REGULATIONS

on the procedure for the development, review, approval, use, and analysis of test items for assessing student knowledge at Jalal-Abad International University

Section 1. GENERAL PROVISIONS

1.1. Purpose and Scope

1. These Regulations establish uniform requirements for the development, review, approval, use, storage, and analysis of test items for ongoing, midterm, and final assessment of students' knowledge, as well as for conducting knowledge surveys and assessing students' residual knowledge at Jalal-Abad International University (hereinafter referred to as the University, JAIU).
2. These Regulations govern the procedure for organizing the assessment of students' knowledge in the University's electronic information and educational environment using the LMS eBilim / AITest system.
3. These Regulations apply to departments, faculties, the Academic and Methodological Council, the Academic and Information Department, and other structural units involved in the development, administration, and analysis of testing.
4. These Regulations apply to all forms of instruction and levels of training offered at the University, unless otherwise specified by separate local regulations.
5. The requirements of these Regulations are mandatory for faculty members, department chairs, deans' offices, members of the Academic and Methodological Council, staff of the Academic and Information Department, administrators of the LMS eBilim / AITest system, and other authorized persons.

1.2. Purpose and Objectives

1. The purpose of these Regulations is to ensure uniform, objective, transparent, and manageable requirements for the assessment of students' knowledge.
2. The main objectives of these Regulations are:
1. ensuring the quality of test items;
2. to enhance the objectivity of knowledge assessment;
3. creating and maintaining an up-to-date electronic bank of test questions;
4. regulating the digital testing procedure in the LMS eBilim / AITest system;
5. ensuring psychometric, procedural, and technical analysis of test results;
6. creating conditions for academic integrity and the protection of test materials.

Section 2. REGULATORY BASIS AND TERMS

2.1. Regulatory Basis

1. These Regulations have been developed based on:
1. the Law of the Kyrgyz Republic "On Education";
2. the state educational standards of the Kyrgyz Republic;

3. the Charter of JAIU;
4. the University's internal regulations governing the organization of the educational process, assessment of knowledge, the use of digital educational systems, and the assurance of educational quality;
5. generally accepted approaches to pedagogical assessment and testing.

2.2. Key Terms

1. The following key terms are used in these Regulations:

Test item — a unit of a test consisting of a question, answer options, and the correct answer key.

Test — a set of test items united by a common assessment objective, structure, administration parameters, and grading rules.

Electronic test item bank — a systematic collection of approved test items stored in the LMS eBilim / AITest IS or in the University's integrated secure repository.

Multiple-choice test item — a closed-ended test item requiring students to select one or more correct answers from the options provided.

MCQ (Multiple Choice Question) — a type of multiple-choice test question requiring the selection of one correct answer from the options provided.

MSQ (Multiple Select Question) — a type of multiple-choice test question requiring students to select several correct answers from the options provided.

Distractor — an incorrect but plausible answer option in a test question.

Importing test questions — uploading tests to the eBilim LMS / AITest IS from a file in a specified format.

Situational (case-based) test item — a test item based on a description of a professional, clinical, laboratory, or other practice-oriented situation, requiring the learner to analyze the presented data and select one or more correct solutions.

A clinical situational test item is a type of situational (case-based) test item based on a description of complaints, medical history, symptoms, physical examination findings, and laboratory and diagnostic test results, designed to assess the learner's clinical reasoning.

Task difficulty (p) — the proportion of students who answered the task correctly.

Test discrimination — the ability of the task to distinguish between students with higher and lower levels of preparation.

Test reliability — a measure of the internal consistency and stability of test results.

Test validity — the degree to which the interpretation and use of test results are justified and correct in accordance with the assessment objectives, the subject matter, and the learning outcomes being assessed.

Test objectivity — the degree to which test results are independent of the instructor's subjective influence, random external factors, and inconsistencies in the administration procedure.

Test content representativeness — the degree to which the test corresponds to the core content of the discipline, the test specifications, and the key learning outcomes.

Test standardization — ensuring uniform testing conditions, including identical instructions, time limits, access procedures, system settings, and grading criteria for all students in a given category.

Summative assessment — the evaluation of learning outcomes at a specific stage, aimed at determining the extent to which students have achieved learning outcomes and mastered a subject, module, or educational program as a whole.

Formative assessment — assessment conducted during the learning process to provide feedback, identify gaps in preparation, and subsequently adjust teaching and student learning activities.

Psychometric analysis — a statistical analysis of the quality of test items and the test as a whole based on students' performance results.

Procedural analysis — an analysis of test administration parameters, including date, time, duration, number of attempts, streams, groups, and other organizational parameters.

Technical analysis — an analysis of the system's log data, including IP address, device, system events, technical failures, and other parameters of the digital testing procedure.

Primary attempt — the first scheduled attempt to take the test within the established schedule.

Reserve attempt — a retake of the test scheduled by decision of an authorized person or commission in the event of a technical failure or for other established reasons.

Knowledge assessment — a form of assessment conducted to determine the extent to which students have mastered key topics, sections, or the discipline as a whole at a specific stage of instruction or upon completion of the discipline.

Retention assessment—a form of testing conducted a certain period of time after the completion of a discipline or module to evaluate students' retention of foundational knowledge, understanding of key connections, and professionally and safety-relevant content elements.

Critically significant (fatal) test items — items in which a student's error could lead to or indicate an inability to prevent a patient's death, severe irreversible disability, failure to recognize a life-threatening condition, or incorrect performance of an emergency procedure or intervention.

LMS eBilim / AITest IS — the University's official digital system used for assigning, conducting, recording, and analyzing tests.

