AUTOMATED QUALITY ASSURANCE OF EDUCATIONAL TESTING

Dr. Rositsa DONEVA
Faculty of Physics and Engineering Technologies
University of Plovdiv "Paisii Hilendarski"
Plovdiv, Bulgaria

Dr. Silvia GAFTANDZHIEVA
Faculty of Mathematics and Informatics
University of Plovdiv "Paisii Hilendarski"
Plovdiv, Bulgaria

Dr. George TOTKOV
Faculty of Mathematics and Informatics
University of Plovdiv "Paisii Hilendarski"
Plovdiv, Bulgaria

ABSTRACT

This paper presents a study on known approaches for quality assurance of educational test and test items. On its basis a comprehensive approach to the quality assurance of online educational testing is proposed to address the needs of all stakeholders (authors of online tests, teachers, students, experts, quality managers, etc.). According to the proposed approach is developed an original software application Test Quality Evaluation (TQE) for the automation of the stakeholders' activities for quality assurance of educational tests throughout the whole lifecycle. The application retrieves and provides analysis of data from online tests conducted and specially designed surveys for quality evaluation of educational tests by students and experts. It allows tracking and evaluating the quality of educational tests in real time and provides the related quantitative data in different levels of generalization – in the level of a separate educational test, of educational tests of an entire course, or educational tests of a subject area. The software application has been put under real-time testing for quality evaluation of educational tests, included in e-learning courses from different subject areas that prove its applicability.

Keywords: **Assessment quality, educational testing, automated quality assurance, online tests and test items quality.**

INTRODUCTION

According to ISO 9000:2015 quality assurance is focused on providing confidence that quality requirements to an object (product, service, process, person, organization, system, resource, etc.) will be fulfilled. To adapt the notion of quality assurance for higher education, the challenge is to determine how to identify whether the level of desired quality is maintained for every 'educational object' (Machado-da-Silva et al., 2015; Mutiara, Zuhairi & Kurniati 2007). Dill (2010), for example, puts the accent of quality assurance in higher education on the student assessment:

> ***"The term quality assurance in higher education is increasingly used to denote the practices whereby academic standards, i.e., the level of academic achievement attained by higher education graduates, are***

maintained and improved. This definition of academic quality as equivalent to academic standards is consistent with the emerging focus in higher education policies on student learning outcomes — the specific levels of knowledge, skills, and abilities that students achieve as a consequence of their engagement in a particular education program."

Similarly, according to the European Standards and Guidelines for Quality Assurance in the European Higher Education Area (EUSHARE, 2015), the assessment of students' knowledge and progression is also a key component of the systems for internal quality assurance in higher education:

"Considering the importance of assessment for the students' progress and their future careers, quality assurance processes for assessment take into account the following:

- ***Assessors are familiar with existing testing and examination methods and receive support in developing their own skills in this field;***
- ***The criteria for and method of assessment as well as criteria for marking are published in advance;***
- ***The assessment allows students to demonstrate the extent to which the intended learning outcomes have been achieved. Students are given feedback, which, if necessary, is linked to advice on the learning process;***
- ***Where possible, assessment is carried out by more than one examiner;***
- ***The regulations for assessment take into account mitigating circumstances;***
- ***Assessment is consistent, fairly applied to all students and carried out in accordance with the stated procedures;***
- ***A formal procedure for student appeals is in place".***

Regardless of whether talking about traditional learning, blended learning or e-learning, the main modern mean for objective knowledge assessment is through conducting online tests, typically using a Learning Management System (LMS). Precisely because the assessment through online tests has become an integral part of modern educational testing activities in all forms of training, their quality assurance is of a prime importance for achieving a high level of educational services, offered by higher education institutions. For this reason we regard the quality of educational testing as quality of e-testing in this paper.

The concept of quality is related to educational testing in two contexts. On the one hand, to have a reliable academic assessment it is significant to ensure the quality of the online tests themselves. The quality assurance of online tests and test items affects all stages of their lifecycle – from the design and development to test conducting and scoring. In this sense, it concerns a relevant group of stakeholders in the education system – test authors, teachers/assessors, methodologists and experts in the test subject area, students (the testees). On the other hand, the quality of online tests is an important component of internal university systems for management and assurance of the educational quality as a whole, i.e. it is of essential interest to another group of stakeholders – quality managers and policymakers in higher education institutions.

This paper aims to propose a comprehensive approach to the quality assurance of online educational testing addressing the needs of all stakeholders (authors of online tests, teachers, students, experts, quality managers, etc.). In the following sections the literature review is presented. A comprehensive approach to the quality assurance of

online educational testing is proposed with its characteristics, stages, models for quality assurance and stakeholders. Next a software application TQE is introduced (developed on the basis of the proposed approach). TQE allows automation of the stakeholders' activities for quality assurance of online tests throughout the whole lifecycle, tracking and evaluating the quality of online tests in real time and provides related quantitative data in different levels of generalization – quality measures in the level of a separate online test, of online tests of an entire course, or of online tests of an academic specialty, etc. TQE is experimented for quality evaluation of a test, included in e-learning courses from 3 different subject areas (physics; informatics; a foreign language).

LITERATURE REVIEW

The development of quality educational tests is a complex task, subject to a lengthy and labor-intensive iterative process (Totkov, Raikova & Kostadinova, 2014). Different authority organizations have published materials, guidelines and standards related to quality assessment to help the improvement of the quality of assessment of learning achievements. As, for example the "Standards for Educational and Psychological Testing" (APA, 2014), published collaboratively by the American Educational Research Association (http://www.aera.net/), the American Psychological Association (http://www.apa.org) and the National Council on Measurement in Education (http://www.ncme.org) since 1966, that represents the gold standard in guidance on testing in the United States and in many other countries.

One other direction in the efforts to improve assessment quality is based on the development of quantitative methods for the evaluation of test quality. The idea for quality evaluation of test items on the basis of the test response analysis originates around the middle of the last century in Item Response Theory (Hambleton, Swaminathan & Rogers, 1991), Classical Test Theory and the so called Rasch Model. The assessment is performed by the authors of the test items and/or by experts in the subject area after testing. The analysis provides empirical data on how individual test items are performed in real test situations. The data obtained is subject to special procedures and the analysis is done in relation to the following test characteristics (Pyrczak, 1973; Mark, 1985; Hambleton, Swaminathan & Rogers, 1991): difficulty, discrimination index, analysis of distractors (for questions with optional answers). The calculated values indicate which test items need to be modified or removed to improve the test quality (Rasch, 2017).

Reliability, validity and fairness are three fundamental properties of a test by which the technical quality of tests is evaluated (Hamilton, Stecher & Klein, 2002). The reliability of a test refers to the degree to which a test scores are free from various types of chance effects (Hamilton, Stecher & Klein, 2002). According to (Saad et al., 1999) there are four ways of estimating reliability: test-retest, alternate or parallel forms, inter-rater and internal consistency. After the reliability is estimated, the information can be reported via a reliable statistic – the reliability coefficient and standard error of measurement. The validity of a test refers to the extent to which the scores on a test provide accurate information for the decisions that will be based on those scores (Cronbach, 1971; Messick, 1989). A test's validity is established in reference to a specific purpose and the test may not be valid for different purposes. There are several ways to estimate the validity of a test including content validity, concurrent validity, predictive validity and face validity (Professional Testing, 2017). The fairness of a test refers to its freedom from any kind of bias. The test should be appropriate for all students irrespective of race, religion, gender, or age. The test should not disadvantage any student, or group of students, on any basis other than the student's lack of the knowledge and skills the test is intended to measure. According to (Professional Testing, 2017) "Item writers should address the goal of fairness as they undertake the task of writing items. Test items should be reviewed for potential fairness problems during the item review phase and any items

that are identified as displaying potential bias or lack of fairness should then be revised or dropped from further consideration".

The evaluation of tests includes also assessment of other important tests properties as test materials, norms, computer generated reports, including a global final evaluation, etc. While in some cases the test quality characteristics could be measured by performing analysis of test results, the measurement of the other is possible only by a subjective evaluation of test by experts, teachers or students on the basis of specially developed quality models. A large number of experiments are conducted by different stakeholders for educational test quality evaluation on the basis of different quality models, for example:

- **Quality evaluation of test items for language learning (CDC, 2017);**
- **Self-evaluation of the quality of test items by teachers (CFATIQC, 2017; CITL, 2017);**
- **Quality evaluation of multiple choice questions as structure and taxonomy (Amouei, 2014);**
- **EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests (EFPA, 2013);**
- **Quality evaluation of multiple choice test items, created with automated processes (Gierl & Hollis, 2013);**
- **Quality evaluation of online tests by students (Legault, 2017; CITL, 2017).**

In spite of the big interest in various aspects of student assessment, the quality assurance issues have not been fully addressed. For example, very slight attention is paid to the needs of educational quality managers and policymakers from having accurate and accessible information to inform the right decisions regarding quality. The above listed experiments where the test quality assurance approaches are devoted to a specific type of tests, subject area, or stakeholder, also show that additional research is needed. The approach and software tool, proposed in this paper, try to overcome these disadvantages by examining the issues in their complexity and integrity and providing quantitative measures of the online test quality in an automated manner.

