UDC 378.146, 371.261

L. I. Hladka, Bogdan Khmelnitsky Cherkasy National University

SYSTEM APPROACH TO KNOWLEDGE QUALITY EVALUATION BY MEANS OF COMPUTER TESTING

Hladka L. I.

System Approach to Knowledge Quality Evaluation by Means Of Computer Testing

Article is devoted to the unified development of the methodical approach to generate tests for qualitative monitoring of students knowledge from all forms of learning. The review of online services designed to form a bank of tasks and knowledge testing. In article the structure of the package tests for control knowledge which agreed with the scale of ECTS first proposed. Using online testing system "Online Test Pad", conducted testing of the proposed method of forming a package of tests on the example of the organization of the control of knowledge with subject "Algorithms and Programming".

Key words: Tests, Types of Tests, Differentiated Evaluation, Online Testing System.

Under present conditions of organizing educational process, which foresee a systematic knowledge control, there arises the need for the knowledge control automation through appropriate software. As it has already been noted by researchers, the computer-based testing is an effective checking method [1 – 4]. One of the advantages of the computer-based testing is an opportunity to organize the control without restricting its frequency thus securing immediate obtaining of accurate results of the control instantly after the test completion. Tests differ from traditional knowledge control by the objectivity of learning results measurements since they rely not on teachers' subjective judgements but on objective criteria. Moreover the results of computerized testing can be better analyzed.

The advantages of measurements and formalized knowledge assessment utilizing computer-based testing systems include: the opportunity for the automation of knowledge control and test results assessment; convenience of test preparation; a

great number of questions which cover the subject on the whole; arbitrary test duration; the integration of testing systems with segregated databases etc. Therefore the choice of software plays an important role in performing computer-based testing, both for the test bank formation and the testing procedure implementation. At present online services are rapidly developing and expanding, and their portability, cross-platformity and opportunity for instant application encourage users to their obvious advantageous use.

The aim of the paper is the implementation of the systematic approach to the knowledge quality assessment in the form of computer-based testing.

The problems of the theory of knowledge quality assessment in the form of computer-based testing, the structure and forms of tests, algorithms for processing test results and their interpretation are studied in works of K. Ingenkamp, J. Raven, V. S. Avanesov, Yu. V. Bogachkov, Ya. S. Brodsky, I. E. Bulakh, A. N. Zemlyakov, Yu. I. Ivanov, I. Alekseychuk, N. Talyzin, V. Bocharnikov, N. Shiyan et al. [4; 5; 6; 7].

Let us consider **the systematic approach** to the computer-based testing organization.

We shall define essential stages of the systematic approach to the computer-based testing organization:

1. Preparatory stage.

✓ To form the system of the knowledge control based on the teaching materials of the subject: the number and forms of tests, grading scale.

✓ To devise packages of tests and questions in accordance with the system of the knowledge control: theoretical questions, practical tasks (different levels), tests (the details of compiling the test package are studied below).

✓ To create the test database for self-study in keeping with the devised test package (the details of compiling the test package for self-study are studied below).

2. Specification stage. The checking procedure and the test contents are specified, the preliminary calculation of the control indices and parameters is carried out.

At the specification stage it is necessary to do the following steps:

- The test approbation by the devising teacher by means of self-testing to discover technical or syntactical errors.
- Refinement of the developed test system by a group of experts (for instance, by applying the Delphi Method).

The Delphi Method consists in an independent and multiple questioning of experts. As experts we suggest members of the education board of the corresponding speciality. The first questionnaire contains general questions. After being completed by experts the questionnaires are returned to the test designer and processed. On the basis of the results of the first questionnaire we make the second questionnaire. In this questionnaire experts range certain elements, agree or disagree to the test assessment parameters of the test designers. The results of the second questioning are the foundation for the third questionnaire. The experts consider their previous answers against the background of the collective voice. The third questionnaire is final.

3. The stage of the knowledge control organization in student groups. After testing it is required to process the test results and calculate the following parameters:

- the coefficient of the test complexity Kc (a ratio of the number of students, who did not solve the corresponding test, to the total number of students in the group). At this stage too easy (the value of Kc is close zero) or difficult (the value of Kc is close to one) test questions must be revised and improved;

- the index of discrimination. The value of the index of discrimination is in the range (-1÷+1). In case of a negative value the tests should be sorted out since the negative value of the index of discrimination means that low-achieving students answer this test question better than high-achieving ones;

- the discrimination coefficient is a coefficient which allows the assessment of the question quality. It is a correlation coefficient between scores which the tested student gets for a certain question and his/her total score for doing the whole test. The negative value of the discrimination coefficient proves ill-posedness of the analyzed

question from the viewpoint of validity of estimating the level of learning material assimilation.

