INNOVATIVE DIDACTICAL MEANS FOR DEVELOPING AND ASSESSING
STUDENTS’ INTELLECTUAL REFLECTION IN THE HIGH SCHOOL
EDUCATION OF GENETICS

Veselin Vasilev, Teodora Kolarova and Isa Hadjiali
University of Plovdiv “Paisii Hilendarski”
Correspondence to: Teodora, Kolarova
E-mail: teokolarova@gmail.com

ABSTRACT
This article has substantiated the need to use innovative didactical means - cognitive tests with a reflective close for the
formation, development and diagnostics of the intellectual reflection in the study of classical and molecular genetics in the 9th
and 10th grade. There are presented authors’ versions of tests with a reflective close to activate and to measure two properties
which characterize intellectual reflection skills - productivity of reflective thinking and awareness of their own mental
activities of 15-16-year-old students. Each of the developed tests is structured in two subtests. The first subtest is composed by
following the model of criteria-oriented tests and it contains tasks designed to update the pupils’ reflection acts over their own
cognitive activity. The second subtest is marked with the term “reflective close” (end) and it includes questions that require
students’ reasoning on their methods of action applied to solving the tasks of the first subtest. With the results from solving the
first subtest we measure the productivity of the reflective thinking, and the results from the second test are used for judging the
level at which students are aware of the reasons of their own cognitive actions. We determined the qualities reliability and
validity of the constructed tests by the expert evaluation and testing them among a sample of 170 students in 9th and 10th grade.
The results of statistical processing of empirical data were performed with the program SPSS 13.00 and they give us the
confidence to say that the overall set of tasks with the additional to them a reflective close are reliable and valid means for
development and diagnostics of the intellectual reflection in school teaching in genetics.

Keywords: intellectual reflection, cognitive tests with a
reflective close, reliability and validity of tests; high school
genetics education.

Introduction
Today, we can confidently claim that in the Bulgarian
pedagogical science there is a strong tradition in the study of
reflective issues. It has already been established a
comprehensive concept for the multi-faced nature and
manifestations of the psychic phenomenon of reflection, as
well as the possible ways in which it can be purposefully
formed and applied in the general educational practice.
According to this concept with its founder V. Vassilev,
reflection is understood as socially and culturally conditioned
thought process, “deliberately targeted (and meaningful)
towards self-knowledge, which manifests itself in different
forms” (34, p. 99-100). When the subject of analysis is ones’
own cognitive activity and the various actions and operations
of its implementation, an intellectual reflection type can be
distinguished (31, 33, 34, 35).

Two more specific forms are described in the literature, in
which the intellectual reflection can be updated - on the one
hand, it manifests itself as a subject’s awareness of the
“grounds and sources of their own thoughts, actions and
knowledge”, i. e. as “a reflective reproduction of the paths
and ways in which knowledge is obtained (“retrospective
reflection”); on the other hand, intellectual reflection is
“constructing the plan, scheme, model, in which a problem
task will be solved (“prospective reflection” ), done with the
careful reading and application of their own cognitive
capabilities by the subject (34, p. 111; 35, p. 52).

To denote the reflection of intellectual type, the English-
speaking writers, who follow the classical ideas of J. Dewey for reflection, prefer the term “reflective thinking” (“thinking about thinking”). He considered it to be a special form of problem solving, thinking to resolve an issue which involved active chaining, a careful ordering of ideas linking each with its predecessors. Within the process, consideration is to be given to any form of knowledge or belief involved and the grounds for its support (9).

Despite the use of different terms, the opinions of both classical and contemporary authors that explore reflection mainly in the cognitive aspect coincide in discovering its essential characteristics. It is presented as an active and deliberative cognitive process especially directed from the subject towards different manifestations and features of the own mental activity. Similarity is also found with regard to the role that the intellectual reflection has for the development of man and in a particular learner. In summarizing the theoretical concepts in this respect its important functions for mental and personal development of learners of all ages are outlined (7, 9, 15, 16, 18, 25, 30).

The recent empirical developments have shown that the formation of intellectual reflection in learning helps understanding the ways through which specific discipline knowledge is acquired; it allows the learner to realize the course of their own reasoning and creates conditions for self-organization of thinking (24, 31); it improves the self-regulation of learning and increases the efficiency of the learning process (5, 36, 38, 39); it enriches students’ mental culture and “turns” them into subjects of independent activity, able to successfully learn a life (24, 34, 37); it creates in young people the preconditions for self-development and self – improvement (14, 16, 20, 26).

