

ASSESSMENT BY CONSTRUCTED-RESPONSE ITEMS IN TEACHING OF BULGARIAN LANGUAGE BASED ON DATA OF NATIONAL STUDY (BULGARIAN RESEARCH) AND INTERNATIONAL STUDY (PIRLS 2006 – PROGRESS IN INTERNATIONAL READING LITERACY STUDY)

Tatyana Angelova
Sofia University

Abstract. The main theme of the text is assessment by constructed response items in Teaching Bulgarian Language based on Data of National Study and International Study. The problems connected with creating of standards settings in school education are discussed.

Keywords: teaching of Bulgarian language; data of study PIRLS 2006

0. A few introductory words

The introduction of state educational standards for the content of teaching of the Bulgarian language as a native language in secondary schools (2014-2020) reflects the strife for modernization in keeping with the changing educational context in unifying Europe. New curricula have been designed and new textbooks have been written for the teaching of the Bulgarian language in accordance with the new state standards. Their implementation into everyday practice breeds the necessity for the assessment and measurement of students' achievements and necessitates the constant control on the quality of teaching of Bulgarian as a native and national language throughout schools in Bulgaria. Invaluable prerequisites for the achievement of high quality teaching of Bulgarian and successful control is the creation of a system for external assessment.

According to the normative documents of the Ministry of Education and Science, assessment is a process for the establishment of the achieved results and for the grading of students. It is personal for everyone assessed. It can be of two types – *internal assessment* accomplished by the teacher and *external assessment*, in which assessors can be: a school committee, officers from the Ministry of Education and Science, officers from the regional inspectorates affiliated to the Ministry of

Education and Science, national committees for the assessment of state school-leaving/matriculation examinations.

The concept of external assessment is well formulated in the normative documents. Good experience has been gained in applying external assessment in educational practice. We will mention here the administration of the two national test examinations after the completion of the seventh grade in specialized schools (2000 and 2001) containing a section *Bulgarian language* and the pilot matriculation exam (2002) in the course of implementing the Modernization of Education in Secondary School (a reform which was financially supported by the World Bank). Evaluation is applied in relation to the nuclear areas of the Cultural-educational field *Bulgarian language and Literature*: socio-cultural competence, linguistic competence, socio-cultural and linguistic competence in oral and written communication.

External assessment of the teaching of Bulgarian is deemed constructive and fully functional as a regulating mechanism for the quality of teaching on condition that:

- it is applied systematically;
- fully standardized assessment procedures are applied (with approbated measuring characteristics);
- new state educational requirements for assessment are established, which will provide the blueprint for criterion orientation of the tasks and items for assessment;
- assessors are trained how to assess open-ended items and questions;
- the procedures for processing the information are digitalized and the data are available to all professionals in the field;
- Full transparency for all interested parties of the whole process is guaranteed.

The transformation of external assessment in the teaching of Bulgarian into a properly functioning and fruitful mechanism presupposes thorough knowledge and full understanding of all foreign experience in the field and requires that practicable ideas be borrowed from that experience which can only be applied in full recognition of the specificity of our educational traditions and needs.

The international contrastive research in education gives new directions in the development of all processes relating to external assessment – theories, instruments, tendencies, etc. A key role in their organization play the International Association for the Evaluation of Educational Achievement <http://www.iea.nl> and Lynch School of Education, Boston College <http://pirls.bc.edu>. The international surveys (TIMSS, PISA, PIRLS, etc.) brought to the fore the understanding that no adequate and objective evaluation of students' achievements at different age levels and grades can be accomplished on the basis only of standardized tests, no matter how perfect, reliable and well-designed they are. The quality of education in the framework of this strategy is measured not only by registering the acquired and demonstrated knowledge of students, but also through measuring the so-called contextual and procedural variables, i.e. those which reflect the quality of the school, "the good school", the climate in the school, the satisfaction of both students and parents with what the school has to offer, the

identification of both teachers and schoolchildren with ‘their school’, the cooperation between teachers and students, team work and so on.

Bulgaria participates for the second time in an international survey of literacy and reading skills of pupils in the fourth grade PIRLS’2006. The first cycle was completed in the period 2001 – 2004. Bulgaria ranked fourth among the participant countries. The second cycle is in its full swing now; it is scheduled for 2004-2008.

