46

Application of expert decision-making technologies for fair evaluation in testing problems

© Hryhorii Hnatiienko[1], © Vitaliy Snytyuk[1], © Nataliia Tmienova[1] and © Oleksii Voloshyn[1]

[1] Taras Shevchenko National University of Kyiv, Ukraine

g.gna5@ukr.net, snytyuk@gmail.com, tmyenovox@gmail.com, olvoloshyn@ukr.net

Abstract. New approaches to evaluation in testing problems are considered. The closed type questions are investigated and the substantiated formulas of the definition of testing results are offered. To formalize testing, the apparatus of decision theory and expert technologies are successfully used. The classification of questions during testing is given. A unified approach to the formalization of different types of test tasks is introduced. It is proposed to use an algebraic approach to determine the evaluation results in the testing problem. The problem statement, the algorithm for evaluating test results with multiple choice, and an example that illustrates this type of testing are given. The problem of evaluation of test results in conformity assessment tasks is considered. The problem statement, formalities for establishing conformity, variants of formulas, and an illustrative example for calculating the degree of similarity between answer variants are described. The problem statement and algorithm for solving the problem of estimating the correctness of the sequence established by the respondents are described.

Keywords: formalization of testing, expert evaluation, decision making, classification of testing problems.

1 Introduction

Testing is ambiguous, debatable, but reliable, powerful, effective, and, for some areas of human life, an irreplaceable tool. In particular, the testing procedure is fruitfully used in programming, engineering, medicine, psychiatry, education, etc. Moreover, for example, pedagogical testing performs several functions simultaneously: educational, diagnostic, evaluative, stimulating, developing, educational, and so on.

Today, there are many opinions about the appropriateness of using tests. On the one hand, tests are seen as a means of positive improvement of the educational process in the direction of its technologicalization, reducing complexity and objectivity. On the other hand, tests are seen as a means of reducing the role of the teacher, and test results are sometimes considered unreliable.

The truth, as always, is somewhere in the middle. It is necessary to use and develop the best features of this approach and reject those that negatively affect the use of the

tool. It should be noted that testing itself is gradually becoming the main form of exams. After all, the tests eliminate the shortcomings of empirical control: the test consists of a number of tasks in a certain direction and a standard known to the teacher that is a sample of complete and correct performance of the task.

At the same time, incorrect approach to the organization of testing or unreasonable assessment can lead to appeals, demotivation of respondents, claims against teachers, allegations of unfair and non-transparent assessment, disputes, and other misunderstandings. Therefore, the study of some non-trivial types of questions used in testing is relevant and necessary.

Despite comprehensive and partly well-founded criticism, testing the educational process has a number of advantages:

• it is a qualitative and objective method of assessment, as procedures for conducting and verifying the quality of test tasks for all respondents in the group are standardized;

• testing is a fairly accurate tool, as the scale of the test depends on the number of questions included in it, and can vary greatly;

• it is a sizable tool that allows you to determine the level of knowledge of the respondent throughout the discipline as a whole and in its individual sections;

• it is a soft tool that puts all respondents on equal conditions, using a single procedure and common evaluation criteria;

• it is a fair method of assessing knowledge, which puts all respondents on an equal footing in the process of control and assessment of the knowledge, significantly eliminating the subjectivity of the teacher: there is information that testing can reduce the number of appeals more than three times;

• is cost-effective because the main costs of this method of evaluation are one-time and are much lower than in written or oral control.

2   Relevance of the research

The problem of studying and improving testing for many years is the focus of researchers of different countries. Various aspects and various problems of knowledge control are studied by many domestic and foreign scientists [1-3]. Today, a significant number of publications are devoted to studying the possibilities and prospects of computer technology testing students' knowledge [4-7]. Models of e-learning development in virtual reality [8], learning through computer games [9] and the concept of lifelong learning are also considered [10, 11]. Fuzzy formulations and heuristics in the assessment [12-15], information communication systems in the educational environment [16], assessment methods in distance learning [17, 18] are investigated. Comprehensive study and improvement of testing processes [6, 14] will help to improve and modernize the quality of the education system as a whole.

3   Classification of question types in test tasks

Many researchers study various aspects of testing. Therefore, there are different classifications of tests that have different degrees of justification. For the purposes of this work, the authors propose a modified classification of test tasks. Test tasks are traditionally divided into two large groups: closed-ended and open-ended.

Today there are about 300 types of test questions. There are different approaches to the classification of test questions. In this article, we will follow this approach to the classification of test questions.