Literature reviews show that in recent years an increasingly clear trend has been formed to specify the abstract knowledge of the reflection in the methodology, technology or diagnostic means with a view to their application in the learning process. However, most of them serve the needs of higher education for closer integration of academic knowledge to real practical situations, to assist the process of professional development and improvement in various spheres of human activity (12, 15, 18, 26, 31).

For comparison, the list of investigations on reflective phenomena in the secondary school education is still very short in the world pedagogic community and in our country as well. Though J. Dewey used to present concrete ideas about the genesis of reflective thinking and to give orientations for its implementation in school education (9), and J. Piaget described the skill of reflection (“reflective abstraction”) in 11/12-year-olds and worked up combinatory tasks for its research (28), there are few contemporary experimental studies on the reflective problems in school science education, particularly in biological education (5, 10, 14, 20, 22, 38, 39). The established programs, methodologies and technologies in this direction, however innovative they are, are not sufficient to explore in depth the characteristics of the intellectual reflection in students, as well as to be more systematically applied in pedagogical practice. We can also add to this fact that so far we have not known any standardized methods for forming and measuring the reflection of intellectual type, which are specially adapted to the genetic educational content, and their data to allow categorization of the respondents in grades or groups.

This paper presents our experience in constructing and standardizing the didactic means that are used for formation and development of intellectual reflection, as well as for its diagnostics in school education in general and molecular genetics in the 9th and 10th grade.

**Materials and Methods**

**Research focus**

Our idea is the scientific understanding of the essential features, the operational structure and ways of forming a reflection of intellectual type to be transformed into tools, in particular in didactic cognitive tests. We plan that they will perform both formational and diagnostic functions in teaching genetics. We assume that both functions can be successfully implemented if the cognitive tests consist of specialized “reflective tasks” and if a “reflective close” (end) is added to them. These tasks should be designed in such a way that to shape actions of students' reflection over their own cognitive activity associated with application of genetic concepts and principles from the educational program material in a conventional and unconventional context. Through including a reflective close we expect to activate and to be expressed externally (in words) students’ thoughts about their own learning actions applied to solving the test tasks.
Having in mind the general signs for identifying a pedagogical innovation – new product or improving an existing one (process, technology, method, etc.), applicability and practical utility (16, 32), we denote the cognitive tests with a reflective close created by us as innovative didactic means. In the bibliographic reference we do not find any data about versions of didactic tests to have been developed and such similar to those that are proposed in our models, and this makes them new. Although we included some parts of the whole format of these tests in our previous studies (13, 21, 22), it can be assessed from their results that the constructed means also match to the two other signs of innovation.

**Research design**

The theoretical basis, on which the created cognitive tests with a reflective close are based on, is the model developed by V. Vassilev of the essential signs and actions of intellectual reflection, described in the language of operations (34, p. 112-113). Part of the operational elements of this model are specified and “turned” in educational tasks for reflection in the teaching material in molecular and classical genetics which is studied in secondary school biological education. The methods underlying in our study are also these of V. Davidov, A. Zack and their followers for the formation of two types intellectual reflection - formal and thoughtful - for students of primary school age. The general intention of these methods is students to understand, with the help of artificial space-combinational tasks, the way of achieving the answers through grouping the tasks by a similarity of the methods used in answering them (8, 37). The idea of A. Zack gives us example cues to construct an additional set of questions (reflective close) to the test material for guiding and helping students to think of how and what way to solve the problems in the test.

When constructing the test material, it is necessary to clarify the basic indicators according to which the acts of intellectual reflection will be monitored and measured in the teaching process. In selecting these indicators we trust to the empirical elaborations on the reflective thinking of learners of various ages, including our earlier studies in that problematic area (10, 21, 22, 24, 31). We fix on the two criteria, which are often preferred by the aforementioned authors – “productivity” and “awareness” of intellectual reflection. The reflection productivity is associated with the transfer of knowledge in conventional and unconventional situations, where the transformation and expansion of their own cognitive experience is required as a reflective act (11). The criterion “awareness of intellectual reflection” refers to the level at which the grounds of one’s own cognitive actions become aware, here we put emphasis on how to solve learning assignments, considered equivalent to a reflective action (37).

In connection with the diagnostics of the selected criteria they need to be operationalized to quantifiable indicators so that they to be representative for the psychic construct of our observation (intellectual reflection) and to allow correct interpretation of the scientific data from its empirical study. Since reflection over cognitive activity can occur with varying intensity in training (12, 17, 18, 25, 27), we assume that according to each of these criteria it can be updated at four possible levels: zero or pre-reflective (L₀), low (L₁), medium (L₂) and high (L₃). Literary review in the discussed aspect gives us reason to believe that these levels are appropriate to be adopted as benchmarks for measuring the productivity and awareness of intellectual reflection. A more complete description can be seen in the evaluation scale that is presented further in the material (Table 3).