METHODS

The methods used in the study include:

- **proposing a comprehensive approach for automated quality assurance of educational tests (conducted online) by all stakeholders;**
- **developing models for educational test quality evaluation;**
- **developing an original web-based software application for automated quality evaluation of educational tests;**
- **verification of the proposed models and software application with real data.**

Comprehensive Approach for Automated Quality Assurance of Online Testing

The approach for the quality assurance of online educational testing, proposed here, possesses the following characteristics that prove its comprehensiveness:

- **The approach provides a possibility to obtain all the possible data of the evaluation of test items and tests as a whole, typical for the primary approaches for quality assurance of educational tests on the basis of:**
  - **A statistical analysis of the test responses after the test probation among a representative group of students or after conducting it in real test situations;**
  - **Specially developed test quality models for evaluation by experts or students.**

- **It enables the process of assuring e-testing quality to be informed by input from representatives of all relevant stakeholders (test authors, teachers/assessors, methodologists and experts in the test subject area, students);**
- **The approach allows evaluation of the test quality during the complete lifecycle of a test (its design stage or usage stage, or even afterwards);**
- **It addresses the needs of quality-related information of all stakeholders (incl. educational quality managers and policymakers);**
- **It supports quality assurance activities in different levels of generalization in the level of separate online test items, of an online test as a whole, of online tests of an entire course, or of online tests of an academic specialty, etc.**

Our model of the comprehensive approach (see Figure 1) demonstrates the basic components of an integrated system for educational test quality evaluation, namely:

- **two perspectives on the problem (contexts) – of the quality of online tests themselves and of the educational quality as a whole;**
- **instruments for educational test quality evaluation (test quality models) – two quality models for test evaluation by questionnaires from experts and students/testees (see Table 1), two quality models for evaluation of the basis of testees' responses of separate test items and educational tests as a whole (see Table 2 and Table 3), quality model for evaluation of the quality assurance process itself of educational tests an entire course, an academic specialty, a professional field, or an area of higher education (see Table 4);**
- **different stages of the test quality assurance (the testing lifecycle) – test design and development, test approbation, test conducting and scoring after test usage;**
- **the categories of significant players (the testing stakeholders) – test authors, teachers/assessors, experts (in didactics, in the test subject area), students, quality managers and policymakers;**
- **mutual relations that reflect the usage of quality models in the testing lifecycle by the stakeholders depending on their role in the process as evaluators or users of the quality measures obtained.**

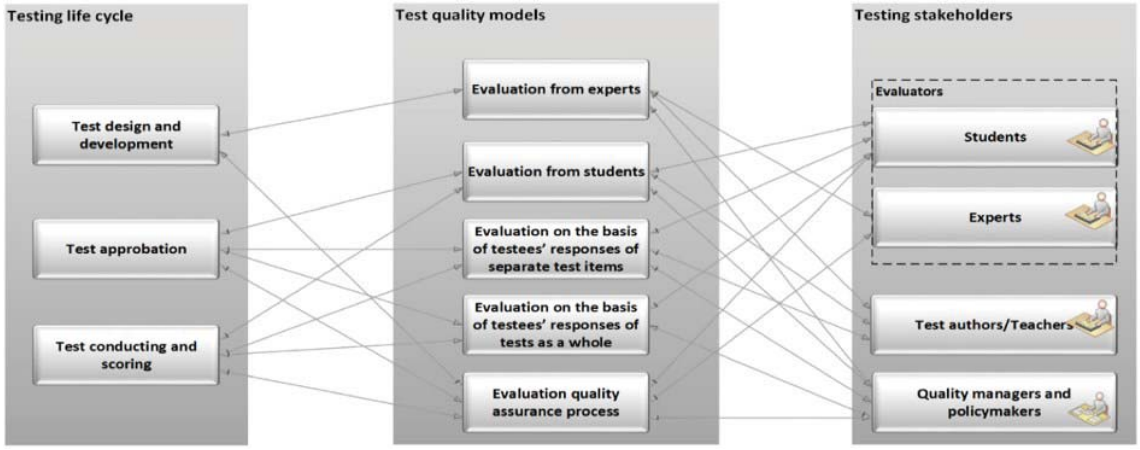


Figure 1. Stages, models for quality assurance, stakeholders

The approach will be easily applied if educational e-testing is organised using an LMS. This is not a limiting condition because the modern LMS provides tools for creating test items, creating online tests from a bank with pre-established test items, conducting online tests. Some systems, including Blackboard (Blackboard Help, 2017) and Moodle (Moodle Documentation, 2017), provide tools for automated analyses of test items. The

LMS also provides tools that allow teachers to organize and conduct surveys within the e-learning courses, the results of which can be used for analysis.

The possibilities for creating and conducting online tests and evaluating the quality of tests by experts, teachers and students allow automated quality evaluation of online tests on the proposed approach. The evaluation of the overall quality of online tests and test items included in them will be carried out on the basis of the proposed models. The next three subsections present the test quality models.

Quality models for test evaluation by questionnaires from experts and students (testees)

The two quality models presented here allow evaluation of the overall quality of online tests by experts and students through filling out questionnaires for the educational test quality evaluation at different stages of an online test lifecycle:

- **the design stage of online tests where the evaluation is done by experts and a representative set of students;**
- **the stages of test approbation and test conducting and scoring where the evaluation is performed by students.**

The models evaluate developed test items, tests, test conducting, test evaluative results, and the test interface design. The results obtained are relevant to the stakeholders and allow:

- **the authors of an online test to improve the quality of the test items and of the overall quality of the online test at the design stage after testing it with a representative set of students.**
- **the authors of online tests to make changes in the test items in order to improve their quality after the test has been taken by a representative set of students and after the test has been conducted in real situations;**
- **quality managers and policymakers to ensure the overall quality of training.**

The questionnaires for online test evaluations by experts and students are developed on the basis of a proposed hierarchical model for quality evaluation (based on Legault, 2017; Totkov, Raikova & Kostadinova, 2014; CITL, 2017; Amouei et al., 2014). The model includes 52 indicators broken down into 5 criteria as follows:

- **test items - 21 indicators (A1 to A21), of which 21 are evaluated by experts and 10 by students;**
- **test - 8 indicators (B1 to B8), of which 6 are evaluated by experts and 7 by students;**
- **test conducting - 7 indicators (from B1 to B7), of which 6 are evaluated by experts and 5 by students;**
- **test evaluative results - 12 indicators (from D1 to D12), of which 8 are evaluated by experts and 8 by students;**
- **test interface design - 4 indicators (from D1 to D4), of which all 4 are evaluated by experts and 2 of them by students.**

The evaluation of each composite indicator is obtained as the sum of the indicators' evaluations (evaluated with a five-point scale). Table 1 presents the questions included in the questionnaires for experts and students. These questionnaires are used in a survey for quality evaluation of online tests.

Table 1. Model for educational test quality evaluation by experts and students

Criteria Indicator	Questionnaire for Experts	Questionnaire for Students
A. TEST ITEMS QUALITY		
A1. Test items are formulated clearly and accurately	✓	
A2. Test items don't allow ambiguous interpretations	✓	✓
A3. Test items don't contain formulations that help students to find the right answer.	✓	
A4. Test items check specific knowledge, ability or skill	✓	✓
A5. The creation and selection of test items follows the informative principle over the full range of variation of the test for cognitive complexity levels of students and not only of individual cognitive knowledge	✓	
A6. A simple but grammatically correct positive form of the test items is used in the form of a sentence of 5-15 words.	✓	✓
A7. Test items don't use words with undefined content such as "sometimes", "often", "always", "all", "never", "big" and "less","more", double negations, excluding "or", quantum negation, and so on (unless the test item intends to understand the listed language constructions).	✓	✓
A8. The answer of a test item doesn't follow from the answer of another test item	✓	
A9. Test items are determinated and don't require further clarification	✓	✓
A10. Test items don't require knowledge beyond the curriculum, program, or educational standard	✓	✓
A11. Test items have a specification in the relevant test bank	✓	
A12. Test items don't require students to do detailed analysis, calculations, or answers	✓	✓
A13. Test items are sufficiently meaningful and comprehensive to achieve the set goals	✓	
A14. Test items are accompanied by specially designed instructions for their use	✓	
A15. Test items are clearly formulated and contain detailed instructions	✓	✓
A16. Test items require original thinking		✓
A17. Test items don't contain contradictory or inaccurate instructions, introductions or explanations	✓	✓
A18. Test items don't contain complex instructions, introductions or explanations	✓	✓
A19. Test items are designed in full compliance with the requirements of the testology	✓	
A20. The text of test items doesn't have excessive verbal and unnecessary information	✓	
B. TEST QUALITY		
B1. Test items are ordered in ascending order of difficulty	✓	✓
B2. The complexity of the test is not "enhanced" by the introduction of multiple additional phrases in the test items' condition	✓	✓
B3. Test items included in the test reflect well the content and purpose of the course	✓	✓
B4. The test contains competent, grammatical and interesting questions and situations causing students to answer and not to choose answers	✓	✓
B5. A sufficient number of test items are provided to determine whether a student has learned the material	✓	
B6. Test items included in the test provoke students' thinking	✓	✓
B7. Tests are designed with a sufficient degree of interactivity to engage students and provide an objective assessment of their knowledge and skills	✓	
B8. The test doesn't contain banal test items.		✓
B9. Tests are developed with an appropriate methodology	✓	

