UOT 37:001

Gulnara Gurban Iskandarova,
doctorial student in the program of doctor of philosophy
of Azerbaijan University of Languages
https://orcid.org/0009-0000-6172-6428
E-mail: giskenderova@yahoo.com
https://doi.org/10.69682/arti.2025.92(4).149-154

THE EDUCATIONAL ASSESSMENT JOURNEY: THROUGH PAST EXPERIENCES TO THE FUTURE

Gülnarə Qurban qızı İskəndərova,
Azərbaycan Dillər Universitetinin
fəlsəfə doktoru proqramı üzrə doktorantı

TƏHSİLDƏ QİYMƏTLƏNDİRMƏNİN YOLU: KEÇMİŞ TƏCRÜBƏLƏRDƏN GƏLƏCƏYƏ DOĞRU

Гюльнара Гурбан гызы Искандарова,
докторант по программе доктора философии
Азербайджанского Университета Языков

ПУТЬ ОЦЕНИВАНИЯ В ОБРАЗОВАНИИ: ОТ ПРОШЛЫХ ОПЫТОВ К БУДУЩЕМУ

Xülasə. Təhsildə qiymətləndirmənin gələcəyini nəzərdən keçirərkən keçmişi, indini və gələcək perspektivləri geniş tarixi baxışla nəzərə almaq lazımdır. Müəllimlərin keçirdiyi dərslərin keyfiyyəti, şagirdlərin öyrənmə təcrübələri, tədris proqramının qiymətləndirilməsi, akademik uğurlar və sistem performansı təhsildə qiymətləndirmədən böyük dərəcədə asılıdır. Qiymətləndirmələr, müəllimlərin, ali təhsil müəssisələrinin yekun hesabat qiymətləndirilməsi ilə yanaşı, tədris və öyrənmənin inkişafını dəstəkləyən davamlı izləmə və təkmilləşdirmə vasitəsi kimi də xidmət edir. Böyük təsir gücünə görə, onun keçmişini, indiki vəziyyətini və gələcək imkanlarını dərindən anlamaq faydalı olardı. Bu təhlilə başlayarkən, son 3000 il ərzində keçirilmiş qiymətləndirmələrdən başlayaraq və artıq bir çox yerlərdə mövcud olan indiki vəziyyətə qədər qısa tarixi tura nəzər salacağıq.
Açar sözlər: *Qiymətləndirmə, təcrübə, perspektiv, inkişaf, imkanlar, şərait*

Abstract. When considering the future of educational evaluation, I want to consider the past, the present, and the horizons from an extensive historical perspective. The quality of teacher-led lessons, student learning experiences, curriculum evaluation, academic achievement, and system performance are all greatly influenced by educational assessment. Assessments serve as a tool for summative accountability evaluations of educators, institutions of higher learning, and administration as well as for formative development of teaching and learning. Because of its immense power, it would seem beneficial to have a sophisticated grasp of its past, present, and future possibilities.

As I start this analysis, I take you on a quick historical tour starting with evaluations from the previous 3,000 years and ending with the present, which is already happening in many settings and circumstances.
Key words: *Assessment, evaluation, experience, perspective, development, possibilities, circumstances*

Аннотация. При рассмотрении будущего оценивания в образовании необходимо учитывать прошлое, настоящее и перспективы будущего с широкой исторической точки зрения. Качество проводимых учителями уроков, учебный опыт учащихся, оценивание учебных программ, академические достижения и эффективность системы в целом, в значительной степени зависят от механизмов оценивания в образовании. Оценивание служит не только для подведения итогов –будь то учителями или высшими учебными заведениями, – но и как средство непрерывного наблюдения и совершенствования, поддерживающий процесс преподавания и обучения. Учитывая его значительное влияние, полезно будет глубоко понять его исторические корни, текущее состояние и возможные

будущие направления. Начав с анализа, мы кратко рассмотрим историю оценивания за последние 3000 лет и перейдём к современному состоянию, которое уже реализовано во многих странах и образовательных учреждениях.

Ключевые слова: *оценивание, опыт, перспектива, развитие, возможности, ситуации*

Research indicates that the initial assessment was conducted in China. Commencing in this nation in 587, the evaluation was referred to as one of the "Chinese imperial evaluations". Exams have to be passed in order to be employed here as a public worker. These examinations, which were first administered verbally and then in writing, persisted in China till 1905.

Records from between 2,500 and 3,000 years ago indicate the beginning of the ancient Chinese Imperial testing system (1). Throughout China's vast empire, brilliance was recognized and rewarded according to that system. It was intended to find pupils with excellent reading and memory proficiency meritocratically in order to run the emperor’s system of rule and authority over a vast populace, as opposed to depending just on referrals, bribery, or nepotism. In order to accomplish these objectives, the system used standardized assignments (such as writing an essay based on Confucian ideas) under monitored conditions to guarantee performance comparison and integrity (2).