Table 1. Classification of test tasks

Types of closed-ended questions	Closed-ended test questions	Types of open-ended questions	Open-ended test questions
Type 1.1	Dichotomous questions (True-false questions): choosing an alternative answer	Type 2.1	Supplement the answer
Type 1.2	Multiple choice questions with the choice of one correct answer	Type 2.2	Short answer question
Type 1.3	Multiple choice questions with a choice of several correct answers	Type 2.3	Essay question
Type 1.4	Matching questions	Type 2.4	Calculating the result
Type 1.5	Ordering questions	Type 2.5	Determining the interval of values

Problems using open-ended questions will not be considered in this paper. We formalize the first five types of questions in different classes of problems and unify the assessments on these questions.

4   Formalization of question types that are used in the control of knowledge

To optimize knowledge control, it is necessary to use the following procedures:
- unification of assessment;
- ensuring the information unity of the evaluation;
- ensuring systematic evaluation;
- unambiguous interpretation of questions;
- unambiguous formation of the assessment on the basis of answers.

The implementation of these procedures is associated with the formalization of question types, procedures for their evaluation, and criteria for determining a comprehensive assessment [19].

Let the total number of questions that cover the subject area, according to the ontology, and form a logical scheme of testing, be equal n.

They are all divided into k types, depending on what type of answer should be matched the question.

The set of questions is described as follows:

$$Q = (Q_1, Q_2, ..., Q_n) = (Q_1, Q_2, ..., Q_{n_1}, Q_{n_1+1}, ..., Q_{n_{k-1}}, Q_{n_{k-1}+1}, ..., Q_{n_k}) \quad (1)$$

Questions of type 1.1 are formally presented as follows:

$$Q^1 = \left\langle Q_i, A_i(\{0,1\}), 1 \right\rangle, \quad (2)$$

where $Q^1 -$ the subset of questions of type 1.1, $Q_i^1 -$ the questions of the first type, $Q_i^1 \in Q^1$, $i = \{1, ..., n_1\}$, $A_i^1 -$ the set of possible answers to $i -$ th question, 1 - the number of answers to be selected from the set of values A_i^1, $i = \{1, ..., n_1\}$.

Questions of type 1.2 are formally presented as follows:

$$Q^2 = \left\langle Q_i, A_i(a_1, a_2, ..., a_{p_i}), 1 \right\rangle, \quad (3)$$

where $i = \{n_1 + 1, ..., n_2\}$, $a_j -$ possible answers to questions of type 1.2, $j = \{1, ..., p_i\}$, $p_i -$ the number of answer options.

Questions of type 1.3 are defined by the following formalism:

$$Q^3 = \left\langle Q_i, A_i(a_1, a_2, ..., a_{p_i}), d_i \right\rangle, \quad (4)$$

where $i = \{n_2 + 1, ..., n_3\}$, $a_j -$ possible answers to questions of type 1.3, $d_i -$ the possible number of answers, moreover $d_i = const \in \{2, p_i\}$ and determined a priori or $d_i -$ a variable, $d_i \in \{1, p_i\}$.

Questions of type 1.4, where the elements of one set should be placed in accordance with the elements of another set, are called matching questions and formally described as follows:

$$Q^4 = \left\langle Q_i, A_i(A_i^1, A_i^2), f(D_i^1, D_i^2) \right\rangle, \quad (5)$$

where $i = \{n_3 + 1, ..., n_4\}$, A_i^1, $A_i^2 -$ the sets of elements between which we should to establish a correspondence, $f: D_i^1 \to D_i^2 -$ the answer in the form of an established correspondence between the sets A_i^1, A_i^2.

Questions of type 1.5, where the correct sequence of actions or words (answer options, etc.) should be established, are used in ordering test tasks and described by the following formalism:

$$Q^5 = \left\langle Q_i, A_i(a_1, a_2, ..., a_{p_i}), R_i \right\rangle, \quad (6)$$

where $i = \{n_4 + 1, ..., n_5\}$, $a_j -$ answers options to questions of type 1.5 of the set of elements for which it is necessary to establish the correct sequence, that is to build a ranking $R_i = a_{i_1} \succ a_{i_2} \succ ... \succ a_{i_{n_5}}$.

Questions of type 1.1 and type 1.2 are trivial, so let us take a closer look at approaches to formalizing the last three types of questions. Each of the questions of type 1.3-1.5 contains ambiguity and uncertainty, so they should be formalized for the unambiguous perception of such questions.

5 Evaluation of test results in the analysis of the answers of type 1.3 (with multiple choice)

Consider a formal description of multiple choice in closed-ended testing questions. Note that models and methods of multiple choice based on the axiom of non-displacement were studied in the monograph [19].