Section 3. REQUIREMENTS FOR THE STRUCTURE AND CONTENT OF TEST ITEMS

3.1. Structure of a Test Item

1. Each test item must include:

1. a clearly formulated task prompt;
2. answer choices;
3. the correct answer key;
4. if necessary—the question identifier, difficulty level, subject area, and other administrative details.

1. The question stem must be:

1. unambiguous;
2. grammatically correct;
3. logically complete;
4. free of hidden clues;
5. appropriate to the students' level of preparation.

1. Answer choices must:

1. be grammatically consistent with the main text of the task;
2. be comparable in form and length;
3. exclude obvious clues;
4. include one correct answer, unless otherwise specified by the test format.

1. The following is not permitted:

1. the use of factually incorrect questions;
2. duplicating answers;
3. having multiple correct answers in a "single answer" format;
4. using distractors that do not work or are clearly absurd;
5. phrasing that allows for multiple interpretations without specific clarification.

3.2. Content Requirements

1. Test items must:

1. align with the course syllabus;
2. reflect learning outcomes, topics, and key content elements;
3. correspond to the students' level of preparation;
4. ensure the assessment of knowledge, understanding, application, and analysis of the course material.

1. When developing tests, the department must ensure that test items align with:

1. the course objectives;
2. the thematic structure of the course;
3. the stated competencies and learning outcomes;
4. the level and type of assessment.

1. The test item bank should not include items that test secondary, random, or methodologically insignificant information unrelated to the main learning outcomes.

Section 4. TEST PLANNING AND SPECIFICATION

1. Tests are developed based on the test specification.
2. The test specification must define:

1. the subject;
2. the module or section;
3. the type of assessment;
4. the number of questions;
5. the distribution by topic;
6. distribution by difficulty level;
7. if necessary, distribution by Bloom's taxonomy levels;
8. test duration;
9. number of attempts;
10. responsible persons.

1. The test specification is approved by the department before the test materials are uploaded to the LMS eBilim / AITest system.
2. The test may not be used without an approved specification or another equivalent document established by the University.
3. For the purpose of assessing knowledge and verifying retained knowledge, tests are developed based on the approved specifications, taking into account the course objectives, learning outcomes, the scope of the material covered, the students' level of preparation, and the need to verify the preservation of the core content of the course.
4. Tests for midterm and final assessments should not completely duplicate the content of ongoing, interim, or final assessments; rather, they should focus not on the rote reproduction of minor details, but on verifying foundational knowledge, well-established cause-and-effect relationships, typical professionally relevant scenarios, and key learning outcomes.

5. For medical disciplines, midterm and final exams must include tasks that assess the recognition of clinically significant conditions, the prevention of gross errors, and an understanding of safety-critical aspects of the discipline within the scope of the students' level of training.
6. When conducting knowledge assessment and residual knowledge assessment, the department is required to ensure subsequent analysis of results by student, group, topic, test item, common errors, and corrective measures.
7. For medical disciplines, midterm and final exams should be primarily aimed at verifying the retention of basic and clinically significant knowledge, rather than the reproduction of random, rare, or methodologically minor details.
8. Every test used at the University must possess sufficient validity, reliability, objectivity, and content representativeness in relation to the objectives of the specific type of assessment.
9. Test validity is ensured by:

1. alignment of the test content with the course syllabus;
2. alignment of the test with the test specifications;
3. alignment of learning outcomes with the competencies being assessed;
4. the alignment of the difficulty level with the students' level of preparation;
5. the alignment of the task format with the assessment objectives;
6. the use of tasks that truly assess the required knowledge, understanding, application, and analysis.

1. Test reliability is ensured by:

1. a sufficient number of tasks;
2. the internal consistency of the test;
3. the consistency of administration parameters;
4. consistency in instructions and scoring;
5. the use of high-quality items with acceptable psychometric properties;
6. exclusion of defective, ambiguous, and randomly functioning questions.

1. The objectivity of the test is ensured by:

1. standardization of the testing procedure;
2. uniform parameters for time, number of attempts, and scoring;
3. the use of automated scoring in the eBilim LMS / AITest IS;
4. prevention of unauthorized interference with test results;
5. compliance with uniform requirements for the organization and conduct of assessments.

1. Summative assessment is used to record learning outcomes for a topic, module, discipline, or other completed stage and is primarily used in modular, midterm, and final assessments.

2. Formative assessment is used to provide feedback to students and instructors, identify gaps in preparation, and subsequently adjust the educational process. Elements of formative assessment may also be used when analyzing test results, even if the test itself is summative in nature.

Section 5. QUANTITATIVE STANDARDS, DISTRIBUTION BY BLOOM'S LEVELS AND TASK FORMATS

5.1. General Principles for Compiling a Test Item Bank

1. The size of the test item bank for a discipline is determined by:

1. the course load in credits;
2. the number of topics and modules;
3. the type of assessment;
4. the need to ensure test variability;
5. the need to update and rotate test items;
6. the use of tests across multiple academic groups, streams, and languages of instruction.

1. The test item bank must be sufficient not only for administering a single test, but also for:

1. generating different versions;
2. reducing the repetition of questions;
3. ensuring the objectivity of assessment;
4. conducting retakes and backup attempts;
5. regularly updating and psychometrically selecting questions.

1. An insufficient number of test items in the bank leads to the following risks:

1. students encountering the same questions repeatedly;
2. a decrease in the objectivity of results;
3. an increased likelihood of students memorizing specific answers rather than demonstrating their knowledge;
4. the inability to rotate tests;
5. limitations in psychometric analysis.