**Development of the tests**

Within this paper we present two versions of cognitive tests with a reflective close. One of them – “Knowledge of the genetic processes in the cell and mechanisms of cell division” (denoted T₁ for short) is adapted to the sections “Genetic processes” and “Cell division” which are taught in the school subject of “Biology and Health Education” for 9th grade. In these themes of the curriculum content there is priority in the fields of molecular genetics and cell biology. The test “Knowledge of the properties heredity and variability of a multicellular organism” (T₂) is specified in the section “Heredity and Variability”, which is the subject of study in teaching the subject “Biology and Health Education” for 10th grade. Here, knowledge in the field of classical and modern genetics dominates.

Both tests are conditionally structured as subtest 1 and subtest 2, (ST₁, ST₂) and they correspond to the selected criteria for the diagnostics of intellectual reflection in the teaching of general and molecular genetics.

Subtest 1 is included in T₁ test and it aims to measure the productivity of intellectual reflection in teaching molecular genetics in the 9th grade, and subtest 1 in test T₂ – the same
quality in teaching general and modern genetics in 10th grade. In their essence these subtests possess the invariant signs of the criterial didactic tests and they comply with the basic requirements for their elaboration described in the pedagogical literature (3, 4).

Criterial subtests are built in accordance with certain cognitive goals, which can be seen in greater detail in our previous publications (21, 22). Major reference points for defining these goals are the main criteria and indicators from the taxonomy of J. Gilford that are applied to measure the general cognitive ability, to which the skill of intellectual reflection belongs, too (11).

To achieve these goals, 8 reflective tasks are planned in subtest 1 in each of the two tests (T₁ and T₂). The common characteristic of this type of learning assignments with examples of most of them in training in biology were extendedly presented in previous works of one of the authors of this paper (22) as well as in our newer joint research (21). Here, we will illustrate with examples two of the tasks on reflection included in subtest 1 of test T₁ (see Appendix 1).

Generally, the reflective tasks in the first subtests of the two tests model cognitive situations that require students to use their controlled knowledge in genetics for solving specific training problems. Some of the tasks involve a transfer of the acquired cognitive experience in familiar (conditional) situations in order to prove/reject a claim, to draw a conclusion or to build their self-judgment, including their own texts (see e.g. Task 1, Appendix 1). In those cases the student is required to convert information about the studied genomic structures (genes, alleles, chromosomes, etc.), or processes (replication, transcription, translation, mitosis, meiosis, etc.) from one form to another (figurative, symbolic or semantic). It is believed that it in the process associated with restructuring the acquired cognitive experience and in the transition to new, more diverse forms for its external expression (diagrams, symbols, words) may be caused, and therefore reflective actions to be measured (11, 38).

Other tasks involve wide transfer of the acquired cognitive experience in unfamiliar(unconditional situations) in order to produce hypotheses and predictions for properties of genomic structures, for development of genetic processes or concerning likelihoods for inheritance of certain characteristics. We use an unknown text for the students as the equivalent of an unfamiliar situation and it contains relatively new data for them. These tasks, as they are associated with a broader transmission of the learned genetic concepts and laws, require the student, on the basis of the information in the text, to develop and justify hypotheses about possible changes of genomic structures under the influence of a certain factor, to predict the future course of a genetic process in response to a certain effect, or to suggest the possible target/result of a biological sample experiment (see e.g. Task 2, Appendix 1). We agree with the opinion that the hypothesis and prediction, in their capacity of forms for fulfillment of the prediction, are a series of reflective actions (23).

Three of the 8 tasks of subtest 1 in the two tests are of multiple choice answers in order to provide technological time for their solution. All other tasks (5 in both subtests) require free construction of students’ answer, as written speech activates and facilitates reflection. Each task is evaluated with a minimum of 1 to 5 points, and the raw score is formed by summing the correct answers.

Subtest 2 in the composition of the test T₁ is planned to measure the level of awareness of intellectual reflection in training in molecular genetics in the 9th grade, and subtest 2 in test T₂—the same variable, but in teaching general and modern genetics in the 10th grade. In their essence, both subtests are a reflective addition to the tasks of the above-mentioned criterial tests. This addition reproduces V. Vassilev’s idea for including a reflective close in various psycho-diagnostic tests, which can usefully serve as a tool for reflection forming (34, p. 156).