5.1 Statement of the problem with multiple choice of options

Suppose that we have a set of answer options $a_i \in A, \; i \in I = \{1,...,n\}$, the number of which is equal to n, $n = |A|$. Part of the answers n_1, $n_1 < n$, are correct and they form a subset A^1, $A^1 \subset A$, and the other part of the answers n_0, $n_0 < n$, are false and they form a subset A^0, $A^0 \subset A$, , moreover $A^1 \cup A^0 = A$. In addition, we assume that all answers to the test $a_i \in A, \; i \in I$, are equivalent.

For many test tasks, this statement is natural and logical. For example, to choose from given variants of numbers those that are divisors of a given number. There are many options for this type of task. That is, such an approach takes place in everyday life and the task of its formalization in testing is relevant. The peculiarity of such tasks is that they reflect the well-known truth: "Many men, many minds." Therefore, the decision must be justified to the measure prompted by the logic of its construction, the evaluation policy determined by the test organizers, common sense, and so on.

For the specified statement of the testing problem, it is expedient to apply the algebraic approach to the definition of results of estimation which is successfully used in the theory of decision-making and at the application of technologies of expert estimation. In the algebraic approach, formalization involves the calculation and justification of all possible answers. The maximum number of points for a reliably selected subset of options is equal to B. The number of points for a correctly selected element of a subset of correct answers $b = B/n_1$.

5.2 Statement of the problem with multiple choice Algorithm for evaluating test results with multiple choice of answer options

Therefore, problems that a priori depend on the subjective component cannot be solved without the use of heuristics. The heuristic formula of determination of a point estimation for a choice of answer variants in the form of the set $V \subset A$ which is generated by answers of the respondent is offered. Moreover, the number of elements $v = |V|$ in the set V can be different: from 0 to n, $0 \le |v| \le n.$ Let the number of correct answers selected by the respondent be v_1, and the number of incorrect answers, which he identified as correct be v_0, $v_1 + v_0 = v \le n$. Accordingly, the number of answer options that are not involved in the respondent's response to the question is equal to $n - v$.

Heuristics H1. The value of the fine k for each mismatch of the answer is entered. It is equal to:

- H1.1 - some reasonable coefficient k that reflects the subjective perception of the testing organizers about the "error price", for example, $k = 2$;
- H1.2 - the value of the expression $k = 1 + p_0$, where $p_0 -$ the probability of incorrect answer;
- H1.3 - the value of some function $k = f_{E1}(n_0, p_0)$, set by the experts, which depends on the number of incorrect answers and their probabilities.

Heuristics H2. For the incomplete answer, that is when $v < n_1$, a partial proportional assignment of points is assumed:

- H2.1 - in accordance with the ratio of the received correct answers $v_1 \leq v$, to the total number of correct answers n_1;
- H2.2 - the value of some function $f_{E2}(n_1, p_1)$, set by experts, where the $p_1 -$ probability of obtaining the correct answer.

Of course, a partially correct answer can be guessed by the respondent with a higher probability, but the scores for it are also attributed proportionally lower.

Heuristics H3. For situations when the respondent did not select any answer $(v = 0)$ or all answers are marked as correct, i.e. $v = n$, the penalty is a zero score for lack of selectivity: $B = 0$.

Important and ambiguous is the situation when the respondent did not identify any correct answer. In this case, a respondent may indicate a different number of incorrect answers. Depending on the policy of planning test tasks and the position of the decision-maker, this situation can be described and regulated by additional heuristics.

Heuristics H4.1. In the absence of correct answers, the score is always zero, regardless of the number of incorrect answers.

Heuristics E4.2. When the number of correct answers is zero, fewer incorrect answers are preferable to more incorrect answers.

To formalize this heuristic, we use the lower limit of the described situation. To do this, consider the decision-making situation, which is formalized by a tabular function $c(v_1 = 1, v_0 = n_0)$, when the respondent identified one correct answer $v_1 = 1$ and all n_0 incorrect answers $v_0 = n_0$. According to the described technology, the value of the estimate is determined by the following heuristics.

Heuristics H5. We will assume that the situation $c(v_1 = 0, v_0 = 1)$ is next to the situation $c(v_1 = 1, v_0 = n_0)$ and worsens the resulting assessment by one step, i.e. $c(v_1 = 0, v_0 = 1) = c(v_1 = 1, v_0 = n_0) = c(v_1 = 1, v_0 = n_0 - 1)$. The following situations of determining the resulting assessment are calculated in one of the ways:

