

Assessment of Test Quality in German for Bachelor Students on the Basis of Common European Framework in Foreign Language

Inna V. Kazantseva

Tatjana I. Zaburdaeva

Anna L. Kolyago

Mari State University, Yoshkar-Ola, Russia
Email: inna-kazanceva@yandex.ru

Doi:10.5901/mjss.2015.v6n3s7p381

Abstract

According to the Concept of modernization of Russian education accepted by the Ministry of education of Russian Federation the main task of modernization of education is to provide the modern education quality. Nowadays there is a problem of objective assessment of the level of academic achievements of students in higher educational institutions of Russia. In the absence of high-quality test materials one can not speak about the objective assessment of the quality of the results obtained in the educational process. Nowadays there are a lot of researches on the assessment of test tasks in English. The actuality of our work is to develop an algorithm for assessment of the quality of test materials in German to identify the level of formation of different language skills of the students of higher educational institutions in accordance with the Common European Framework in knowledge of foreign languages. The article involves the development, experimental testing in forming the mode of the algorithm, which allows to assess the quality of test tasks in German, developed with the requirements of the Common European Framework, verifying its effectiveness on the basis of monitoring data. The results of the research will have practical value: the developed algorithm can be widely used in educational practice. The obtained algorithm will allow improving tests and developing new quality tests, which will be able to assess the level of language skills of students at the end of study of discipline "German language" in different courses of study. Normative documents and certification of test materials will be used to assess the quality of test tasks used in the final assessment of students of higher educational institutions.

Keywords: education quality, assessment, competence, algorithm, specification, test

1. Introduction

The change of the priority of education due to the adoption of the federal state educational standard of general education has a significant impact on the training of teachers in pedagogical high school. This task is now one of the most important directions of updating of the content of the higher pedagogical education, which gives it a new quality. The study represents one of the modern trends of modernization of education having developed theoretical basis in pedagogical science but having a number of difficulties in real educational process. The aim of the article is to analyze the principles of the development, experimental testing in the forming operation of the algorithm, which allows to evaluate the quality of the test tasks in the German language, meeting the requirements of Common European Framework (Weir, 2005), to check its effectiveness on the basis of monitoring data.

Despite the rather large existing theoretical groundwork, so far it was not offered any unambiguous and justified method of the procedure of examination of quality of test materials in German as a foreign language. Therefore one of the tasks of this work is to build a unified estimation algorithm for the examination of test materials, where it should be offered safe approach to the examination and evaluation of each task submitted to the bank. The methodological bases of the research are the works on the theory test, dedicated to the general laws of development of test materials and to the processing of test results (the works of A. Anastazi, V.S. Avanesov, V.I. Vasilyev, V.P. Besspalko, M.B. Chelyshkova and others), as well as work on the principles of expert analysis, of quality control and statistical data (works of G.G. Azgaldov, B.G. Litvak, V.Y. Pereverzev, V.S. Cherepanov and others).

2. Results and Discussions

The proposed formulation of the problem involves the study of an integrated approach to the evaluation of the quality of the test material in the German language, acting as methodological regulative in the quality assessment procedures and ensuring an integrated quality assessment of test materials. The developed algorithm allows to forecast the availability of low quality tasks in different banks of test materials in German language for the levels A1-C2, and taking into account the structural features of the test tasks in general and the individual subtests to assess the quality at the stage of development (Hennig, 1982, 1987).

The empirical basis of the formulated principles are the specific properties of integrated activities in science. These properties are achieved by deliberately interdisciplinary, the pursuit of comprehensive and integrated analysis of multi-level facilities, the target determinism of specific problems and high level of cognitive process. These properties constitute the empirical base of formulation of the basic principles of integrated studies.

Comprehensive and systematic approaches are aimed at the regulation of complex objects and they are close due to similarity of solved problems. It should be remembered that the basic provisions of the system approach are the direct theoretical basis of comprehensive research methodology. On this basis, the transitional nature of comprehensive methodology is revealed that is reflected in the presence in it of an invariant core, consisting of a general scientific nature of the standards and scientific regulators determined by the specific features of a particular object under study. Basic principles of integrated studies constitute indicated invariant core and they are an interpretation of important systemic laws in the area of scientific activities, with the given earlier properties of complex research.