Let us define the main provisions of the suggested approach to the test packages formation:

1. Firstly, it is advised to determine the student's knowledge model.
2. It is essential that while designing test packages general requirements to the design of education tests and four test quality constituents should be observed [8].
3. The tests are to have various and multilevel form and take into account an opportunity for their usage during different kinds of the student's actual knowledge control: placement, formative (self-examination and module control) and summative.
4. The design of the test package should be guided by age norms and other standards stated in documents.
5. The application of content validity principles is obligatory. Content validity is recognized if 70% of the course content includes factual knowledge, so accordingly 70% of the test questions must check factual knowledge.

It is recommended to adopt a five-constituent model of the student's knowledge. This model is to recognize the division of knowledge into declarative and procedural.

The content of the test package must be in a full conformity with the purpose of its development as well as the documents which determine the content of the subject.

To take into account the test quality we shall have a look at every component separately.

The intellectual component of the test quality includes:

- a complete coverage of all provisions of the academic course;
- the number of errors (logical, speech) in stating questions and answers;
- coverage by the questions of the subject assimilation levels: terminology proficiency, understanding of laws, decision of practical tasks.

To take the intellectual component into consideration it is necessary:

- to put questions to every provision of the lecture;

- to evaluate the tests before their implementation;
- to make up questions checking terminology, laws, and ability to do practical tasks.

The statistic component of the test quality enables us to consider the possibility of a chance finding of the correct answer: for example, from 10 multiple-choice questions with 5 choices it is possible to choose correct answers in two questions. In order to take into account the statistic component of the test quality it is needed to use no less than four-five choices; to use multiple-response tests, arrangement tests, matching tests, fill-in-the-blank tests; to keep in mind the possibility of a chance finding of the correct answer by setting the lowest passing result at 60% of the possible maximum. For example:

1. How to make macros, that will change SIZE to 5 during precompilation? Task for 1 point, basic level, section “Elements of C language”.

- #define SIZE = 5
- **#define SIZE 5**
- const SIZE = 5;
- #const SIZE 5
- define SIZE 5

2. Which line of code contains errors? Task for 3 points, high level, section “Datatypes in Ci”

```
struct A {
  char a[];  // 1
};
struct B {
  char a;
  char b[];  // 2
};
struct C {
  char a[];  // 3
}
  char b;
};
int main() {
  char a[];  // 4
  return 0;
}
void test(char a[]) { // 5
```

- **// 1**
- // 2
- **// 3**
- **// 4**
- // 5
- in other lines
- no errors

The psychological component of the test quality allows for the term repetition in formulating questions and correct answers; easiness of formal remembering correct answers. Hence, a test package per module is to exceed 100 questions (a selection for every student is about 25 questions); when formulating wrong answers it is advisable to use “correct” phrases. For example:

1. What code can replace postfix increment in line 1?

```
#include <stdio.h>
main() {
  int a = 1;
  int b = 2;
  int * array[] = { &a, &b };
  int ** pp = array;
  (*++pp)++; // 1
}
```

   - a+=1;
   - b+=1;
   - array[0]+=1;
   - **array[1]+=1;**
   - pp+=1;

The discipline and organization component makes it possible to consider the possibility of cheating or prompting. Therefore, there is to be a time limit for test taking (<1 minute for an easy question).

Existing computer testing systems should use as many questions of different types as possible [9]:

- multiple-choice;
- multiple response;
- matching;
- arrangement;
- ranking;
- fill-in-the-blank;
- dealing with images.

It allows to adequately assess the student’s knowledge and, accordingly, establish the level of his/her competence under the requirements of one of the essential world standard constituents – the education and vocational program for

specialist training with the application of various types of control which are carried out for the purpose of:

- placement control – to identify the level of the students' competence in those subjects which have been studied before a particular course or a total level of students' competence acquired during previous period of study;
- formative control – to establish the actual level of students' competence and assess the assimilation of the material for self-study;
- summative control – to determine the level of student's competence in theoretical and practical material of a particular subject.