5.2. Standards for the size of the test item bank depending on the complexity of the discipline

1. The minimum recommended size of the test item bank for a discipline is determined as follows:

Course workload	Number of hours	Minimum recommended size of the test item bank	Recommended number of questions per test

Course workload	Number of hours	Minimum recommended size of the test item bank	Recommended number of questions per test
	30 hours	100–150	20–25
	60 hours	200–250	25–30
	90 hours	250–350	30
	120 hours	300–400	30–40
	150 hours	350–500	30–40
	180 hours or more	at least 500	determined by the test specifications

1. For disciplines with multiple modules, multiple assessment stages, or instruction in multiple languages, the size of the question bank is increased based on:

1. the need for separate coverage of topics;
2. the need for equivalent language versions;
3. testing in multiple groups;
4. the need for backup and retesting.

1. For courses with a large number of students, it is recommended to maintain a ratio of at least 5:1 between the total number of items in the question bank and the number of questions assigned to a single student on a single test.
2. A ratio of test questions in the bank to a single test version ranging from 5:1 to 10:1 is considered optimal.
3. When creating a test in the eBilim LMS / AITest IS, it is recommended to ensure:

1. random selection of questions;
2. variety in test sets;
3. question rotation;
4. reduction in the repetition of identical test combinations.

5.3. Justification of Standards for the Size of the Test Item Bank

1. The number of credits reflects the overall workload of the course, the scope of the content, the number of topics, and the depth of mastery of the material. Consequently, an increase in the course's workload must be accompanied by a proportional increase in the size of the test item bank.
2. Establishing a minimum size for the test item bank is necessary to:

1. cover all sections of the discipline;
2. distribute questions by level of difficulty;
3. include questions at different cognitive levels;
4. create separate sets for modular, midterm, and final assessments;
5. preventing excessive repetition of questions.

5.4. Distribution of test items according to Bloom's taxonomy levels

1. When developing tests, the department must ensure that tasks are distributed across cognitive levels in accordance with the course objectives, the type of assessment, and the students' level of preparation.
2. For modular and midterm assessments, the following indicative distribution of tasks across Bloom's taxonomy levels is recommended:

Level	Content	Recommended proportion
		20–25%
		25–30%
		20–25%
		10–15%
		5–10%

1. For final assessments, it is recommended to shift the focus toward higher cognitive levels by increasing the proportion of tasks involving application, analysis, and professionally oriented decision-making.
2. In medical, engineering, and professionally applied disciplines, the proportion of tasks at the Apply and Analyze levels should be no less than the proportion of simple reproductive tasks.
3. For final assessments in medical disciplines, particularly in the "General Medicine" specialty, the following approximate distribution of tasks according to Bloom's taxonomy levels is recommended:

Level	Recommended proportion
	no more than 10–15%
	15–20%
	30–35%
	25–30%
	10–15%

5.4.1. Distribution of Test Items by Difficulty Level

1. When developing tests, the department is required to ensure that test items are distributed by difficulty level: easy, medium, and difficult.
2. The distribution of questions by difficulty level is determined by the type of assessment, the students' level of preparation, the course content, the number of credits, and the test specifications.
3. Easy questions are those primarily designed to assess basic knowledge, terminology, structures, fundamental concepts, and standard facts.

4. Medium-level tasks are defined as those designed to assess understanding, interpretation, comparison, and application of knowledge in a typical educational, professional, or clinical situation.
5. Difficult tasks are defined as those designed to assess analysis, integration of knowledge, clinical or professional decision-making, recognition of complex cause-and-effect relationships, and the resolution of highly complex situational problems.
6. For modular assessment, the following distribution of tasks by difficulty level is recommended:

1. easy — 25%;
2. moderate — 40%;
3. difficult — 20%;
4. independent study assignments — 15%.

1. For the final assessment, the following distribution of tasks by difficulty level is recommended:

1. easy — 15%;
2. medium — 40%;
3. difficult — 30%;
4. Self-study assignments — 15%.

1. For the midterm exam, the following distribution of questions by difficulty level is recommended:

1. easy — 20%;
2. medium — 45%;
3. difficult — 20%;
4. Self-study assignments — 15%.

1. For the residual knowledge assessment, the following distribution of tasks by difficulty level is recommended:

1. easy — 25%;
2. medium — 45%;
3. difficult — 15%;
4. Self-Study Program (SSP) questions — 15%.

1. With 40 questions in a single test, the approximate distribution is as follows:

for module assessments:

1. easy — 10;
2. medium — 16;
3. difficult — 8;
4. SRS assignments — 6.

For the final assessment:

1. easy — 6;
2. medium — 16;
3. difficult — 12;
4. Self-study assignments — 6.

For the midterm exam:

1. easy — 8;
2. medium — 18;
3. difficult — 8;
4. Self-study assignments — 6.

For residual knowledge assessment:

1. easy — 10;
2. medium — 18;
3. difficult — 6;
4. Independent study assignments — 6.

1. Self-directed learning (SDL) assignments are included in the test specifications as a separate content category and are distributed across difficulty levels in accordance with the course objectives.
2. For medical disciplines, the final assessment should not be limited to predominantly easy tasks; the main content of the test should consist of medium and difficult tasks focused on application and analysis.
3. The specific distribution of tasks by difficulty level and by SRS is approved by the department in the test specification.

5.5. Test Question Formats

1. The following main formats may be used for testing at the University:

1. MCQ — multiple-choice question with one correct answer;
2. MSQ — a task with multiple correct answers;
3. case-based test questions;
4. clinical case-based test questions;
5. other formats, if supported by the eBilim LMS / AITest IS and approved by local University regulations.