Criteria Indicator	Questionnaire for Experts	Questionnaire for Students
C. TEST CONDUCTING		
C1. There are formulated clear criteria to evaluate the test	✓	✓
C2. The process of computing testing provides a user-friendly and interactive multimedia interface	✓	✓
C3. The process of computer testing provides students with the opportunity to return to unresolved tasks	✓	✓
C4. The testing environment provides the ability to update the content of test items	✓	
C5. The student has information about upcoming testing (test structure, time to solve, etc.)	✓	✓
C6. Tests are planned to be conducted at appropriate intervals	✓	
C7. There is enough time to solve the test		✓
D. TEST EVALUATIVE RESULTS		
D1. The final grade is well-founded, categorical and impartial	✓	✓
D2. Timely feedback is provided for (self-)assessment, allowing students to track their learning progress	✓	
D3. Each test is scheduled to end with a grade	✓	
D4. Each test will be completed by result analysing, determining the level of training and the quality of the testing conducted	✓	
D5. The assessment criteria have been published in advance	✓	✓
D6. The assessment methodology has been published in advance	✓	✓
D7. The assessment allows students to show the extent to which the learning outcomes are achieved	✓	
D8. There is an official student complaint procedure	✓	✓
D9. The feedback is timely and allows students to track their learning progress		✓
D10. The feedback includes explanations of mistakes and personal comments		✓
D11. The feedback gives new knowledge		✓
D12. The evaluation is carried out in accordance with the established procedures		✓
E. TEST INTERFACE DESIGN		
E1. The interface allows students to track their learning progress	✓	✓
E2. All parts of the test items are located on the same page	✓	✓
E3. The presentation of the different types of test items is consistent	✓	
E4. The placing of too many test items on one page is avoided	✓	

Quality models for evaluation on the basis of testees' responses of separate test items and tests as a whole

In the proposed overall approach to quality assurance of online tests, the quality of test items is evaluated on the basis of an analysis of the responses to tests carried out at different stages of their life – at the test approbation stage after carrying out the test with a representative set of students and at the test conducting and scoring stage after the test is conducted in real test situations. The quality of each test item and of the test as a whole is evaluated on the basis of the calculated statistic data (Moodle Documentation, 2017; Thompson & Levitov, 1985; Pyrczak, 1973; Mark, 1985; Hambleton, Swaminathan & Rogers, 1991; Hamilton, Stecher & Klein, 2002; Cronbach, 1971; Messick, 1989; Professional Testing, 2017). The quality of each test item is evaluated on the basis of the calculated facility index, standard deviation and discrimination and the values obtained (see Table 2). The values obtained at the design stage after testing among a representative set of students allow the authors of the online test to determine which test items should be processed or excluded from the online test prior to conducting the test in real situations. The values obtained after the real conducting of online tests allow their authors to make changes to the test items in order

to improve test quality and to increase their reliability after the real testing. The evaluation results are also important for the quality managers and policymakers. They allow them to take measures to improve the quality of tests that contain test items with unsatisfactory quality and thus to provide higher quality of testing and training.

Table 2. Quality evaluation of test items on the basis of testees' responses

Index	Definition	Values	Evaluation
Facility index – FI	The percentage of students that answered to the test item correctly	<5%	Extremely difficult test item
		6 % - 10%	Very difficult test item
		11% - 20%	Difficult test item
		20% - 34%	Moderately difficult test item
		35% - 64%	Neither difficult nor easy test item (About right for the average student)
		66% - 80%	Fairly easy test item
		81% - 89%	Easy test item
		90% - 94 %	Very easy test item
		>95%	Extremely easy test item
Standard deviation – SD	A measure of the spread of scores about the mean and hence the extent to which the question might discriminate	<33%	Unsatisfactory test item
		>33%	Satisfactory test item
Discrimination index – DI	The percentage of correct answers by students who have scored highly on the other parts of the test	<0%	Invalid test item
		0% - 19%	Very weak discrimination
		20% - 29%	Weak discrimination
		30% - 50%	Adequate discrimination
		>50%	Very good discrimination

The overall quality evaluation of online tests carried out in higher education institutions (in the chosen course, academic specialty, professional field, field of higher education or all electronic tests carried out at university) allows the quality managers and policymakers to take measures to improve the quality of unsatisfactory quality tests in order to ensure a higher quality of training. The overall test quality on the basis of the answers given by the students and the results obtained is evaluated through a calculation of average grade, standard deviation, skewness, kurtosis, coefficient of internal coherence, standard and relative error. Table 3 presents the values and their interpretation used within the proposed evaluation model.

Table 3. Educational test quality evaluation on the basis of testees' responses

Measure	Definition	Values	Evaluation
Average grade	The average of students' scores	<50%	Unsatisfactory result
		50% - 75%	Satisfactory result
		>75%	Unsatisfactory result
Standard Deviation – SD	A measure of how widely values are dispersed from the average grade	<12%	Unsatisfactory result
		12% - 18%	Satisfactory result
		>18%	Unsatisfactory result
Skewness	A measure of the asymmetry of the distribution of scores	<-1	Lack of discrimination between students who do better than average
		[-1,-1]	Perfectly symmetrical distribution
		>1	Lack of discrimination near the pass fail border

Measure	Definition	Values	Evaluation
Kurtosis	A measure of the flatness of the distribution of scores	0-1	The test is discriminating very well between very good or very bad students and those who are average
		>1	The test is not discriminating very well between very good or very bad students and those who are average
Coefficient of internal consistency – CIC	A measure of the reliability of the assessment scales	>90% 75%-90% 64% - 74% <64%	Perfect result Satisfactory result Unsatisfactory result Completely unsatisfactory result
Error ratio – ER	It estimates the percentage of the standard deviation which is due to chance effects rather than to genuine differences of ability between students	<50% >50%	Satisfactory result Unsatisfactory result
Standard error – SE	It estimates how much of the standard deviation is due to chance effects and is a measure of the uncertainty in any given student's score	<7% 8% >8%	Perfect assessment Good assessment Substantial proportion of the students will be wrongly graded

Quality model for evaluation of the quality assurance process of online tests

The model allows evaluation of the quality assurance process itself (of testing in an entire course, an academic specialty, a professional field, an area of higher education) during all stages of the online test lifecycle. The overall process of the quality evaluation of online tests is evaluated on the basis of the data stored from the filled questionnaires and test results. The evaluation model is hierarchical and includes three levels – 5 objects, 10 criteria and 20 indicators (Table 4). The evaluation obtained enables the quality managers and policymakers to monitor the evaluations and receive a summary of the courses in which the evaluation of online tests is being conducted at all times in which they want to monitor the process.

Table 4. Evaluation of the quality assurance process

Object *Criteria* Indicator
1. Learning course
1.1. The quality of all tests in the course is evaluated by students and experts
1.1.1. Data for all tests in the course:
- Information: Test;
- Number of respondents (completed questionnaires) of each evaluated online test in the course;
- Number of respondents of all evaluated online tests in the course.
1.1.2. Summarized results of the survey by the evaluated characteristics of tests in the course:
- Question (evaluated characteristic);
- Percentage of experts/students that answered 1-5 to the evaluated characteristic in all evaluated online tests in the course.
1.1.3. Summarized results of the survey for all tests in the course:
- Information: Tests;
- Average grade of each evaluated online test in the course;
- Average grade of all evaluated online tests in the course.
1.2. The quality of all tests in the course is evaluated on the basis of the testees' responses of the tests

Object
Criteria
Indicator

1.2.1. Calculated statistic values: Average grade, Standard Deviation, Coefficient of internal consistency, Error ratio and Standard error of all online tests in the course.