In accordance with the foregoing the main methodological principles of comprehensive research are formulated:

- the principle of plurality of subject research bases,
- the principle of construction of the complex subject,
- the principle of interdisciplinary problem orientation,
- the principle of target-oriented multidisciplinary organization.

These principles express the specific properties of the complex activities in science, they are an interpretation of the system of norms in relation to research, its action covers both cognitive and socio-organizational aspects of the research process in modern science.

Expertise of the quality of the test materials must meet the general methodological principles of consistency, objectivity, comprehensiveness, professional competence and continuity of all stages of the examination and the need for the enforcement of expertise.

To solve the given tasks it is expected to use methods of testology, qualimetry, probability theory and mathematical statistics.

Assessments of the quality of variant tests are conducted on their compliance with the following criteria:

- the first criteria - the completeness of coverage requirements for the training of graduates by each option of the test and all the options together;
- the second criteria - the correct proportions of the content of the test;
- the third criteria - to check compliance of the content of the reviewed system of tasks of test's specification. (Zaburdaeva, Serebryakova, Kazantseva, Kolyago, Ivanova, 2015).

Depending on the used methods of collecting primary information studies can be divided into:

- quantitative;
- qualitative.

The practical implementation of scientific research requires an integrated approach - sharing quantitative and qualitative methods.

Quantitative research - is the primary tool for getting necessary information for planning and decision-making. The basis of the methodology of quantitative research are always clear mathematical and statistical models, which allows as the result to have a precise quantitative (numerical) values of the studied parameters (components of the specification to the source - Input and tests - Output). The main advantage of quantitative research is that they reduce the risk of making wrong decisions and choice of an inaccurate planning parameters and quality assessment. Quantitative research is the most appropriate way of numerical assessment of:

- prospects for the development of the algorithm of assessment of the quality of test materials;
- effectiveness of the various activities of the university to support and promote quality assessment algorithm test materials;
- trends in the development of the algorithm and its individual components;

- efficiency of the system specifications of the test tasks;
- reaction of the test-students for possible action of testers.

Qualitative research, in contrast to the quantitative focus not on statistical measurements, and are based on the understanding, explanation and interpretation of empirical data and are a source of formation of hypotheses and productive ideas. In qualitative research projective and stimulating techniques are widely used that help researchers uncover the motives, preferences, values, respondents regarding the content of the test material. The use of qualitative research at the stage of strategic development of the estimation algorithm allows to:

- generate ideas on complex algorithm;
- assess algorithm;
- assess the elements of the algorithm evaluation of the quality of test materials.

Qualitative methodology can be used during tactical research to select the most successful version of the estimation algorithm of the quality of test materials. For testing alternative materials to the specific elements of already established bank of test materials can be offered.

The main sources of information are:

- Check (monitoring);
- Experiment;
- Expert assessment.

Experiment is a study of the effect of one factor on another while controlling extraneous factors. Expert assessment is an assessment of the research processes by qualified specialists – experts. This assessment is especially needed when you can not get the favor of unmediated information about any process or phenomenon.

In the course of processing and analysis of the survey data the first stage is the frequency analysis. A description statistics of studied indications is the next.

The second stage of data processing and analysis study is the description of the correlations between the studied variables.

Testing the hypotheses put forward by the research is performed by the correlation, variance or factor analysis. As a result of the analysis of data the hypothesis is confirmed or rejected, which in any case says about the getting result.

Analyzing the content of individual test tasks one should:

- to assess the selected right answers of the tests;
- to assess the content of tasks in terms of the requirements for the educational achievements. To analyze the expert needs to compare the contents of each task and the list of numbered test requirements, which includes materials for expertise and is given an expert together with tests. As a result of comparison the expert reveals a number of skills requirements, which is mainly focused on the contents of this task;
- to determine the level of the base into three groups: basic, advanced, complex;
- to determine the level of significance of the contents of each test task. Assessment of the significance is connected with the need to include in the test only those content items that are the most important to the development of a training course within the requirements for the training's level. In this case, the focus must be placed on that measures the task without taking into account the intermediate elements of knowledge demanded in the process of the assignment of the test;
- to determine the expected percentage of the test done by the examinee with satisfactory training;
- to determine the expected percentage of the sample by the examinee;

It is recommended to focus on the student with an average level of training. As a result of this analysis, the expert should identify bad task and give out recommendations for their revision in terms of content.