The exams under the system became ever more difficult and rigorous, and in the end, no one could receive the highest grade – that was saved only the emperor alone. This sophisticated technology was justified in part by the consequences of the selection process; selected candidates not only gained lucrative positions but they were also identified publicly on lists of examinees and granted the privilege to wear badges or colors that denoted the level of the examination they had passed.

It should come as no surprise that there was a thriving industry creating cheat materials (such as miniature books that resembled Confucian classics) and apprehending cheaters. Scholarship carried with it enormous prestige and the potential for social advancement.

Its history dates to approximately the same period since judgment is a crucial component of education. Research indicates that the initial assessment was conducted in China. The evaluation was first conducted in this nation in 587 and is often referred to as the "Chinese imperial exams". It was a requirement to successfully complete an assessment in order to be employed here as a public servant. These examinations, which were first administered verbally and then in writing, persisted throughout China until 1905.

On the other hand, as noted in Encyclopedia Brittanica (2010a), the literary and rhetorical heritage of the Roman Empire, including grammarians' and rhetoricians' schools, gave rise to European educational evaluation. Schools were established at the same period throughout Europe in a variety of cathedrals, monasteries, and bishop schools. Church clerics had more sophisticated theological and intellectual training under Charlemagne because they had to learn Latin to properly interpret the Bible. Early in the Renaissance, as European civilization advanced, schools were established under the direction of a bishop, a cathedral official, or even secular guilds for those judged capable of teaching. Early Renaissance Europe saw the opening of institutions run by bishops, cathedral officials, or even secular guilds for those judged capable of teaching. To promote safe travel and unrestricted thought, certain privileges and responsibilities were granted to the instructors and students at these institutions. Many of the clerical traditions of reading significant books and scholars assessing the caliber of education by comparing how students perform in oral arguments, debates, and conflicts to the opinions of more senior experts were adopted by European universities in the 1100s. In the centuries that followed, performances and assignments for writing were added to the oral arguments as a means of evaluating the caliber of learning results. Nevertheless, evaluation was predicated, as in the Chinese Royal system, on the experience and discretion of more qualified academics or officials (3).

These systems were established to satisfy the demands of either religion or society requiring numerate and literate academics, intellectuals, and administrators. Traditionally, access to higher education – or even just basic education – was restricted. Standardized tests were used to choose applicants in a mostly

competitive manner, even if they were merely the procedure and not the task as well as the score. Families and children participated in these programs because achieving academic achievement offered them hope for a life free from hardship and manual labor. As a result, schooling served as a means of educating students for the final test, and assessments given along the schooling process served only as a facsimile of the final evaluation. Assessments were, at their core, summative judgments of the students' skills.

As education grew through the 1800s, especially in higher education, more effective ways to reduce the burden of memorized recitations were sought for (3). As a result, leaving exams was imposed as a prerequisite for admission to university programs, government employment, and learned professions like teaching. In the 1800s, as a greater number of students attended colleges, more effective methods of gathering data were developed. These methods included the essay test and the method of responding to problems in writing independently without assistance. Examining students still follow this custom by arranging their desks in neat rows and completing their test papers within time constraints and scrutiny.

Since the very beginning of educational programs in Europe, researchers have considered and tested a variety of techniques to gauge pupils' level of understanding. Exams used to be administered verbally. The teacher graded the Latin comments that the students delivered in response to the questions. The first written examinations were introduced at Cambridge University in 1792 (4, p.7-9).

However, it became clear by the early 1900s that there were significant issues with how these significant intellectual exercises were scored. Scores for the same job varied because markers were inconsistent with each other or remained constant within themselves over time or between assignments or objects. As a result, multiple-choice question exams were created early in the 20th century so as to guarantee speed in administration and accuracy in scoring (5). It's also important to note that the use of multiple-choice exam techniques resulted in substantial reductions in both time and expenses. This factor contributed to the widespread use of standardized, machine-scored exams for graduate school enrollment, university admission, and even high school evaluation during the course of the 20th century. Within classical test theory statistical modeling, the method of scoring items separately (i.e., right or incorrect) produced simple and recognizable values (5).

Educators categorize the history of assessment into many phases. Prior to the beginning of the First World War, exams were used to evaluate pupils' knowledge, with a focus primarily on subject-matter proficiency. New concepts for evaluating student accomplishments emerged at the very beginning of the twentieth century as educational development patterns became more pronounced. The "concept of progressive education" originated in the US at that time and subsequently extended to other nations. The fundamental tenet of the "innovative philosophy of education" was that life skills should be developed in line with the directions that industry and technology are developing, rather than education being founded on scientific knowledge (4, p.7-10).

