LANGUAGE EDUCATION AND TESTING

СОВРЕМЕННОЕ ЯЗЫКОВОЕ ОБРАЗОВАНИЕ И ТЕСТИРОВАНИЕ

Sergey Y. Trapitsyn
Doctor of Education, Professor, Head of the Department of Educational Management and Human Resources Management, A.I. Herzen State Pedagogical University of Russia, 48 Moika River Embankment, Saint Petersburg, 191186, Russian Federation, trapitsin@gmail.com
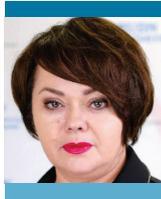ORCID: 0000-0003-3752-8848
SPIN-code: 1582-2439

Marina N. Kunovski
PhD in Philology, Associate Professor, Head of the Russian Language Department No. 2, Russian Language Institute, RUDN University, 10 Miklukho-Maklaya St, bldg. 3, Moscow, 117198, Russian Federation, kunovski-mn@rudn.ru
ORCID: 0000-0002-5872-2052
SPIN-code: 7029-6172

Natalya V. Pomortseva
Doctor of Education, Associate Professor, Head of the Department of Russian Studies, Ethno-Oriented Pedagogy, and Digital Didactics, Russian Language Institute, RUDN University, 10 Miklukho-Maklaya St, bldg. 3, Moscow, 117198, Russian Federation, pomortseva-nv@rudn.ru
ORCID: 0000-0003-4224-8138
SPIN-code: 5744-8264

Pedagogical test as a tool for assessing the quality of education

Sergey Y. Trapitsyn[1] ✉ **, Natalia V. Pomortseva**[2] **, Marina N. Kunovski**[2]

[1]Herzen State Pedagogical University of Russia, *Saint Petersburg, Russian Federation*

[2]RUDN University, *Moscow, Russian Federation*
✉ trapitsin@gmail.com

Abstract. The results of the analysis of diagnostic capabilities and areas of application of pedagogical tests as a tool for assessing the quality of education are presented. Special attention is paid to the requirements for tests that ensure objective and reliable measurement of educational outcomes, and criteria for evaluating the quality of test systems

are formulated. It is shown that the effectiveness of pedagogical tests is achieved through a clear definition of testing goals, a conscious and reasoned definition of the subject of assessment, compliance with the rules of selection and design of test tasks, as well as the use of relevant methods of processing and interpreting test results. The typology of texts is presented, the characteristics of the target purpose, the specifics of construction and the areas of application of various types of tests are given. The prospects for the development of pedagogical testology are determined.

Introduction

Education quality management cannot be carried out without complete, reliable, and up-to-date information on educational processes, their results, and the implementation conditions. The importance and relevance of this task is growing with the development of digital technologies and automated monitoring tools. In this regard, pedagogical testology is a promising area for the development of pedagogical science and practice. The use of tests as a means of assessing academic achievement, despite their long history of use, still evokes draws mixed reactions in the teaching community and the general public, giving rise to numerous myths and interpretations. However, it appears that the negative reaction to the increasing prevalence of testing as a simple and effective tool for assessing educational quality is often due to unclear and incomplete understanding of the methodology, methods, and tools of pedagogical testing, as well as the poor quality of the tests themselves and the failure to adhere to the basic principles of their development and application. Another source of misconceptions and criticism stems from a failure to understand that no form of assessment, even the most sophisticated and objective, can

change, or improve, the quality of education, just as a thermometer cannot change temperature. Any pedagogical assessment, including a test, serves merely as a tool for obtaining information about the degree to which students have mastered specific content. A test is a far more accurate tool than the evaluation criteria favored by some teachers, such as "gave a sufficiently complete and detailed answer," "deeply understands the essence of the question," or "allows minor errors and inaccuracies," etc., allowing them to act as objective and dispassionate (or so it seems?) arbiters, assessing the quality of a student's knowledge (but, alas, not their own ability to convey it).