1. The primary format for ongoing and midterm assessments is the MCQ, as it ensures:

1. unambiguous assessment;
2. automated processing;
3. convenience of statistical analysis;

4. comparability of results.

1. The MSQ format is used in cases where the content of the question objectively implies the presence of more than one correct answer and when it is necessary to verify:

1. comprehensive understanding of the topic;
2. the ability to distinguish between several correct characteristics;
3. systematic thinking;
4. analysis of a clinical or professional situation.

1. The use of MSQ is permitted only if:

1. the question clearly indicates the need to select multiple answers;
2. all correct and incorrect options are formulated correctly;
3. the assessment method is established in advance;
4. students are familiar in advance with the rules for completing this type of task.

1. At the University, multiple-choice questions are defined as closed-ended questions requiring the selection of one or more correct answers from the options provided.
2. The main types of multiple-choice questions include:

1. MCQ — a task requiring the selection of one correct answer;
2. MSQ — a task requiring the selection of several correct answers.

1. The use of the general term “multiple-choice question” is permitted only as a general designation for a group of test questions. The test specification, test item bank, grading criteria, and instructions for students must specify the specific type of question: MCQ or MSQ.

5.6. Features of MCQ and MSQ Use

1. MCQ-format questions are used primarily to assess:

1. knowledge;
2. understanding;
3. standard application of material;
4. selection of the single most correct solution in a professional or clinical situation.

1. In MCQ questions:

1. only one correct answer is allowed;
2. the remaining options must be plausible distractors;
3. it is not permissible for the correct option to be easily guessed.

1. MSQ questions are used to assess:

1. comprehensive understanding;
2. the ability to identify a set of correct characteristics;
3. the ability to eliminate partially incorrect statements;
4. the professional selection of several correct solutions.

1. In MSQ questions:

1. it must be explicitly stated: “Select several correct answers”;
2. the number of correct answers may be fixed or variable, if specified in the instructions;
3. the scoring method must be consistent within a single test.

1. When using MSQ, it is recommended to use one of the following scoring methods:

1. full credit only if all correct answers are selected and there are no incorrect ones;
2. partial scoring for a partially correct selection;
3. penalty points for selecting an incorrect option, if specified in the test guidelines.

1. The MSQ scoring model used must be approved in advance by the department and communicated to the students.
2. The MSQ format should not be overused. In large-scale midterm exams, the test should be based on the MCQ format, supplemented by a limited number of well-designed MSQs.

5.7. Mandatory Use of Situational (Case-Based) Test Items for Final Assessments in Medical Disciplines

1. At the University, given the predominant cohort of students majoring in “General Medicine,” situational (case-oriented) test items must be included in final assessments for medical disciplines.
2. Situational (case-based) test items are the primary form of assessment in final examinations for medical disciplines, as they evaluate:

1. the ability to apply theoretical knowledge in a clinical setting;
2. the ability to analyze complaints, medical history, symptoms, and examination results;
3. the ability to select a diagnostic, therapeutic, preventive, or tactical strategy;
4. readiness to make professionally sound decisions.

1. In final assessments in medical disciplines, tests must not be limited to reproductive tasks aimed solely at recalling terms, definitions, and isolated facts.
2. In final exams for medical disciplines, the main content should consist of tasks at the Apply, Analyze, and, if necessary, Evaluate levels of Bloom’s taxonomy.
3. Situational (case-based) test items in the final assessment may be used in the following formats:

1. a situational task in the MCQ format—selecting the single most correct answer;
2. a situational task in the MSQ format—selecting multiple correct answers, if the content of the case objectively requires a comprehensive selection.

1. For medical disciplines, it is recommended that situational (case-oriented) test items constitute at least 60% of the final exam, and for specialized clinical disciplines, at least 70–80% of the total number of items.
2. For courses in which final assessment is linked to the development of clinical competencies, the department is required to ensure that the test specifications mandate the inclusion of situational (case-based) test items.
3. Situational (case-based) test items must be aligned with:

1. with the learning outcomes of the discipline;
2. with the list of professional competencies to be developed;
3. with the content of the practical skills to be assessed;
4. with the structure of the OSCE, OSPE, and other forms of final assessment, if such forms are provided for.

1. In disciplines that include OSCE and OSCE, the use of multiple-choice test items of two main types is permitted:

1. MCQ — questions with a single correct answer;
2. MSQ — questions requiring the selection of multiple correct answers.

1. In subjects with OSCE and OSCE-P, the MCQ format is used primarily to assess:

1. the most likely diagnosis;
2. the single most correct clinical decision;
3. the most appropriate next step;
4. the interpretation of a specific test result.

1. In subjects with OSCE and OSCE-P, the MSQ format is primarily used to assess:

1. the ability to identify a set of clinical signs;
2. the selection of several indications or contraindications;
3. selecting several steps in an algorithm;
4. the identification of several correct actions in a clinical or practical situation.

1. For disciplines with OSCE and OSCE-P, MCQ and MSQ tasks serve as a supplementary assessment tool and do not replace practical stations, checklists, evaluation sheets, or other forms of practical skills assessment.

5.8. Psychometric Requirements

1. The following indicative standards are applied when analyzing the quality of test items:

Item difficulty (p):

1. 0.30–0.80 — optimal;
2. 0.20–0.29 — acceptable, but the item is difficult;
3. 0.81–0.90 — acceptable, but the item is easy;
4. < 0.20 — needs revision;
5. 0.90 — needs to be revised.