- H5.1 - descending function: $c(v_1 = 0, v_0 = i) = c(v_1 = 0, v_0 = i - 1)/k$ for $i = 2, ..., n_0$.
- E5.2 - the situation $c(v_1 = 0, v_0 = n_0)$ is equivalent to the situation $v_1 + v_0 = n$: $c(v_1 = n_1, v_0 = n_0)$ i.e. its consequence is a zero assessment of the respondent. In this

case, estimates for different numbers of incorrect answers $(v_0 = 1,...,n_0)$ with zero number of correct answers $(v_1 = 0)$ are determined as follows:

$$c(v_1 = 0, v_0 = i) = c(v_1 = 1, v_0 = n_0) - i \cdot c(v_1 = 1, v_0 = n_0)/n_0, \quad i = 1,...,n_0.$$

5.3 Example of testing with multiple choice of options

Let us consider the situation of constructing a test task with the following parameters: $n = 5, \ n_1 = 3, \ n_0 = 2.$ Without reducing the generality, we will assume that in the options for answers to the test question, the first three answers are correct, and the last two are incorrect. That is, the question is posed in such a way that true and false answers can be presented in the form of a vector: (1,1,1,0,0).

In the vectors that correspond to the answers of the respondent, the elements will be marked as follows: "1", if the respondent chose the correct answer, "0", if the respondent chose the wrong answer, and "*", if there is no answer.

Heuristics H1.1, H2.1 and H3 are used in the construction of this illustrative test assessment. Thus, heuristics H1, H2, H3 are transformed into such answer options and such estimates are calculated for them.

Answer 1: $a^1 = (1,*,*,*,*) \vee (*,1,*,*,*) \vee (*,*,1,*,*) \Rightarrow c(a^1) = 1/3$;

Answer 2: $a^2 = (1,1,*,*,*) \vee (*,1,1,*,*) \vee (1,*,1,*,*) \Rightarrow c(a^2) = 2/3$;

Answer 3: $a^3 = (1,1,1,*,*) \Rightarrow c(a^3) = 1$;

Answer 4: $a^4 = (1,*,*,0,*) \vee (*,1,*,0,*) \vee (*,*,1,0,*) \vee (1,*,*,*,0) \vee (*,1,*,*,0) \vee (*,*,1,*,0)$
$$\Rightarrow c(a^4) = 1/3/2 = 1/6;$$

Answer 5: $a^5 = (1,*,*,0,0) \vee (*,1,*,0,0) \vee (*,*,1,0,0) \Rightarrow c(a^5) = 1/3/4 = 1/12$;

Answer 6: $a^6 = (1,1,*,0,*) \vee (*,1,1,0,*) \vee (1,*,1,0,*) \vee (1,1,*,*,0) \vee (1,*,1,*,0)$
$$\Rightarrow c(a^6) = 2/3/2 = 1/3;$$

Answer 7: $a^7 = (1,1,*,0,0) \vee (*,1,1,0,0) \vee (1,*,1,0,0) \Rightarrow c(a^7) = 2/3/4 = 1/6$;

Answer 8: $a^8 = (1,1,1,0,*) \vee (1,1,1,*,0) \Rightarrow c(a^8) = 1/2$;

Answer 9: $a^9 = (1,1,1,0,*) \vee (1,1,1,*,0) \Rightarrow c(a^9) = 1/2$.

When constructing test content, you can use heuristics that will have different patterns and correspond to other configurations of answers. Depending on the choice of heuristics, the sensitivity of the function that determines the value of the resulting estimate changes. But the problem of selecting such heuristics is not the subject of this work.

6 Evaluation of test results in the analysis of the answers of type 1.4 (matching questions)

Statements of testing problems in the analysis of the answers of type 1.4 can be varied and their comprehensive review is not the subject of research in this paper. Let us consider only some aspects of research problems that may arise when using this type of testing.

6.1 Statement of the problem of matching

Suppose we have some set $A = \{a_1, ..., a_n\}$ of n elements, which we will call the set of definitions, and a set $B = \{b_1, ..., b_m\}$ of k elements, which we will call the set of values. The test (true, reliable, known, correct, ideal, etc.) correspondence of the elements of the set of definitions A to the elements of the set of values B given by the mapping $f^0 : A \to B$. The task of the respondent is to establish the mapping f^1 between the sets A and B, and which will be as close as possible to the test mapping f^0. Based on the differences between the test reflection f^0 and the reflection f^1 given by the respondents, a reasonable and fair assessment should be determined.

It is clear that this type of test task can have different relations: injection, surjection, and ideally bijection. The type of relation in the test task can be reported to the respondent in advance, before testing, or not be reported, then the student's task becomes more complicated.

6.2 Formalisms for establishing matching

At such statement of a problem various configurations of initial data can take place:
- all elements of the set A must be matched to all elements of the set B;
- all elements of the set A must be matched to some elements of the set B', $B' \subset B$;
- some elements of the set A', $A' \subset A$, must be matched to all elements of the set;
- some elements of the set A', $A' \subset A$, must be matched to some elements of the set B', $B' \subset B$.