Inspired by the valuable experience in external assessment of both native and foreign institutions and the needs of our educational practice, a team of specialists from three faculties of Sofia University (Faculty of Slavic Studies, Faculty of Mathematics and Informatics and Faculty of Pre-school and Primary Education) set on to develop a project “The development of a strategy for external assessment of the degree of achievement of the State Educational Requirements for Educational Content (SER for EC) at the end of the primary educational level (IV grade) and at the end of the basic educational level (VIII grade) in the cultural-educational areas “Bulgarian language and literature” and “Mathematics, informatics and information technologies”.

The team which won the management of the project in the contest launched by Open Society Institute included various experts from Faculty of Mathematics and Informatics; Faculty of Preschool and Primary Education (FPPE), Faculty of Slavic Studies); international experts: Anica Alexova (Macedonia), Johanna Crighton (The World Bank), Sauli Tacala (Finland).

The serious and profound work which was accomplished for the period 2004-2005 was summarized and published in a volume of more than 1000 pages. Special parts of this report were devoted to the current state of external assessment in Bulgaria, to the international practices of external assessment, to the methodology of test design, to the Framework for assessment and test specificity of the test in the Bulgarian language and Bulgarian literature, in Mathematics, Informatics and Information Technologies, including also the designing of sub-tests in each of the described school subjects. The attention was focused on the creation of instruments which could be fruitfully applied to the survey of the educational milieu (school, schoolchildren, etc) - by questionnaires.

The topic of the current report focuses on open-ended test items as constituent elements of national and international research on assessment.

1. Nature of the test items used in the assessment of the education in the Bulgarian language

Before we describe open-ended test items, let us make a little clarification to help us in our argumentation. The clarification concerns *a test item* and *a non-test item or task* in the teaching of the Bulgarian language.

A test item in Bulgarian:

– is included as an element of a system of test items with a clearly defined educational objectives. The objectives are described in a specification containing the areas of study and fields of knowledge to which each test items applies. A test item

is part of a test variant which assesses the component *language* (communicative competence in Bulgarian) or the component *metalanguage* (skills for the description of the communicative competence in Bulgarian);

- is an instrument for the measurement of a given skill in Bulgarian (for example, detecting a spelling mistake);

- is used to measure and assess what has been achieved by each student at a specified level of proficiency in accordance with the specification which includes the area of knowledge and respectively the levels of analysis, synthesis, application and so on, relating to the subject the Bulgarian language.

- meets the requirements set for the subject *Bulgarian* which is assessed and measured;

- meets certain test requirements (for reliability, validity and difficulty);

- is fitted to a scale of measurement;

- is easily subject to standardization and digitalization

2. Test items are designed to measure and assess different levels of receptive expertise and different skills for producing short texts. On the basis of this distinction test items are grouped into large sets:

A. Test items which offer pre-given answers among which the pupil chooses only one correct answer, the others being distracters. They are best-suited for a certain level of receptive expertise: the skill to recognize linguistic facts at different levels (morphological, syntactic, etc.); skills to detect errors arising from deviations from the established norms of the literary language; skills to detect and define different linguistic means which are used to achieve cohesion and coherence, etc. Such items are generally termed **multiple-choice** tasks/items.

B. Test items which require that the pupil being assessed produce an answer in the form of a short text. Such items are usually called open-ended test items and are also known in Anglo-Saxon literature as constructed response questions.

3. Characteristics of constructed response questions in view of their application in the assessment of education in the Bulgarian language¹⁾

a. As was already mentioned, constructed response questions (CRQ) require the *construction* of a short answer, used to measure the level to which a certain skill has been applied to the completion of the task, relating to certain aspects of the content of what is taught in the subject Bulgarian.

The instruments in the national survey (NS 2004-2005) covered the following content areas corresponding to the State Educational Requirements for the Cultural-educational area *Bulgarian language and literature*:

- Socio-cultural competence – 15%²⁾;

- Written communication – 55%;

- Linguistic competence – 30%;

- Literary competence – 20 %;

- Interaction with a literary text – 25%.

All constructed response questions related to the nucleus *Written Communication* since it is of paramount importance to assess the skills of schoolchildren to use the written medium of their native language.

The international survey (IS 2006) includes two groups of texts: five informative texts and five literary texts. On the basis of these texts both constructed response questions and multiple-choice questions have been designed. The questions are designed so as to measure and assess the students' reading and comprehension skills or generally speaking their literacy, i.e. the level of literacy of students in the fourth grade. (See C for details).

Unlike the writing of a study composition, constructed response questions require the expression and justification of personal opinion, the formulation of ideas and conceptions relating to the content of educational material, its depth of knowledge, understanding, etc.

