LOOK ACADEMIC PUBLISHERS
OPEN ACCESS

INTERNATIONAL JOURNAL OF ENVIRONMENTAL & SCIENCE EDUCATION
2016, VOL. 11, NO. 14, 7129-7144

Innovative Approaches to Increasing the Student Assessment Procedures Effectiveness

Evgenij M. Dorozhkin[a], Marina B.Chelyshkova[b], AlexeyA. Malygin[c], Irina A. Toymentseva[d] andTatiana Y. Anopchenko[e]

[a]Russian State Vocational Pedagogical University, Ekaterinburg, RUSSIA; [b]State University of Management, Moscow, RUSSIA; [c]Ivanovo State University, Ivanovo, RUSSIA; [d]Samara State University of Economics, Samara, RUSSIA; Southern Federal University, Rostov-on-Don, RUSSIA.

ABSTRACT

The relevance of the investigated problem is determined by the need to improving the evaluation procedures in education and the student assessment in the age of the context of education widening, new modes of study developing (such as blending learning, e-learning, massive open online courses), immediate feedback necessity, reliable and valid assessments. The purposes of the article are multistage adaptive measurements validation and testing for increasing the student assessment procedures effectiveness and getting immediate feedback, reliable and valid assessments. Multistage adaptive measurements mentioned above are based on modern test theory Item Response Theory (IRT). The main research methods are math models and measurements on the basis of IRT models, mathematical-statistical methods (descriptive statistics, Bayesian models and maximum likelihood method) and the systematic analysis of the developing practices for student evaluation during the assessment procedures, opinion polls and questionnaires of learning process participants at university. The article presents validation and results of multistage adaptive measurements application, description of adaptive measurement algorithm (leading to the increase of effectiveness in student assessment procedure due to the selection of optimal task difficulty for each student, creating a situation of success during computer-based test session with the tasks accomplishable at individual pace, increase of assessment accuracy and cutting of labor input. Multistage adaptive measurements, as one of the innovative approaches increasing the student assessment effectiveness, admit of individualization principle, actualization in education and getting immediate feedback for improving learning process and the content of education. Multistage adaptive measurements can be applied in blending learning, massive open online courses and e-learning. The article can be of interest for teaching staff and experts in developing the effective methods of learning outcomes assessment.

KEYWORDS
Adaptive measurement algorithms; assessment; multistage adaptive measurements; effectiveness

ARTICLE HISTORY
Received 07 May 2016
Revised 19 June 2016
Accepted 12 August 2016

Introduction

Currently it is a vital task for vocational education to conduct students and their learning outcomes assessment in the context of a changing educational

CORRESPONDENCE Evgenij M. Dorozhkin ✉ evgeniy.dorozhkin@rsvpu.ru

landscape. There are several reasons supporting this fact. First, nowadays massive open online courses are gaining traction, and achieving mastery of which requires immediate feedback and flexible assessment of learning outcomes in on-line mode. Second, it is a new approach to treatment of learning outcomes in terms of competency building approach. Finally, it is the necessity of correlation between educational programs of vocational education, being developed on the basis of Federal State Education Standards, and occupational standards and labor market requirements (Dorozhkin, Zaitseva & Tatarskikh, 2016). In the age of rapidly developing technologies in education and the accumulation of big data, it is necessary to see the way of conducting effective student assessment and obtaining reliable results with high predictive validity (Chelyshkova, 2002).

Since competencies are of meta-latent nature and vitally important for effectiveness in professional activities, as well as reveal themselves after learning process, they can be defined as a deep stable behavioral properties of a human person, predicting the effectiveness of professional activities on the basis of acquired knowledge, skills and work methods. Despite the fact that knowledge, skills and work methods are also of latent nature, they, as compared to competencies, are still closer to the surface layer of new growths in a personality of an individual. (Spencer &Spencer, 2005; Boyatzis, 2008) Under these circumstances, it is necessary to develop a model assessment (leading to the increase of effectiveness) for the competencies formation level as learning outcomes.

The effectiveness of learning outcomes assessment procedures can be achieved by choosing the method of combining quantitative and qualitative approaches in pedagogical measurements. This two-paradigm approach involves putting the results of quantitative and qualitative educational measurements on the same level scale (Zvonnikov, 2006). The application of this approach requires multistage measurements conduction, including several stages, on each of those the measuring instruments for knowledge, skills or competences assessment are used. Each stage of such measurements should correspond to a specific range of competence scale level, the measuring instruments of which become more complicated from the first to the next stage. Such multistage measurements are based on a modern test theory Item Response Theory (IRT) (Hambleton, Swaminathan& Rogers, 1991; Crocker & Algina, 2006).

