions. The questions were not considered to be a prerequisite or basis for another test, as shown in Table 2.

The item characteristic curve test is intended to check whether the form of the item characteristic curve fits the actual response and data of the subjects. The relationship between the ability level and the probability of correct responses can be observed visually through the curve, and if it can be approximated as a monotonically increasing curve, it means that the items satisfy the hypothesized monotonicity principle and can use the relevant parameters obtained through the item response theory model, as shown in Figure 3. The parameters ensure the scientific validity of the postadaptive test.

The above matrix shows that most of the test questions developed in this study can satisfy the monotonicity principle of the item characteristic curve, and the quality of the test questions is high enough to meet the test requirements of the later high school IT adaptive academic level test. Since updating the test questions requires a lot of labor and material resources and will cause some disturbance to the normal teaching arrangement of the test schools, after meeting the basic needs of the experiment, some test questions with unscientific item characteristic curves will not be updated, including deletion and addition.

3.3. Feasibility Analysis of Automation System. In the framework of item response theory, $p_i(\theta)$ is used as the subject response function for item i (θ denotes the subject's ability), and the reliability of a single item at a single subject level is

obtained with the help of the item information function $I_i(\theta)$, in which it can be seen from the equation that the amount of item information derived from different ability values is different. The total information of the test is obtained through the summation of the item information function, which is the concept of reliability of the evaluation of individual subjects at the whole test level, which is the test information function $I(\theta)$, as in Equation, and the standard error SE of the test is also recorded for each subject's ability level by the machine test, as in Equation. In order to be able to initially realize the analysis report of personalized chemistry, enrich the information of the subjects obtained from the test, construct the corresponding information literacy ability map based on the information literacy ability framework, and finally realize the generation of the report sheet of personalized chemistry based on the analysis of different students' response results, where S_x is the standard deviation of test scores, and r_{xx} is the reliability coefficient. When the test scores are standardized, the mean is 0, and the standard deviation is 1, as shown in the following equation:

$$\begin{aligned} I_i(\alpha) &= \frac{1}{\text{VAR}(\alpha_i/\alpha)}, \\ I(\alpha) &= \sum_{i=1}^n \frac{1}{\text{VAR}(\alpha_i/\alpha)}, \\ SE(\alpha) &= \frac{1}{\sqrt{\sum_{i=1}^n I_i(\alpha)}}, \\ SE &= \frac{1}{\sum_{i=1}^n I(\alpha)}. \end{aligned} \quad (6)$$

By analyzing the standard errors of the two tests, the overall reliability coefficients of the two tests were obtained separately, and the reliability coefficients of such a large-scale and risky test were above 0.9, which had high reliability and stable and consistent measurement results. The reliability coefficients of the tests are above 0.9, and the results are stable and consistent.

Considering the feasibility of a system from many different perspectives before developing and designing it brings great convenience for later system improvement, optimization, and maintenance. Feasibility analysis is an essential step in the design and development of any system, mainly to analyze whether the set goals can be achieved under the existing conditions and whether the whole system can be completed successfully.

4. System Optimization Test

First of all, the traditional paper-and-pencil test is a 75-question test with a fixed time of 75 minutes. On the one hand, in order to ensure that the average response time for each question is 1 minute, so that everyone can complete the test, everyone faces the same test paper, but not all the 75 questions are meaningful to the test taker, for example, questions that are particularly easy for the more able students provide less information, and questions that are

TABLE 2: Potential prerequisites and foundations for the knowledge test.