The second direction of the work of an expert is connected with the analysis of the quality of the content of the test, having a plurality of parallel variants. Analysis of the quality of execution for each variant separately and then all the information is combined on the group of variants. Assessment of the quality of test variants is carried out on their compliance with the following criteria:

- The first criterion is the completeness of the coverage requirements for the training of graduates by each variant of the test and by all the variants together.

The need to assess the completeness is connected with the fact that usually not all requirements can be displayed in the test. The fuller display, the higher the content validity of the test, the more confidence in the validity of the marks got by the students doing the test.

- The second criterion is the correct proportions of the content of the test.

This criterion allows to assess the correctness of the display of the contents of the model of educational discipline

in the test. To assess the correct proportions of the test the expert calculates the percentage of tasks in the test oriented to material of each section (line content). Then he gives his vision of the optimal ratio of sections and calculates the difference, characterizing the deviation to the developers from his assessments.

- The third criterion is the inspection of compliance of the content of the reviewed system of the specification test's tasks.

Inconsistency may arise as with a decrease as increase in the projected number of tasks in the test. The degree of non-compliance is determined by counting the percentage of tasks not provided by the specification of content or aspect activity aspects, which were planned, but were not included in the test.

The third direction of the expert's work is intended to prepare general conclusions and recommendations for improving the content of the test. In the third section of the review the expert gives his general impression of the content of the test. All doubts and wishes of the expert should be expressed here, his recommendations for improving the content. You can assess the ratio of tasks, checking the knowledge of the theory of subject and its practice.

Statistical analysis of the results of empirical approbation test is carried out to test the quality of the quantitative study. At this stage, according to the selected mathematical and statistical theory characteristics of test's tasks and test are calculated.

Development of testing technology contributed to the process of integration and adaptation of mathematical and statistical data processing apparatus to the theory of pedagogical measurement. Today, for the purposes of processing the results of testing are used four basic theories: Classical Test Theory (KTT), Generalizability Theory (GT), Item Response Theory (IRT), Equating/Linking. In Russian testology are often used Classical Test Theory и Item Response Theory. During the mass large-scale studies there is a need to attract several theories simultaneously. So, for example, in the practice of national testing US KTT is used to calculate coefficients of typical reliability, GT is used to describe a variation of diagnosed characteristic, IRT to generate an adaptive test of the test items, and the model Equating / Linking -In order to get the final assessment of educational achievements of students on the results of its tests pass from admission to graduation in order to identify the dynamics of personal development and for final testing at the end of training.

The applicability of a particular mathematical and statistical theory of processing test results is determined not so much by its capacity to reflect the diversity of the phenomena and processes in the assessed object as the usefulness of the getting results. Let us consider the main features of the application of these theories for the expertise of the quality of test materials on the results of approbation test.

Application of the methodology KTT for the expertise of the quality of tests and test tasks allows to get reliability coefficients of the measurement, to evaluate the consistency and uniformity of the tasks in the test, to assess the creation validity of the test in a whole.

Classical theory of tests is based on linear transformations of the raw test scores, which improves the comparability of the test results of different groups of examinees, but at the same time, the nature of an ordinal scale of the observed results of the test does not change.

Generalizability Theory is considered as the expansion and liberalization of the KTT. The methodology GT is based on the analysis of variance (ANOVA) procedures for measurement of errors, thus allowing to identify multiple sources of measurement errors. Application of GT is the most effective for the analysis of parallel test variants.

Item Response Theory is designed to assess latent examinees' parameters and parameters of the test tasks. IRT is a part of a general theory of latent structure analysis, but unlike the last, IRT examines the distribution of measured parameters as continuous variables. The most significant benefits of IRT are usually referred to the stability and objective assessments of the difficulty setting tasks, their independence from the properties of the sample of examinee doing the test. Furthermore, IRT provides the ability to change the parameters of examinee and the tests in the same scale having properties of interval scale. The last advantage is very important because the conversion of the initial values of various origins in a standard scale allows you to correlate the level of knowledge of any examinee with the measure of difficulty of each task test. In Equating / Linking methodology an eclectic mix of measurement procedures is used - scaling, alignment, conversion of primary points in the test, etc. Thus, these procedures are able to assess the quality of test task (Item) in heterogeneous tests, which is practically impossible in previous theories. Furthermore, equating procedures are necessary by using heterogeneous multi-disciplinary panels of tests used for the final certification test.