In this regard, researchers, specialists, and practicing teachers are faced with the question: what constitutes a pedagogical test and what is the assessment mechanism used within it? An attempt to answer this question is **the goal of study**.

Pedagogical testing has attracted the attention of researchers both in terms of its origins, formation, and development in Russia (Ganin, Ganina, 2012) as a pedagogical phenomenon and as a scientific field, beginning with the introduction of the term "test" itself, the development and evolution of testing as a method of assessing learning, and leading

up to the emergence of the Unified State Exam as a mandatory state method of assessing knowledge. It has also attracted the attention of researchers in terms of research into test-based diagnostics of educational quality, analyzing the current state of pedagogical testing, and its advantages and disadvantages (Verbitsky, Kreslavskaya, 2012). Pedagogical testing has been adapting to new formats for presenting tasks. In 2005 V.Yu. Pereverzev described innovative approaches to creating test tasks for computerized delivery, methods for selecting content, and the technology for developing specifications for pedagogical tests for various disciplines[1].

M.B. Chelyshkova and V.I. Zvonnikov in their work devoted to the modern means of evaluating learning outcomes present theoretical and methodological approaches to the design of pedagogical tests, an analysis of their functions and applicability, and the use of tests at various levels: in school assessment systems, in final exams, and in education quality monitoring systems[2]. M.B. Chelyshkova considers testing to be a scientifically sound method for assessing student knowledge. She emphasizes the importance of theoretical foundations and scientific methods in the design of pedagogical tests, in selecting content, and in developing assignment forms in accordance with the requirements of testing technology. A key step is the analysis of test results: the obtained empirical data is used to statistically substantiate the test's quality and further improve it. When selecting assignments, it is important to focus primarily on the level of students' knowledge and on maintaining motivation for learning[3] (Chelyshkova, 2002).

I.L. Paskhover provides a definition of the term "pedagogical test" and describes the key criteria for their classification, examining the stages of test development and the principles for selecting the content and format of test items (Paskhover, 2011). A classification of tests by assessment types, test item content, assessment methodology, presentation format, and scope of application is presented in the works of S.V. Pashnin (Pashnin, 2009).

V.S. Avanesov analyzes the phenomenon of the probability of a correct answer to test items, emphasizing that using the same system of items for different purposes and levels of test-takers is unacceptable. This approach leads to an increase in measurement error and, consequently, to a decrease in their quality; achieving absolute objectivity in pedagogical testing is impossible, since measurements inevitably include subjective components. To increase the level of objectivity while developing test materials, it is necessary to deliberately reduce subjective factors (Avanesov, 2008). The researcher emphasizes the educational potential of tests as a tool for intellectual development, stimulating the activation of the learning process and improving the quality of knowledge (Avanesov, 2004). K. Ingenkamp (Ingenkamp, 1991) also considers pedagogical diagnostics as a prerequisite, condition, and outcome of learning, as well as a factor in motivating learning, discussing the role and function of pedagogical control in the educational process.

Researchers emphasize aspects of test development and application in educational practice such as standardization and normalization, parameterization accuracy, statistical justification of testing hypotheses, analysis of regression and correlation dependencies, and scaling of test results. These elements are crucial for the effective use of tests in student assessment (Mayorov, 1997; Neiman, Khlebnikov, 2000). B.P. Bitinas identifies

---

[1] Pereverzev, V.Yu. (2005). *Technology of developing test tasks: a reference guide a teaching aid for students of secondary vocational educational institutions*. E-Media. (In Russ.). EDN: QUVHKP

[2] Zvonnikov, V.I., & Chelyshkova, M.B. (2008). *Modern means of evaluating learning outcomes : training manual for students of higher education institutions training in pedagogical specialties*. (In Russ.). EDN: QWPIVR

[3] Chelyshkova, M.B. (2002). *Theory and practice of designing pedagogical tests : training manual*. Logos. (In Russ.).

several key functions of pedagogical diagnostics: feedback, which provides teachers with information for adjusting their teaching trajectory; evaluating the effectiveness of teaching activities; and predicting future work based on the data obtained. Pedagogical diagnostics is aimed at improving the quality of the educational process, with a special emphasis on the development of the personality of students, as tests have educational potential (Bitinas, 1993).