	Grade	Number of individuals	Average value	Standard deviation (SD)	T-value
Traditional exam	Speaking	47.62	48.13	50.11	2.12**
	Reading	39.74	40.49	40.97	
Computer exam	Speaking	23.78	24.83	26.11	2.69**
	Reading	75.74	60.03	62.92	

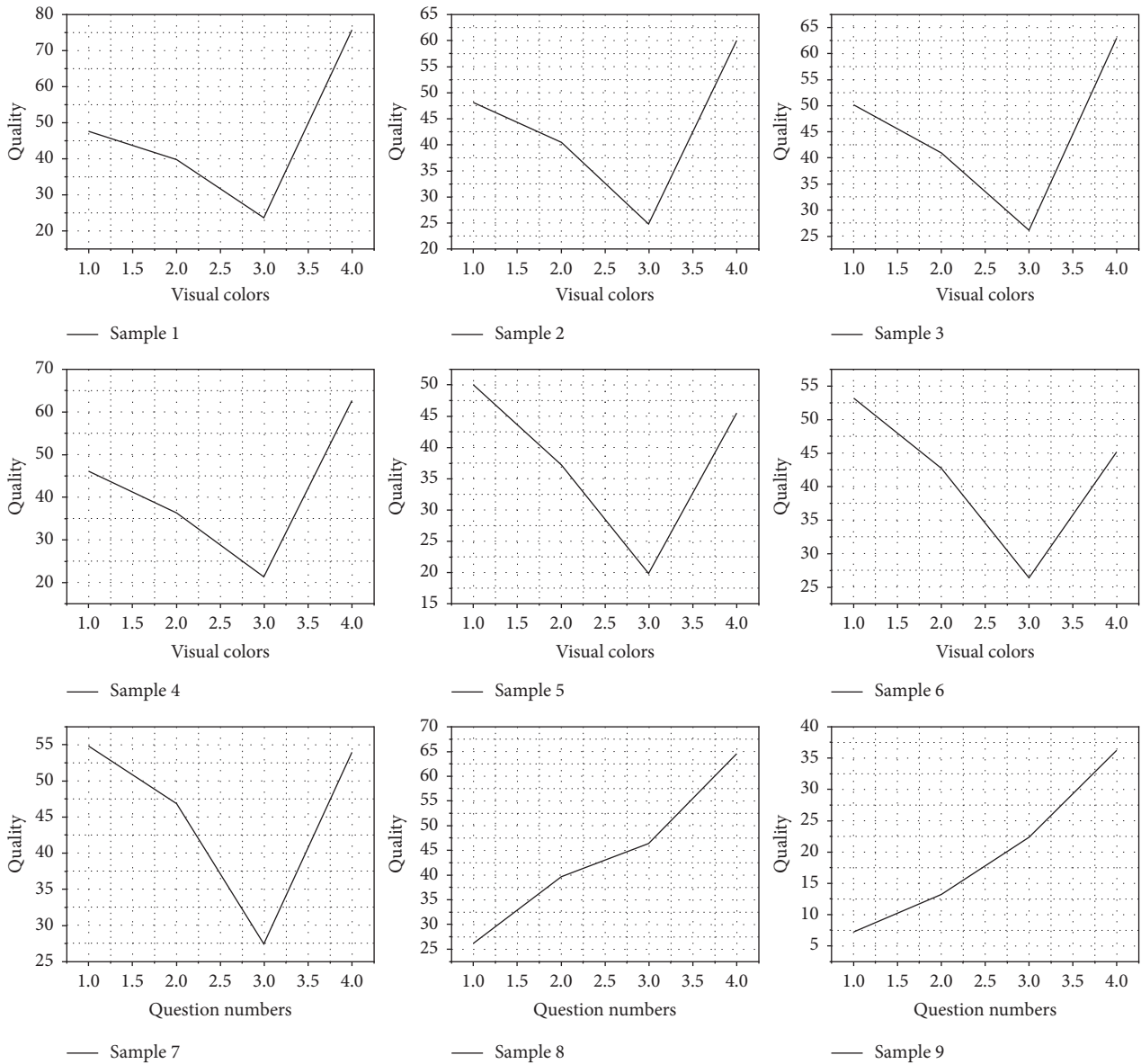


FIGURE 3: Matrix of project characteristic curves.