Discriminatory power:

1. ≥ 0.30 — good task;
2. 0.20–0.29 — acceptable;
3. 0.10–0.19 — requires revision;
4. < 0.10 — subject to revision;
5. negative value — requires immediate verification of the answer key and content review.

Effectiveness of distractors:

1. A distractor is considered weak if it is selected by fewer than 5% of students, provided the sample size is sufficient.

Test reliability:

1. ≥ 0.85 — high;
2. 0.70–0.84 — acceptable;
3. 0.60–0.69 — low;
4. < 0.60 — insufficient.

1. The department is required to ensure:

1. the creation of a bank of test questions in a volume no less than that recommended for credits;
2. the distribution of questions by topic and cognitive level;
3. the use of predominantly correct MCQs;
4. the inclusion of MSQs when methodologically justified;
5. the mandatory inclusion of situational (case-based) test questions in the final assessment for medical disciplines;
6. regular rotation and updating of the question bank;
7. revision of items based on the results of psychometric analysis.

5.9. Critically Important (Fatal) Test Items

1. For medical disciplines, tests for modular and final assessments, midterm exams, and residual knowledge assessments must include critically important (fatal) test items designed to assess the recognition of life-threatening conditions, the

prevention of gross diagnostic and tactical errors, and adherence to safety-critical principles of professional practice.

2. Critically significant (fatal) test items are those in which a student's error may lead to or indicate an inability to prevent:

1. the patient's death;
2. severe irreversible disability;
3. failure to recognize a life-threatening condition;
4. incorrect performance of an emergency procedure or manipulation.

1. Critically significant (fatal) test items are identified separately by the department and are not considered a distinct difficulty level alongside easy, medium, and difficult items.
2. When planning a test, the department must consider critically significant (fatal) items as a separate safety category that is superimposed on the general distribution of items by difficulty level.
3. A critically important (fatal) test item may fall into the easy, medium, or difficult difficulty level depending on its content, but must be accounted for separately in the test specification and in the subsequent analysis of results.
4. The recommended proportion of critically important (fatal) test items is:

1. for modular assessments — up to 10% of the total number of items given to the student;
2. for final assessments — 10–15%;
3. for a knowledge assessment — up to 10%;
4. for residual knowledge assessment — up to 10%.

1. With 40 questions in a single test, the recommended number of critical (failing) items is:

1. for module-based assessment — 2 to 4;
2. for final assessment — 4 to 6;
3. for a midterm exam — 2 to 4;
4. for residual knowledge assessment — 2 to 4.

1. For lower-level courses, including fundamental medical disciplines, the department is required to align the number and content of critically significant (fatal) test questions with the students' level of preparation and prevent their unjustified complexity.
2. Critically significant (fatal) test items are subject to mandatory separate analysis following the completion of testing.
3. Based on the results of the analysis of critically significant (fatal) tasks, the department is required to:

1. identify typical critical errors;
2. identify problem areas;

3. take corrective action;
4. if necessary, schedule a retest on the relevant sections.

Section 6. PROCEDURE FOR THE DEVELOPMENT, REVIEW, PILOT TESTING, AND REVISION OF TEST ITEMS

1. Test items are developed by the department's faculty in accordance with the course syllabi and the approved test specifications.
2. Each test item is subject to internal content and technical review prior to use.
3. The department reviews test items based on the following criteria:

1. alignment with the course content;
2. correctness of wording;
3. presence of a single correct answer or a correctly formulated multiple-choice format;
4. quality of distractors;
5. absence of errors;
6. alignment with the students' level of preparation.

1. Based on the results of the review, the test item:

1. is approved for use;
2. is sent for revision;
3. is rejected.

1. The pilot testing of new test items is conducted in accordance with the procedures established by the department, the Academic and Methodological Council, and the University's local regulations.
2. Based on the results of the pilot testing and psychometric analysis, a decision is made:

1. to keep the item in the current bank;
2. revise;
3. replace it;
4. to archive it;
5. to delete it.

1. Each task can have one of the following statuses:

1. draft;
2. under review;
3. approved;
4. archived;
5. deleted.

Section 7. PROCEDURE FOR IMPORTING AND USING TESTS IN LMS eBilim / AITest

1. Test questions can be entered into the system manually or uploaded to LMS eBilim / AITest IS by importing from a file in the specified format.
2. Tests are imported based on materials approved by the department.
3. Before uploading a test to the system, the following must be checked:

1. the correctness of the question text;
2. the correctness of the answer choices;
3. the correctness of the answer keys;
4. compliance with specifications;
5. the file's technical suitability for import.

1. After importing the test into the eBilim LMS / AITest IS, the following must be checked:

1. the number of questions uploaded;
2. correctness of the test structure;
3. correct display of text and answer choices;
4. correctness of the answer keys;
5. the correctness of the test's association with the subject, module, language, and other parameters.

1. A test is approved for use only after substantive and technical verification has been completed.
2. The electronic test item bank is maintained, stored, and used in the eBilim LMS / AITest IS or in the University's integrated secure electronic repository.

Section 8. TESTING PROCEDURE

1. For each test in the eBilim LMS / AITest IS, the following must be configured:

1. start date and time;
2. end date and time;
3. duration;
4. number of attempts;
5. number of questions;
6. order of questions;
7. order of answer options;
8. grading criteria;
9. test completion rules.

1. If the system's technical capabilities allow, it is recommended to use:

1. random selection of questions;
2. random shuffling of answer choices;
3. time limits;
4. automatic recording of results;

5. logging of user actions.

