

EEIA 2019
**International Conference "Education Environment for the
Information Age"**

**PERSPECTIVE APPROACHES TO STUDENT'S COMPETENCE
ASSESSMENT IN MODERN UNIVERSITY**

Zhanna Sizova (a), Tatyana Semenova (b), Natalia Naydenova (c)*, Viktor Zvonnikov (d),
Marina Chelyshkova (e), Alexey Malygin (f)

*Corresponding author

(a) Doctor of Medicine, Professor, Director of Methodological Center for Accreditation of Specialists in First Moscow State Medical University named after I.M. Sechenov, 8/2 Str. Trubetskaya, MOSCOW 119991, Russia, sizova-klinfarma@mail.ru

(b) PhD (Medicine), Docent, Director of Department for Medical Education and Personnel Policy in Health Ministry of RF, 3 Lane Rakhmanovsky, GSP-4, MOSCOW 127994, Russia, SemenovaTV@rosminzdrav.ru

(c) PhD (Education), Deputy Head of the Comparative Education Center, Institute for Strategy of Education Development of the Russian Academy of Education, 5/16 Str. Makarenko, MOSCOW 105062, Russia, naydenova@instrao.ru, naydenova@my.com*, tel.: +7 926 2128867

(d) Doctor of Education, Professor, Professor of Department for Continuing Psychological and Pedagogical Education in Ivanovo State University, 39 Str. Ermak, IVANOVO 153025, Russia, zvonnikov@mail.ru

(e) Doctor of Education, Professor, Chief Specialist of Methodological Center for Accreditation of Specialists in First Moscow State Medical University named after I.M. Sechenov, 8/2 Str. Trubetskaya, MOSCOW 119991, Russia, mchelyshkova@mail.ru

(f) PhD (Education), Docent, Vice-Director for Education Affairs in Ivanovo State University, 39 Str. Ermak, IVANOVO 153025, Russia, malygin@ivanovo.ac.ru

Abstract

The article is devoted to topical issues of evaluating the competence of graduates from medical universities. The theoretical basis for such an assessment is built on the base of several approaches to the assessment of learning success: a) multiplicity; b) interactivity; c) multistage; d) adaptability. It was found that the assessment of educational achievements of standardized tests is less informative than the use of situational tasks (cases). The model of multistage adaptive measurements is presented in the article and it was stated that an accreditation mode is more convenient to use the modular adaptation principle instead of branching strategies. The authors clarified the concept of "interactivity", having examined in detail two versions of the definition, and settled on the fact that interactivity is an activity of the student in the information and communication environment that does not require special digital skills, and only the presence of basic skills in dialogue with a computer is necessary. The levels of interactivity are of three types: 1) passive; 2) active; 3) imitation. Testing was carried out on a representative sample of universities in Russia, which was conducted in several stages. As a recommendation for the modernization of case-tasks, it is supposed to include imitating audio and video plots in order to reach higher levels of interactivity. Nevertheless, approbation of such case-tasks was carried out not only in the synchronous interaction of a computer and a student, but also in an asynchronous one, which already contains specific elements of adaptability.

© 2019 Published by Future Academy www.FutureAcademy.org.UK

Keywords: Case, multiplicity, interactivity, adaptability, sampling, interdisciplinarity.



1. Introduction

People around us assess what we have already done

H. Longfellow

By the beginning of the second decade of the 21st century, an understanding was established among the community of specialists in the field of educational measurement that traditional tests are not enough to assess the level of competence of university graduates or specialists even when the tests contain tasks with constructed answers. Other assessment tools are needed to identify the level of development of a whole range of requirements, presented in the form of professional competencies or labor functions in the standards, which generally characterize the ability of students to practice in professional fields. This potential is provided by specially designed multiple case-studies involving elements of interactivity and allowing to assess the professional readiness of graduates of higher educational institutions to perform work functions (Sizova, Semenova, & Chelyshkova, 2017). That is, as indicated in the epigraph, the case studies the student's activity in solving a particular situation or how he is already able to perform various work functions of the future medic.