particularly difficult for the less able students provide less information and are not very meaningful. On the other hand, traditional paper-and-pencil tests require all students to take the test at a certain time, and even if the test takers are required, studies have shown that the longer the test, the greater the impact on test-taker accuracy, and the greater the potential for test-taker anxiety. On the one hand, the system provides students with questions that match their ability level based on their responses, and these 40 questions are all

around the students' ability level, which is more scientific for estimating their ability and obtaining their test scores. After statistical analysis, the submission time of the quiz students gathered around 13 minutes, with an average time of 835.7 seconds, or about 14 minutes, as shown in Figure 4, where the vertical coordinate is the number of students, and the horizontal coordinate is the response time (in seconds), and the test takers were able to answer in a short time and were able to answer the more informative questions in a highly

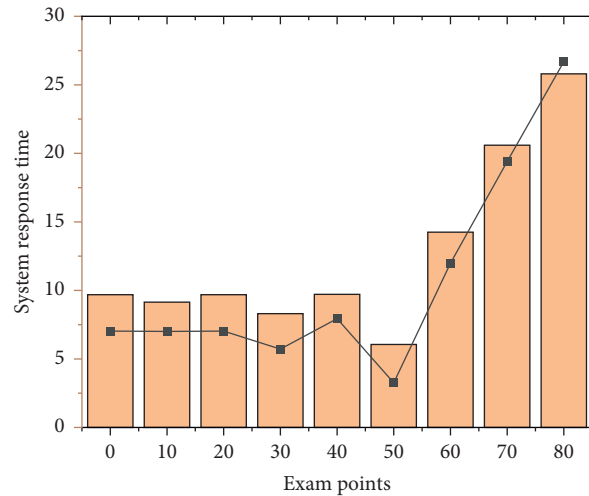


FIGURE 4: System response time distribution.

focused manner. Thus, preliminary indications are that they are significantly better than the traditional paper-and-pencil test, both in terms of test length and response time. Then, from the perspective of items, partial independence is mainly to determine whether there is interdependence or association between items. In order to avoid using the expert analysis method later to ask relevant IT teachers and teaching researchers to analyze the items again, so before the construction of the question bank, the author communicated with relevant measurement experts and IT teachers and teaching researchers to take the local independence between items into consideration in the test preparation process, so that two things can be achieved at once.

The implementation of the “one paper per thousand” approach does not take into account the actual ability of students, which intuitively gives the impression of fairness, but the theory and practice of testing tell us that students can be reliably measured only when the difficulty of the questions matches their true ability level. However, with IRT-based computer-adaptive tests, the questions in the test bank have fixed difficulty, discrimination, and guessing factors that do not change from sample to sample. The computer-adaptive test system provides the questions that match the ability value of the test taker for each response, and according to the principle of maximum information, each question presented is guaranteed to provide the maximum amount of information, so that only a small number of questions are needed to achieve an accurate measurement. Therefore, IRT-based computerized adaptive tests outperform traditional paper-and-pencil tests in terms of the coverage group of test accuracy. Then, whether the accuracy of traditional paper-and-pencil tests and computer-based adaptive tests is significantly correlated for the same group of test takers’ abilities requires quantitative analysis rather than empirical guesses alone. In this study, 190 students who participated in both the traditional paper-and-pencil test and the computer-based test were selected, and the 190 students who participated in the traditional paper-and-pencil test were first analyzed with the help of Bilog-MG, a

statistical analysis software, to obtain their ability level values based on their responses, which were based on classical measurement theory. The 190 students’ proficiency levels obtained in the machine test system were based on item response theory. Therefore, the normality of the students’ ability levels was verified for both test formats as shown in Figure 5.

The histogram with a normal curve clearly shows that the high school IT academic level test in the form of machine test can also meet the normal distribution approximately, and most of the students’ ability values are around zero, which shows that the use of machine test technology in the high school IT academic level test has a certain scientific basis. To further investigate the degree of correlation between the two tests, descriptive statistics were conducted as shown in Figure 6, and then a two-sided test was conducted using Pearson correlation type. In IRT, however, the amount of test information actually reflects an internal consistency reliability of the test; that is, it reflects the magnitude of the effect of the quality of the measures of the questions on the degree of certainty of the measurement results. Classical measurement theory evaluates subjects based on total test scores and does not pay enough attention to the individual items of the test.

