The OSCAR PUBLISHING Services
International Journal of Pedagogics

VOLUME Vol.05 Issue03 2025
PAGE NO. 150-154
DOI 10.37547/ijp/Volume05Issue03-42

Analysis of test systems used in classical and modern test theories for assessing students' knowledge

Rasulov Ulugbek Murodulloyevich
Assistant at the Uzbekistan-Finland Pedagogical Institute, Uzbekistan

Received: 27 January 2025; **Accepted:** 25 February 2025; **Published:** 23 March 2025

Abstract: The article presents an analysis of test systems used in classical and modern testing theory in the world. Based on classical theory and widely used systems, the objectives of the iSpring, Google Forms, SAT, GRE, GMAT, LSAT, MCAT tests, the level of academic orientation, structure, cases of computer adaptation, duration of testing, information about the coverage of subjects in the test are presented. The main purpose of the article is to show how important it is to create test systems based on modern testing theory. For this purpose, modern testing systems StartExam, Proaction, TOEFL, and IRT tests were analyzed. It is shown that it is advisable to use the modern IRT test system in higher education institutions.

Keywords: Test, system, task, result, testing, assessment, modern, model, theory.

Introduction: Currently, test technologies in the higher education system are widely used to monitor the academic performance of students during the learning process. Therefore, let us examine the test systems based on classical and modern test theories used in foreign and Commonwealth countries.

Test systems based on Classical Test Theory include iSpring, GRE, SAT, GMAT, TOEFL, and others. Let us review these test systems.

The iSpring test system typically operates based on classical (traditional assessment) methods and is aimed at automating the learning and assessment processes. iSpring functions as part of a platform for creating e-learning and educational materials. It provides convenient tools for creating, evaluating, and analyzing various types of tests and surveys. iSpring is mainly used for conducting online tests. It simplifies the creation of tests, adding questions, and assessing them. This system can be used on computers or mobile devices. iSpring automatically grades tests and analyzes the results in real time. The results and statistics from the tests include correct and incorrect answers, time tracking, and final scores. iSpring also enables the creation of interactive learning materials and allows the integration of video and audio. This helps engage students and make educational content more interesting. The system also offers adaptive testing capabilities, which means questions can be adjusted during the test based on students' responses. This helps better identify the individual knowledge level of each student. iSpring is a convenient platform for creating, evaluating, and analyzing modern and interactive tests. It offers various test types, statistical tools, and real-time assessment capabilities. Adaptive testing and computer-based administration make iSpring a modern test system. This system is used for creating tests and assessment systems in schools, universities, and online courses, for evaluating staff knowledge, conducting training and certification processes, and measuring knowledge level or work efficiency through testing and analysis. iSpring is a modern and interactive platform for creating, assessing, and analyzing tests, providing various test types, statistical data, and real-time evaluation features.

The GRE (Graduate Record Examination) is a test designed to assess analytical, mathematical, and verbal skills acquired during undergraduate studies and is considered one of the classical tests. The test is divided into two types: General (for business schools) and Subject (for scientific and technical fields). The GRE General test is a general exam that assesses a student's knowledge of algebra, mathematics, geometry, and English. This version of the test is required for admission to graduate programs abroad. The GRE

Subject test evaluates a student's knowledge in a specific subject area. The GRE assesses verbal and mathematical skills, written expression, and critical thinking abilities. This comprehensive and rigorous exam not only evaluates a student's analytical abilities but also determines their readiness for graduate-level studies. Globally, the GRE is available in two formats: a traditional paper-based standardized exam and an online version conducted via computer. In the latter, the difficulty of the questions depends on the test-taker's previous answers. Thus, the accuracy of a student's responses in the verbal and math sections determines the complexity of the questions in subsequent sections. The GRE is not a fully traditional test because it is considered an adaptive test system. However, it is not fully adaptive either, as it adapts only at the section level, not at the individual item level. Therefore, the GRE is considered a partially adaptive test. The GRE is based on the Item Response Theory (IRT) and adapts not at the individual item level but at the section level.

The SAT (Scholastic Assessment Test) is a standardized test used in the United States to assess the knowledge of applicants and high school students. In order to be admitted to prestigious U.S. universities, one must achieve a high score on the SAT, and in most cases, American students take this test before graduating from high school. Every student who dreams of studying abroad is advised to take the SAT test.

Since its introduction in 1926, the SAT has undergone several changes, primarily in its language and scoring system, as well as in its name—eventually being renamed the Scholastic Assessment Test. Today, SAT is recognized as an independent brand name and is widely used.

The SAT test consists of two components:

- SAT Reasoning Test (Thinking Test): assesses mathematical skills, reading comprehension, and writing ability.

- SAT Subject Test: includes 20 tests divided into 5 key blocks—sciences, mathematics, world and U.S. history, English literature, and foreign languages.

Although the SAT has recently transitioned into a partially adaptive format, it still remains closer to a classical testing approach. Starting in 2023, the SAT moved to a digital format and became a semi-adaptive test. This means the SAT is no longer a fully traditional test but, like the GRE, it adapts to a test-taker's skill level in a limited way.