Recently, more attention is paid to the improvement of educational tests as a means of measurement to assess the formation of different kinds of skills. In connection with the transition to a two-tier system of training and educational programs development according to the new Federal State Educational Standards it is necessary to improve measurement and control tools. In terms of foreign language guide in terms of measurements may be Common European Framework, which involves the assessment of formation of skills into 6 levels (A1 - C2). As for the German

language, the guideline for teachers may be standard «Profile Deutsch», involving the proper use of lexical and grammatical material, which the student must have at a certain level, and containing self-assessment scale for different types of speech activity (Glaboniat, Müller, 2003). The level of knowledge of German is an integral part in the preparation of specifications for tests. Specification is a document that serves as the basis for the development of tests with the specified quality indicators. Specification of the test is developed to determine the extent of further validity and relevance of the developed test.

In accordance with the standard in German language «Profile Deutsch» systemic functional grammar and is only checked within the levels A1-B2.

Technology of creation of tests for German as a foreign language pursues 4 main stages:

1. The first point in compiling a grammar test is to determine the level of the degree of development that we want to test. The level is mastered if the student has coped with 80% of jobs.
2. Selection of the authentic text. The texts and grammar tests should always have a communicative function. Therefore, the grammar test should consist not of separately taken sentences, it should be presented in a coherent text, because testable lexical and grammatical structures must conform to the required level.
3. Preparing any test it is important to formulate the task of what we want to get as a result (eg, if is offered 20 words to fill in gaps 10 (Schüttelkasten), it should be clearly written in the task that only 10 words are to fit as correct variant).
4. Relevance of any test can only be tested according to the specifications in the instructions (what is the purpose of the test, to whom is it, the weight of each item, how many points are in the end, etc.) (Hennig, 1982).

Possible variants of the task to test the formation of grammatical skills are:

- to write the text from the sentences/ to write the sentence from the words;
- multiple choice (number of distractors at least 4);
- text with gaps / frame with the words (Schüttelkasten), the number of distractors should be 2 times larger than the gaps;
- text with gaps of open form can be used to check the levels of development of the B1-B2;
- word formation;
- paraphrase;
- correction of mistakes (in each line to the left is wrong variant in the right hand column under the number line - a place for correct variant) (Zaburdaeva, 2013).

Each level requires a certain grammar test (A1 - flexion, A2 - coupling elements (Konnektoren), etc. Full final grammar test should consist of different types of tasks, depending on the stage and contains at least 20 to 30 tasks (Items).

General rules for making the grammar test are:

- one mustn't give the gap in the first sentence;
- If the task is to fill the gaps in the text, you can not finish the first sentence with a gap and to start the next sentence with a gap;
- In the tests with an open response option one can not verify the vocabulary because there are always synonymous;
- gaps should be no more than on each 7word; the optimal variant when the gap is between 7-10th words;
- no more than 2 gaps in the sentence;
- all answers must be in alphabetical order;
- before any task an example how to do the task must be given (Glaboniat, Müller, 2003).

Most teachers prefer multiple choice tests, making of which, according to the rules of the German testology suggests keeping the following points:

1. Every task (Item) should determine the relevance of the element;
2. Items should not be dependent on each other, i.e. the answer to one item should not affect the other (especially in the case of the previous wrong decision);
3. there must be only 1 right answer;
4. the correctness of distractors. If one distractor is obviously wrong, it must be acceptable, i.e., so that low-level students can choose it as correct, but not "filter out" immediately;
5. all distractors should look the same;
6. distractors must be the same length;

7. distractors must have the same degree of difficulty;
8. avoiding negatives;
9. the task must not contain lexical clues eg, in the form of recurring component of a compound word (to avoid repeated verbatim from the text in the task) (Hennig, 1982).

Taking into account that the tests are worked out on the basis of authentic texts, one should pay attention to the forbidden topic in the text: disaster, war, divorce, death, violence, religion, murder, rape, politics.

The most difficult task is to work out a test that allows objectively evaluate the level of compliance or non-compliance of students' knowledge of model of the State Standard.

Creating the test materials for any form of control (computer or paper variant) must meet certain principles, worked out requirements. The scaled tests should solve this task.