Based on the mean and standard deviation of the descriptive statistics, the following conclusions can be drawn: the mean and standard deviation of the ITAT using the machine test are smaller than those of the traditional paper-and-pencil test, and the mean of θ_1 is relatively smaller than θ_2 , which means that the general level of the subjects’ ability obtained by the machine test is higher; that is, the machine test is more sensitive to students’ progress. θ_1 is larger than θ_2 , which indicates that there is a smaller gap in ability between subjects obtained by means of the machine test. In order to clarify the degree of correlation between the two, SPSS and Pearson correlation coefficients were used to analyze the correlation between the two. In general, the reliability of a large standardized test must reach 0.9 to be convincing, but the correlation coefficient between

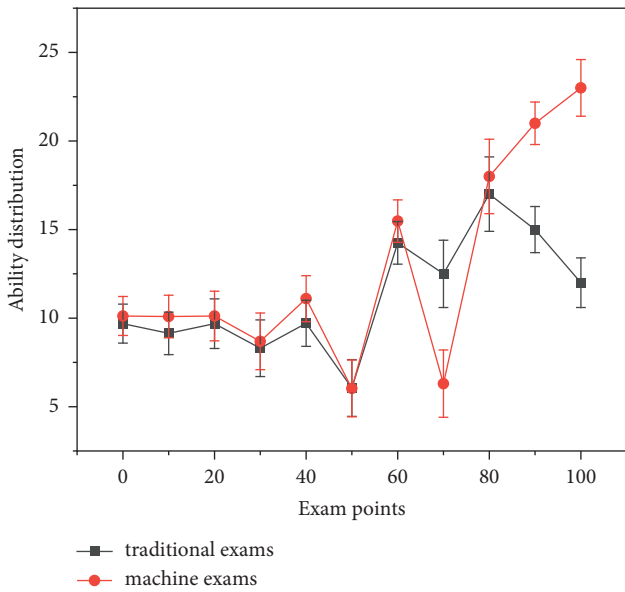


FIGURE 5: Comparison of the ability distribution of subjects in the traditional test and the machine test.

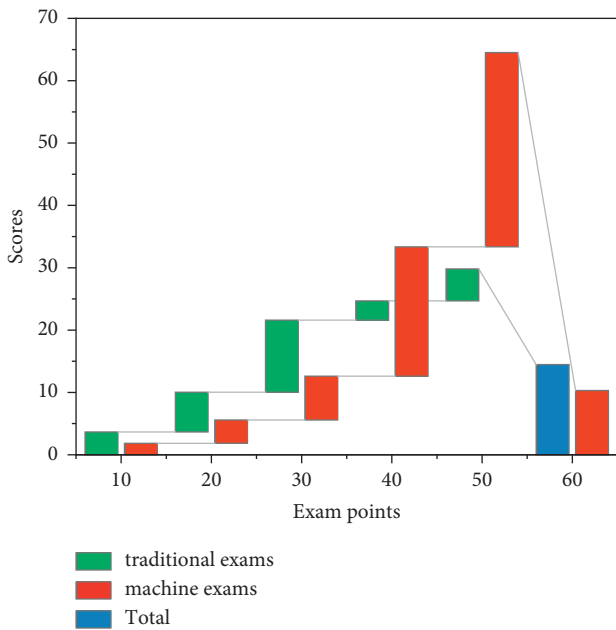


FIGURE 6: Correlation between traditional paper-and-pencil tests and machine tests.

the two tests reached 0.913 in this study, which is certainly a satisfactory correlation for a large-scale test. The correlation between the traditional paper-and-pencil test and the machine test is significant at the 0.01 level.