Methodology and Methods of pedagogical testing

According to the principles of interrogative logic, the structure of any test question (or task) should contain two main parts: the substantive (knowledge) part, and the interrogative part. The first part can be a true, incomplete (requiring further explanation), or false statement; the second which is the actual question (task), defines the action the test-taker must perform. This structure of a pedagogical test establishes its basic requirements:

• *substantive validity* — correspondence between the substantive (knowledge) part of the test, and the subject (section, topic) being studied;

• *functional validity* — correspondence between the test and the assessment objective (recognition, demonstration of knowledge, comprehension, explanation, etc.);

• *certainty* — relates more to the interrogative part of the test, setting clear requirements for the action the test-taker must perform;

• *simplicity* — the test's substantive part should be formulated concisely, its presentation should clearly reflect the content of the subject area being tested, and the defining characteristic should be necessary and sufficient;

• *unambiguity* — the test should not contain ambiguous interpretations, allow for multiple interpretations, and should include clear criteria and assessment rules;

• *reliability* — the accuracy, reproducibility, and consistency of results when testing the same

student (group) repeatedly, as well as resistance to random errors and measurement errors;

• *differentiation ability* — the test should be able to detect statistically significant differences in both the level of mastery of individual sections of the assessed content and/or actions performed, and the level of student preparation.

The diversity of purposes, users, methods, and data processing techniques determine the specifics of testing systems while still meeting the general requirements:

1. The test must be consistent with the nature of education quality improvement decisions that will be made based on its results.

2. Test results must be assessed simultaneously with an analysis of the students' abilities in mastering the content being tested, with equal attention to both assessing educational achievements and the opportunities to achieve them.

3. The characteristics of the data obtained must be consistent with the subject of assessment.

All content elements declared as measurable must actually be measured, and tasks must be valid in relation to the learning outcomes being assessed. Assessment tasks must produce consistent results. They must be authentic, reflecting situations that students will encounter in their professional and personal lives. Results must be reliable. During the testing process, students must have equal opportunities to demonstrate their achievements.

4. Assessment results must be fair for all students, regardless of gender, place of residence, nationality, etc.

5. Conclusions regarding the achievement of planned results and their achievability must be substantiated not only by the data obtained but also by an analysis and description of the conditions under which this becomes possible. For example, higher results for a certain group of students may be explained by the fact that they had a high level of motivation to learn, were taught by outstanding teachers, had more academic hours to master the subject, etc.

Research Results

The rapid development and widespread use of tests in connection with the development of online courses and distance learning, has led to a situation where a significant number of them do not meet the specified requirements and are, in fact, not tests, increasing the number of opponents and critics of testing. The situation is exacerbated by the fact that in many cases, not only the developers of testing programs and tools, but also the teachers who use them, do not fully understand the purposes of testing, the conditions for its use, its limitations, and diagnostic capabilities.

An analysis of current educational quality assessment practices shows that its implementation is significantly complicated by the lack of clear formulations and measures of the goals planned for achievement during the learning process. Undoubtedly, educational standards set certain benchmarks for personal, meta-subject, and subject-specific outcomes, such as: mastery of social norms, rules of conduct, roles, and forms of social life; mastery of the basics of self-control, self-assessment, and decision-making; Understanding, interpreting, and commenting on texts, etc. However, such general formulations leave ample room for interpretation by each individual teacher: "demonstrated good knowledge," "demonstrated the ability to apply knowledge in practice," "can solve non-standard problems," "shows interest in learning," etc.

With such goal setting, it's not entirely clear which specific test tasks allow for a clear, complete, and accurate assessment of the above-mentioned knowledge quality characteristics; the content, difficulty level, and sequence of questions are determined intuitively by each teacher, and often not in the best possible way. There is also no consensus on the number of questions required to test a given volume of educational material. Accordingly, different indicators of knowledge quality are assessed, which reduces the diagnostic value of the assessment system.