The first term "multiple" is synonymous with multidimensionality. It means improving the quality of case content and simultaneous evaluation of several variables defined by a variety of competencies or labor functions from standards. Due to the multidimensionality, each multiple case-study allows one to evaluate the mastery of several competencies by the students that increases the validity of the results from assessment procedures based on the increase in the number of measured variables (Dorozhkin, Chelyshkova, Malygin, Toymontseva, & Anopchenko, 2016).

The second term "interactivity", which is extremely popular nowadays, can be associated with several levels of manifestation, when various audiovisual means are introduced into the structure of case-task in the mode of dialogue between the student and the computer (Sedig, Parsons, & Babanski., 2012). The increase in the level of case's interactivity is provided by the solution of a whole range of tasks, including the creation of innovative concepts and layouts of interactive cases; the development of innovative software products that provide for the possibility of further development in the direction of increasing the level of interactivity; processing and analyzing large-scale approbation case studies to improve their qualities. Separate results on the implementation of these works in the Methodological Center for Accreditation of Specialists, created on the basis of the First State Medical University, are discussed in the article.

2. Problem Statement

Especially important are multiple interactive case-studies for making administrative and managerial decisions in evaluation procedures with high stakes (certification, accreditation, certification of qualifications, etc.). These case-tasks assess the compliance of examinees to professional activity based on the requirements of educational or professional standards. The next problem is that usually case-tasks, even in the shortest version, take a long time to complete (at least 20-30 minutes). At the same time, the case-study is interdisciplinary in terms of assessing readiness for further medical practice. That is, the

case-items in every case-task belonged to different doctor's work functions. Therefore, the case-study had interdisciplinarity according to this criterion (Klaassen, 2018). Therefore, during certification or accreditation, it is necessary to minimize the case-tasks for each test subject by optimizing the difficulty of cases without loss of reliability and validity of measurement results (Haug, Solfjeld, Ranheim, & Bardsen, 2018). These possibilities are opened up by an adaptive approach to measurement, which involves the use of modern testing theory (IRT).

3. Research Questions

In accordance with the problem posed and the goal chosen, a number of research questions were raised and resolved:

1. What type of adaptive measurements is better to give preference? The answer to the question required a comparative analysis of various types of adaptive measurements, based on the computer presentation of evaluation tools and involving the use of IRT (Ushakov & Romanova, 2010).

2. How will the content of multiple case-tasks be simulated in order to evaluate the entire required set of variables in accreditation (levels of mastering the labor functions of professional standards)?

3. How and by what means should the ideas of interactivity in multiple case-tasks be implemented?

4. How will case-items be grouped for testing multiple interactive cases?

The posed research questions found their solution in the theoretical and methodological results of the study.

4. Purpose of the Study

The aim of this article is to present methodical approaches to creating interactive multiple case-tasks, tested during accreditation of graduates of Russian medical schools in 2018-2019. There are outline theoretical prospects for the development of accreditation on the basis of multistage adaptive measurements in subsequent years.

5. Research Methods

In answering the first question, various formats of computer adaptive measurements involving the use of IRT were analyzed. These include multi-stage adaptive measurements and varying branching strategies, when assessment procedures form a unique set of tasks for each student. The varying strategy has undoubted advantages due to the high level of optimization of the case-item difficulty. Each item is selected step by step from the bank in accordance with the results of the performance of the previous item (successful or unsuccessful). Since the student's score is recalculated after each answer to the next task using Bayesian methods and the IRT theory, the number of tasks is minimized without loss of measurement reliability.

However, along with the advantages, the varying branching strategies for adaptive measurements carry with them certain risks. They are associated with the need to use expensive software. In addition, time costs are increased in order to ensure substantial comparability of measurement results. There is

necessary to achieve high reliability and fairness of accreditation results. The multi-stage adaptive strategy has several advantages in evaluation procedures with high stakes.

Multistage adaptive measurements individualize the selection of tasks not for individual student, but for the subgroups, into which the whole group of students is divided when entering the measurement procedure (Yan, von Davier, & Lewis, 2014). All students undergoing accreditation, first perform the same set of tasks, and then divided into two subgroups in accordance with the results of their implementation. For each of the two subgroups the next set of tasks is issued, optimized for difficulty and the subsequent division is made. The idea of such a division is shown in Figure 01, where the modules correspond to the tasks for the various subgroups of students, and module 1 is issued to the whole group.