The following problem can be formalized using known formalisms:
- injective mapping, when a relationship is established between the elements of two sets, in which two different elements of the set A are never compared the same element of the set B;
- surjective mapping, when for each element of the set B there is at least one element of the set A, such that $f(a) = b$;
- bijective reflection, which is both injective and surjective.

It is clear that the bijection between sets A and B can be established only if they are of equal power.

6.3 Variants of formulas for calculating the similarity measure between the options for matching

For further presentation, we introduce additional notations. Let $A^{0(s)}$ be a subset of objects of the set A, $A^{0(s)} \subset A$, the elements (element) of which correspond to the subset $B^s \subset B$ in the test mapping f^0, $s = 1, 2, \dots$. The value of the index $s = 1, 2, \dots$ depends on the specific statement of the problem and, in particular, on what mapping it is formalized: injection, surjection or bijection. The corresponding subsets given by the respondents will be denoted by $A^{l(s)} \subset A$, $l = 1, \dots, k$, where $k -$ the number of respondents.

The formulas for determining the similarity measures between the test mapping and the mapping performed by the $l -$ th student, $l = 1, \dots, k$, which are proposed in [20], are as follows:

$$\mu_l^{(1)} = 2 \cdot \sum_{s=1}^{S_l} \left(\left| A^{l(s)} \cap A^{0(s)} \right| / \left(\left| A^{l(s)} \right| + \left| A^{0(s)} \right| \right) \right) / S_l \quad , \tag{7}$$

$$\mu_l^{(2)} = 2 \cdot \sum_{s=1}^{S_l} \left(\left| A^{l(s)} \cap A^{0(s)} \right| / \max \left(\left| A^{l(s)} \right|, \left| A^{0(s)} \right| \right) \right) / S_l \quad , \tag{8}$$

$$\mu_l^{(3)} = 2 \cdot \sum_{s=1}^{S_l} \left(\left| A^{l(s)} \cap A^{0(s)} \right| / \left(A^{l(s)} \cup A^{0(s)} \right) \right) / S_l \quad , \tag{9}$$

where the number of elements of the set A^l is denoted by $\left| A^l \right|$; S_l, $l = 1, \dots, k$, $-$ the number of subsets of the set B used by the $l -$ th student in answering the test task.

The fourth formula proposed in [20]:

$$\mu_l^{(4)} = \sum_{s=1}^{S_l} \lambda^{l(s)} \quad , \tag{10}$$

where the values $\lambda^{l(s)}$, $s = 1, \dots, S_l$, $l = 1, \dots, k$, are defined as follows:

$$\lambda^{l(s)} = \begin{cases} 1/S_l, & if \ l - th \ respondent \ set \\ & correct \ mathing \ of \ s - th \ element, \\ 0, & if \ l - th \ respondent \ was \ mistaken. \end{cases}$$

6.4 Example of calculating the degree of similarity between the options for matching

There are different interpretations of the described problem. Consider a situation where the set of definitions A consists of 9 elements: $A = \{a_1, \dots, a_9\}$, and the set of

values B consists of 3 elements: $B = \{b_1, b_2, b_3\}$. The test mapping f^0 looks like $\{a_1, a_3, a_7\} \rightarrow \{b_1\}$, $\{a_2, a_5, a_6, a_8\} \rightarrow \{b_2\}$, $\{a_4, a_9\} \rightarrow \{b_3\}$.

The answers of the four respondents will be presented in the form of a table:

Table 2. Results of the survey on matching

Indices of respondents	Indices of the set B		
	b_1	b_2	b_3
0 (test mapping)	$\{a_1, a_3, a_7\}$	$\{a_2, a_5, a_6, a_8\}$	$\{a_4, a_9\}$
1st respondent	$\{a_1, a_2, a_3, a_7\}$	$\{a_5, a_6, a_8, a_9\}$	$\{a_4\}$
2nd respondent	$\{a_1, a_3, a_5, a_7, a_8\}$	$\{a_2, a_4, a_6, a_9\}$	$\varnothing$
3rd respondent	$\{a_2, a_3, a_4, a_5\}$	$\{a_1, a_6, a_8\}$	$\{a_7, a_9\}$
4th respondent	$\{a_2, a_3, a_5, a_7, a_8\}$	$\{a_1, a_4\}$	$\{a_6, a_9\}$

The cells of the table contain subsets of the elements of the set A, which each of the respondents put in accordance with the subsets of the set B, each of which in this example consists of one element. The order of the elements of sets is insignificant and indicates only the fact of assignment to a subset, not its importance. The sign of an empty set means that the 2nd respondent did not refer to the third subset of the set B any element of the set A.