2. Academic specialty

2. 1. The quality of all tests in an academic specialty is evaluated by students and experts

2.1.1. Data for all tests in the academic specialty:
- Percentage of courses in the academic specialty with conducted surveys;
-Information: Course, Test;
- Number of respondents (completed questionnaires) of each evaluated online test in the course;
- Number of respondents (completed questionnaires) of all evaluated online tests in the course;
- Number of respondents of all evaluated online tests in the academic specialty.
2.1.2. Summarized results of the survey by the evaluated characteristics of tests in the academic specialty:
- Question (evaluated characteristic);
- Percentage of experts/students that answered 1-5 to the evaluated characteristic in all evaluated online tests in the academic specialty.
2.1.3. Summarized results of the survey for all tests in the academic specialty:
- Information: Course;
- Average grade of all evaluated online tests in the course;
- Average grade of all evaluated online tests in the academic specialty.

2.2. The quality of all tests in the academic specialty is evaluated on the basis of testees' responses of tests

2.2.1. Calculated statistic values: Average grade, Standard Deviation, Coefficient of internal consistency, Error ratio and Standard error of all online tests in the academic specialty.

3. Professional field

3.1. The quality of all tests in the professional field is evaluated by students and experts

3.1.1. Data for all tests in the professional field:
- Percentage of academic specialty in the professional field with conducted surveys;
- Percentage of courses in the professional field with conducted surveys;
- Information: Academic specialty, Course, Test;
- Number of respondents (completed questionnaires) of each evaluated online test in the course;
- Number of respondents (completed questionnaires) of all evaluated online tests in the course;
- Number of respondents of all evaluated online tests in the academic specialty;
- Number of respondents of all evaluated online tests in the professional field.
3.1.2. Summarized results of the survey by the evaluated characteristics of tests in the professional field:
- Question (evaluated characteristic);
- Percentage of experts/students that answered 1-5 to the evaluated characteristic in all evaluated online tests in the professional field.
3.1.3. Summarized results of the survey for all tests in the professional field:
- Information: Academic specialty, Course;
- Average grade of all evaluated online tests in the course;
- Average grade of all evaluated online tests in the academic specialty;
- Average grade of all evaluated online tests in the professional field.

3.2. The quality of all tests in the professional field is evaluated on the basis of testees' responses of tests

3.2.1. Calculated statistic values: Average grade, Standard Deviation, Coefficient of internal consistency, Error ratio and Standard error of all online tests in the professional field.

4.Area of higher education

4.1.The quality of all tests in the area of higher education is evaluated by students and experts

4.1.1. Data for all tests in the area of higher education:
- Percentage of academic specialty in the area of higher education with conducted surveys;
- Percentage of professional fields in the area of higher education with conducted surveys;
- Percentage of courses in the area of higher education with conducted surveys;
- Information: Professional field, Academic specialty, Course, Test;
- Number of respondents (completed questionnaires) of each evaluated online test in the

Object
Criteria
Indicator

course;
- Number of respondents (completed questionnaires) of all evaluated online tests in the course;
- Number of respondents of all evaluated online tests in the academic specialty;
- Number of respondents of all evaluated online tests in the professional field;
- Number of respondents of all evaluated online tests in the area of higher education
4.1.2. Summarized results of the survey by the evaluated characteristics of tests in the area of higher education:
- Question (evaluated characteristic);
- Percentage of experts/students that answered 1-5 to the evaluated characteristic in all evaluated online tests in the area of higher education.
4.1.3. Summarized results of the survey for all tests in the area of higher education:
-Information: Professional field, Academic specialty, Course;
- Average grade of all evaluated online tests in the course;
- Average grade of all evaluated online tests in the academic specialty;
- Average grade of all evaluated online tests in the professional field;
- Average grade of all evaluated online tests in the area of higher education.

4.2. The quality of all tests in the professional field is evaluated on the basis of testees' responses of tests

4.2.1. Calculated statistic values: Average grade, Standard Deviation, Coefficient of internal consistency, Error ratio and Standard error of all online tests in the academic area of higher education.

5. University

5.1. The quality of all tests is evaluated by students and experts

5.1.1. Data for all tests in the university:
- Percentage of areas of higher education with conducted surveys;
- Percentage of professional fields with conducted surveys;
- Percentage of academic specialty with conducted surveys;
- Percentage of courses with conducted surveys;
- Information: Area of higher education, Professional field, Academic specialty, Course, Test;
- Number of respondents (completed questionnaires) of each evaluated online test in the course;
- Number of respondents (completed questionnaires) of all evaluated online tests in the course;
- Number of respondents of all evaluated online tests in the academic specialty;
- Number of respondents of all evaluated online tests in the professional field;
- Number of respondents of all evaluated online tests in the area of higher education;
- Number of respondents of all evaluated online tests.
5.1.2. Summarized results of the survey by the evaluated characteristics of all tests:
- Question (evaluated characteristic);
- Percentage of experts/students that answered 1-5 to the evaluated characteristic in all evaluated online tests.
5.1.3. Summarized results of the survey for all tests in the university:
- Information: Area of higher education, Professional field, Academic specialty, Course;
- Average grade of all evaluated online tests in the course;
- Average grade of all evaluated online tests in the academic specialty;
- Average grade of all evaluated online tests in the professional field;
- Average grade of all evaluated online tests in the area of higher education.
- Average grade of all evaluated online tests.

5.2. The quality of all tests in the professional field is evaluated on the basis of testees' responses of tests

5.2.1. Calculated statistic values: Average grade, Standard Deviation, Coefficient of internal consistency, Error ratio and Standard error of all online tests.

Software Application for Automated Educational Test Quality Evaluation

The main purpose of the proposed original application for automated evaluation of the quality of online tests on the basis of the proposed approach is to enable stakeholders (authors of online tests and quality managers and policymakers) to generate documents evaluating the quality of online tests in real time which allow the quality of tests to be

improved. The application that allows automated quality evaluation of online tests on the proposed quality assurance approach should provide the following basic functionalities:

- **retrieving results from online tests conducted in testing environments and/or in LMS used by the higher education institution;**
- **analysing results of the conducted online tests;**
- **analysing results of the conducted surveys for quality evaluation of online tests by students and experts;**
- **generating documents for quality evaluation of online tests by different users.**

The software prototype of the application for automated quality evaluation of online tests TQE has been developed on the basis of previous studies in the field of automated quality evaluation in higher education (Doneva & Gaftandzhieva, 2015; Gaftandzhieva, 2017; Gaftandzhieva, 2016; Totkov, Gaftandzhieva & Doneva, 2016). During the development of the software prototype, part of the application's functionalities were realized. The realized functionalities allow:

- **retrieving results from online tests conducted in LMS Moodle;**
- **performing analysis of results of online tests conducted in LMS Moodle;**
- **analysing the results of surveys conducted in LMS Moodle for quality evaluation of online tests by students and experts in the field;**
- **generating documents for quality evaluation of electronic tests by two stakeholders – authors of online tests and quality managers and policymakers.**

The prototype of the TQE application includes two panels for each context to enable users (authors of online tests and quality managers and policymakers) to generate evaluation documents in the form of reports. The TQE application is written in PHP and uses Jasper-Soft BI Suite (JasperSoft, 2017) capabilities for creating reports and analysis by retrieving data from different information sources, for storing and organizing reports in a repository, and for presenting them in the form preferred by the user. The application is developed in 4 steps:

Step 1. Studying the information context of an online test and surveys in LMS Moodle;
Step 2. Integration between the JasperSoft BI Suite tool for development of report templates (JasperSoft Studio) and an LMS Moodle database, which is set as a data source for data retrieving and creating documents (reports) that reflect online test quality;
Step 3. Development of templates of analytical reports in JasperSoft Studio, which can be used later to generate the real reports containing summarised data related to online test quality;
Step 4. Compiling of templates of analytical reports (developed in Step 3) in a special internal format, storing them in the JasperReports Server repository and Integration of the JasperReports Server with TQE through JasperServer REST API and PHP wrapper.

As a result of the study carried out in step 1, the Moodle Feedback activity is chosen to be used for carrying out surveys among students and experts. For each of the two quality evaluation models questionnaire templates were created, which were included as a part of the learning activities in each e-course, so that they are completed by at least three experts during the online test design and by participating students after completion of the training. As a result of the study 18 tables from the Moodle databases, which store data related to the online test and the surveys conducted, have been studied in detail. Data stored in the tables is identified which can be used to accumulate dynamic online test quality evaluation.