1. Test results recorded in the eBilim LMS / AITest IS are considered official records of students' knowledge assessment results, unless otherwise determined by a technical failure report, a decision of the appeals committee, or a university order.
2. When developing the final exam in medical disciplines, the department is required to include situational (case-based) test items as a mandatory component of the assessment.
3. Midterm exams and final exams are conducted according to an approved schedule determined by the University, the faculty, the department, or an authorized unit, with mandatory recording of testing parameters in the LMS eBilim / AITest system.
4. When conducting knowledge assessments and residual knowledge checks, it is recommended to use a random selection of questions from the approved test item bank, random shuffling of questions and answer options, as well as uniform time limits and the number of attempts for students in the same category.

Section 9. USE OF LMS eBilim / AITest IS DATA FOR PSYCHOMETRIC, PROCEDURAL, AND TECHNICAL ANALYSIS

9.1. General Provisions

1. Data generated by the eBilim LMS / AITest IS is used for:

1. psychometric analysis of tests;
2. procedural analysis of test administration;
3. technical analysis;
4. reviewing appeals;
5. recording technical malfunctions;
6. ensuring academic integrity;
7. improving the test item bank.

9.2. Data Sources

1. The following are used for analysis:

1. individual student test protocols;
2. summary of test results;
3. distribution of answers for each question;
4. data from the electronic test item bank;
5. information on the number of attempts;
6. information on the date and time of the test;
7. information on the duration of the test;
8. device data, IP address, and other system parameters;
9. user activity logs, if generated by the system.

9.3. Psychometric Analysis

1. Based on the system data, the University is entitled to calculate:

1. the difficulty of the task;
2. the frequency of answer choices;
3. the effectiveness of distractors;
4. the proportion of ineffective distractors;
5. the mean, median, minimum, and maximum scores;
6. measures of dispersion;
7. pass rate;
8. discriminatory power;
9. reliability indices;
10. analysis of high- and low-performing groups;
11. other test quality indicators.

1. Psychometric analysis is conducted:

1. after final and interim testing;
2. at the discretion of authorized personnel—after modular testing;
3. when revising the test item bank;
4. upon the detection of anomalies or the receipt of substantiated complaints.

1. Based on the results of the analysis, decisions are made:

1. to keep the question in the question bank;
2. revise the wording;
3. correct the distractors;
4. to check the answer key;
5. archive;
6. exclude.

1. Psychometric, procedural, and content analysis must be conducted based on the results of the final assessment, the midterm assessment, and the residual knowledge assessment, as well as, at the department's discretion, based on the results of the module and interim assessments.
2. Based on the results of the midterm assessment and the residual knowledge assessment, the department prepares an analytical report that includes a general description of the results, results by group, results by topic, an analysis of the quality of test items, typical student errors, conclusions, as well as corrective and preventive measures.
3. Based on the test results, the department conducts a separate analysis of the performance on critically important (fatal) test items, identifies typical student errors, determines the level of mastery of safety-critical sections of the discipline, and, if necessary, develops corrective measures.
4. Test validity is verified by:

1. comparing the test with the specification;

2. comparing the tasks with the course syllabus;
3. verifying coverage of key topics and learning outcomes;
4. expert evaluation of the content accuracy of the tasks;
5. verifying that the question formats are appropriate for the type of assessment;
6. analyzing whether the test is overloaded with secondary and methodologically insignificant questions.

1. Test reliability is assessed by:

1. calculating internal consistency indices;
2. analyzing the stability of results;
3. analyzing the distribution of scores;
4. assessing the number of defective items;
5. revising the test if reliability indicators are low.

1. Test objectivity is verified by:

1. monitoring the consistency of testing conditions;
2. analyzing parameters such as time, attempts, flows, and groups;
3. verifying the accuracy of automated grading;
4. analyzing system logs, IP addresses, login times, and other procedural and technical parameters.

1. The representativeness of the content is verified by:

1. analyzing the distribution of tasks by topic;
2. analyzing the distribution by cognitive levels;
3. verifying the presence of questions on key, rather than random, sections of the discipline;
4. verifying whether the test corresponds to the core content of the discipline and the objectives of the specific type of assessment.

1. The summative function of the test is verified by ensuring that the test:

1. is administered at the end of a topic, module, semester, or discipline;
2. is designed to summarize learning outcomes;
3. has a predetermined passing threshold;
4. is used to make a final determination of the level of mastery of the discipline or a part thereof.

1. The formative function of the test is verified by analyzing whether the test results are used:

1. to provide feedback to the student;
2. to identify gaps;
3. to adjust teaching;

4. to re-explain topics;
5. to develop corrective and preventive measures.

9.4. Procedural Analysis

1. To monitor the testing procedure, the following parameters are used: date, time, duration, number of attempts, group, test, stream, and other organizational parameters.
2. Procedural analysis is conducted for the purpose of:

1. monitoring adherence to the schedule;
2. distinguish between primary and backup attempts;
3. analyze results by streams, groups, and languages of instruction;
4. identify circumstances affecting the objectivity of testing.

9.5. Technical Analysis

1. For technical analysis, information regarding the device, IP address, start and end times, session interruptions, re-logins, and other technical parameters is used.
2. This data is used to:

1. confirmation of completion of testing;
2. review of technical incidents;
3. analysis of disputed situations;
4. preparing technical failure reports;
5. operational analysis.

9.6. Academic Integrity

1. Data from the eBilim LMS / AITest IS may be used to identify signs of testing violations.
2. The decision regarding a violation is made by authorized personnel or a committee based on the totality of the data, not automatically by the system.