As a result of calculating the similarity measures given by formulas (7) - (10), we obtain table 3.

Table 3. Values of similarity measures for test mappings and mappings specified by respondents

Indices of respondents	Index of similarity measure			
	$\mu^{(1)}$	$\mu^{(2)}$	$\mu^{(3)}$	$\mu^{(4)}$
0 (test mapping)	0	0	0	0
1st respondent	0,669	0,583	0,617	0,778
2nd respondent	0,417	0,367	0,311	0,556
3rd respondent	0,397	0,333	0,3	0,444
4th respondent	0,278	0,217	0,222	0,333

Based on the similarity measures calculated in this way, it is possible to determine the estimates of the respondents based on the test results. To do this, let us enter additional heuristics.

Heuristics H6. Choosing the similarity measure among those described by formulas (7) - (10).

Heuristics H7. Choosing a formula to translate the similarity measure into a score. For example, $c_l = \mu^{l(\tilde{t})} \cdot M$, where $c_l -$ the score, $\mu^{l(\tilde{t})} -$ selected similarity measure for the $l -$ th respondent using the heuristics H6, $M -$ the maximum value of the rating scale.

7 Evaluation of test results in the analysis of the answers of type 1.5 (ordering questions)

This paper considers a closed-ended problem, namely the problem of arranging list elements in a certain sequence, i.e. determining the order of elements (objects, alternatives, entities), sequence of actions, operations, processes, calculations, chain of events, judgments, etc. In this case, the respondent is offered a list of concepts, phenomena, dates, words, etc., which he must arrange in the correct sequence. Such test tasks occur in various fields, for example:
- establish a chronological sequence of events;
- determine some logical sequence;
- to formulate some definition from a set of randomly given words;
- arrange some numbers in ascending or descending order;
- to restore the order of proof of some theorem;
- write a sequence of calculations when writing program code, which provides the definition of the value of a given formula, etc.

Such tasks help to formulate algorithmic thinking in students, consolidate the relevant knowledge and skills.

Consider a formal description of the ordering of elements in closed-ended questions during testing. Note that the models and methods for determining the competence of respondents on the basis of the axiom of immutability in the ranking of alternatives were studied, in particular, in the monograph [21].

7.1 Statement of the problem of assessing the correctness of the established sequential order

Suppose a set of elements of a complete answer $a_i \in A, \ i \in I = \{1,...,n\}$, is given, the number of these elements is equal to n, i.e. $n = |A|$. The respondent must build a linear (complete) order on this set, i.e. a strict ranking of the given elements of the answer. We will indicate the correct order of elements, which is known to the teacher, and for which the respondent receives the maximum score, by R^0, $R^0 = a_{i_1} \succ a_{i_2} \succ ... \succ a_{i_n}, \ i_j \in I, \ j \in I.$ Thus, the testing procedure can be formalized in the class of ranking problems.

Note that the ability to guess the answer is the main reason for the negative attitude of teachers to the closed-ended form of tasks. To eliminate this shortcoming, even the correction of test scores on guess is used, the essence of which is that from the total

score obtained by each respondent the number that can be guessed is subtracted in accordance with the provisions of probability theory.

Since the number of possible answers for the problems of ranking elements is equal to $n!$, then even if $n = 5$, for example, $n! = 5! = 120$, and it increases significantly with an increasing number of elements to be arranged. That is, the probability of guessing the correct answer is extremely low. Therefore, educators' warnings about the possibility of guessing are unfounded.

7.2   Solution algorithm

When evaluating the tasks of ordering the set of given elements, the dichotomous evaluation of the task is most often used: "Yes" - "No", 0 or 1. Some heuristic evaluation rules are used less often. For example, a correctly completed task is evaluated with three points, the error at the end of the task is evaluated at 2 points, the error in the middle of the task is evaluated at 1 point, and the error at the beginning of the task entails zero evaluation value. It should be noted that sometimes in these tasks it is advisable to establish only a dichotomous, binary assessment. But a large number of test tasks allow for variation of estimates in a wide range.

We will apply the algebraic approach where it is appropriate and justified. That is, the value of the respondent's assessment $C(R*)$ will proportionally depend on the distance of his answer $R*$ to the correct (or - ideal, reference) answer R^0, and symbolically indicate this in this way: $C(R*) = B \cdot (1 - d(R^0, R*) / d^M)$,

where $B -$ the maximum possible score for the answer, $d^M -$ the maximum possible distance, i.e. the distance from the correct answer to the completely incorrect (the opposite to the ideal) answer. For example, when the correct answer is $R^0 = a_1 \succ a_2 \succ a_3 \succ a_4$, then the farthest from it, the "worst" answer is $R* = a_4 \succ a_3 \succ a_2 \succ a_1$.