This study conducts reliability analysis in two aspects; the first aspect is homogeneity reliability, also called internal consistency coefficient, presenting the correlation between all question scores to understand the degree of consistency between questions. In classical measurement theory, the degree of consistency of measurement results is often

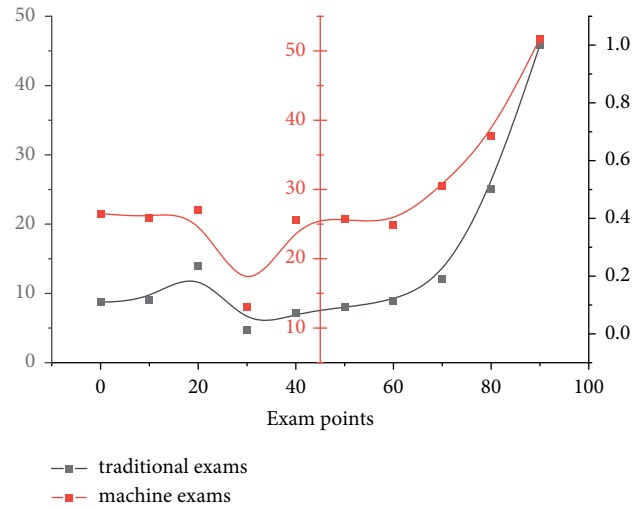


FIGURE 7: Consistency analysis of traditional exams and machine exams.

expressed by calculating the correlation coefficient between multiple results. However, intuitively and logically, the development of a measurement theory should begin with an analysis of the individual items of the test and then analyze the role of the entire test in evaluating subjects. Therefore, this study first focuses on the internal relationships of the test items and analyzes the internal consistency of each of the four test sets based on classical measurement theory, as shown in Figure 7, that is, expressed by the Cronbach alpha coefficient, where a coefficient close to 1 indicates a high internal consistency of the questions, and in general, a reliability of at least 0.80 is acceptable in basic research, while a reliability below 0.35 is low and must be rejected. This ensures that the construction of the machine examination question bank has good internal consistency items.

5. Conclusion

The whole research process involves the identification of test questions and their attributes in the construction of the question bank, the construction of the adaptive test system, the comprehensive analysis of the mock test results, and the thinking about the application of the high school adaptive academic level test. As the most critical part of the whole process of constructing the high school IT adaptive academic level test system, it involves the cooperation of the expert staff of the preparation group, the construction of the two-way itemized list, and the design of the test questions grouped according to the proportion of the itemized list and the anchor test, and in addition to the goal of verifying the effectiveness of the application of the high school IT adaptive academic level test system in this study, in order to show the individualized chemistry report simultaneously, it is necessary to further ask relevant subject experts, educational measurement experts, and front-line teachers to expand the question attributes; that is, the question attributes extend to module specific knowledge points, cognitive levels, and information literacy ability levels. On the one hand, through

the feasibility analysis, we completed the design of functional modules, database, and development process based on the system design objectives, principles, and related core elements to build a complete framework for the system development; on the other hand, the system was compiled entirely in PHP, relying on the local area network of the computer room to realize the interaction between all clients and the server, and the calculation of the ability values in the process of the subjects' responses was done using the computer adaptive test theory. On the other hand, the system is fully implemented in PHP, and all clients interact with the server based on the local area network of the server room, and the computation of the ability value during the test is based on the great likelihood estimation in the computer adaptive test theory.

In the future, we will obtain the information quantity of each subject's answer through each subject's standard error and then obtain the sum of all subjects' information quantity, at which time we could obtain the overall standard error of the test and then obtain the validity of the whole machine test through Equation.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Department of Flight Training and Management, Luoyang Flight College, Civil Aviation Flight University of China.