To achieve the aim of the application in step 2 of the development process of TQE, relevant report templates are designed according to the specific parameters (e.g. e-course, professional field, and area of higher education). The choice of templates suitable for generating documents for the educational test quality evaluation according to the proposed models for quality evaluation is done on the basis of the analysis of the institution's information infrastructure and subsequent systematization which of the data stored in the LMS Moodle can be accumulated automatically for each of the model criteria.

For complete educational test quality evaluation by students and experts, templates of reports are designed (see Column 1 of Table 5) depending on the specific parameters (see Column 2 of Table 5). They allow users (See Column 4 and Column 5 of Table 5) to retrieve evaluation data and generate evaluation reports that contain aggregated evaluation information on the online test quality (see Column 3 of Table 5).

Table 5. Templates – educational test quality evaluation by experts/students

Template	Parameter	Returned information	Author	Quality managers and policymakers
Summarized results of the survey by the evaluated characteristics of an online test	Online test	• Question (evaluated characteristic) • Percentage of experts/ students that answered 1-5	✓	✓
Summarized results of the survey by the evaluated criteria of an online test	Online test	• Criteria • Average grade	✓	✓
Summarized results of the survey by professional fields and courses		• Professional field • E-course • Average grade		✓
Summarized results of the survey by a professional field		• Professional field • Average grade		✓
Summarized results of the survey by an area of higher education	Area of higher education	• Professional field • Average grade		✓
Summarized results of the survey by professional fields and courses in an area of higher education	Area of higher education	• Professional field • E-course • Average grade		✓
Summarized results of the survey by evaluated characteristics of an online test in an area of higher education	Area of higher education	• Question (evaluated characteristic) • Percentage of experts/ students that answered 1-5		✓

To evaluate the quality of test items on the basis of an analysis of the responses and the overall quality of an online test on the basis of the statistical results after testing from TQE application, a document template is designed dependent on the specific parameter (an online test). The template presents in tabular form the calculated values of the facility index, standard deviation and discrimination of each test item included in the evaluated online test, and the answers given by the students are extracted dynamically from the Moodle database. The quality of test items is evaluated on the basis of the calculated values of the facility index, standard deviation and discrimination, average score, standard deviation, asymmetry of distribution, internal coherence coefficient, standard and relative error. According to the range in which the value falls (see column 3 of Table 2) dynamic evaluation is given (see column 4 of Table 2). The template provides an option for a dynamic generation of texts with a value

analysis on the basis of the range in which the value falls, and the alternative evaluations in Table 3.

To retrieve evaluation data for the indicators of the model for evaluation of the quality assurance process of an online test (see Table 4) 35 templates of reports are designed (15 for summarizing an experts' grade and 20 for summarizing students' grades and the result) depending on the specific parameters. They allow quality managers and policymakers to retrieve evaluation data and generate reports that contain summary data for the ongoing quality evaluation of online tests.

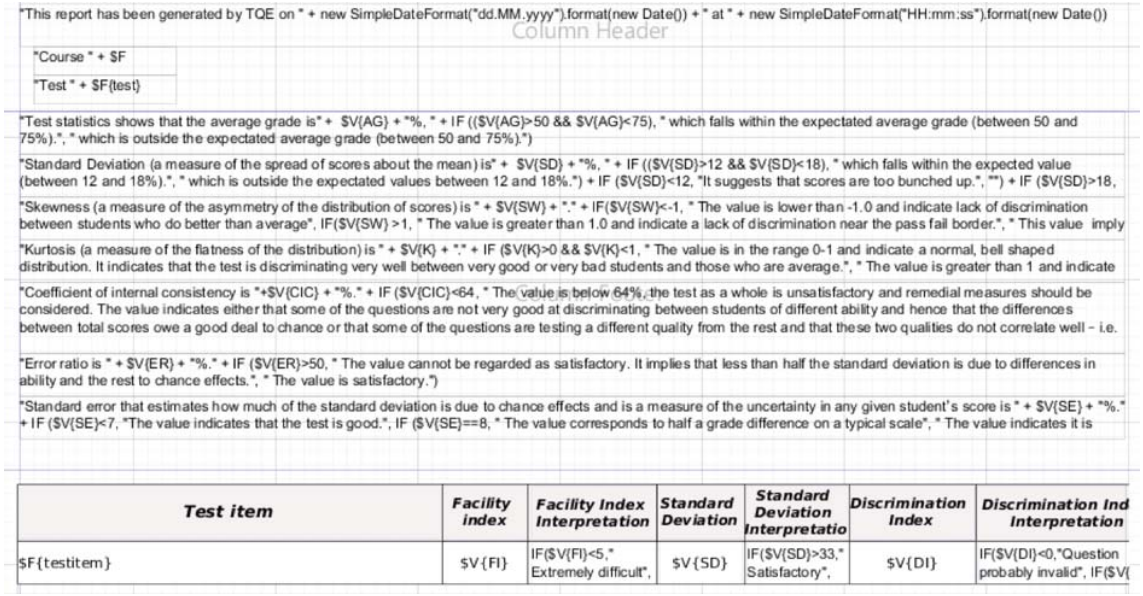


Figure 2. Template (created with JasperSoft)

A total of 43 software models of accumulating templates are developed through JasperSoft's template design tool in Step 2. Users can apply them to generate real documents that contain aggregated results from the ongoing quality evaluation of an online test by students and experts, or results from the analysis of test items. Figure 2 presents the developed model for analysing the quality of test items on the basis of students' results. The developed templates are compiled in a special internal format and are stored in the Jaspersoft repository, which is realized in Step 4. In this way, they can be used both by the level of the JasperSoft system, TQE and another external applications for the generation of evaluation reports that are filled with data from the given data source (Moodle Database).

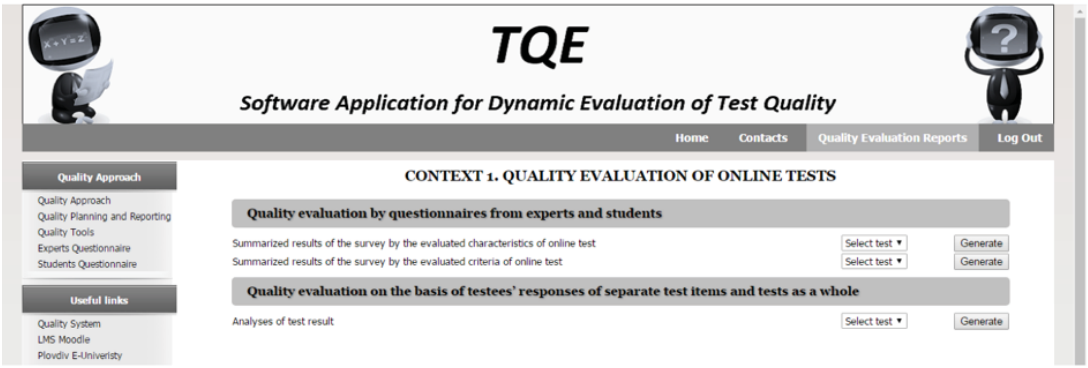


Figure 3. Reports for quality evaluation of online tests – authors of tests (screenshot from TQE)

Users of TQE can use it to generate dynamic evaluations in the form of reports by selecting from the proposed capabilities (a dynamically generated list of available templates of reports for both groups of users in two evaluation contexts, see Figure 3 and Figure 4) at any time they want to monitor the ongoing evaluations and the results of the test quality analysis on the basis of student responses and results. Authors of online tests have access to functionalities that allow them to analyse the results of surveys, monitor the process of surveys and evaluate the quality of their test items included in an online test conducted in Moodle. TQE enables quality managers and policymakers to analyse the results of all surveys automatically, monitor the process of all surveys and evaluate the quality of all test items. The monitoring of the quality assurance process and the analysis of the results can be obtained by an online test, an e-course, a professional field and an area of higher education. Besides selecting the type of report that will be generated in real-time, the user must set values for the necessary local parameters of the report. The alternative parameters and their values between the user can choose are retrieved from the data source. This limits the user's choice and thus eliminates the possibility of introducing incorrect data.