On the one hand, for obtaining reliable results during student assessment procedure, it is necessary to use sufficient number of measuring instruments with stable parameters of their difficulties and differentiation ability at each stage, which will result in high accuracy of the values and the duration of assessment tests for each student. On the other hand, for providing construct and predictive validity, assessment of learning outcomes in the frames of competency building approach involves the use of case studies, the implementation of which requires a considerable amount of time. Thus, it is necessary to search the possibilities of compliance with the above mentioned conditions and the transition to multistage adaptive measurements.

Methods

During the research the following methods were used: theoretical (analysis, synthesis, specification, generalization, modeling); diagnostic (questionnaires,

interviews, testing, method of tasks and assignments); empirical (studying teaching staff work experience, normative and educational documents; pedagogic supervision); experimental (ascertaining experiment); method of descriptive statistics and maximum likelihood method.

Results and Discussions

The base of the research was the Ivanovo State University.

The research was conducted in three stages:

At the first stage, the theoretical analysis on assessment activities under the conditions of new learning modes formation and development was carried out. The methodological approach to the issue, theories and methodic of educational researches were determined; the issue, the purpose and the methods of research were pointed out; the plan of research was set up.

At the second stage, the choice of multistage adaptive measurements strategy was justified, the implementation of multistage adaptive measurements and measuring instruments (including their testing) were developed, the pilot-testing was conducted, the results obtained in the course of experimental work were analyzed, reviewed and defined.

At the third stage, experimental work was completed, the theoretical and practical conclusions were defined, the results were summarized and systematized.

The measurement is theoretically understood as a process of establishing a correspondence between the evaluated characteristics and points on the scale, in which the ratio between the different marks is expressed in numerical properties series (Stevens, 1946). The process of educational measurement, providing the gaining of the unbiased and comparable information, includes measurement object (one or more latent characteristics), measurement procedures, measuring instruments (items, test and scale for fixation of the measuring object's marks), the analysis and interpretation of measurement results. These components of the measurement process have their analogues in the traditional educational control, but these procedures are more of intuitive nature there. In the case of educational measurements, each component is in the process of scientific substantiation of quality. It is particularly important, if it is about the summative assessment, results of which are used for management decision making. In this case, the objects of measurement are knowledge, skills and competences formed, the structure and level of which are compared to the standards stated in the educational requirements as a result of their learning outcomes.

Proceeding to educational measurements as to the most reliable and valid method of obtaining information on the learning outcomes is due to the need of increasing the objectivity, accuracy and effectiveness of the assessment processes. For obtaining the most accurate and unbiased results, final assessment procedure can lead to unnecessary expenses (time, financial, human resources), so the proceeding to multistage adaptive measurements seems the best and most effective way. It minimizes measurement error, and therefore, increases its accuracy and test duration, maximizes the validity of marks.

Justification of multistage adaptive measurements as an effective method of learning outcomes assessment takes place through approaches to modeling. The

selection of a strategy and developing an algorithm of giving a measuring instrument, acquiring the effectiveness of educational measurements, precedes the beginning of multistage adaptive measurements. Thus, adaptive measurements can be classified as two-stage and multistage, according to which different strategies and algorithms are developed (Chelyshkova, 2001; Malygin, 2011).

Multistage measurement represents such measurement organization, in which the testee moves along his individual path in the process of accomplishing sets of items, different in number and difficulty at each stage. Selection and items presentation algorithm is based on the principle of feedback: after the testee selects the right answer, his next task is more difficult, but if the answer is wrong, then the next item is easier than the previous one. Thus, the adaptive multistage measurement is based on context-depended algorithms, when the next step depends on the previous one and is made only after the assessment of its results.

In return, the multistage adaptive measurement strategies are divided into fixed strategy and flexible strategy, depending on how measuring instruments for multistage adaptive measurement are designed. If one and the same set with fixed position of measuring instruments on the axis of the difficulties is used for all students, but each of them moves through a set depending on the results of his previous step, then the adaptive measurement strategy is deterministic-branching. Difficulty measuring instruments in set are usually placed at equal distances from each other or choose the descending step, consistent to the increase of difficulty, adjusting the pace of implementation for a student. Here is the description of the two most common strategies, related to the fixed strategy.

Pyramid strategy

The essence of the pyramid strategy is that all students begin with average difficulty items. If the student's answer is correct, then he is given a item with the next degree of difficulty. If the student's answer is wrong, he is given a less difficult item. The procedure repeats as long as the student passes the necessary amount of items. For the implementation of pyramid strategy the number of items for each difficulty level in test is to be determined with predetermined number of measurement stages (it coincides with the number of difficulty levels).