Here, the basic idea of the reflective close, with the aid of specially formulated questions, is the student to reflect the specific way (strategy, method) in which the problem is solved, i.e. to realize the grounds of their own mental action by taking out (exteriorize) in an external plan the applied method of solution. Although the idea of the criterial subtests is also to stimulate the reflexive thinking of learners on the grounds of their own knowledge and actions, the tasks listed can register mainly the products from the transformation of that knowledge and actions. Speaking in the style of D. Schon, the process and methods through which this transformation is done remain “hidden” in mind in the form of “tacit knowledge” (30). We look for an opportunity for its realization and expression as an external form in the
formation of a reflective close in the criterial subtests.

In each of the two tests (T₁ and T₂) the reflective close includes three questions (see ST₂, Appendix 1). The learners present the answers to them after completion of work on the criterial subtest (ST₁ of T₁/T₂). The first question requires you to group tasks of the first subtest depending on the presence/absence of similarity (formal or meaningful) in their solution. Basing on V. Davidov and A. Zack (8), we accept as formal the similarity in which the student relies on private, non-essential reference points (e.g. ways of presenting the information in the statement of the task, a form for submitting results thereof, etc.). If the student is based on the common, essential guidance for the classification of the tasks (e.g. basic algorithm, a set of similar cognitive actions of solution, etc.), this similarity is considered to be meaningful. The second question puts the requirement students to analyze their own cognitive actions implemented to resolve ST₁ through grouping them by similarity. The third question involves comparing, selecting and evaluating different versions for grouping the learning tasks (Appendix 1).

The questions that present the reflective close are of a multiple choice, and the overall content of those forming the subtest 2 of the test “Knowledge of the genetic processes in the cell and the mechanisms of cell division” (T₁) are given in Appendix 1. Subtest 2 of the test, “Knowledge of the properties heredity and variability of a multicellular organism” (T₂) is similar to the previous, and it differs only in the thematic focus of selected subjects, given in the contents of the included tasks. The answers to each question are coded with points from 0 to 3 which correspond to the assumed by us four levels of reflection (L₀ to L₃). The test score is formed by summing the points proposed for evaluation of the answers to the questions (Appendix 1).

Methods
The assumption, which was expressed a priori, that these tests can function as effective means of diagnostics and development of intellectual reflection in school education in genetics, has been verified with a statistical survey. It aims to verify the developed means for their reliability and validity, in order to make well-founded decisions that concern the application of the tests in the teaching practice in biology.

In accordance with the general recommendations for standardizing psycho-diagnostic and criterion-referenced didactic tests (1, 3, 4, 26), our statistical survey was carried out in the following directions:

1. Output of a standard for successfulness and measurement of the parameters “reliability” and “validity” of the criterial subtests distinguished in each of the two tests T₁ and T₂, where they are labeled with the term “subtest 1”.

2. Measurement of the parameters “reliability” and “validity” of each of the two tests T₁ and T₂, in their full format presented by subtest 1 (ST₁) and the subtest 2 (ST₂).

In both directions, according to the specificity of the test material, we search various indicators to assess its properties reliability and validity. The empirical data necessary for this purpose are result from the applied method of expert assessment (a priori analysis) and of testing the test instruments in the school year 2008/2010 in a sample of 170 students from the 9th and 10th grade (a posteriori analysis). Statistical processing of the data from the a priori analysis and the approbation of the test material was carried out using the program SPSS 13.00 for Windows.

Results and Discussion
Reliability of the criterion-referenced subtests
An important requirement in criterion-oriented measure is to establish the minimum limit, or so-called standard for successfulness (U), by which to judge whether the students have achieved the objectives of the test. Most frequently in the pedagogical practice, this standard is determined by the way of expert evaluation (1), (3, p. 222), on which we also based the norms of both subtests. To determine what should be the minimum number of tasks solved correctly, we give to the experts K. Klauer’s binomial model, which is detailed in the table applied to formalize the standards in criterial tests (3, p. 256), (19). By comparing the proposed standards for successfulness in this table of the test material evaluation, experts believe that the student must have solved at least 5 from 8 reflective tasks in each of the two subtests so that to accept that the student has achieved their cognitive objectives (see Table 1). The experts can assume a task as correctly solved when it can be assessed with at least three of five points. The test score is formed by summing the correct answers of the tasks (see more detail about them in 13).