Section 10. STORAGE, PROTECTION, AND ACCESS TO TEST MATERIALS

1. The electronic test item bank must be stored in the University's secure digital environment.
2. Access to the test item bank, test results, and administrative data is granted only to authorized personnel within the scope of their functional responsibilities.
3. The following is prohibited:

1. transferring usernames and passwords to third parties;
2. unauthorized copying of test materials;
3. changing test keys and parameters without the appropriate authorization;
4. downloading materials outside the established procedure.

1. Test materials and test results must be backed up in accordance with the procedures established by the University.
2. The retention periods for electronic data and documents are determined by the University's local regulations and document management rules.

Section 11. TECHNICAL FAILURES, INCIDENTS, AND APPEALS

1. A technical failure is defined as a confirmed malfunction of the eBilim LMS / AITest IS, equipment, network, or other technical component that prevents the start, continuation, or proper completion of testing.
2. In the event of a technical failure, a report in the prescribed form shall be drawn up.
3. The report shall specify:

1. date, time, and location;
2. subject and type of assessment;
3. the last name, first name, and patronymic of the student or students;
4. a description of the incident;
5. system information;
6. decision made.

1. Based on the report, a decision may be made to:

1. to restore access;
2. to grant a backup attempt;
3. postpone the test;
4. to invalidate the incorrect result;
5. to conduct an additional review.

1. The following are used when reviewing an appeal:

1. the individual test protocol;
2. question-by-question results;
3. information on the date, time, and duration;
4. device and IP address data;
5. user activity log;
6. report on a technical malfunction, if one was prepared;
7. other materials confirming the circumstances of the case.

1. The procedure for filing and reviewing appeals is determined by the University's local regulations.

Section 12. ALLOCATION OF RESPONSIBILITIES

1. Department:

1. organizes the development of test questions;

2. conducts content review;
3. reviews the results of psychometric analysis;
4. makes decisions regarding the revision, replacement, archiving, or exclusion of items.

1. Instructor:

1. develops test items;
2. ensures their alignment with the syllabus;
3. prepares materials for upload;
4. participates in verifying the accuracy of the test after import.

1. Department Chair:

1. organizes the review and approval of materials at the department level;
2. monitors the quality of test items;
3. ensures the timely transfer of materials for upload.

1. Academic and Methodological Council:

1. reviews issues related to the methodological support of test administration;
2. discusses the results of the analysis of test item quality;
3. develops recommendations for improving the structure and content of the test item bank;
4. makes proposals for updating testing approaches at the University.

1. The Academic Information Department and/or the LMS eBilim / AITest IS administrator:

1. ensure the technical upload of tests;
2. configure testing parameters;
3. ensure data export;
4. support technical analysis;
5. participate in the documentation of technical incidents.

1. Dean's Office:

1. monitors compliance with testing schedules;
2. participates in the resolution of disputes within its jurisdiction.

1. Vice Rector for Academic Affairs:

1. exercises general oversight of the implementation of these Regulations.

Section 13. FINAL PROVISIONS

1. These Regulations shall enter into force upon approval by the University Rector.

2. Amendments and additions to these Regulations shall be made in accordance with established procedures upon the recommendation of departments, faculties, the Academic and Methodological Council, the Academic and Information Department, and other authorized bodies of the University.
3. In the event of any conflict between these Regulations and the laws of the Kyrgyz Republic, the laws of the Kyrgyz Republic shall prevail.
4. The Vice Rector for Academic Affairs is responsible for overseeing the implementation of these Regulations.

Adopted by the Academic Council of JAIU
Minutes No. ____ dated "_" _____ 20

Approved
Rector of Jalal-Abad International University
_____ M.R. Narbayev
"_" _____ 20

APPENDIX 1. Form for Expert Evaluation of a Test Assignment

Subject: _____
Topic / module: _____
Question ID: _____
Author: _____

Criterion	Yes	No	Note

Conclusion:
[ ] Accept
[ ] Revise
[ ] Reject

Expert: _____
Date: _____

APPENDIX 2. Summary Table of Psychometric Norms

Indicator	Standard	Solution
Difficulty 0.30–0.80	Optimal	Keep
Difficulty 0.20–0.29	Difficult	Monitor / refine
Difficulty 0.81–0.90	Easy	Monitor / refine
Difficulty < 0.20	Too difficult	Review
Difficulty > 0.90	Too easy	Review
Discriminability ≥ 0.30	Good	Keep
Discriminatory power 0.20–0.29	Acceptable	Monitor
Discriminatory power 0.10–0.19	Low	Revise
Discriminatory power < 0.10	Insufficient	Review
Discriminatory power < 0	Unacceptable	Check answer key / exclude
Distractor < 5%	Weak	Review
Reliability ≥ 0.85	High	Satisfactory
Reliability 0.70–0.84	Acceptable	Acceptable
Reliability 0.60–0.69	Low	Needs improvement
Reliability < 0.60	Insufficient	Review

APPENDIX 3. Technical requirements for preparing a file for import into the eBilim LMS / AITest IS

1. Each question begins with a serial number and a period.
2. Each answer choice is placed on a new line.
3. Answer options are designated as: A), B), C), D), and E) if necessary.
4. The correct answer is marked with a "+" sign before the text of the answer option.
5. Only one correct answer is allowed per question, unless otherwise specified by the system settings.
6. For MSQ-format questions, the technical format provided by LMS eBilim / AITest and approved by the University is used.
7. The following are not permitted:

1. violation of the question structure;
2. missing answer key;
3. duplicate answer choices;
4. technically incorrect characters;
5. violation of the template that prevents import.