References

- [1] P. Abedi, M. H. Keshmirshekan, and E. Namaziandost, "The comparative effect of flipped classroom instruction versus traditional instruction on Iranian intermediate EFL learners' English composition writing," *Journal of Applied Linguistics and Language Research*, vol. 6, no. 4, pp. 43–56, 2019.
- [2] M. Alaa, I. S. M. A. Albakri, C. K. S. Singh et al., "Assessment and ranking framework for the English skills of pre-service teachers based on fuzzy delphi and TOPSIS methods," *IEEE Access*, vol. 7, pp. 126201–126223, 2019.
- [3] Y. Bin and D. Mandal, "English teaching practice based on artificial intelligence technology," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 3, pp. 3381–3391, 2019.
- [4] B. Cope, M. Kalantzis, and D. Sears, "Artificial intelligence for education: knowledge and its assessment in AI-enabled learning ecologies," *Educational Philosophy and Theory*, vol. 53, no. 12, pp. 1229–1245, 2021.
- [5] A. Rzhuevskiy, O. Kutyuk, V. Vysotska, Y. Burov, V. Lytvyn, and L. Chyrun, "The architecture of distant competencies analyzing system for it recruitment," in *Proceedings of the 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT)*, pp. 254–261, IEEE, Lviv, Ukraine, 17–20 Sept. 2019.
- [6] A. Kiv, V. Soloviev, and S. Semerikov, "CTE 2018-How cloud technologies continues to transform education," in *Proceedings of the 6th Workshop on Cloud Technologies in Education*, Kryvyi Rih, Ukraine, December 21, 2018.
- [7] C. Korkmaz and A.-P. Correia, "A review of research on machine learning in educational technology," *Educational Media International*, vol. 56, no. 3, pp. 250–267, 2019.
- [8] A. Lin, Q. Wu, A. A. Heidari et al., "Predicting intentions of students for master programs using a chaos-induced sine cosine-based fuzzy K-nearest neighbor classifier," *IEEE Access*, vol. 7, pp. 67235–67248, 2019.
- [9] V. Lytvyn, Y. Burov, P. Kravets et al., *Methods and Models of Intellectual Processing of Texts for Building Ontologies of Software for Medical Terms Identification in Content Classification*, pp. 354–368, IJDM, 2019.
- [10] A. M. Marouf and S. S. Abu-Naser, "Intelligent tutoring system for teaching computer science I in Al-azhar university, gaza," *International Journal of Applied Agricultural Research*, vol. 3, no. 3, pp. 31–53, 2019.
- [11] L. S. Marques, C. Gresse von Wangenheim, and J. C. R. Hauck, "Teaching machine learning in school: a systematic mapping of the state of the art," *Informatics in Education*, vol. 19, no. 2, pp. 283–321, 2020.
- [12] N. Peng, "Research on the effectiveness of English online learning based on neural network," *Neural Computing & Applications*, vol. 34, no. 4, pp. 2543–2554, 2022.
- [13] C. Perrotta and N. Selwyn, "Deep learning goes to school: toward a relational understanding of AI in education," *Learning, Media and Technology*, vol. 45, no. 3, pp. 251–269, 2020.
- [14] A. V. Ryabko, O. V. Zaika, R. P. Kukharchuk, T. A. Vakaliuk, and V. V. Osadchyi, "Methods for predicting the assessment of the quality of educational programs and educational activities using a neuro-fuzzy approach," in *Proceedings of the CEUR Workshop Proceedings*, pp. 154–169, Bamberg, Germany, 2022.
- [15] M. Farhan, S. Jabbar, M. Aslam et al., "IoT-based students interaction framework using attention-scoring assessment in eLearning," *Future Generation Computer Systems*, vol. 79, pp. 909–919, 2018.
- [16] S. Sarwar, Z. U. Qayyum, R. García-Castro, M. Safyan, and R. F. Munir, "Ontology based E-learning framework: a personalized, adaptive and context aware model," *Multimedia Tools and Applications*, vol. 78, no. 24, pp. 34745–34771, 2019.
- [17] Z. Sun, M. Anbarasan, and D. Praveen Kumar, "Design of online intelligent English teaching platform based on artificial intelligence techniques," *Computational Intelligence*, vol. 37, no. 3, pp. 1166–1180, 2021.
- [18] O. N. Yarygin, A. A. Korostelev, L. G. Akhmetov, and A. Maselino, "Modeling of competence as a tool of goal setting for education in modern society[J]," *International Journal of Recent Technology and Engineering*, vol. 7, no. 6, pp. 72–77, 2019.
- [19] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education-where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–27, 2019.
- [20] A. D. Zubkov, "MOOCs in blended English teaching and learning for students of technical curricula," *Integrating Engineering Education and Humanities for Global Intercultural Perspectives*, vol. 131, pp. 539–546, 2020.