Criterial subtests reliability was measured in the aspect of internal consistency between the items (reflective tasks), which are conditionally divided into two equal subgroups
Each of these subgroups unites 4 tasks within each of subtests – ST1 of T1 and ST1 of T2. Cronbach’s coefficient alpha is chosen as an indicator of reliability and it monitors the presence/absence of congruence between the general scores of the two halves that form the subtests (6). The empirical values of the coefficient α, obtained after statistical processing of the results from probation of the compound complex of tasks, vary in the range between 0.5 and 0.6 (Table 1). Having in mind the relatively small number of items included in the two subtests, the values obtained show good reliability of the combination of reflective tasks.

The validity of criterial subtests is analyzed in terms of two of its varieties - content validity and internal criterion validity. The content validity is judged by the correspondence between each of the tasks and the cognitive goals of the subtest. Each expert assesses it on a scale where “+1” means that the task corresponds to a certain goal (or set of goals); “0” – it can not be considered explicitly; “-1” - the task does not match to any goal (see detail about the tasks and objectives of the test in 21).

### Table 1
Summarized results from the expert evaluation and aprobation of subtest 1 in test T1 “Knowledge of the genetic processes in the cell and mechanisms of cell division” (ST1/T1) and subtest 1 in test T2 “Knowledge of the properties heredity and variability of a multicellular organism” (ST1/T2)

<table>
<thead>
<tr>
<th>Standards and statistic parameters for reliability and validity of ST1 in T1 and T2</th>
<th>Methods and evaluation coefficients</th>
<th>Subtest ST1/T1</th>
<th>Subtest ST1/T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of experts (n)</td>
<td>9</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Number of tasks (N)</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Standard of successfullness (U)</td>
<td>0 ≤ α ≤ 1</td>
<td>0 ≤ α ≤ 1</td>
<td></td>
</tr>
<tr>
<td>Split – half – reliability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cronbach’s coefficient α</td>
<td>α = 0.5891</td>
<td>α = 0.5233</td>
<td></td>
</tr>
<tr>
<td>Content validity</td>
<td>CVR = 0.910</td>
<td>CVR = 0.880</td>
<td></td>
</tr>
<tr>
<td>Lawshe’s content validity ratio coefficient (CVR)</td>
<td>−1 ≤ CVR ≤ +1</td>
<td>−1 ≤ CVR ≤ +1</td>
<td></td>
</tr>
<tr>
<td>Internal criterion validity</td>
<td>Chi-square test (χ²), ϕ correlation coefficient (ϕcorr)</td>
<td>χ² = 3.84</td>
<td>χ² = 3.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ϕ ≈ 0.382, ϕcorr = 0.843</td>
<td>ϕ ≈ 0.778, ϕcorr = 0.797</td>
</tr>
</tbody>
</table>

The quantification of the assessed conformity was calculated with S. Loshi’s coefficient CVR (3, p. 230), (19, p. 223). After mathematical processing of the data from experts’ evaluation, the values of Loshi’s coefficient CVR were close to 1, which gives us grounds to assert that the content validity of the survey subtests is very high (Table 1). As an indicator of internal criterion validity we used the degree of coherence between the results of each subtest and the total test score as a whole. The numerical calculation of internal congruence of the answers between ST1 and T1 on one side, and on the other side – between the answers of the ST1 and T2 was done using χ² criterion and the correlation coefficient ϕcorr proposed by G. U. Koule (19, p. 244). The data from the applied statistical procedures (Table 2) clearly show that in both subtests χ² values are higher than the critical value χ²₀.₀₅₁ = 3.84 (at level of significance α = 0.05 and error probability p < 0.05), and the absolute values of the coefficient ϕcorr are higher than 0.5.

We can reasonably assume that the criterion validity of both subtests is very good. **Reliability of the complete tests**

The second direction in which we conducted the statistical study involves an empirical examination of the properties reliability and validity of each test T1 and T2, in their overall configuration represented by subtest ST1 and subtest ST2. The statistical parameters, which we used to assess the measured properties, are systematized in Table 2.

The reliability of the formed two tests T1 and T2 was
measured in two different aspects, united in the more general concept of “consistency”. On one hand, we evaluated the internal consistency between all units (items) that constitute each test – the reflective tasks and the questions in the “reflective close”. On the other hand, we monitored whether there is congruence between items within each test, previously divided into two approximately equal subgroups. The first subgroup comprises 6 tasks of subtest 1, the second combines 2 tasks of subtest 1 and three questions of subtest 2.

In the first aspect (noted with the term “internal consistency reliability”), we use Cronbach’s coefficient alpha as an indicator of reliability and it shows the presence/absence of correlation between the answers of test units, as well as the strength of this correlation (6). The second aspect of reliability (noted with the term “split-half reliability”) was measured by the Spearman – Brown’s coefficient r_{ab}, which is preferable for testing of linear correlation between the raw scores of test items divided into two halves. In its two dimensions, reliability was determined after statistical analysis and processing of the results of testing of the test material in that sample. The empirical values of Cronbach’s coefficient α for both tests were between 0.7.