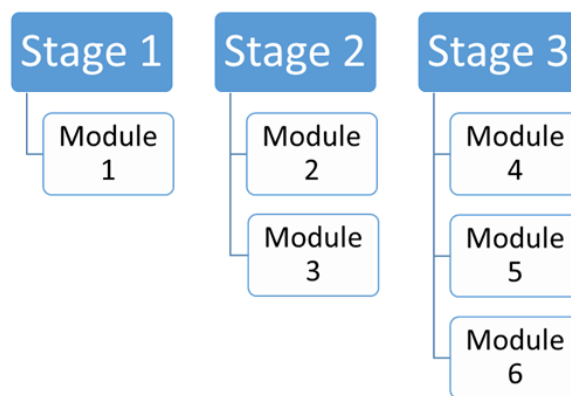


Figure 01. Model of multistage adaptive measurements

The advantages of multi-stage adaptive measurements are associated with relative ease of implementation and higher comparability of measurement results achieved by dividing the students into subgroups. In the branching strategy the individualization of the task selection is carried out for each test group. These advantages make it possible to regard multi-stage adaptive measurements as a promising direction for the development of accreditation of graduates in medical universities.

When answering the second question, it became obvious that a simple case structure for the accreditation of graduates in medical universities will not be enough. When accrediting, it is required to assess the graduate's mastery of the set of labor functions that are part of the professional standard of a medic. For this reason, it is necessary to refer to the development of multiple case-tasks, within which there are several subsets of questions. Each subset of questions is intended for a separate variable, and there should be at least two such variables, but no more than five - six. If we transfer this idea to the measurement language, then we can say that we should talk about the method of creating multidimensional cases that require careful planning of the content, its clear structuring and reference to factor analysis to ensure high validity of the measurement results. The presence of several measurement variables that provide an assessment of the level of mastering a greater number of labor functions significantly increases the reliability of conclusions when making management decisions in the accreditation of health professionals.

The answer to the third question related to the implementation of interactivity in multiple cases requires an analysis of the approaches to the interpretation of the term "interactivity", which is rather ambiguous in the scientific and methodological literature. In the most general case, we can assume that

interactivity is the ability of the information and communication system to actively and diversely respond to user actions (Malygin, 2012). Based on this general definition, it can be argued that working with a computer, which is the interaction of a user with some device, is inherently interactive. Such interaction can be directed to the search for content or consist in its management, include navigation through the content or any other actions with its elements.

In a narrower sense, only electronic content is considered interactive in which operations with its elements are possible, with the obligatory inclusion of manipulations with objects represented in audiovisual form. Moreover, this second definition also includes interference in the dialogue processes between the student and the computer, which are characterized by modeling typical user reactions to external influences or when the conditions of the processes change (Crisp, 2019). Although the second definition is fairly accurate, it is clearly overloaded with signs of interactivity. Therefore, the authors of this article believe that the most productive approach to interpreting interactivity in education is to adopt the first general definition and highlight several levels of interactivity that do not require special skills and knowledge of specific computer programs from users, except for having basic, widely used programs or skills, working with websites (text editor, Internet, social networks, etc.).

The levels of interactivity are determined by the degree of activity of interaction between users and electronic educational resources in the mode of training or evaluation. In this article, in relation to the task of creating interactive cases, it is proposed to distinguish 3 levels of interactivity, which have an increasing character of interaction as the level increases and are characterized by the number and type of manipulations with objects presented in audiovisual form. The first level can be called passive, when the reaction of the evaluative means to the actions of the student is one-step in nature, and the student himself performs only the navigation through the case content and performs the simplest actions with its elements, not transforming them, but only choosing the right answers in the tasks. It is this one that is characteristic of case-tasks developed in 2018 at the Methodological Center for Accreditation, since the one-sided influence of the student on the case material occurs in them.