Nevertheless, in the 1950s and 1960s of the 20th century, this notion became obsolete. The vast majority of nations have altered their educational standards since the dawn of the space era, and new methods of student evaluation have begun to emerge. The teaching of precise subjects and its more efficient assessment received all of the focus. This kind of evaluation was used to compare students' levels of mastery in addition to gauging each student's unique knowledge and abilities. Subsequently, specialists started to consider unique initiatives that would remove these comparisons from the classroom and enable them to be made between schools, states, and regions. This demonstrates the ongoing and continuous character of experimenting with and improving assessment-related processes (4, p.7-10).

In order to ensure the reliability of results and their interpretation, the ideas of validity have expanded over the 20th century, and the techniques of validation have gotten more intricate and multifaceted (6). These comprised, among other things, equating, norming, item response theory, component analysis, scale reliability, and standard setting (7). It is

important to note that statistical techniques for analyzing test results originated in the early phases of psychology. Educational testing turned to resemble statistics more and more as psychometric techniques grew more sophisticated. It's interesting to note that assessment is widely used in education, yet only a few educators are skilled in the statistical techniques required to analyze examinations, particularly ones administered by government agencies. It is true that very few instructors receive training in statistical test analysis procedures, even fewer comprehend them, and almost none apply them. Of course, an evaluation consists of more than simply a test or a series of questions. It is a valid effort to put into practice a sample inside a curriculum area, construct, or subject. The difficulty in assessing comes from the fact that the most straightforward or worthwhile content in any field is typically the one that is easiest to test and grade. Learning objectives for K–12 education, not to mention higher education, need students to do more than just memorize, recall, and repeat lists of terms, facts, or data points. Deep processing requires the retention of specific information, yet just recalling those facts is insufficient. Pupils must demonstrate critical thinking, creativity, problem-solving skills, analysis, and synthesis. Evaluating these kinds of talents is really challenging and complex (4, p.3-5).

Scriven's (1967) work on the establishment of the assessment philosophy and standards was a significant advance in the second decade of the 20th century. He recognized that one important component of the assessment was the differentiation between summative evaluation, which assessed the quantity, caliber, or value of the outputs generated by the process, and formative evaluation, which occurred early enough in the process to affect its endpoints. In education, the term "formative evaluation" was immediately adopted to refer to tests that instructors employed in the classroom to determine which students needed to learn what subject matter (8, p.39).

However, since the late 1980s, the distinction between both formative and summative learning has changed from one of time to one of a kind, primarily because of Sadler's work in 1989. Formative evaluations were limited to informally conducted classroom interactions rather than official examinations. As a result, testing for learning has evolved into a kid-friendly strategy (9, p.285) for integrating learners into their education and producing deep, meaningful results without the stressful demands of testing.

As a result, there are a lot of crucial factors that guide the effective implementation of educational assessment that are separate from one another as we move forward.

The enormous quantities of data that are currently accessible through online learning behaviors and evaluations have captured the attention of those working in the field of educational statistics. With the advent of computing technology, it is now feasible to automate the creation of test items (10, p.15), score performance (11), and tailor test material based on exam-taker achievement (12, p.25). Online and remote testing have become popular due to the Covid-19 epidemic, yet questions have been expressed regarding how the technology is utilized to ensure the quality of student achievement (13, p.128). Many educators and students now use virtual or remote instruction, learning, and evaluation, particularly in view of our pedagogical approaches to the COVID-19 epidemic. It's obvious that some families value this as it allows their kids to advance quickly without being constrained by their peers or the teacher. For these families, pursuing digital education would represent a promising future. However, depending just on an automated interface to conduct assessments or deliver instruction might dehumanize the essentially human process of educating and learning (14, p.125-126).

There is already a lot of interest in using "big data" from computer-based exams to investigate in greater depth how learners complete assignments. Test-taker involvement is gathered, used, and reported on by several extensive computer-based testing programs as a component of their method of gathering information. This will entail using technology to supplement human labor in the classroom as opposed to using it to replace teaching and testing personnel (15, p.189-201). Thankfully, there are some instances of educational technology used for evaluation. Instead of replacing the work that educators and

students must perform to determine what they need to acquire or teach and to track their progress, these technologies assist it. Above all, they have a strong relationship to the curriculum and the instruction that instructors are providing (16, p.446).

Technology that helps instructors make accurate and helpful judgments about what pupils should learn next in the classroom is necessary for them to succeed in overcoming their own mistakes. Bennett (17, p.7-15) noted that while technology would advance, there might not be an increase in computer-based exams. This evaluation offers us the possibility of a better future by supporting teachers rather than removing them (18; 19).

This article offers a comprehensive exploration of the evolution of educational assessment, drawing from diverse historical traditions including the ancient Chinese imperial exams, the rhetorical and scholastic practices of medieval Europe, and the development of modern measurement and evaluation techniques. The scientific novelty lies in integrating these varied perspectives to illustrate how assessment has continually adapted to meet changing educational, social, and technological needs.