Figure 1 shows an example of 10-step measurement with 55 items. At the beginning a student is given a item of average difficulty (level 5). At the second step he may be given either 5th or 6th level item. It is obvious that at each step different items can be given, the difficulty level of which coincides with the number of the step taken. If the test has got items of 10 difficulty levels, then in general each testee is given 10 items out of the 55 included into the test.

It should be noted that the pyramid strategy at each difficulty level requires a certain number of tasks (Figure 2). The largest number of items (equal to the number of difficulty levels) may be involved at the secondary level. Only one item is used at the highest level. On the adjacent levels the number of items is different by 2 (except the levels adjacent to the average level). It can be seen that on the first level 2 items are required, on the second - 4, on the third – 6 etc. On the last level 1 item is used, on the last but one level– 3 items etc.

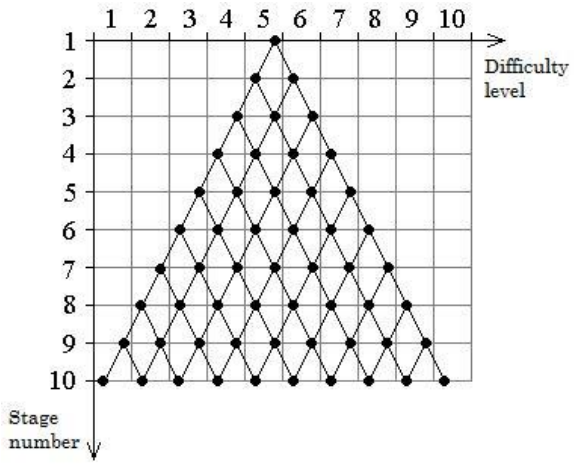

Figure 1. Distribution of items for 10-stage measurement

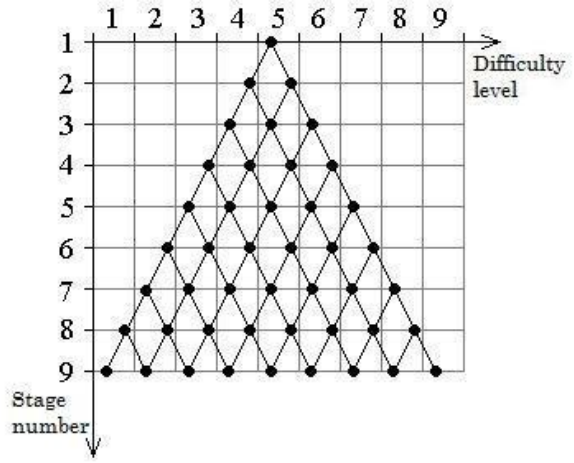

Figure 2. Distribution of items for 9-stage measurement

The general formula for determining the number of assignments on any difficulty level can be used.Let us assume that we have K difficulty levels.Average difficulty level number is defined as the integral part of the number of levels division in 2: obviously, if K is even, it is $K/2$, and toobtainan odd K we get $(K+1)/2$. At the level of difficulty with number, lower than $K/2$, $2i$ tasks are used, were i is the number of levels. For levels of difficulty with numbers, larger than $K/2$, the number of tasks is equal to $2(K-i)+1$, where i is the number of levels. In total, in the test $\frac{(1+K)K}{2}$ will be used.

Table 1 shows the number of items on each level and in test in general for different values of measurement stage number.It is clear that the pyramid strategy can be used only if a large amount of items of different difficulty levels is presented. However, it corresponds to a simplified understanding of the multistage adaptive measurements.

Table 1. The number of items on each level of the pyramid strategy

Number measurement stage	*Items in total*	*Items on level number*																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	15	2	4	5	3	1															
6	21	2	4	6	5	3	1														
7	28	2	4	6	7	5	3	1													
8	36	2	4	6	8	7	5	3	1												
9	45	2	4	6	8	9	7	5	3	1											
10	55	2	4	6	8	10	9	7	5	3	1										
11	66	2	4	6	8	10	11	9	7	5	3	1									
12	78	2	4	6	8	10	12	11	9	7	5	3	1								
13	91	2	4	6	8	10	12	13	11	9	7	5	3	1							
14	105	2	4	6	8	10	12	14	13	11	9	7	5	3	1						
15	120	2	4	6	8	10	12	14	15	13	11	9	7	5	3	1					
16	136	2	4	6	8	10	12	14	16	15	13	11	9	7	5	3	1				
17	153	2	4	6	8	10	12	14	16	17	15	13	11	9	7	5	3	1			
18	171	2	4	6	8	10	12	14	16	18	17	15	13	11	9	7	5	3	1		
19	190	2	4	6	8	10	12	14	16	18	19	17	15	13	11	9	7	5	3	1	
20	210	2	4	6	8	10	12	14	16	18	20	19	17	15	13	11	9	7	5	3	1