The scaled tests are integration tests for all types of speech activities and are intended to determine the level of knowledge of foreign language on the European scale of assessment (A1-C2). They are very popular because they are very economical in time, have lower financial costs and provide the opportunity to immediately check all students with different levels of knowledge in the same group. One scaled test can be created for all 6 levels, but it is usually created for 2-3 levels. The control of formation of lexical and grammatical skills is checked on the basis of texts for reading or listening.

In assessing the results of tests the main task is the need to find out how high-level students and low-level students made the test. To analyze the tasks that make up a test and to determine the level of difficulty / ease of tasks to the examinee, at least 200 people must take part in the testing.

Since the late 90s of XX century up to date each year number of tutorials, articles and methodical works on educational measurement and testing in education increases. In general, modern studies Russian and foreign scientists on the problems of testing in education are carried out in four directions:

- 1) the development of scientific approaches to improve the content of measuring instruments in the context of modern interpretations of the quality of educational achievements;
- 2) the development of theoretical and methodological support of designing of reliable and valid tests;
- 3) scientific justification for the use of tests;
- 4) scientific justification of scaling the test data and their interpretation for the application of the results of educational measurements in the management of the quality of education.

3. Conclusion

The purpose of this research was to give the scientific justification for a comprehensive approach in evaluating of the quality of test materials in the German language and the offer of the relevant methods for estimating of the integral quality of test materials for the certification test materials that have been created and will be created by teachers of German. These methods involve the assessment of the substantive aspect of tests of all speech activities and tests to check the formation of lexical and grammatical skills of students in levels A1-C2. This allows you to generate funds of standardized assessment tools that provide reliable assessment of educational achievements of students in the establishments of vocational education.

The research process includes the systematization of the results and the consistent realization:

1. Synthesis of existing in the world and Russian science experience in the development and assessment of tests in foreign languages in the context of training of future teachers in higher education institutions.
2. Description of the sequence of actions aimed at determining of the expected percent of the test making by the examinee.
3. Identify of tools and the development of criteria for the diagnosis of available test tasks in the German language.
4. Adaptation of identified tools to the process of training of future teachers of German.

3.1 Stage

1. The study of the works of German testers Jürgen Quetz, Karin Vogt, Christin Krekeler, Sibylle Bolton, a fundamental source "ALTE-Handreichungen für Testautoren" (4 modules).
2. Development of algorithm of assessment of quality of test materials from the standpoint of comprehensive assessment, carrying out an experimental work for the testing of individual elements of an algorithm of comprehensive quality assessment of test materials during the mass testing of students.

3. Expertise of the quality of available test materials in high school in German language for linguistic specialties.
4. Organization of 1 scientific-methodical seminar for university teachers.

3.2 Stage

1. Development of 2 banks of test items for the German language as a foreign language for students of high schools on the level of A2-C1 on aspects of listening, reading, speaking and writing.
2. The development of the certification program of tests in German language for language high schools and secondary educational institutions.

The resulting algorithm will allow in future improving the existing tests and creating new quality tests that can objectively assess the level of knowledge of students at the end of the course of study discipline "German language" in different courses.

References

- Zaburdaeva, T.I., Serebryakova, M.A., Kazantseva, I.V., Kolyago, A.L., Ivanova, O.V. (2015). Control of Competences Formation of Foreign Language Bachelor Students on the Basis of LMS Moodle. *Review of European Studies*, Vol. 7 (8), pp. 295-307. <http://dx.doi.org/10.5539/res.v7n8p295>
- Glaboniat, M., & Müller, M. (2003). *Profile deutsch*. Berlin, München: Langenscheidt.
- Hennig, G. (1982). *ALTE-Handreichungen für Testautoren. Itemtypen*. New York: Newberry House.
- Hennig, G. (1987). *A Guide to Language Testing*. New York: Newbury House. <http://dx.doi.org/10.1177/026553229301000308>
- Weir, Cyril J. (2005). *Limitations of the Common European Framework for developing comparable examinations and tests*. Oxford: University Press. <http://dx.doi.org/10.1191/0265532205lt3090a>
- Zaburdaeva, T.I. (2013). The algorithm of working out grammar tests on the basis of the handbook – standard "Profile Deutsch". *XXI century: Results of the Past and Problems of the Present+*, (8), pp.187-193. Penza: Penza State Technological Academy.