A great challenge for home educators is the question of ranging summative test questions by the level of difficultness and their compliance with the parameters stated in the corresponding education standard. These problems can be avoided by means of keeping to the following rules:

1. Test questions are to be divided into three categories each determining a particular level of knowledge in the subject: basic, intermediate, advanced.
2. Follow the recommendations of the Ministry of Education and Science of Ukraine on the summative knowledge assessment.
3. Correlate the assessment structure with the credit based modular learning, which is founded on combining modular learning and measurements of the students' academic workload in ECTS credits. ECTS classifies students according to their level of knowledge and, as a result, makes ranging easier. Employing ECTS grading scale for checking the level of knowledge it is necessary to use tasks of different levels while forming the test package (as well as the selection). For this purpose we suggest that the package should be divided into three groups:

- 60% – basic level tasks measured at 2% per every task out of the total score which equals 100% (DE level);
- 25% – intermediate level tasks measured at 6% per every task out of the total score which equals 100% (BC level);

- 15% – advanced level tasks measured at 10% per every task out of the total score which equals 100% (A level).

For instance, if the selection includes 40 test questions, then 24 questions equal 1 point each, 10 questions – 3 points each, and 6 questions – 5 points each.

The main results of the research. To prepare students for testing, the test database for self-study has been formed. While compiling the database we conducted a survey of online services aimed at organizing knowledge control through testing.

We categorize online services for knowledge control through testing into two groups:

1) services which provide for the formation of your own test bank;

2) systems with an open-source test bank.

Table 1 shows the most popular online services to design tests and exercises and organize different types of control (placement, formative, thematic, periodic, summative, self-examination) which may be conventionally referred to the first group.

Table 1

Online Services for the first category for knowledge testing

Service address / Service purpose	Service Possibilities
http://www.classmarker.com/ test design service with an English interface	publishing of tests in the service, test printing
http://learningapps.org the learning support service with interactive modules, the interface supports several languages, including Russian	building of different charts with the in-built Mindmap tool or exporting ready-made charts, forming audio/video content; the possibility for the development of interactive exercises; the possibility to transfer tests to your html page; the possibility for the exchange of interactive tests
http://master-test.net/uk test design service with a Ukrainian interface	has a constructor with a standard functional
http://quizlet.com/ service to design and use flash-exercises and leaning games	the possibility to devise flash-cards of three forms: 1) Flashcard Mode – these flash-cards are similar to the traditional paper form; 2) Learn Mode –

	these flash-cards reflect correct/incorrect answers and omitted questions; 3) Test Mode – this type of flash-cards allows choosing the types of questions: fill-in-the-blanks, multiple-choice, matching, yes/no answers.
http://www.quia.com/ an international English study program, online test constructor	the possibility to design tests using audio files and pictures, learning games
http://onlinetestpad.com a multifunctional online test constructor «Online Test Pad»	a flexible test scoring; the possibility to build graphs and bar charts on the basis of the test results.
http://www.banktestov.ru/ a test design service with a Russian interface	has a constructor with a standard functional
http://www.qreature.ru a test design service	test design using audio/video files and pictures which are stored on other servers

Table 2 contains a few services of the second group which facilitates test design for a wide range of subjects pertaining to programming, and a test bank to ensure different types of control.

Table 2

Online Services for the second category for knowledge testing of subjects related to programming