GMAT (Graduate Management Admission Test) is a computer-adaptive test based on the principle of sequential analysis. In such tests, the selection of the next question depends on the test-taker's response to the previous one. The GMAT is required for admission to business schools and is used worldwide. It assesses students' analytical, verbal, and mathematical skills, as well as their writing and reading abilities. GMAT results are valid for up to five years.

The GMAT consists of four parts:

1. Analytical Writing Assessment – a written task with a 30-minute limit, where the test-taker must write an essay based on analysis and reasoned critique.

2. Integrated Reasoning Section – includes 12 questions with a 30-minute time limit.

3. Quantitative Section – consists of 37 questions with a 75-minute limit.

4. Verbal Section – includes 41 questions, also with a 75-minute time limit.

Unlike standardized tests, the GMAT is a computer-adaptive test. This means the system adjusts the difficulty and content of the questions based on the student's abilities. The test begins with medium-difficulty questions. If the test-taker answers correctly, more difficult questions follow; if the answers are incorrect, the computer adapts by offering easier questions. This process continues until the system accurately determines the test-taker's ability level.

Another globally recognized test is the LSAT (Law School Admission Test), which is used for admission to law schools. This test evaluates reading skills and verbal logical reasoning. In the U.S., Canada, and many other countries, LSAT results are considered a key indicator for law school admissions.

The LSAT consists of five sections:

- Logical reasoning

- Reading comprehension

- Analytical reasoning

- Writing sample

- Experimental section (this section is unscored)

Each section is allotted 35 minutes. As mentioned, only four of the five sections contribute to the final score; the experimental section is used to pre-test new questions and formats. Each section includes three types of questions: reading, analytical, and logical. Thus, the LSAT assesses whether a student is ready for legal education.

The LSAT is scored on a scale from 120 to 180, with each question carrying equal weight. To apply to law schools, applicants must submit their total score across all test sections. The LSAT can be taken up to three times per year.

Another globally recognized test is the MCAT (Medical

College Admission Test), which is used for admission to medical universities. The MCAT is extremely important for future doctors and veterinarians. Primarily, this test is necessary to evaluate the knowledge level of those applying to educational institutions in the medical field. In other words, the goal of the MCAT is to assess the student's knowledge and analytical thinking skills to determine their readiness for studying in medical institutions.

The MCAT consists of four sections: reasoning, physics, biology, and essay writing. The full duration of the test is 7 hours. A student can earn up to 15 points in each section, making the total maximum score 60 points. A distinctive feature of this test is that no points are deducted for incorrect answers.

The validity of the MCAT certificate and scores depends on the specific medical school the student is applying to. Typically, the certificate is valid for 5 years. If a student fails to take the MCAT twice, they must request special permission to take it a third time. Although the MCAT is considered a classical test, it has been adapted to meet the demands of modern medical education. In its current form, it can be classified among modern tests.

In addition to the classical tests discussed above, modern test theory plays a particularly important role. In 2023, four popular test tools were used for staff assessments: Google Forms, iSpring, StartExam, and Proaction. Even during global economic crises, many companies did not cut back on employee training budgets. Research has shown that even small businesses value their staff. For companies, it is crucial to select candidates not only based on professional qualifications but also on how well they align with the company culture.

To achieve this, many organizations found it appropriate to use tests in the hiring process. There are several reasons for this:

1. Expert-based design: Unlike other selection methods, tests are well-studied tools developed by sociologists, psychologists, and specialists. Employers can set the number of questions, testing time, result presentation format, and more based on expert recommendations.

2. Transparency: Testing ensures a higher level of transparency— all employees receive the same set of questions with the same difficulty level, reducing the influence of human bias.

3. Proven validity: While tests are not the most precise tools, their simplicity and cost-effectiveness have led to their widespread use. Moreover, they are easy to conduct—tests can be created in spreadsheets or on paper. When the number of test-takers is large, automated computer-based test systems are used.

Now let's explore in detail modern tests based on IRT (Item Response Theory), which are widely used in higher education.

The Proaction testing system is one of the modern systems built on Item Response Theory (IRT), distinguishing it from classical test theory. Proaction uses IRT methodology to evaluate tests, allowing for individual analysis of each question and each student. IRT models make tests more adaptive and precise.

Proaction applies a personalized approach in test analysis and focuses on the individual evaluation of every student. In this system, each test response, its difficulty level, and the reliability and validity of the administered tests are assessed using IRT methodology. The system is developed using advanced technologies and modern methodologies.

Compared to classical test theory, Proaction offers a more nuanced and individualized approach. It evaluates each student's responses and knowledge level separately, and thus provides personalized test results for each individual.