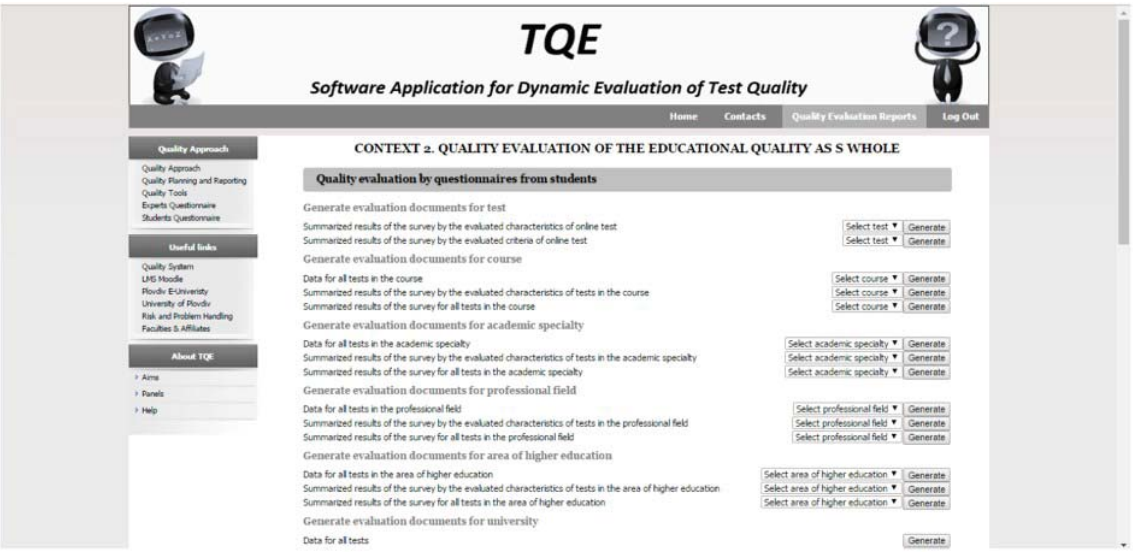


Figure 4. Reports for quality evaluation of the educational quality as a whole – quality managers and policymakers (screenshot from TQE)

The generated report contains data (in the form of a table and chart) that the user can use to evaluate to what extent the students have participated in surveys for quality evaluations and analyse the results from conducted surveys. The generated report can be displayed on the screen and the user has the possibility to download it in his or her preferred format.

FINDINGS AND DISCUSSIONS

The software prototype TQE is experimented for quality evaluation of a test, included in e-learning courses from 3 different subject areas (physics; informatics; a foreign language):

- **English – A2/B2 (professional field 1.3. Pedagogy of teaching in …);**
- **Physics (professional field 4.1. Physical sciences);**
- **Web Programming (professional field 4.6. Informatics and computer science).**

The quality of the tests is evaluated through the automated test quality analysis of the test items on the basis of the students' responses and statistical data for the test.

Quality Evaluation Report: Analysis of test results

This report has been generated by TQE on 04.12.2016 at 20:28:17

Course English A2/B1

Test English test- Pre

Test item	Facility index	Facility Index Interpretation	Standard Deviation	Standard Deviation Interpretatio	Discrimination Index	Discrimination Index Interpretation
1. Identify what part of speech are the following words: movement; relationship; negotiation.	83.33	Easy	38.92	Satisfactory	2.4	Very weak discrimination
2. Choose the group of state verbs	83.33	Easy	38.92	Satisfactory	-37.14	Question probably invalid
3. Choose which category contains extreme adjectives	58.33	About right for the average student	51.49	Satisfactory	-3.65	Question probably invalid
4. Which of these groups contain words with silent letters?	91.67	Very easy	28.87	Unsatisfactory	-18.46	Question probably invalid
5. Choose the correct preposition of the verb depend	91.67	Very easy	28.87	Unsatisfactory	-18.46	Question probably invalid
6. Choose the correct preposition of the verb "spend"	66.67	Fairly easy	49.24	Satisfactory	38.95	Adequate discrimination
7. Choose the correct preposition of the verb "fond"	83.33	Easy	38.92	Satisfactory	-11.61	Question probably invalid
8. Determine whether sentences I can't find my keys. Have anyone seen them? Carrie's really a close friend. We know	58.33	About right for the average student	51.49	Satisfactory	-3.65	Question probably invalid
9. Select the correct compliance to the phrase "keep my mind ..."	33.33	Moderately difficult	49.24	Satisfactory	34.11	Adequate discrimination
10. Select the correct way to complete the phrase "mind your ..."	66.67	Fairly easy	49.24	Satisfactory	38.95	Adequate discrimination
11. Select the most appropriate phrase to complete the sentence: Since that quarrel they haven't been on	66.67	Fairly easy	49.24	Satisfactory	71.67	Very good discrimination
12. Select the correct words to complete the phrase "my mind....."	8.33	Very difficult	28.87	Unsatisfactory	-15.47	Question probably invalid

Test statistics shows that the average grade is 65.97%, which falls within the expectated average grade (between 50 and 75%).

Standard Deviation (a measure of the spread of scores about the mean) is 13.97%, which falls within the expected value (between 12 and 18%).

Skewness (a measure of the asymmetry of the distribution of scores) is -0.1219. This value imply a perfectly symmetrical distribution.

Kurtosis (a measure of the flatness of the distribution) is 0.1245. The value is in the range 0-1 and indicate a normal, bell shaped distribution. It indicates that the test is discriminating very well between very good or very bad students and those who are average.

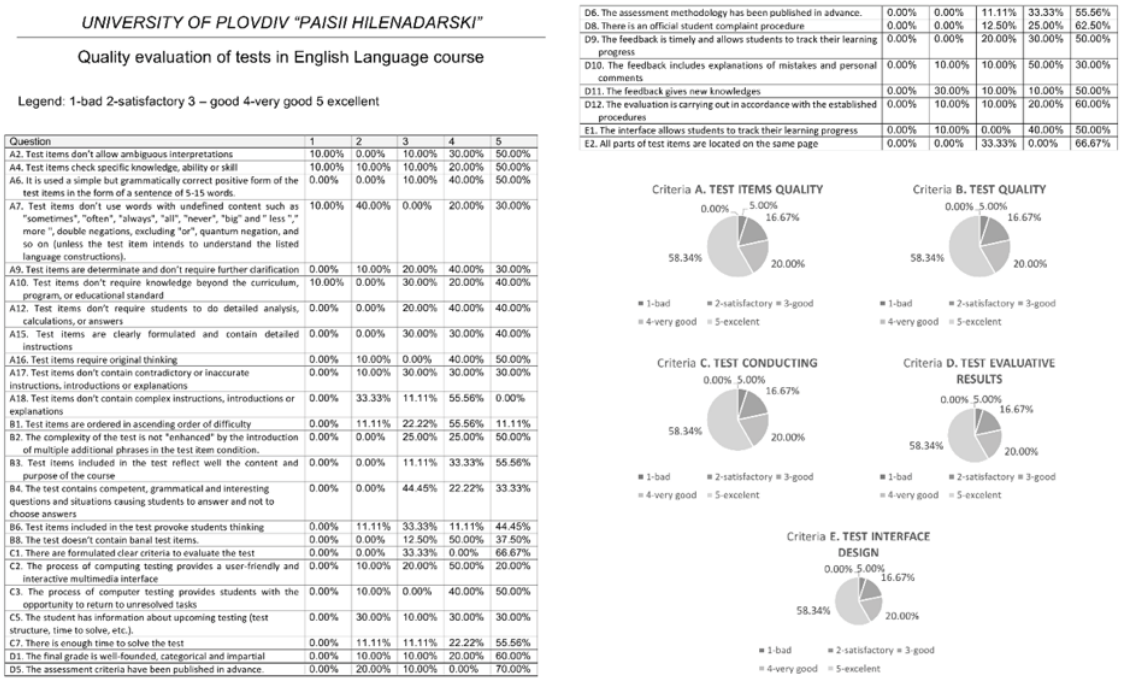
Coefficient of internal consistency is 23.52%. The value is below 64%, the test as a whole is unsatisfactory and remedial measures should be considered. The value indicates either that some of the questions are not very good at discriminating between students of different ability and hence that the differences between total scores owe a good deal to chance or that some of the questions are testing a different quality from the rest and that these two qualities do not correlate well – i.e. the test as a whole is inhomogeneous.

Error ratio is 87.45%. The value cannot be regarded as satisfactory. It implies that less than half the standard deviation is due to differences in ability and the rest to chance effects.

Standard error that estimates how much of the standard deviation is due to chance effects and is a measure of the uncertainty in any given student's score is 12.22%. The value indicates it is likely that a substantial proportion of the students will be wrongly graded in the sense that the grades awarded do not accurately indicate their true abilities.

Figure 5. Analysis of the results of an English pre-test

Questionnaires for evaluating the students' satisfaction are included in each of the courses with added tests. Students fill them in after they complete their training. Figure 6 presents a report of summarised results from the evaluation, carried out as part of the experiment, according to the characteristics of the test for assessing knowledge gained in the English language course.