Having in mind the relatively small number of items, these values are an indicator of good reliability (Table 2). The coefficient of correlation r_{ab} is in the range between 0.5 and 0.7 (for α = 0.01, p < 0.01), which describes the strong dependence between the results of the two halves of each test and thus demonstrate their very good reliability.

**TABLE 2**
Summary of statistical evaluation of the selected indicators that characterize the reliability and validity of test T₁ (“Knowledge of the genetic processes in the cell and mechanisms of cell division”) and test T₂ (“Knowledge of the properties heredity and variability of a multicellular organism”)

<table>
<thead>
<tr>
<th>Statistical indicators for reliability and validity of T₁ and T₂</th>
<th>Methods and coefficients for evaluation</th>
<th>Test T₁</th>
<th>Test T₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency reliability</td>
<td>Cronbach's coefficient α</td>
<td>α = 0.7473</td>
<td>0 ≤ α ≤ 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 ≤ α ≤ 1</td>
<td>α = 0.7244</td>
</tr>
<tr>
<td>Split – half – reliability</td>
<td>Spearman – Brown’s coefficient (r_{ab})</td>
<td>r_{ab} = 0.546</td>
<td>0 ≤ r_{ab} ≤ 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−1 ≤ r_{ab} ≤ +1</td>
<td>0 ≤ r_{ab} ≤ 1</td>
</tr>
<tr>
<td>Content validity</td>
<td>Lawshe’s content validity ratio coefficient (CVR)</td>
<td>CVR = +1</td>
<td>1 ≤ CVR ≤ 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−1 ≤ CVR ≤ 1</td>
<td>CVR = 0.846</td>
</tr>
<tr>
<td>Convergent validity</td>
<td>Chi-square test ($\chi^2$), ϕ correlation coefficient (ϕ_{cor.})</td>
<td>$\chi^2 = 36.131 &gt; \chi^2_{0.01/2} = 6.64$</td>
<td>$\phi = 0.461$, $\phi_{cor.} = 0.751$, 0 ≤ $\phi_{cor.}$ ≤ 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2 = 63.091 &gt; \chi^2_{0.01/2} = 6.64$</td>
<td>$\phi = 0.609$, $\phi_{cor.} = 0.813$, 0 ≤ $\phi_{cor.}$ ≤ 1</td>
</tr>
</tbody>
</table>

The validity of the tests was analyzed in two more private aspects: regarding the question how much the combination of the items (tasks and questions) is representative for the research construct, i.e. for the intellectual reflection (content validity); from the point of view to what extent the test results correlate with the results of other standardized tests that measure similar constructs (convergent validity).

Content validity was determined through expert evaluation. For this purpose, a group of teachers, university professors in genetics and biology teachers with high professional qualifications (a total of 9 experts for T₁ and 13 for T₂) assessed to what extent each test meets the author’s intention they to measure the productivity of intellectual reflection and the level of its awareness. The quantitative expression of this congruence was calculated with the coefficient of Lawshe CVR (3, p. 230), (19, p. 223). After statistical processing of data from the expertise, the measured values of this coefficient were equal or close to 1 (Table 2),
which indicates a high content validity of the test material in both its variants.

As a basis for measuring the convergent validity we used results of the same sample of students after solving a test, which we had standardized earlier, and it was designed to diagnose the productivity of intellectual reflection in students from the 9th grade (22). We did the numerical calculation of the dependence between the answers of this test and the results of the newly-constructed tests with the help of the same coefficients used to demonstrate the convergent validity of subtest 1 in T1 and T2. The obtained values of $\chi^2$ criterion (see Table 2) are higher than the critical area $\chi^2_{0.01} = 6.64$, while the absolute values of $\phi_{corr}$ are higher than 0.5 (for $\alpha = 0.01$, $p < 0.01$). These data show a statistically significant correlation between the results of the standardized test that measures the productivity of reflective thinking and the total score in each of the two tests with a reflective close (T1 and T2), which is a sign of their good construct validity.

The next step in standardization of the cognitive tests with a reflective close is related to resolving the issue of how reliable the evaluation system is and it was developed by the authors for intellectual reflection diagnostics. Earlier on, we put the hypothesis that four of the levels at which it is possible to update the reflection - from zero ($L_0$) to high ($L_3$) and they could serve as benchmarks to measure its productivity and awareness. Here, these levels are specified for the selected criteria for evaluation of reflective thinking in solving the constructed tests (see Table 3).