It should be noted that this applies not only to tests; any of the methods and forms of knowledge quality assessment used by teachers has its own advantages and disadvantages, as well as its own limitations. A disadvantage of an oral examination is the high degree of uncertainty and randomness; the assessment obtained on the exam is often subjective. At the same time, this form of assessment is sufficiently flexible and adaptable; the teacher has the opportunity to quickly respond to the student's response, ask clarifying and/or additional questions, and maintain direct contact with the student, creating a motivating emotional environment. Written assessment, although more objective than an oral exam, does not allow the teacher to directly monitor the progress of the assignment. Furthermore, written assessment, in addition to subject matter knowledge, requires the student to have written language skills, express their thoughts competently, and possess the ability to search, select, structure, and critically analyze information, among other skills.

There is a problem when a teacher assesses the quality of knowledge: students' subjective perception of the fairness of the assessment. They often attribute low grades to teacher bias. Research shows that for high-performing students, the agreement between their own assessment and the teacher's assessment is 46%, while for low-performing students, the agreement is only 11%. Solving this problem requires developing students' reflective skills and the ability to adequately self-assess.

When discussing the purposes of using pedagogical tests as a means of assessing educational quality, it is important to clearly define the object and subject of assessment. According to ISO 9000:2015, quality is the degree to which a set of characteristics of an object meets requirements. Clearly, the object of pedagogical testing is a student (or a group of students), characterized by a large number of attributes (properties). It is also clear that the

subject of assessment is, generally speaking, the level of their learning (the quality of learning outcomes). Moreover, the set of diagnosed properties (features) may vary, particularly in terms of the substantive characteristics of knowledge (completeness, depth, strength, consistency, etc.), subject area (mathematics, natural sciences, humanities, etc.), and learning activities (information search, analysis, comparison, classification, argumentation, etc.).

Another goal of pedagogical monitoring may be to identify problem areas in the educational process, for example, sections and topics that pose the greatest learning difficulties, as well as differences in the quality of teaching, technologies used, and learning conditions.

Many researchers argue that the very fact of pedagogical monitoring influences student motivation. Contrary to the claims of testing critics, studies have shown that this is also true for pedagogical tests. These studies have yielded interesting results, including increased self-esteem and self-worth, interest in learning, a desire for more thorough preparation, and so on. According to research by P. Broadfoot, M. Crossly et al, school teachers significantly improved their professional skills through the process of constructing test tasks (Broadfoot, 1979; Crossley, Broadfoot, Schweisfurth, 2007). P. Broadfoot also cites data showing that, for example, tests assessing students' thinking skills in problem-solving and finding innovative solutions had a positive impact on their academic achievement and led to increased cognitive interest in the learning process itself.

Finally, the purpose of testing may be to evaluate the quality of the tests themselves, their diagnostic and differentiating ability, validity, reliability, etc.

Thus, the goals of pedagogical testing may be:
• determining the degree to which learning outcomes comply with the requirements of the Federal State Educational Standard;
• comparing individual academic performance with similar group indicators;
• assessing the dynamics of change in individual and group academic performance;
• creating a replicable database of high-quality test assignments to facilitate teachers' assessment procedures;
• conducting comparative studies of educational quality at the intercountry, national, regional, and institutional levels.

Depending on these goals, basic approaches to testing can be distinguished: norm-oriented, criterion-oriented, and personality-oriented. Norm-based testing is typically used in selective education systems when it is necessary to select a specific group of students (usually those demonstrating the best results), for example, for university entrance exams, subject Olympiads, etc., or, conversely, to identify a so-called "risk group" — students demonstrating consistently low learning outcomes. This approach is also used to compare the academic achievements of one student with others in the content area defined by the test tasks, or between groups of students (for example, those studying the same subjects at different universities, control and experimental groups in a pedagogical experiment, etc.). Furthermore, norm-based testing can be used to compare the difficulty of test tasks, the degree to which students master various sections of the educational material, etc. The results of norm-based testing are typically based on statistical data obtained from a sufficiently representative sample of test subjects.