UNIVERSITY OF PLOVDIV "PAISII HILENADARSKI"

Quality evaluation of tests in English Language course

Legend: 1-bad 2-satisfactory 3 – good 4-very good 5 excellent

Question	1	2	3	4	5
A2. Test items don't allow ambiguous interpretations	10.00%	0.00%	10.00%	30.00%	50.00%
A4. Test items check specific knowledge, ability or skill	10.00%	10.00%	10.00%	20.00%	50.00%
A6. It is used a simple but grammatically correct positive form of the test items in the form of a sentence of 5-15 words.	0.00%	0.00%	10.00%	40.00%	50.00%
A7. Test items don't use words with undefined content such as "sometimes", "often", "always", "all", "never", "big" and " less ", " more ", double negations, excluding "or", quantum negation, and so on (unless the test item intends to understand the listed language constructions).	10.00%	40.00%	0.00%	20.00%	30.00%
A9. Test items are determinate and don't require further clarification	0.00%	10.00%	20.00%	40.00%	30.00%
A10. Test items don't require knowledge beyond the curriculum, program, or educational standard	10.00%	0.00%	30.00%	20.00%	40.00%
A12. Test items don't require students to do detailed analysis, calculations, or answers	0.00%	0.00%	20.00%	40.00%	40.00%
A15. Test items are clearly formulated and contain detailed instructions	0.00%	0.00%	30.00%	30.00%	40.00%
A16. Test items require original thinking	0.00%	10.00%	0.00%	40.00%	50.00%
A17. Test items don't contain contradictory or inaccurate instructions, introductions or explanations	0.00%	10.00%	30.00%	30.00%	30.00%
A18. Test items don't contain complex instructions, introductions or explanations	0.00%	33.33%	11.11%	55.56%	0.00%
B1. Test items are ordered in ascending order of difficulty	0.00%	11.11%	22.22%	55.56%	11.11%
B2. The complexity of the test is not "enhanced" by the introduction of multiple additional phrases in the test item condition.	0.00%	0.00%	25.00%	25.00%	50.00%
B3. Test items included in the test reflect well the content and purpose of the course	0.00%	0.00%	11.11%	33.33%	55.56%
B4. The test contains competent, grammatical and interesting questions and situations causing students to answer and not to choose answers	0.00%	0.00%	44.45%	22.22%	33.33%
B6. Test items included in the test provoke students thinking	0.00%	11.11%	33.33%	11.11%	44.45%
B8. The test doesn't contain banal test items.	0.00%	0.00%	12.50%	50.00%	37.50%
C1. There are formulated clear criteria to evaluate the test	0.00%	0.00%	33.33%	0.00%	66.67%
C2. The process of computing testing provides a user-friendly and interactive multimedia interface	0.00%	10.00%	20.00%	50.00%	20.00%
C3. The process of computer testing provides students with the opportunity to return to unresolved tasks	0.00%	10.00%	0.00%	40.00%	50.00%
C5. The student has information about upcoming testing (test structure, time to solve, etc.).	0.00%	30.00%	10.00%	30.00%	30.00%
C7. There is enough time to solve the test	0.00%	11.11%	11.11%	22.22%	55.56%
D1. The final grade is well-founded, categorical and impartial	0.00%	10.00%	10.00%	20.00%	60.00%
D5. The assessment criteria have been published in advance.	0.00%	20.00%	10.00%	0.00%	70.00%
D6. The assessment methodology has been published in advance.	0.00%	0.00%	11.11%	33.33%	55.56%
D8. There is an official student complaint procedure	0.00%	0.00%	12.50%	25.00%	62.50%
D9. The feedback is timely and allows students to track their learning progress	0.00%	0.00%	20.00%	30.00%	50.00%
D10. The feedback includes explanations of mistakes and personal comments	0.00%	10.00%	10.00%	50.00%	30.00%
D11. The feedback gives new knowledges	0.00%	30.00%	10.00%	10.00%	50.00%
D12. The evaluation is carrying out in accordance with the established procedures	0.00%	10.00%	10.00%	20.00%	60.00%
E1. The interface allows students to track their learning progress	0.00%	10.00%	0.00%	40.00%	50.00%
E2. All parts of test items are located on the same page	0.00%	0.00%	33.33%	0.00%	66.67%

Figure 6. Example of a generated evaluation report for quality of test items and tests

Quality managers at the University of Plovdiv generated TQE documents on the relevant proposed templates for evaluating the quality assurance process. They did so on the basis of the survey data for quality evaluation of test items as part of the experiment. The aim was to obtain summary information in real-time, which allows:

- **monitoring of the planned surveys of professional fields and fields of study;**
- **monitoring of the conducted surveys of professional fields and fields of study;**
- **monitoring of the results of the conducted surveys of professional fields and fields of study.**

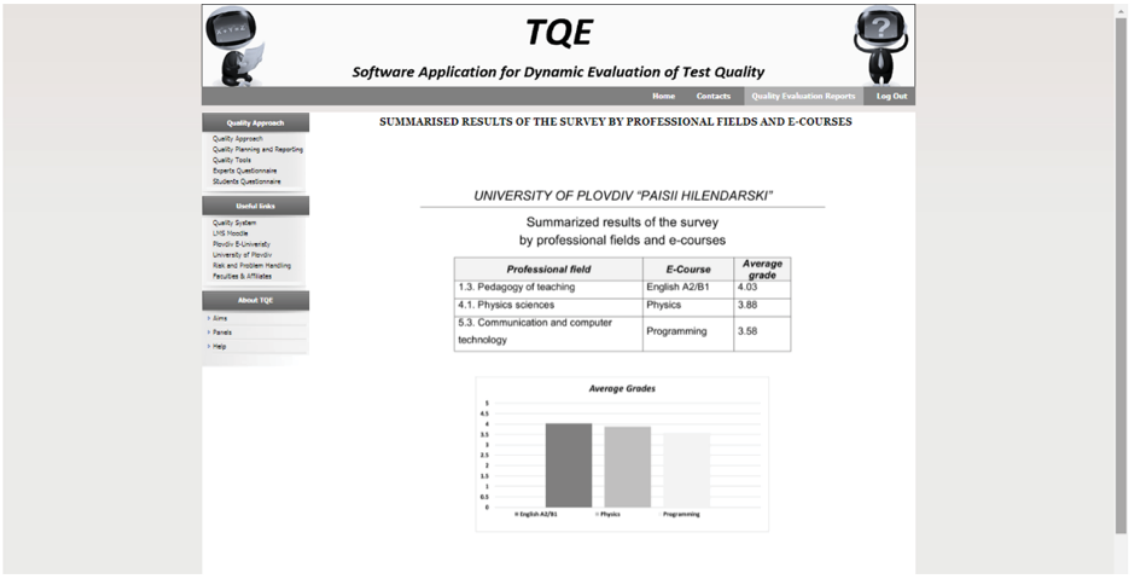


Figure 7. Generated evaluation report for summarised results of the survey (screenshot from TQE)

The report generated through TQE (see Figure 7) shows that the students have evaluated highly the quality of test items and tests conducted within the e-learning courses in the following professional fields: Pedagogy of teaching in..., Physical sciences, Communication and computer technology. Students have given all professional fields an average grade of above 3.

Proposed models and software application cover known approaches for development of high quality educational tests (e.g. Legault, 2017; Totkov, Raikova & Kostadinova, 2014; CITL, 2017; Amouei et al., 2014; Thompson & Levitov, 1985; Pyrczak, 1973; Mark, 1985; Hambleton, Swaminathan & Rogers, 1991; Hamilton, Stecher & Klein, 2002; Cronbach, 1971; Messick, 1989; Professional Testing, 2017). In addition, they provide the opportunity for automated evaluation of the quality assurance process, address the needs of quality-related information of all stakeholders and support quality assurance activities at different levels of generalization in the level of separate online test items, of an online test as a whole, of online tests of an entire course, or of online tests of an academic specialty. TQE proves the possibility of automated quality assurance of educational tests at each stage of their lifecycle from all stakeholders' point of view.

CONCLUSION

The main theoretical contribution of this paper is the proposed comprehensive approach to the automated quality assurance of online tests from all stakeholders' point of view (teachers, students, experts, quality managers, etc.) by assessing the tests' quality at different stages of their lifecycle - from their creation and pre-evaluation to their conduction. The approach is directed towards finding an integrated approach for automated quality evaluation of tests, which leads to reduction of efforts for manual quality evaluation. The proposed comprehensive approach for quality assurance and developed models are applicable for any education institution.

The software application TQE developed according to the proposed approach was put under real-time testing. The experiments are carried out to prove the practical significance and applicability of the created software application.

The current study is limited because TQE retrieves and analyses data stored only in the databases of LMS Moodle. The plans for further implementations are TQE to be developed to be used by each higher education institution, regardless of the type of the relevant university information systems and the diversity of the used LMS.