The second level of interactivity can be considered active. In addition to numerous background information, including various formats of analysis, graphical information, photographs and other documentary descriptions of the patient's condition, taken at the first level, cases of the second level imply active actions from both the computer interface and the person tested. In the course of carrying out the tasks of the case, depending on the need, images of a teacher asking questions, to which the students give answers in oral or written form, may appear on the computer screen. Often, instead of simulating the teacher's dialogue with the student, one or several video imitations of the patient's reception situations are used with the ability of the student to take active actions, for example, in the form of moving around the screen of various elements, palpating the patient with a mouse or a touch screen with the accompanying video switching, etc. (Fu, Kayumova, & Zakirova, 2017).

The third level of interactivity, which can be called imitative, is based on adaptive flexible interaction between the students and the computer, completely freeing teachers from participating in teaching or assessment processes (Papanthymou, 2018). His place is taken by models that imitate human activities, including the behavior of the patient. With this interaction, the assessment system is able not only to independently verify the correctness of the answers of the students, but also to adapt to these

answers, taking into account their individual characteristics, taking into account the attendant factors. Models of machine learning, which are characteristic of the third level of interactivity, are capable of creating individual training programs, taking into account the characteristics of the student in possession and development of educational information. It is quite obvious that such a degree of freedom causes a situation of fuzzy decisions based on the results of case studies, therefore the third level of interactivity is unacceptable in accreditation and is suitable only for the training mode.

To answer the fourth question about grouping a sample of students for testing multiple interactive cases, a sampling approach was chosen in several stages (Lavrakas, 2008). The priority of the choice of a specialty and discipline for approbation was determined by their importance for the professional training of Russian healthcare specialists (Fahim & Negida, 2018). Based on this reasoning, case studies in the specialties of General Medicine and Pediatrics were chosen for testing.

6. Findings

In accordance with the chosen theoretical approaches, a typical structure of multiple cases was created in the Methodological Center for Accreditation, within which several subsets of questions were singled out. Each of these subsets was intended for assessing the readiness to perform a certain work function, the number of which was chosen in accordance with professional standards in the specialties of Pediatrics and General Medicine. Each multiple case for these specialties included at least 5 variables, assessed using 12 problem questions with a slightly different distribution of questions by variables in different cases.

For example, in the structure of the case study in the specialties "Pediatrics", there were allocated blocks of questions on:

- performing a survey of children for the purpose of establishing a diagnosis (1st function);
- the purpose of treatment for children and control of its effectiveness and safety (2nd function);
- implementation of rehabilitation programs, the implementation of preventive measures, the organization of medical personnel activities, record keeping (3rd, 4th and 5th functions).

Tasks on the first two functions were an obligatory part of all cases both in the specialty "Pediatrics" and in the specialty "General Medicine". Multidimensionality is a great advantage of the structure of the discussed cases, which increases the substantive and prognostic validity of the results of accreditation. The coverage of all functions of professional standards in the relevant specialty and a clear orientation of the content of tasks to the problems of professional activity serve to improve the quality of the substantive characteristics of the case, contributing to the growth of its validity (Epstein & Hundert, 2002).

However, multidimensionality always causes problems when creating evaluation tools that include interdisciplinary tasks, since it is difficult to figure out exactly which variables they evaluate and which variable dominates each case task? In other words, you have to look for an answer to the question of the extent to which the content of each task is related to the variable for which it was intended to be evaluated. To answer, one has to turn to factor analysis, which can provide the rationale that the task set actually measures the necessary variables in accordance with the specification of the evaluation tool. The

use of factor analysis apparatus to improve the content of multiple cases is planned for the next stage of work in 2019.

Another problem that arose in the course of discussions during the work on multiple cases is related to the identified lack of local independence of the questions necessary for the requirements of the theory of educational measurements in mass assessment procedures. The presence of dependence means that the correct answer to one of the questions is a condition for answering subsequent questions of the case. In particular, in cases developed in the specialties "Pediatrics" and "General Medicine", the possibility of appointing the correct treatment, monitoring its effectiveness, etc., depends on the establishment of the correct diagnosis. In this regard, it was decided to correct the answers of the students in the event of a wrong diagnosis by giving them the correct answer with the corresponding correction of the points accrued for answering the questions of the case. Thanks to this decision, all the students were able to reach the end of the case and try to answer all his questions.