Criterion-based testing is much more commonly used in pedagogical assessment and is primarily intended to assess students' mastery of educational material, determine compliance with standards, assess students, and assign academic grades. Each value on the assessment scale (level of academic success) corresponds to a specific test result, usually recorded as the percentage of tasks completed correctly. Interpreting test scores requires

a detailed specification of the content area being assessed. Another possible model for constructing a criterion-based test is one in which each position on the assessment scale corresponds to test tasks of varying difficulty, allowing one to assess the student's compliance with a given level (criterion requirement).

Designing tasks for criterion-referenced tests requires more detailed and precise rules and specification of the content area being tested. In international literature, these are often referred to as domain-referenced tests. So-called mastery tests also exist, which are criterion-referenced tests. They are used to differentiate between students who have mastered and those who have not. Thus, the ability of a test to provide information about students' mastery of a specific content area is what makes it criterion-referenced, rather than the use of a threshold or other differentiation criteria.

Student-based testing aims to assess student progress and is widely used in education quality monitoring systems. This may involve re-taking the same test to assess the strength of learning (so-called "retained knowledge") or a test with increasingly challenging tasks to assess a student's progress from one level of learning to a higher level, for example, from reproductive (memorization and recall) to productive (understanding and application) or creative (creation of new knowledge, solving non-standard problems).

Overall, pedagogical testing is an effective means of informing all stakeholders about the quality of education. The assessment of educational achievements through testing, a systematically repeatable process, includes four components: the purpose of testing, the subject and object of assessment, data processing methods, and the users of the test results.

While the effectiveness of testing systems for assessing educational quality is generally recognized, researchers and specialists note a number of problems that reduce the quality of test-based assessment:

1. Excessive reliance on testing significantly increases the workload of teachers, which often leads to a negative attitude toward the very idea of testing and causes latent or overt resistance to the introduction of test-based assessment systems.

2. Difficulties exist with validating tests, evaluating them, and aligning their diagnostic capabilities. Difficulties in validating assignments and tests in general also manifest themselves in the tests' low adaptability to rapidly changing conditions.

3. Tests are often criticized for deviating from a criterion-based approach to determining knowledge quality and instead implementing a norm-oriented test design model, i.e., they serve a supervisory rather than a developmental and supportive function.

4. The difficulty of interpreting test results, i.e. what should be used to assess educational quality — either raw test scores or a complex value-added indicator based on a multi-level analysis of student (group, or educational organization) results along with their contextual characteristics.

5. There are problems with comparing and accounting for data obtained from ongoing monitoring and final testing data.

One of the main problems with pedagogical testing and the main argument of critics of the widespread use of tests in educational quality assessment is the question of trust in the results and the test's ability to meaningfully assess the quality of knowledge. These test characteristics directly depend on their reproducibility and can be statistically assessed; i.e., test quality requirements dictate the need for statistical analysis of test items. Such analysis is warranted in situations where a well-prepared student may, for various reasons, fail a test item, or, conversely, when a weak student randomly answers all test

questions correctly. Furthermore, each test item, as it is used, acquires certain statistics, allowing conclusions to be drawn about its content and functional validity and reliability, depending on the proportion of test takers who successfully or unsuccessfully complete it. Generally, these statistics should conform to a normal distribution. Failure to meet this condition can (should) serve as grounds for a serious and in-depth analysis of the test item's quality. It should be noted that with this approach, the probability of traditional errors in pedagogical assessment ("errors of generosity," "flexible standard," "halo error," "central tendency," etc.) is quite low, although it is not excluded (for example, intentional or unintentional simplification (complication) of test tasks by the teacher, the use of tasks with "quasi-choice," subjective determination of the rule for converting test results into an academic grading scale, etc.).