**TABLE 3**
Evaluation scale for the diagnostics of intellectual reflection in solving the test T1 “Knowledge of genetic processes in the cell and mechanisms of cell division” (ST1/T1) and T2 test “Knowledge of the properties heredity and variability of a multicellular organism”

<table>
<thead>
<tr>
<th>Levels of reflection (L)</th>
<th>Criteria of intellectual reflection diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reflection productivity</td>
</tr>
<tr>
<td>$L_0$ (zero)</td>
<td>Less than 5 solved tasks in $ST_1/T1$ and $ST_1/T2$</td>
</tr>
<tr>
<td>$L_1$ (low)</td>
<td>5 solved tasks in $ST_1/T1$ and $ST_1/T2$</td>
</tr>
<tr>
<td>$L_2$ (medium)</td>
<td>6-7 solved tasks in $ST_1/T1$ and $ST_1/T2$</td>
</tr>
<tr>
<td>$L_3$ (high)</td>
<td>8 solved tasks in $ST_1/T1$ and $ST_1/T2$</td>
</tr>
</tbody>
</table>

Note: The number of points for answering the questions in ST2/T2 is given in Appendix 1 of this article. However, they are not presented in the sheets that students get to work on at the test. Answers to questions in ST2/T2 are valued similarly.

It is necessary to specify that according to the constructed evaluation scale the productivity levels of intellectual reflection are determined depending on whether the standard for successfulness of both criterial subtests in T1 and T2 has been achieved (at least 5 tasks solved correctly). The higher productivity ($L_2$ to $L_3$), is recorded by solving a greater number of reflective tasks beyond the verge for successfulness. The level of awareness of intellectual reflection is assessed with the intensity of the ratiocinations given by the students on their own cognitive actions in solving the test tasks. Each of these levels of awareness is measured in dependence on whether the student has achieved the standard of successfulness in the first subtest and on the average number of points for answering the questions in subtest 2. Here, those answers in which a student is aware of the substantial similarity between the ways of solution of tasks and gives reasons for their significance to students’ learning are estimated higher (levels $L_2$ to $L_3$).

As an indicator of reliability of the presented scale we used the consistency between the experts and authors’
assessments of the ways for measuring the indicators that characterize the proposed criteria for diagnostics of intellectual reflection. For this purpose, the experts are required to give their agreement (“+1”), hesitation (“0”) or disagreement (“-1”) with the authors’ assessment for recording the indicators in numerical values. This aspect of reliability, noted with the term “inter-rater-reliability”, was determined using Holsti’s composite reliability coefficient CR, which shows what part of experts have a positive assessment for a given category. The coefficient is calculated with the formula: \( N \times \text{average number of positive assessments} / (1 + (N-1) \times \text{average number of positive assessments}) \), where \( N \) is the number of experts. The values of the coefficient can vary in a range between 0 and 1. After replacement of the data of expert assessments in the above formula, the values of CR, which have been measured for the indicators of the criterion “productivity” of reflective thinking are getting closer to 1 (CR for \( T_1 = 0.996 \), CR for \( T_2 = 0.998 \)). Regarding the indicators recorded for the criterion “awareness of reflection” we also found a high degree of congruence between the experts’ and authors’ assessments for diagnostics of the selected value (CR for \( T_1 = 1 \), CR for \( T_2 = 0.993 \)). The data obtained give us the confidence to say that the reliability of the evaluation system, which has been developed to measure the productivity of intellectual reflection and awareness of their own mental actions of learners in training in genetics, is high.

Conclusions

This study presents the results of an innovative attempt to be highly operationalized the abstract knowledge for intellectual reflection in didactic means – means for its purposefully formation and diagnostics in school education in genetics. The developed versions of cognitive tests with a reflective close have been verified through expert evaluation and pre-tested in a representative sample of students in 9th and 10th grade. Statistical processing of the recorded data shows that the constructed means are good indicators of reliability and validity. Therefore, the whole set of tasks together with the added to them reflective close can function as an effective instruments for activating the reflection of intellectual type in training in classical and molecular genetics. There is reason to assert that besides forming potential they also possess diagnostic capabilities. The evaluation scale made for this purpose allows the tested subjects to be grouped into categories according to productivity of their reflective thinking, and depending on the level at which they understand the ways of performing their own cognitive actions. In our view, such grouping allows an analysis of correlations between the selected indicators that characterize the acts of intellectual reflection, and on this basis, the development of specialized technology oriented to its active formation and development in various degrees of biological education.