According to works [21, 22], distances in ordinal (rank) scales, in particular, in rankings, are measured using various metrics, in particular:

- Cook metrics of mismatch of ranks (places, positions) of list elements

$$C^K(R^0, R*) = \sum_{i \in I} \left| r_i^* - r_i^0 \right|$$

(11)

where $r_i^0 -$ the rank of the $i -$ th element of the list in the reference ranking of the elements of the list R^0, $r_i^* -$ the rank of the $i -$ th element of the list in the ranking of the elements R^* specified by the respondent;

- Hamming metrics $d^H(R^0, R*) = \sum_{i \in I} \sum_{j \in I} \left| b_{ij}^0 - b_{ij}^* \right|$,

(12)

where $b_{ij}^0 = 1$, $i, j \in I$, if and only if in the correct answer R^0 there is a relation $a_i \succ a_j$, $i, j \in I$;

$b_{ij}^0 = -1$, $i, j \in I$, if in the correct answer R^0 there is a relation $a_i \prec a_j$, $i, j \in I$;

$b^*_{ij} = 1$, $i, j \in I$, if and only if in the answer of the respondent R^* there is a relation $a_i \succ a_j$, $i, j \in I$;

$b^*_{ij} = -1$, $i, j \in I$, if the respondent defined the following order in his answer R^*: $a_i \prec a_j$, $i, j \in I$.

- Euclid metrics
$$d^E(R^0, R^*) = \left(\sum_{i \in I} \left(r_i^0 - r_i^* \right)^2 \right)^{1/2}, \quad i \in I;$$

- preference vector, the elements of which are the number of alternatives that precede each alternative in the ranking [23].

The maximum possible distances between the standard and the worst ranking:

- for Cook metrics of the form (11) $d^{KM} = n^2 / 2$ for even n; and $d^{KM} = (n^2 - 1)/2$ for odd n;

- for the Hamming metric of the form (12) the maximum distance is $d^{HM} = n(n-1)/2$.

It should be noted that the partial answers of the respondent should also be perceived and fairly assessed. It is clear that this procedure must be justified and formalized. That is, the approach described in this paper can be generalized in case of incomplete answers (when the respondent could not complete the test for technical reasons, did not have time to complete the test, does not know the complete correct answer, but is sure of its fragments, does not want to give a full ranking of a given set of elements). This situation can be considered as a case of incomplete rankings.

8   Conclusions

The paper proposes new approaches to calculating the assessment in testing using different types of questions:
- multiple choice questions;
- matching questions;
- ordering questions.

The approaches proposed by the authors are reasonable and formalized, so they can be applied in different subject areas. A positive feature of the proposed approaches is the transparency of a priori of testing rules set by the organizers, the absence of situations of uncertainty during the evaluation procedure, monotony of the behavior of the function, which reflects the integrated evaluation of the respondent. In addition, the described approaches allow the possibility for further development and improvement. Approaches previously used in practice have been proposed primarily because of their simplicity. However, due to the development of soft computing [24], such approaches can be supplemented, as it is necessary to distinguish, for example, a completely incorrect answer from a partially incorrect one.

References

1. Avanesov V.S. Form of test tasks: textbook / V.S.Avanesov. – M.: Testing Center, 2005. – 153p.

2. Balykina E.N. Questions of construction of test tasks: ucheb.-method. allowance / E.N. Balykina, V.D. Skakovsky // Fundamentals of pedagogical change. Issues of development and use of pedagogical tests; edited by V.D. Skakovskiy. – Minsk: RIVSH, 2009. – Pp. 128–155.

3. Efremova N.F. Test control in education: a textbook / N.F. Efremova. – M.: Litagent "Logos" publishing house, 2007. – 217 p.

4. Povidaychyk M.M. Modern computer technologies for testing students' knowledge / M.M. Povidaychyk, O.V. Povidaychyk // Scientific Bulletin of Uzhhorod National University. Series: Pedagogy. Social work. – 2011. – Vol. 21. – Pp. 160-163.

5. Avanesov V.S. Premises of test forms in e-learning with distraction analysis // Educational technologies. № 3.– 2013. – Pp.125-135.

6. Snytyuk V.E., Gnatienko G.M. Optimization of the evaluation process in conditions of uncertainty based on the structuring of the subject area and the axiom of immutability // Artificial Intelligence. – 2008. – №3. – Pp.217-223.

7. Tsyganok, V.V., Kadenko, S.V., Andriichuk, O.V. Simulation of expert judgements for testing the methods of information processing in decision-making support systems. // Journal of Automation and Information Sciences. 43(12), 2011, pp. 21-32.