1. A test import is required before bulk uploading tests.

APPENDIX 4. Form for Reporting Technical Malfunctions During Testing

Jalal-Abad International University
REPORT on a technical failure during testing

Date: _____
Time: _____
Location: _____
Subject: _____
Type of assessment: _____
Last name, first name, middle name of the student(s): _____
Class: _____

Description of the failure: _____
System data: _____
Decision made:
[ ] Restore access
[ ] Grant a backup attempt
[ ] Reschedule the test
[ ] Void the result
[ ] Other _____

Signatures:
Instructor _____
System Administrator / Educational Institution _____
Dean's Office Representative _____

APPENDIX 5. Recommended template for test titles

Template:
Subject_language_course_module_assessment_type_academic year

Example:
Faculty Therapy_EN_4th year_Module 1_Assessment_2025-2026

Additional examples:
Normal_Clinical_Anatomy_EN_2nd_year_3rd_semester_Module_1_2025-2026
Normal_Clinical_Anatomy_EN_2nd_year_3rd_semester_final_2025-2026
Normal_Clinical_Anatomy_EN_2nd_Year_3rd_Semester_Knowledge_Assessment_2025-2026

APPENDIX 6. Test Specifications

Subject: _____
Department: _____
Instructor: _____
Language: _____
Course: _____
Module / Section: _____
Type of assessment: _____

Assessment type:
[ ] ongoing
[ ] midterm
[ ] final
[ ] knowledge assessment
[ ] residual knowledge assessment

Assignment format: MCQ / MSQ / mixed
Presence of critically important (fatal) questions: yes / no
Number of critical (fatal) questions: _____
Percentage of critical (fatal) questions: _____
Number of questions in the question bank: _____
Number of questions presented to the student: _____
Number of questions: _____
Duration: _____
Number of attempts: _____
Random selection of questions: yes / no

Random shuffling of answer choices: yes / no
Approval date: _____
Date uploaded to the system: _____
Person in charge: _____
Test status: _____

APPENDIX 7. Test specification form for knowledge assessment / residual knowledge assessment

Course Title: _____
Department: _____
Educational program: _____
Course: _____
Semester: _____
Testing language: _____
Type of assessment:
[ ] Knowledge assessment
[ ] Retention assessment
Purpose of the test:
Learning outcomes/topics being assessed:
Number of questions in the bank: _____
Number of questions given to the student: _____
Test duration: _____
Number of attempts: _____
Passing score: _____
Distribution of tasks by type:
1. Easy: _____
2. medium: _____
3. difficult: _____
4. Independent study assignments: _____
5. situational / case-based: _____
6. MCQ: _____
7. MSQ: _____

Critically important (fatal) tasks:
1. Presence: yes / no
2. number: _____
3. Proportion of the test: _____
4. Brief justification for inclusion: _____
5. Subject to separate analysis: yes / no

Questions on clinically and professionally significant sections:
[ ] Yes
[ ] No
Test construction characteristics:
[ ] Random selection of questions
[ ] Random shuffling of answer choices
[ ] Generation of an individual test
[ ] Other _____
Lead developer: _____
Department Chair: _____
Date of approval: _____

APPENDIX 8. Form for an analytical report on the results of a knowledge assessment / residual knowledge assessment

Jalal-Abad International University
ANALYTICAL REPORT
on the results of the knowledge assessment / residual knowledge assessment

Subject: _____
Department: _____
Educational Program: _____
Course: _____
Group / Section: _____
Date: _____
Test language: _____

Type of assessment:
[ ] Knowledge assessment
[ ] Retained knowledge assessment

1. General description of results

Number of students on the list: _____
Number of participants: _____
Percentage of participation: _____
Average score: _____
Median score: _____
Minimum score: _____
Maximum score: _____
Success rate: _____

2. Analysis of results by group

3. Analysis of results by topic

4. Analysis of test item quality

1. Number of tasks that were too easy: _____
2. Number of tasks that were too difficult: _____
3. Number of questions requiring revision: _____
4. Items with ineffective distractors: _____
5. Items with low discriminative power: _____

5. Analysis of critically important (fatal) items

1. Number of critically important items in the test: _____

2. Number of items successfully completed by the majority of students: _____
3. Number of problematic critically important items: _____
4. Typical dangerous errors: _____
5. problematic safety-critical topics: _____

6. Typical student errors

7. Identified problematic topics

8. Conclusions

9. Corrective and preventive measures

[ ] a review of difficult topics
[ ] additional consultation
[ ] Revision of test questions
[ ] Revision of the test bank
[ ] Retest
[ ] Other _____

Deadline for implementing corrective measures: _____
Responsible persons: _____

Prepared by: _____
Department Chair: _____
Date: _____

CHANGE LOG

Change No.	Basis for Amendment	Pages	Summary of the amendment	Revision	Signature	Date
1						
2						
3						

Edition:

Effective date: “ ” _____ 20

APPROVAL SHEET

№	Position / Role	Full Name	Signature	Date
1	Developed by	Kanetova D.E.		29.12.25
2	Approved: head of the responsible department	Ташиева Т.С		29.12.25
3	Approved: Head of the Educational and Informational Department	Kanetova D.E.		29.12.25
4	Approved: leading specialist for quality	Kalmuratova A.		29.12.25
4	Approved: head of the legal affairs and human resources department / lawyer	Sydykova B.J.		29.12.25
5	Approved: vice-rector for academic affairs	Sadyrova N.A.		29.12.25
6	Approved: vice-rector for science, SR and GE	Asilova Z.A.		29.12.25
7	Endorsed / considered in the established manner	JASU Scientific Council		29.12.25.