BIODATA and CONTACT ADDRESSES of AUTHORS

Rositsa DONEVA is a Professor at the University of Plovdiv "Paisii Hilendarski", Faculty of Physics and Engineering Technologies. Dr. Doneva gained her Ph.D. at October, 1995. She has led/taken part in more than 50 national and international projects in the area of computer science, electronic and distance learning, applications of IT in education, etc. The areas of her academic interest are Intelligent Systems, Conceptual Modelling, Software Engineering, Quality Assurance and Evaluation (of Higher Education, e-Learning, Software, Projects, etc.), Object-oriented Programming, Systems and Technologies for distance and mobile learning. Prof. Doneva is the author of over 110 scientific publications and 40 textbooks and learning materials with over 300 citations.

Rositsa DONEVA
ECIT Department, Faculty of Physics and Engineering Technologies
University of Plovdiv "Paisii Hilendarski", 24 Tzar Assen Str., 4000, Plovdiv, Bulgaria
Phone: +359 32 261 440
E-mail: rosi@uni-plovdiv.bg

Silvia GAFTANDZHIEVA is an Assistant Professor at the University of Plovdiv "Paisii Hilendarski", Faculty of Mathematics and Informatics. Dr. Gaftandzhieva gained her Ph.D. at February, 2017. She has taken part in more than 10 national and international projects in the area of e-learning and distance learning, applications of IT in education, etc. Her research areas include e-learning and distance learning, automated evaluation of quality in higher education, distance learning. Dr. Gaftandzhieva is an author of 30 scientific publications in the field of quality assurance (of HE, e-Learning, Projects, etc.), e-Learning, m-Learning, etc. with over 70 citations.

Silvia GAFTANDZHIEVA
Department of Computer Science, Faculty of Mathematics and Informatics
University of Plovdiv "Paisii Hilendarski", 24 Tzar Assen Str., 4000, Plovdiv, Bulgaria
Phone: +359 886 939 820
E-mail: sissiy88@uni-plovdiv.bg

George TOTKOV is a Professor at the University of Plovdiv "Paisii Hilendarski", Faculty of Mathematics and Informatics. He gained him PhD at January, 1979 and DSc. at January, 2005. Prof. Totkov's main scientific interests are in the sphere of computer science (e-learning, computer linguistics, information modelling, etc.) and computational mathematics (approximation, mathematical statistics, etc.). He is the author of more than 200 publications with over 600 citations. He has been a leader of or taken part in more than 50 national and international projects on the mentioned research fields.

George TOTKOV
Department of Computer Science, Faculty of Mathematics and Informatics
University of Plovdiv "Paisii Hilendarski", 24 Tzar Assen Str., 4000, Plovdiv, Bulgaria
Phone: +359 32 261 240
E-mail: totkov@uni-plovdiv.bg

REFERENCES

Amouei, A., Barari, R., Naghipour D., Mortazavi Y., & Hosseini S. Reza (2014). Evaluation of Multiple Choice Questions Quality Trend as Structure and Taxonomy. *FUTURE of MEDICAL EDUCATION JOURNAL*, 4(3), 26-30.

APA (2014). *The Standards for Educational and Psychological Testing*, Retrieved July 2, 2017, from http://www.apa.org/science/programs/testing/standards aspx#overview

Blackboard Help (2017). *Running Item Analysis on a Test*, Retrieved July 2, 2017, from https://en-us.help.blackboard.com

CDC (2017). *Checklist to Evaluate the Quality of Questions*, Retrieved April 12, 2017, from http://www.cdc.gov/healthyyouth/ evaluation/ index.htm

CFATIQC (2017). *Common Formative Assessment Test Item Quality Checklist*, Retrieved July 2, 2017, from https://docs.google.com/document/preview?hgd=1&id= 1vmkf0UAU21u8bGRRwtFb__4kw6NThEDM8WD9NyMgSFU&pli=1

CITL. (2017). *Improving Your Test Questions*. Retrieved July 2, 2017, from http://cte.illinois.edu/testing/exam/test_ques3.html

Cronbach, L. J. (1971). ***Test validation*****. In R. L. Thorndike (ed.). Educational Measurement, 2nd ed., 443–507. Washington, D.C.: American Council on Education.**

Dill, D. D. (2010). "Quality Assurance in Higher Education: Practices and Issues." In P. P. Peterson, E. Baker, and B. McGaw, (eds.). ***International Encyclopedia of Education. Third Edition*****. 377-383.**

Doneva, R., & Gaftandzhieva, S. (2015). Automated e-learning quality evaluation. ***Proceedings of the International Conference on e-Learning*****. Berlin, Germany. ISSN 2376-6698, 156-162.**

EFPA (2013). ***Revised EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests*****, Test Review Model Version 4.2.6, Retrieved July 2, 2017, from http://www.efpa.eu/download/650d0d4ecd407a51139ca44ee704fda4**

EUSHARE (2015). ***European Standards and Guidelines for Quality Assurance in the EHEA*****. Belgium: EUSHARE.**

Gaftandzhieva S. (2016). Automated Evaluation of Students' Satisfaction, ***International Journal of Information Technologies and Security (IJITS)*****, 8(1), 31-40.**

Gaftandzhieva S. (2017). ***A Model and System for Dynamic Quality Evaluation in Higher Education*****. (Doctoral dissertation). Available from University of Plovdiv.**

Gierl, M.J., & Hollis, L. (2013). Evaluating the quality of medical multiple-choice items created with automated processes, ***Medical Education 2013*****, 47(7), 726–733, doi: 10.1111/medu.12202.**

Hambleton, RK, Swaminathan, H, Rogers, HJ (1991). ***Fundamentals of items response theory*****. Newbury Park (California): Sage Publications. 174 p.**

Hamilton, L. S., Stecher, Br. M.,& Klein St.P. (2002). ***Making Sense of Test-Based Accountability in Education*****. RAND.**

ISO 9000:2015 (2015). ***Quality management systems — Fundamentals and vocabulary*****, Retrieved July 2, 2017, from https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en**

JasperSoft (2017). ***Business Intelligence Solutions*****. Retrieved July 2, 2017, from http://www.jaspersoft.com/business-intelligence-solutions**

Legault N. (2017). ***Post-Course Evaluations for e-Learning: 60+Questions to Include*****, Retrieved July 2, 2017, from https://community.articulate.com/articles/post-course-evaluations-for-e-learning-60-questions-to-include**

Machado-da-Silva, F, Meirelles, F , Filenga, D , Filho, M . (2015). Student Satisfaction Process In Virtual Learning System: Considerations Based In Information And Service Quality From Brazil's Experience. ***Turkish Online Journal of Distance Education*****, 15 (3), 122-142. DOI: 10.17718/tojde.52605.**

Mark, D. R. (1985). The Difficulty of Test Items That Measure More Than One Ability, ***Applied Psychological Measurement*****, 9(4), 401 – 412.**

Messick, S. V. (1989). ***Educational Measurement*****. 3rd ed., 13–103. New York: Macmillan.**

Moodle Documentation (2017). ***Quiz statistics report*****, Retrieved July 2, 2017, from https://docs.moodle.org/29/en/Quiz_statistics_report.**

Mutiara, D., Zuhairi, A., & Kurniati S. (2007). Designing, Developing, Producing And Assuring The Quality of Multi-Media Learning Materials For Distance Learners: Lessons Learnt From Indonesia's Universitas Terbuka. ***Turkish Online Journal of Distance Education-TOJDE*****, ISSN 1302–6488, 8(2), 95-112.**

Professional Testing (2017). ***How do you Determine if a Test has Validity, Reliability, Fairness, and Legal Defensibility?.*** **Retrieved July 2, 2017, from http://www.proftesting.com/test_topics/test_quality.php**

Pyrczak, F. (1973). Validity of the Discrimination Index as a Measure of Item Quality. ***Journal of Educational Measurement*****, 10(3), 227-231.**

Rasch (2017). ***Rasch Software*****, Retrieved July 2, 2017, from https://www.rasch.org/rmt/rmt114d.htm**

Saad, S., Carter, G.W., Rothenberg, M., & Israelson, E. (1999). T*esting and Assessment: an Employer's Guide to Good Practises*, Employment and Training Administration (DOL). Washington, DC. Office of Policy and Research.

Thompson, B., & Levitov, J.E. (1985). Using microcomputers to score and evaluate test items. ***Collegiate Microcomputer*****, 3, 163-168.**

Totkov, G., Gaftandzhieva, S., & Doneva, R. (2016). Dynamic Quality Evaluation in Higher Education (with application in e-Learning). ***Proceedings of the First Varna Conference on E-learning and Knowledge Management: Bridging the Gap between Secondary and Higher Education*****, 8-23.**

Totkov, G., Raikova, M., & Kostadinova, H. (2014). ***The test in e-learning*****. Plovdiv: "Rakursi" LTD.**