8. Morozov, V., Shelest, T., Proskurin, M. Create the model for development of virtual reality E-learning // 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering, UKRCON 2019 - Proceedings, 2019, pp. 1265–1270, 8879876

9. Biloshchytskyi, A., Kuchansky, A., Andrashko, Y., Biloshchytska, S., Danchenko, O. Development of Infocommunication System for Scientific Activity Administration of Educational Environment's Subjects // 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 - Proceedings, 2019, pp. 369–372, 8632036

10. Gogunskii, V., Kolesnikov, O., Kolesnikova, K., Lukianov, D. Lifelong learning is a new paradigm of personnel training in enterprises // Eastern-European Journal of Enterprise Technologies, 2016, 4(2-82), pp. 4–10.

11. Gnatienko G.M. Continuing education as the basis of national wealth of countries // Proceedings of the V International Conference "Modern (e) learning" (MeL), Kyiv, September 9-10, 2010. – Kyiv: Taras Shevchenko National University of Kyiv. – Pp.23-30.

12. Mulesa O.Yu. Methods of taking into account the subjective nature of the input data for the task of voting / O.Yu. Mules // East European Journal of Advanced Technologies. Series: Management processes. – 2015. – Volume 1. – №3 (73). – Pp.20-25. DOI: 10.15587 / 1729-4061.2015.36699 (http://dx.doi.org/10.15587/1729-4061.2015.36699). http://dspace.uzhnu.edu.ua:8080/jspui/handle/lib/1462

13. Mulesa, O. Designing fuzzy expert methods of numeric evaluation of an object for the problems of forecasting / Oksana Mulesa, Fedir Geche // *Eastern-European Journal of Enterprise Technologies*. – 2016. – Vol. 3, N 4(81). - P. 37-43. – Way of Access: DOI : 10.15587/1729-4061.2016.70515. http://dspace.uzhnu.edu.ua/jspui/handle/lib/8849

14. Samokhvalov, Y.Y. Development of the Prediction Graph Method Under Incomplete and Inaccurate Expert Estimates / Cybernetics and Systems Analysis, 54(1), pp. 75-82. (2018).

15. Gnatienko G.M., Snytyuk V.E. Mathematical and software tasks for processing expert information during exams // Artificial Intelligence. – 2010. – №3. – Pp.638-647.

16. Biloshchytskyi, A., Kuchansky, A., Andrashko, Y., Bielova, O. Learning space conceptual model for computing games developers // 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018 - Proceedings, 2018, 1, pp. 97–102, 8526719.

17. Oleg Barabash, Andrii Musienko, Spartak Hohoniants, Oleksandr Laptiev, Oleg Salash, Yevgen Rudenko, Alla Klochko. Comprehensive Methods of Evaluation of Efficiency of Distance Learning System Functioning. International Journal of Computer Network and Information Security (IJCNIS), IJCNIS Vol. 13, No. 1, Feb. 2021. pp.16 - 28, Pub. Date: 2021-02-08. DOI: 10.5815/ijcnis.2021.01.02 http://www.mecs-press.org/ijcnis/v13n1.html

18. Tolyupa, S., Nakonechnyi, V., Tereshchenko, I., Tereshchenko, A. Branch Information Technologies of Quality Management // 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 - Proceedings, 2019, pp. 783–788, 8632105

19. Snytyuk V.E., Yurchenko K.N. Intelligent knowledge assessment management. – Cherkasy, 2013. – 262 p.

20. Gnatienko G.M. Determining the degree of similarity of expert distributors of objects to provide clusters // Bulletin of the University of Kiev. Phys.-Math. science. – 2001. – № 3. – Pp.220–223.

21. Gnatienko G.M., Snytyuk V.E. Expert decision-making technologies: Monograph. – K.: Ltd «McLaut», 2008. – 444p.

22. Voloshin O.F., Mashchenko S.O. Models and methods of decision making: textbook for students of higher educational institutions – 2nd edition. – Kyiv: Kyiv University Publishing and Printing Center, 2010. – 336 p.

23. Hnatiienko H., Tmienova N., Kruglov A. (2021) Methods for Determining the Group Ranking of Alternatives for Incomplete Expert Rankings. In: Shkarlet S., Morozov A., Palagin A. (eds) Mathematical Modeling and Simulation of Systems (MODS'2020). MODS 2020. Advances in Intelligent Systems and Computing, vol 1265. Springer, Cham. https://doi.org/10.1007/978-3-030-58124-4_21. Pp. 217-226.

24. Hnatiienko H., Snytyuk V. A posteriori determination of expert competence under uncertainty / Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019), pp. 82–99 (2019).