7. Conclusion

Although the presentation of the situation in case studies in 2018 year provides for extensive graphic material borrowed from the practice of the professional activities of doctors and closely related to its contemporary tasks, the assessment tool itself is passive. His reaction is manifested only in the only case where the diagnosis is incorrect and the student is informed of the correct answer for correcting his actions when answering subsequent questions from the case. Therefore, it can be argued that the created cases have the first level of interactivity, but they imply a further increase in the level due to the inclusion of imitational audio and video plots. Thus, it can be said that at the achieved level of interactivity in multiple cases, linear interaction between the computer and the student is carried out in a synchronous or asynchronous mode, with elements of adaptability. It should be noted that interactivity does not provide high efficiency of evaluation procedures in accreditation and, most importantly, further interpretation of their results (Jarr, 2012).

A promising direction for the development of multiple interactive cases created in the Methodological Center for Accreditation involves the introduction of a second level of interactivity, further work to improve the quality of their content using factor analysis methods and the transition to multistage adaptive procedures for presenting them to graduates of medical universities during their accreditation. A necessary condition for the implementation of all these areas of work is the calibration of bank assignments for accreditation, which is now carried out on representative samples of sixth-year students of medical universities in Russia.

References

- Crisp, G. (2019). *Interactive e-Assessment: moving beyond multiple-choice questions*.
- Dorozhkin, E., Chelyshkova, M. B., Malygin, A. A., Toymentseva, I. A., Anopchenko, T. Y. (2016). Innovative approaches to increasing the student assessment procedures effectiveness. *International Journal of Environmental and Science Education*, 11(14), 7129-7144.
- Epstein, R.M., & Hundert, E.M. (2002) Defining and assessing professional competence. *Journal of American Medical Association*, 287(2), 226-35.

- Fahim, N., & Negida, A. (2018). Sample Size Calculation Guide. Part 1: How to Calculate the Sample Size Based on the Prevalence Rate. *Advanced Journal of Emergency Medicine*, 2(4), e50. <https://doi.org/10.22114/ajem.v0i0.101>
- Fu, L., Kayumova, L., & Zakirova, V. (2017). Simulation Technologies in Preparing Teachers to Deal with Risks. *EURASIA Journal of Mathematics, Science and Technology Education*, 13(8), 4753-4763.
- Haug, S., Solfeld, A., Ranheim, L., & Bardsen A. (2018). Impact of Case Difficulty on Endodontic Mishaps in an Undergraduate Student Clinic. *Journal of Endodontics*, 44(7), 1088-1095.
- Jarr, K. (2012). *Education practitioners' interpretation and use of assessment results* (Doctoral Dissertation). University of Iowa. <https://doi.org/10.17077/etd.35vh2oc1>
- Klaassen, R. (2018). Interdisciplinary education: a case study. *European Journal of Engineering Education*, 43(6), 842-859. <http://dx.doi.org/10.1080/03043797.2018.1442417>
- Lavrakas, P. (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage Publications, Inc. <http://dx.doi.org/10.4135/9781412963947>
- Malygin, A. (2012). *Adaptive testing in distance learning*. Ivanovo. [in Rus].
- Papanthymou, A. (2018). *Student Self-Assessment in Higher Education: The International Experience and the Greek Example*, 8, 130-146.
- Sedig, K., Parsons, P., & Babanski, A. (2012). Towards a characterization of interactivity in visual analytics. *Journal of Multimedia Processing and Technologies, Special Issue on Theory and Application of Visual Analytics*, 3(1), 12–28.
- Sizova, Zh., Semenova, T., & Chelyshkova, M. (2017) Assessment of professional readiness of health professionals during accreditation. *Medical Bulletin of the North Caucasus*, 12(4). <https://doi.org/10.14300/mnnc.2017.12127> [in Rus].
- Ushakov, A., & Romanova, M. (2010). Adaptive testing in the structure of educational control. *Scientific Notes of P.F. Lesgaft University*, 5(63), 87-93.
- Yan, D., von Davier, A., & Lewis, C. (2014). *Computerized multistage testing: Theory and applications*. New York, NY: CRC Press.