Test Development Technology

Test development technology is of particular interest to researchers and specialists. Since the Law on Education requires every graduate to demonstrate achievement of educational standards in order to receive a certificate or diploma, a simple and reliable system for assessing the degree to which these standards have been achieved is required. A balance must be found between authenticity and the effectiveness of assessment. To ensure this balance, various approaches within classical and modern test theory can be used, such as common anchor techniques, discriminant analysis (DIF — Differential Item Functioning), and others.

The main approaches to test development are the following:

1. Correspondence to the program. Test content should cover the program content as closely as possible.

2. Focus on best teaching and learning practices.

3. Verification (assessment) of all possible levels of achievement.

4. Consideration of real-world conditions.

One of the key requirements for a test as a measure of academic achievement is its ability to differentiate between students of different levels of preparation based on test results. To reliably establish differences in levels of academic success, the distribution of test item scores should be as even as possible across the entire scale. To achieve this, the test includes both difficult and easy items. Generally, the difficulty of test items can range from 10 to 90%; it is advisable to also include very easy (< 10%) and very difficult (> 90%) items. Easy items maintain student motivation during test completion, while difficult items, provided they meet program requirements, reveal unmastered elements of educational standards. However, their number should not be too large, as their inclusion reduces the test's effectiveness as a measurement tool.

To maintain uniform standards (to ensure equal conditions), the following rules must be observed when designing tests:

1. Test items must be selected in strict accordance with established assessment criteria.

2. Test difficulty must be balanced.

3. A thorough analysis of test content and its relevance to the material covered is necessary.

4. Setting a passing score (if necessary) must be based on test performance statistics for each level.

Achievement levels description is the starting point for selecting the content to be tested, developing individual items and their assessment scheme, and establishing achievement standards. A description of achievement levels, representing what a "typical" student of a given age knows and can demonstrate in a given subject, must be provided in the test specifications.

A way to improve test reliability is to use a model in which, given a sufficiently large

number of test items, they can be divided into two parts. In this case, a consistent decline in test scores in the second part of the test may indicate student fatigue, decreased concentration, etc. Therefore, a more productive method would be to evaluate the statistical results of even and odd tasks. For students with similar knowledge, the results of these two groups of tasks should not show statistically significant differences, while they should differ for students of different levels.

Conclusion

The strategic priorities of modern education and its development trends require changes in the system of educational quality assessment. A promising direction for improving the system of academic achievement assessment is undoubtedly the development of pedagogical testing as a scientific discipline and area of pedagogical practice. The following principles should serve as paradigmatic foundations for the development of effective pedagogical testing systems:

The paradigmatic basis for the development of effective pedagogical testing systems should be the transition

• from assessing individual knowledge and skills to comprehensive and interdisciplinary assessment, from assessing subject-specific knowledge to measuring of meta-subject and sub-subject outcomes and universal learning activities;

• from a one-time assessment to assessing the dynamics of academic achievement (added value, progress);

• assessing individual academic results to assessing the ability to work in a team and reach a collective decision;

• assessing learning outcomes exclusively to analyzing the entire set of factors influencing them;

as well as the shift to authentic assessment, using contextualized tasks, tasks with multiple correct answers, and providing students with the opportunity to actively construct the content of their answers.

In conclusion, we would like to present some conceptual ideas that should be considered when implementing pedagogical testing:

1. No single method for assessing the knowledge quality is universal; a promising approach in this regard is the principle of complementarity.

2. In the process of pedagogical measurement, it is difficult to assess the entire set of qualitative characteristics, so it is important to determine exactly which educational achievement characteristics are the subject of assessment. Test results provide only limited (incomplete) information about the quality of knowledge, and therefore, although a test ensures relative assessment objectivity, it cannot be used to draw definitive conclusions about the level of academic success.

3. The quality of knowledge is a learning outcome, which is largely determined by the teacher's personality, the conditions of the educational process, and other factors. Interpreting test results without taking these factors into account cannot be truly objective.

4. Any measurement inevitably includes measurement error, which requires an analysis of the test results statistics, both for the test and for the student.