Adaptive Testing System

An adaptive testing system is an intelligent test system that adjusts the test questions according to the test-taker's level of knowledge. Unlike traditional tests, this system changes the difficulty of each subsequent question based on the response given to the previous one. The adaptive testing system operates on the basis of Computerized Adaptive Testing (CAT) technology. This technology analyzes the test-taker's responses and automatically selects questions that match their level of knowledge during the test process. It allows for the individual assessment of students' knowledge and provides them with questions tailored to their abilities.

In higher education, adaptive testing systems can be highly effective in assessing students' knowledge, enhancing personalized learning, and optimizing the educational process. The benefits of adaptive testing in higher education include:

Individualized approach – Questions are tailored to each student's knowledge level.

Accurate assessment – Results reflect the student's actual proficiency more precisely.

Time efficiency – Reliable outcomes can be achieved with fewer questions.

Identification of strengths and weaknesses – Helps determine areas where the student excels or struggles.

Optimization of learning – Enables instructors to identify which topics need to be explained in greater

depth.

Compared to traditional tests, adaptive testing provides more accurate and efficient assessments while allowing for personalized instruction. Though traditional tests are simpler and more familiar to most, adaptive tests are more effective in today's modern education systems.

TOEFL (Test of English as a Foreign Language) is a standardized academic English test, primarily taken by those applying to universities in the United States. In addition to the U.S., TOEFL scores are recognized in many other countries as proof of English language proficiency. The TOEFL includes several formats, with the most common being TOEFL iBT and TOEFL Essentials.

The TOEFL iBT (Internet-Based Test) uses Item Response Theory (IRT) to adapt the difficulty of questions according to the test-taker's responses, allowing for more precise measurement of English proficiency. Other TOEFL versions include TOEFL ITP (Institutional Testing Program), TOEFL Junior, and TOEFL Primary, which are designed for school-level students

The TOEFL iBT test consists of four sections: Reading, Listening, Speaking, and Writing. It lasts approximately 3 hours and is intended to assess a student's readiness to study in an English-speaking academic environment. Students typically prepare for the TOEFL iBT over 3 to 6 months. In contrast, the TOEFL Essentials test is a simpler and more affordable version with a lower academic focus.

TOEFL is considered a modern test system because it primarily uses the IRT model. Its adaptive version changes question difficulty and type based on the test-taker's responses, which requires the use of IRT.

IRT (Item Response Theory), often referred to as the modern theory of testing, theory of responses to items, or parameterization and modeling theory of pedagogical tests, is a collection of methods that estimate the probability of a test-taker answering various difficulty-level items correctly.

IRT is used to eliminate uninformative questions from questionnaires, assess the relationship between latent traits and observable variables, and optimize recommended tasks for each test-taker. In the IRT model, neither the test items nor the test itself are evaluated in isolation—instead, the interaction between the respondent and the item is modeled.

In psychometrics, IRT is considered a foundational framework for analyzing and evaluating tests, questionnaires, and measurement tools. It assumes a statistical relationship between responses and the underlying traits being measured. Various statistical models are used to estimate both task and respondent parameters. Unlike traditional tests where all questions are considered to have equal difficulty, IRT recognizes that each item has unique properties that must be included in the model.

This differs from methods like Likert scaling, where all tasks are assumed to be interchangeable replications. In contrast, IRT treats each item as a data point with unique parameters that must be incorporated into the model.

CONCLUSION

In summary, IRT models the probability of a test-taker responding correctly to each item. A core characteristic of modern testing theory is the separation of parameters for both test-takers and items—meaning that the probability of a correct response depends on the interaction between the respondent's and the item's latent traits. The specific form of this interaction is determined by the researcher's assumptions and is described through precise mathematical functions, forming the IRT model.

Modern IRT models draw from techniques such as factor analysis, generalized linear models for mixed effects, network models from statistical physics (e.g., Markov fields and the Ising model), and various methods from data science (e.g., collaborative filtering models and restricted Boltzmann machines).

Today, IRT is regarded as one of the most advanced and theoretically grounded testing approaches that accounts for human nature during the test process and is widely used in data-driven educational and psychological assessments.

REFERENCES

Miller, K. (2011). The TOEFL Test: A Study Guide for English Language Learners. McGraw-Hill

The College Board (2016). The Official SAT Study Guide. College Board

Educational Testing Service (ETS) (2018). The Official Guide to the GRE General Test. ETS

Kaplan (2019). Kaplan GMAT Complete 2020. Kaplan Publishing

Kelley, K. (2017). The LSAT Trainer: A Comprehensive Self-Study Guide for LSAT Preparation. LSATMax

Kaplan (2019). Kaplan MCAT Complete 7-Book Subject Review. Kaplan Publishing

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991) Fundamentals of Item Response Theory

Baker, F. B., & Kim, S.-H. (2017)The Basics of Item Response Theory Using R

Wainer, H., & Dorans, N. J. (2000). Computerized Adaptive Testing: A Primer. Lawrence Erlbaum Associates

Davies, R. (Ed.) (2006). Computerized Testing and the Internet. Cambridge University Press.

Pashley, P. J., & Jodoin, M. G. (2014). Handbook of Automated Essay Evaluation. Information Age Publishing.