A wide range of real world stimuli is used, including authentic texts from advertisements, announcements, etc. Designed educational texts are also used such as study texts, tables, graphs, schemes, pictures, etc.

These items have been graded (classified) according to specific criteria – specific rubrics for assessment are used (see under criteria).

b. This type of item is based on a certain stimulus and includes a series of questions organized in an ascending order of complexity.

c. Test items of this type are used to assess skills of high recognition level such as:
– Comparison, contrast, establishment of cause-and-effect relations, identification of exemplar cases or identifying contradictory points of view;

– categorizing or summarizing information;

– constructing schemas, data diagrams;

– Deduction, drawing conclusions, explaining a supposition, inference;

The national survey contains in the form of content areas the following:

– Recognition and understanding of facts (level A);

– Application following a pre-given model (level B);

– Application without a pre-given model (level C);

– Communication. Argumentation (level D)

The items are correlated with reference to the number of points scored for them:
A – 15%, B – 15%, C – 20 % и D – 50 %.

For example the item from NS 2004 (see figure1) about juxtaposition of the two literary texts relates to level D and assesses skills for comparison and contrast and inference-making.

The international survey is designed to measure and assess the basic reading skills:

– Focus on and Retrieve Explicitly Stated Information (facts);

– Make Straightforward Inferences;

– Interpret and Integrate Ideas and Information (interpretation);

– Examine and Evaluate Content, Language and Textual Elements (analysis).
PIRLS 2006 (p.p. 13 – 17).

d. Open-ended test items might require:

- Short answer – a single expression or a few expressions;
- Long answer- a few sentences in a coherent and cohesive text.

e. The assessment of the completion of the task is carried out in accordance with special guidelines and specific rubrics. The rubrics allow for the meaningful hierarchization of the different levels in the completion of the task. Depending on this the answers can be: totally acceptable, partially acceptable and unacceptable, respectively indicating full comprehension, partial comprehension and non-comprehension of what has been read. The rubrics are supplied with a sufficient number of illustrative examples and samples on the basis of which the assessors take their decision as to how many points to score for a given answer. (See assessment criteria).

4. Assessment criteria for test items of the constructed response questions type. Depending on the relative weight of each item students are awarded marks which can easily be scaled and graded, which necessitates the description of criteria for scoring and grading the whole system of items and each item separately.

a. In the instruments comprising the national survey in relation to task of choosing one of two texts depending on the moral of the text, the type of characters depicted in the text, the nature of the events described in the text, etc., the following have been pointed out:

Assessment criteria:

0 – the schoolchild has not given an answer,

1 – One text-based reason has been pointed out (adequate argument),

2 – Two text-based reasons have been pointed out (adequate arguments).

The following data are fed in: 0 pts., 1pt., 2 pts.

Note 1: Spelling mistakes and punctuation errors are not taken into account.³⁾

The appendix contains an answer with a maximum score of two points.

b. In the international survey constructed response questions are assessed in accordance with a well-established set of criteria with a maximum score ranging from 0 to 3 points depending on the depth of textual comprehension or the level and nature of argumentation which the answer requires. Students are duly informed about the score they can get on each question – next to each question there is a drawing of a pencil and a number indicating the maximum score.

Assessment criteria:

A. Assessment of items with a maximum score of **one point**:

Acceptable answer – 1 point

Unacceptable answer – 0 points

B. Assessment of items with a maximum score of **two points**:

Complete comprehension of the material read – 2 points

Partial comprehension – 1 point

No Comprehension – 0 points

C. Assessment of items with a maximum score of **three points**:

Extensive Comprehension of the material read – 3 points;

Satisfactory Comprehension – 2 points;

Minimal Comprehension – 1 point;

Unsatisfactory Comprehension – 0 points (PIRLS' 2006)

Each answer given by a student is scored and assessed independently of all other answers given, without the possibility of any prejudice on the part of the assessors arising from answers to other questions. Some answers can correspond to all criteria defined in the instructions for the task, but could also contain additional information which is not correct or entirely unrelated to the respective question and/or text. The general rule is for such information to be ignored. The only exception to this rule is the case in which the unnecessary information is contrary to the part required in the instructions of the task and fully corresponding to the specified criteria.

It is necessary to emphasize that assessors are trained – they undergo special training in which full coordination of the guidelines, approaches and criteria for assessment and scoring is achieved. Ultimately, at least half of the test items are scored by two assessors thus applying *reliability scoring*. The correspondence in the assessment of the two independent assessors should not fall below 85-90%. In cases in which the marks of the assessors coincide in a percentage lesser than the prescribed one, the two assessors undergo the training again until the required correspondence is achieved on the basis of gained experience.