Bisection strategy

Bisection strategy is based on scientific foundations of modern construction methods and use of modern test theory Item Response Theory (Lord, 1980; Hambleton, Swaminathan & Rogers, 1991). In Item Response Theory (IRT) the values of students' level of training and items difficulty in test are expressed in the same units of measurement - logits, and therefore they can be placed on a standard scale, which allows to correlate the level of any student with the measure of each item difficulty. The absolute value of the difference $(\theta - \beta)$ is the distance, where a student with θ level of ability from the item with β difficulty (Figure 3).

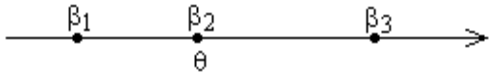

Figure 3. Geometric interpretation of θ and β ratio on the logit scale

If a negative difference is large in absolute value $(\theta - \beta_3)$, it means that student's level of ability is way lower then item difficulty, and most likely he will give the wrong answer. Large positive difference values $(\theta - \beta_1)$ also indicate

level of ability and item difficulty discrepancy, but the other way round. So, in this case the testee will successfully fulfill the item.

If items are arranged in order of increasing difficulty, the following can be stated:

– if a student has successfully fulfilled the item, then most probably he will fulfill the items of a lower difficulty level;

– if a student has not fulfilled the item, then most probably he will not be able to fulfill the item of a higher difficulty level.

These statements arrange tasks in such a way, when a studentis not givenitems, the fulfillment result of which is predictable with a high probability.

Let us assume that all the items are arranged in order of increasing difficulty, with all the items different in difficulty. Let a studentbe given an average difficulty level item. There are two possible situations: a student has not fulfilled the item, a student has fulfilled the item. In the first case most likely it can be assumed that a student will not fulfill more difficult items, in the second case he probably will fulfill easier ones. Thus, the items are divided into two groups. One half consists of the items, fulfillment of which can be predicted, so there is no need in giving them. The other half consists of the items, fulfillment of which can not be predicted: they can be both easier or more difficult than previous ones, so for the next step we have twice as little items than before.

By repeating the following stage of measurements, the set of items is divided into two parts (that is why the strategy is named as multistage adaptive measurements). Thus "the area of ambiguity" narrows. When only one itemis given at a stage, the measurement process is over.

Thus, if there are K number of items, then after the first test stage $K/2$ items are left, after the second one - $K/4$ items, after the third - $K/8$ etc. The measurement process continues till the condition of inequality $(K/2^i)>1$ is observed, where i – is the stage number. This inequality also lets determine the maximum number of stages required by amount of items – it is $[\log_2(K+1)]$, where $[\cdot]$ denotes the integral part of a number (Table 2). Figure 4 shows the strategy for 16 items.

Table 2. The number of stages in the bisection strategy

The number of items	8–15	16–31	32–63	64–127	128–255	256–511	512–1000
The number of stages	4	5	6	7	8	9	10

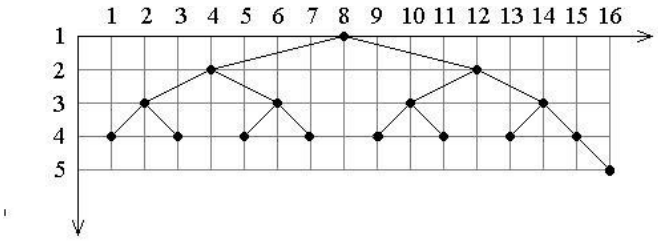

Figure 4. Bisection strategies for 16 tasks scheme

So, bisection strategy can be described as follows. Let us assume we have K number of items.

1) The item with the number $j = [K/2]$ is given, the stage number is i=1.
2) If $(K/2^i)>1$, then we pass to line 3, otherwise the test is over.
3) Increase the number of i stage by 1. If the item j is successfully fulfilled, then item $[j+K/2^i]$ is given, if not – then item $[j–K/2^i]$. Proceeding to line 2.

With the number of tasks available, the strategy described allows to get results within a minimum number of stages, which leads to a decrease in measuring instrument length, but there is a loss of measurement accuracy, and thus of the reliability of the adaptive measurement.