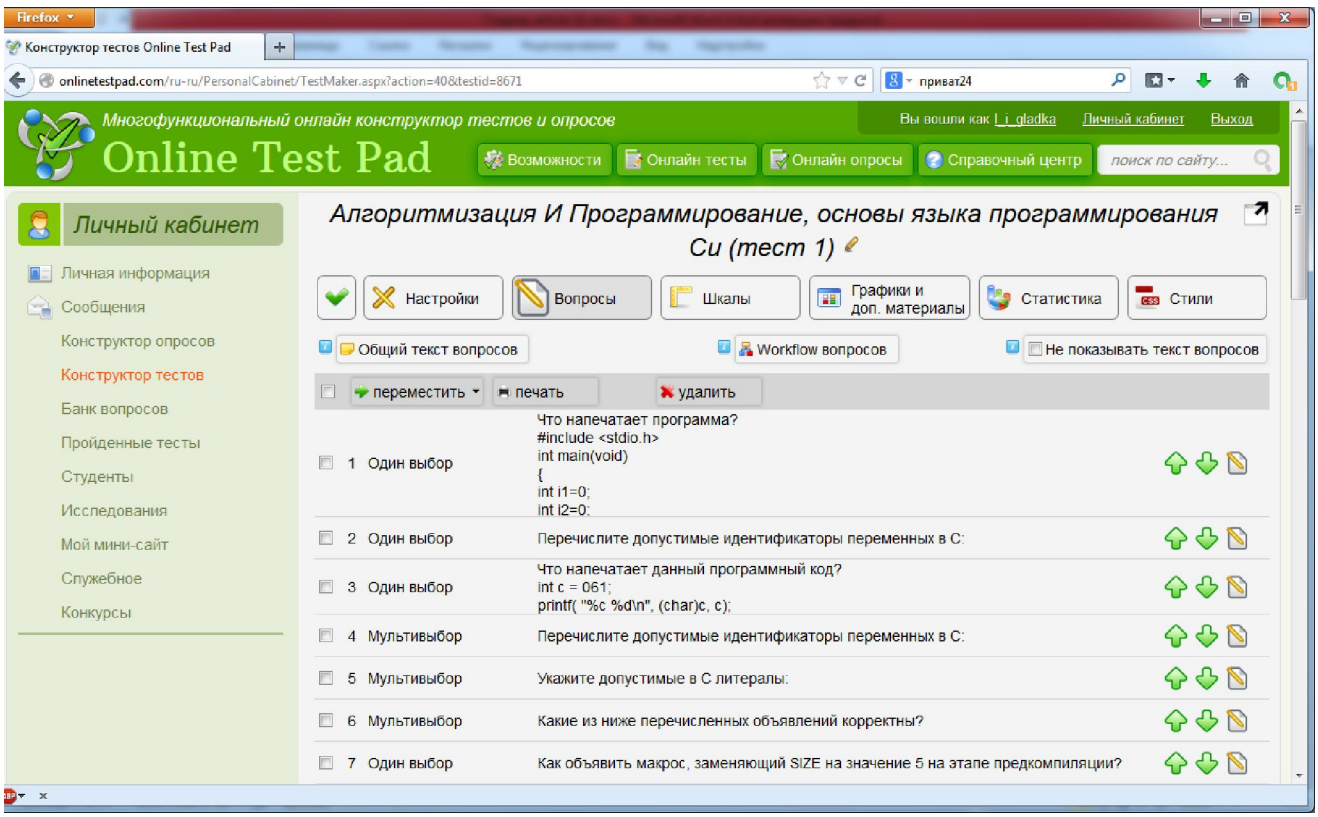
Service address / programming languages with available tests	**Features**
http://www.cquestions.com/ the C programming language	tests classified according to the theme
http://www.pskills.org/c.jsp the C/C++ and C# programming languages, Java, HTML	tests to check the knowledge of programming language basics, not classified according to the theme
http://www.quizful.net/test the C/C++ and C# programming languages, Java, JavaScript, HTML, Ruby, etc.	tests to check the knowledge of programming language basics, not classified according to the theme
https://www.expertrating.com the C/C++ and C# programming languages, Java, JavaScript, HTML, Perl, Ruby, etc.	an international expert knowledge assessment system of IT specialists, free testing only for U.S. residents
www.brainbench.com the C/C++ and C# programming languages, Java, JavaScript, HTML, Perl, Ruby, etc.	an international expert knowledge assessment system of IT specialists, most tests are chargeable

The principles of using these services are similar, that is why every teacher has a chance to choose the best way to provide possibilities for self-study.

In this work testing is organized by means of a multifunctional online test constructor “Online Test Pad”:

1. The test bank has been developed. To approbate the procedure of forming the structure of test packages which is described in the paper, a test package for “Algorithmization and Programming” which includes 60 questions (out of 250 test bank questions) has been developed.

2. Connect the test bank using online service “Online Test Pad” (pic. 1).



Pic. 1 Test design by means of online service “Online Test Pad”

3. Set up the access to the test. Testing is carried out.

4. The test results are analyzed.

The obtained test results are correlated with the actual academic performance. The possibility for building graphs and bar charts on the basis of processed test results in the “Internet Test Pad” allows automation of building every student’s progress path.

So, having studied the basic provisions of the systematic approach to the knowledge quality assessment through computer-based testing the following conclusions may be drawn:

1) the test design needs a systematic approach and consists of respective stages.

2) the development of a perfect test package requires that the teacher has not only links with the course context and academic aims via specific operational terms, but also bears in mind the student's knowledge model to adequately estimate individual abilities.

3) a wide range of online services facilitates preparation of the summative control through self-study.

4) the systematic approach to the test design ensures objective assessment.

References

1. **The Bologna** Declaration on the European space for higher education: an explanation. Retrieved from: http://ec.europa.eu/education/policies/educ/bologna/bologna.pdf.