5. Possibilities for the application of constructed response questions. Multiple-choice questions are a means for measuring and assessing students' skills to follow the norms of the literary language and to distinguish between basic and subsidiary or peripheral in the extracted information. Unlike open-ended questions, constructed response items allow for the assessment of students' abilities and skills to generate written texts, to formulate and defend their own opinion relating to the material they have read and to construct a narrative. In the approbation of the instruments in the national survey it was established that the content prescribed has been acquired by the majority of students – 50% and that constructed response questions have been successfully and acceptably answered by almost 70% of the students. It could safely be concluded that the instruments are adequate for assessing the designed tasks. The results from the international survey are still being processed. For the data from the first international survey PIRLS' 2001 see in the bibliography - Bizhkov 2004. In it Bulgaria ranked fourth among the countries participating in the survey as regards the level of achievement of schoolchildren in the fourth grade.

The following advantages figure prominently among the possibilities for the application of constructed response questions:

- Necessary and sufficient conditions are created for performing formative assessment or assessment which stimulates the linguistic, cognitive and communicative development of schoolchildren;

- Results from the teaching of Bulgarian at both national and regional scale are presented which allows for the comparison of results in different regions of the country;
- On the basis of these data the process of acquiring the standard norms of the literary language by schoolchildren is constructively debated;
- New assessment models and procedures are developed and new guidelines for interpreting scores are designed.
- Management and monitoring of the quality of the teaching of Bulgarian are performed;
- Better control of the educational process in the Bulgarian language and literature is achieved;
- The prerequisites are set for the establishment of State Educational Requirements for the assessment of the educational process in the Bulgarian language and literature;
- Constant feedback on the achievements and difficulties of schoolchildren in the educational process in the Bulgarian language and literature is achieved.

6. A few concluding words. The experience of including constructed response questions in tests in the Bulgarian language is still highly restricted. We need to broaden and deepen this experience so we can fruitfully apply in both internal and external assessment. It is necessary that such questions be applied at school level and later as a standardized procedure. Thus we can take advantage of various models of assessing the achievements and faults in the acquisition of the Bulgarian language. Thus it is possible for us to apply not only integral but also formative assessment. The potential of assessment on the basis of multi-componential procedures – tests and texts – has not been fully realized.

The problems identified here can serve as the basis for future research projects, the better coordination of efforts among specialists in the field of methodology, teachers, linguists and test-makers in order to provide reliable information concerning the state of the education in the Bulgarian language in view of managing the quality of this process better.

NOTES

1. The discussion of constructed response questions is based on information extracted from English sources from Internet (see the end of the bibliography) and the Bulgarian experience in applying the instruments from the national survey and international research focused on fourth-graders (see bibliography).
2. The percentage values express the relative weight of separate items in the overall structure of the test.
3. In the national survey such errors are accounted for in multiple-choice items.

REFERENCES

- Bizhkov, G., F. Stoyanova et al. (2004). *Literacy Diagnostics*. Part three. Reading Comprehension. Sofia.

IS (2006). *International survey of literacy and reading skills of schoolchildren, in which Bulgaria took part for the second time: PIRLS'2006 Progress in International Reading Literacy Study*. The Survey is conducted by International Association for the Evaluation of Educational Achievement. Internet addresses:

<http://isc.bc.edu/pirls2006/> Last visit Jan 14 2018

<http://isc.bc.edu/pirls2006/brochure.html> Last visit Jan 14 2018

Mullis Ina V.S., Ann M. Kennedy, Michael O. Martin, Marian Sainsbury (2006). *PIRLS 2006 Assessment Framework and Specifications*, 2nd edition, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, International Association for the Evaluation of Educational Achievement, Amsterdam, Netherlands.

NS (2004 – 2005). *External Assessment strategy. The creation and approbation of a framework for external assessment of the completion of the primary educational level 1- 4 grade for the Cultural-educational areas Bulgarian language and literature and Mathematics*. Project launched and conducted by a team of professors and lecturers from three faculties of Sofia University: Faculty of Slavic Studies, Faculty of Mathematics and Informatics and Faculty of Pre-school and Primary Education. Report of the Bulgarian Association for Assessment and Measurement in Education. Sofia (2004). Unpublished.

✉ **Prof. Dr. Tatyana Angelova**

Department of Bulgarian Language Teaching

Faculty of Slavic Studies

Sofia University

Sofia, Bulgaria

E-mail: angelova@slav.uni-sofia.bg