The rationale for the study. The relevance of this work is underscored by the growing prevalence of digital learning environments and remote education, which pose new challenges for maintaining the integrity and human-centered nature of assessment. A thorough understanding of assessment's historical development and scientific principles enables educators and policymakers to effectively design and implement assessments that support meaningful learning outcomes in today's complex educational landscape.

Innovative aspect of the study. This study distinguishes itself through its trilingual format – Azerbaijani, Russian, and English – enhancing accessibility and encouraging scholarly dialogue across linguistic and cultural boundaries. Presenting the content in three languages promotes inclusivity and reflects the diverse educational contexts in which assessment practices are applied.

Another innovative feature is the study's integrated approach to exploring educational assessment. It not only revisits historical and philosophical foundations but also connects them with contemporary developments, such as the influence of psychometric theories, digital tools, and artificial intelligence. By linking traditional assessment models with emerging educational technologies, the study offers a comprehensive and forward-looking perspective, valuable for both academic researchers and classroom practitioners.

Implications for practice. The practical significance of this study is reflected in its analysis of how historical and contemporary assessment methods influence current educational systems. It highlights the critical role of valid, reliable, and fair evaluation tools in promoting social mobility, improving instructional strategies, and harnessing advancements such as computerized testing and data analytics.

References

1. China: Five thousand years of history and civilization // - China Civilization Centre: City University of Hong Kong Press, -2007, -pp. 167-185.
2. Feng, Y. From the imperial examination to the national college entrance examination: the dynamics of political centralism in China's educational enterprise. -1995, pp. 28–56.
3. Encyclopedia Britannica //-Encyclopedia Britannica, Inc. Western education in the 19th century. In Encyclopedia Britannica, -2024, -pp.12-14.
4. Mardanov. M. Assessment // Baku State University Press, - 2006, -pp. 1–11.
5. Croft.M. Development and evolution of the SAT and ACT. The history of educational measurement: Key advancements in theory, policy, and practice/ Croft.M, Beard. J, Routledge.J. -2022, -pp. 47–65.
6. Zumbo.B. D. Validity and validation in social, behavioral, and health sciences // Zumbo.B.D, Chan.E. Springer Press. -2014, -pp. 5.
7. Kline, R. B. Psychometrics. SAGE Research Methods Foundations // 2020, -pp. 88–103.
8. Bloom, B. S. Handbook on formative and summative evaluation of student learning / Bloom, B. S, Hastings, J. T, Madaus, G. F. -1971, -pp. 24-95
9. Stobart, G. Fairness in multicultural assessment systems. Assessment in Education: Principles, Policy & Practice. -12(3), -2005, -pp. 275–287.
10. Gierl, M. J and Lai, H. A process for reviewing and evaluating generated test items. Educational Measurement: Issues and Practice, -2016, -35(1), -pp. 6–20.

11. Shin, J. Automated essay scoring using deep learning algorithms. In D. B. A. M. Khosrow-Pour (Ed.), Handbook of research on modern educationaltechnologies, applications, and management / J. Shin, Q. Guo, M.Gierl. -2021, -pp. 410–432.
12. Linden, W. J. V. D, and Glas, G. A. W. Computerized adaptive testing: Theory and practice. - Kluwer Academic Publishers, -2000, -pp.3-27.
13. Dawson, P. Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education. -2021, -pp. 75-130.
14. Asimov, I. Doubleday. Fantasy Sci. Fiction. - 1954,-Vol. 6, -pp. 125–127.
15. Hattie, J. A, and Brown, G. T. L. Technology for school-based assessment and assessment for learning: Development principles from New Zealand. -Journal of Educational Technology Systems, -36(2), -2008, -pp. 189–201.
16. Kingston, N. M. Editorial: The use of organized learning models in assessment. Frontiers in Education / N. M. Kingston, A. C. Alonzo, H. Long, R. Swinburne Romine - 2022, -pp. 446.
17. Bennett, R. E. Educational assessment: What to watch in a rapidly changing world. Educational Measurement: Issues and Practice, -2018, -37(2), -pp.7–15.
18. Firuza A.R. Independent work of students in the process of learning the english language / A.R. Firuza // Педагогика и психология в современном мире: теоретические и практические исследования: сб. ст. по материалам XXV Международной научно-практической конференции «Педагогика и психология в современном мире: теоретические и практические исследования». – № 7(25). – М., Изд. «Интернаука», -2019, – стр. 98-102.
19. Firuza A.R. “Conceptual approaches to the organization of independent work of students in higher educational institutions”. -Studies in English Language Teaching 8 (№3), -2020, -pp.157-165.

Redaksiyaya daxil olub: 10.06.2025