5. The lack of a clearly defined unit of measurement for knowledge quality can lead to misinterpretation of test results (e.g., receiving the same percentage of correct answers (points) for completing tasks of varying difficulty levels).

6. The indicators of knowledge quality measured by the test should correlate with other observable characteristics of the student (e.g., their current grades, class activity, etc.).

7. Test results are significantly influenced by the student's motivation and their psychological

state. This requires the creation of a means of assessing the student's readiness for testing before it begins.

8. Pedagogical monitoring, based on psychometrics and test theory, should eventually be replaced by a new type of measurement that is more focused on assessing the student's progress, providing assistance and academic counseling, rather than on verifying their compliance with a specific standard.

References

Avanesov, V.S. (2004). Fundamentals of pedagogical measurement theory. *Pedagogical Measurements*, (1), 15–21. (In Russ.).

Avanesov, V.S. (2008). The problem of the objectivity of pedagogical measurements. *Pedagogical Measurements*, (3), 3–40. (In Russ.).

Bitinas, B.P. (1993). Pedagogical diagnostics: essence, functions, prospects. *Soviet Pedagogy,* (2), 10–15. (In Russ.).

Broadfoot, P. (1979). *Assessment, Schools and Society (Contemporary Sociology of the School)*. Methuen Publishing Ltd. https://doi.org/10.4324/9780203808702 EDN: SIZRWX

Crossley, M., Broadfoot, P., & Schweisfurth, M. (Eds.). (2007). *Changing Educational Contexts, Issues and Identities: 40 Years of Comparative Education* (1st ed.). Routledge. https://doi.org/10.4324/9780203961995

Ganin, V.V., & Ganina, N.V. (2012). The formation of pedagogical testing in Russia (the end of the XIX -beginning of the XXI century). *Knowledge. Understanding. Ability,* (4), 127–132. (In Russ.). EDN: PXADFD

Ingenkampm, K. (1991). *Pedagogical diagnostics*. Pedagogika publ. (In Russ.).

Mayorov, A.N. (1997). *Tests of school achievements: design, implementation, use.* Education and Culture. (In Russ.). EDN: FBDTPM

Neiman, Yu.M., & Khlebnikov, V.A. (2000). *Introduction to the theory of modeling and parameterization of pedagogical tests*. Prometheus. (In Russ.).

Pashnin, S.V. (2009). Pedagogical tests in the modern educational process, their classification. *Modern Higher Education: An Innovative Aspect*, (4), 34–41. (In Russ.). EDN: LPBQNX

Paskhover, I.L. (2011). Pedagogical test as a tool for the system of assessment and quality control of education. *Magister Dixit*, (4), 175–185. (In Russ.). EDN: OXAPON

Verbitsky, A.A., & Kreslavskaya, E.E. (2012). Testing in education: problems and prospects. *Pedagogy,* (8), 3–13. (In Russ.). EDN: PGWLYB

Научная статья

Педагогический тест как инструмент оценки качества образования

С.Ю. Трапицын[1] ✉ , **Н.В. Поморцева**[2], **М.Н. Куновски**[2]

[1]Российский государственный педагогический университет им. А.И. Герцена, *Санкт-Петербург, Российская Федерация*

[2]Российский университет дружбы народов, *Москва, Российская Федерация*
✉ trapitsin@gmail.com

Аннотация. Представлены результаты анализа диагностических возможностей и областей применения педагогических тестов как инструмента оценки качества образования. Особое внимание уделено требованиям к тестам, обеспечивающим объективное и надежное измерение результатов образования, сформулированы критерии оценки качества тестовых систем. Показано, что эффективность применения педагогических тестов достигается благодаря четкой постановке целей тестирования, осознанному и ар-

гументированному определению предмета оценивания, соблюдением правил отбора и конструирования тестовых заданий, а также применением релевантных методов обработки и интерпретации результатов тестирования. Представлена типология тестов, дана характеристика целевого предназначения, специфики построения и областей применения различных типов тестов. Определены перспективы развития педагогической тестологии.