Finally, we would like to emphasize that the proposed variants of tests with a reflective close are specially made for 15-16-year-old students and adapted to the content in molecular and general genetics in 9th and 10th grade. We tend to believe that the extension of the Bulgarian pedagogical tradition in studying the reflective phenomena is very likely to have more and more diverse means of formation and measurement of intellectual reflection in other subject-matter, in another stage of learning in biology. We hope that this idea will find more adherents among the teaching community, because reflection is a quality and efficient mechanism for improving, and this is what modern education needs.

Acknowledgment

This work is supported by NSF, Grant № D002-345..

REFERENCES

APPENDIX 1

TEST "KNOWLEDGE OF THE GENETIC PROCESSES IN THE CELL AND THE MECHANISMS OF CELL DIVISION (T1)

SUBTEST 1 (fragments)

Task 1. Compose a short text, which reflects the relationship between those listed in A. – F. cellular components and their participation in the transcription process:

A. mRNA  D. Nucleus
B. Activated ribonucleotides  E. RNA polymerase
C. Division of DNA (structural gene)  F. ATP

Write your text (at least 5 sentences) in the answer sheet.

Task 2. The text reflects data for a scientific experiment conducted by researchers M. Meselson and F. Stahl. Based on the information contained in the text and Figure 1, suppose (flattering suggestion for) the goal of the described experiment?

Text:

In 1958 M. Meselson and F. Stahl performed the following experiment. Initially they grow bacteria Escherichia Coli in medium containing heavy isotope of nitrogen /15N/, which includes the synthesis of nitrogen bases in DNA. Then carry the bacteria in medium containing conventional /chat/ isotope of nitrogen /14N/. After replication, Meselson and Stahl subjected to density centrifugation on daughter DNA from the first generation and it appears that it is homogeneous and has a density intermediate between heavy and light DNA. Therefore, one daughter DNA chain is heavy (contains 15N), and the other lighter (from 14N).

A. To prove the role of isotopes of nitrogen for the synthesis of purines and pyrimidines.
B. To prove the mechanism of replication in prokaryotes.
C. To prove the genetic role of DNA for storage and transmission of hereditary program.
D. To prove the semi-conservative mechanism of replication.
E. To prove the role of nitrogen’s isotopes as a method for genetic studies.

In answer sheet circle the letter of the responses are most appropriate.

SUBTEST 2

Question 1. Groups according to how you can divide the tasks?

A. two - those who are elected and require a response encirclement of true statements, and those with free response (1 pt);
B. The four - according to similar actions, which require for their solution (formulation of assumptions (hypotheses), forecasting a result, aggregating information and extracting dependencies) (3 pts);
C. five - according to the processes which are reflected in terms of tasks (replication, transcription, translation, mitosis and meiosis) (2 pts);
D. two - according to the way information is presented for objects and processes in tasks (text and figures) (1 pt);
E. tasks can be divided into groups (0 pts).

In answer sheet circle the letter of response that you consider most appropriate.

Question 2. Which of the tasks included in the test in resolving their similar actions?

A. no similarity in solving tasks (0 pts);
B. 1st and 8th include the formulation of assumptions (hypotheses), 2nd and 3rd call forecasting result of the biological experiment, 4th and 5th - formulate questions requiring integration of information blocks and 6th and 7th - drawing up the text for displaying the relationships between appropriate.
cell structures and processes (3 pts);
C. 1st, 5th and 8th include an assessment of a range of answers given and the choice of true statement, and the remaining tasks of constructing free-response (1 pts);
D. 1st and 7th provide necessary information mainly symbolic (by figures), and other tasks - using a text (1 pt);
E. 1st and 5th include the formulation of assumptions (hypotheses), 2nd and 8th - removal of the arguments to prove the allegations, 3rd and 4th - analysis of the connections (relationships) between individual elements and 5th and 7th - aggregation of information (2 pts).

In answer sheet circle the letter of response that you consider most appropriate.

Question 3. I think the most important way that can be grouped educational problems in biology is:
A. thematic focus as reflected in terms of tasks: genetic processes in the cell (replication, transcription and translation) and cell division (mitosis and meiosis) (2 pts);
B. according to what is the way to present the correct decision (choice of correct answers among those proposed construction or self-response) (1 pt);
C. according to a similarity in the way you (the group of actions or algorithms) for a decision (3 pts);
D. I do not know (0 pts);
E. according to how information is presented in condition of tasks (mostly figurative or through text) (1 pt).

In answer sheet circle the letter of response that you consider most appropriate.