2. **Obuchenie** i iskusstvennyy intellect ili osnovy sovremennoy didacticheskoy vishey shkoly [Education and artificial intellect or basis of modern didactic higher school]. Donetsk, *DOU*. 2002. 504 p. (rus)

3. **Efimov E.N.,** Denisov M.Y., Zhilina E.V. Sravnitel'nyy analiz obrazovatel'nyh system testirovaniya po kriteriyu funktsional'noy polnoty [Comparative analysis of testing educational systems by functional completeness]. *Upravlenie ekonomicheskimi sistemami*. 2012. № 4. Retrieved from: http://www.uecs.ru/instrumentalnii-metody-ekonomiki/item/1236-2012-04-09-07-23-03. (rus)

4. **Ingenkamp K.** Pedagogicheskaya diagnostika [Pedagogical diagnostics] Moscow, 1991. 240 p. (rus)

5. **Ivanov Y. I.** Razrabotka testov i metody kontrolya znaniy po distsiplinam [Development of tests and methods of control of knowledge in disciplines]. *Kachestvo, innovatsii, obrazovanie.* 2008. № 7. Pp. 11 – 17. (rus)

6. **Mayorov A. N.** Teoriya i praktika sozdaniya testov dlya sistemy obrazovaniya [Theory and practice of creating tests for educational system]**.** Moscow, *Intellect-tsentr*. 2001. 296 p. (rus)

7. **Nathan A.** Thompson, David J. Weiss. A Framework for the Development of Computerized Adaptive Tests. *Practical Assessment, Research & Evaluatio.* Vol. 16. No 1. Retrieved from: http://pareonline.net/pdf/v16n1.pdf.

8. **Tan Jin,** Barley Mak, and Pei Zhou. Confidence scoring of speaking performance: How does fuzziness become exact? *Language Testing*. 2012. No 29 (1). Pp. 43 – 65. (eng)

9. **Hossien** Fallahzadeh New software for computation sensitivity analysis to detect hidden bias for partially order set test statistic in observational studies. *Procedia Technology*. 2012. Vol. 1. P. 225 – 229. (eng)

10. **Sua'rez J.,** Franco A., Santos R. A. Order Statistics and Item Bank Analysis in Computer Adaptive Testing. *Procedia Technology*. 2013. Vol. 7. Pp. 273 – 281. (eng)

11. **Lefransua Gi.** Prikladnaya pedagogicheskaya psikhologiya. Psikhologiya uchitelya [Applied pedagogical psychology. Psychology of a teacher]. SPb, *Praym-EVROZNAK*. 2005. 416 p. (rus)

12. **Hensley G.** Creating a Hybrid College Course: Instructional Design Notes and Recommendations for Beginners. *Journal of Online Learning and Teaching.* 2005. Vol. 1, No. 2. Pp. 66 – 78. (eng)

Гладка Л. І.

Системний підхід до оцінки якості знань у формі комп'ютерного тестування

Робота присвячена реалізації системного підходу до організації процесу тестування для забезпечення якісного контролю знань студентів. Проведено огляд онлайн сервісів, призначених для формування банку завдань та організації тестового контролю знань. Вперше запропонована структура пакету

тестових завдань для контролю знань, яка узгоджена зі шкалою ECTS. Використовуючи онлайн систему тестування “Online Test Pad”, проведено апробацію запропонованої методики організації процесу тестування на прикладі організації контролю знань з дисципліни „Алгоритмізація та програмування”.

Ключові слова: тести, типи тестових завдань, диференційована оцінка, онлайн система тестування.

Гладка Л. И.

Системный подход к оценке знаний в форме компьютерного тестирования

Работа посвящена реализации системного подхода к организации процесса тестирования для обеспечения качественного контроля знаний студентов. Проведен обзор онлайн сервисов, предназначенных для формирования банка заданий и организации тестового контроля знаний. Впервые предложена структура пакета тестовых заданий для контроля знаний, которая согласована со шкалой ECTS. Используя онлайн систему тестирования “Online Test Pad”, проведена апробация предложенной методики организации процесса тестирования на примере организации контроля знаний по дисциплине “Алгоритмизация и программирование”.

Ключевые слова: тесты, типы тестовых заданий, дифференцированная оценка, онлайн система тестирования.

Information about the author

Hladka Ludmyla Ivanivna, Candidate of Physical and Mathematical Sciences, docent of the Department of Automation and Computer-Integrated Systems of Bogdan Khmelnitsky Cherkasy National University. The range of scientific interests includes information and communication educational technologies, database design, data mining.

The article was received by the Editorial Office on 03. 02.2014
The article was put into print on 25.02.2014
Peer review: Guzhva V. M., Candidate of Economic Sciences, Associate Professor