Flexible strategy is the presentation of items in algorithm, which predicts optimized difficulty of the item subsequent on the basis of fulfillment the previous one by a student. The main feature and at the same time advantage of flexible strategy is a student's level of ability stepwise revaluation after each fulfillment of the item. As a result, a peculiar sequence of θ values is defined, in relation to which β item difficulty values are selected. The item difficulty and the step vary in such adaptive measurement strategy. The step is determined by the difference between the difficulty of two adjacent item sequences. Implementation of this strategy is possible only with the application of IRT of chance models, which allow to predict the success of the next item fulfillment at a fixed θ value.

The choice of the mathematical model, describing the relationship between the empirical results of the measurement and the values of θ and β latent parameters, is core in IRT. The basic assumption in IRT is an existence of a math model for the relationship between the empirical results and the values of θ and β latent parameters. Researches conducted by A. Birnbaum (1968), F. M. Lord (1980), G. Rasch (1980) on the regression line analysis of the test fulfillment results on θ latent variable come up to a conclusion that the relationship between testees and their true scores is of non-linear nature.

The relative invariance of latent variables values of a particular measurement, certain frequency stability of their values occurrence were the basis for the concept of event probability as a measure of its occurrence possibility. As such event the researchers chose the correct answer of i testee to j item. The transitional probability of i testee's with θ_i level of ability successful fulfillment of different in difficulty items can be considered, when θ_i is i testee parameter, and β *is* independent variable. Then the transitional probability P_i is a function of latent variable β:

$$P_i\{x_{ij}=1|\theta_i\}=f(\theta_i-\beta), i=1, 2, \ldots, N. \quad (1)$$

Similarly, the transitional probability of successful j item fulfillment with β_j difficulty level by different testees is introduced. In this case, the independent variable is θ, and β_j is a parameter that determines the j task difficulty. Then

$$P_j\{x_{ij}=1|\beta_j\}=F(\theta-\beta_j), j=1, 2, \ldots, n, \quad (2)$$

Where x_{ij}=1, if i testee answer to j item is correct, or x_{ij}=0, if i testee answer to j item is wrong; N - number of testees, n - the number of items in the test.

In IRT functions (1) and (2) are denoted as$P_i=f(\beta)$ and $P_j=F(\theta)$respectively, and are referred to as Item response functions (IRF). The graph of the first function is a decreasing individual curve of a testee (Figure 5), and the second increasing function is response curve of a item (Figure 6).

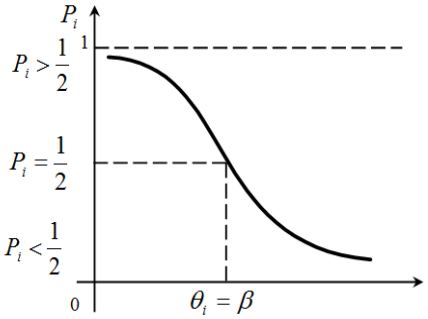

Figure 5. The graph of $P_i=f(\beta)$ (individual curve of a testee)

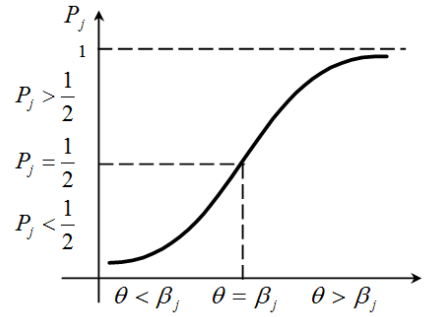

Figure 6. The graph of $P_j=F(\theta)$ (response curve of item)

The number of parameters, involved in analytic function setting, is the basis for the division of IRF family into classes. Among the logistical functions they distinguish several ones, most suitable for practical usage – one-parameter G. Rasch model (1980), two- and three-parameter A. Birnbaum models (1968). The stated models are below.

One-parameter G. Rasch model (1980):

$$P_j(\theta) = \frac{e^{1,7(\theta-\beta_j)}}{1+e^{1,7(\theta-\beta_j)}} \quad (3)$$

$$P_i(\beta) = \frac{e^{1,7(\theta_i-\beta)}}{1+e^{1,7(\theta_i-\beta)}} \quad (4)$$

where θ and β are independent variables for the first and second functions respectively.

Two-parameter A. Birnbaum model (1968) :

$$P_j(\theta) = \frac{e^{1,7a_j(\theta-\beta_j)}}{1+e^{a_j 1,7(\theta-\beta_j)}} \quad (5)$$

$$P_i(\beta) = \frac{e^{1,7a_i(\theta_i-\beta)}}{1+e^{1,7a_j(\theta_i-\beta)}} \quad (6)$$

where a_j is a parameter of aitem differentiation ability, indicating the item differentiation ability upon measuring θlevel of training different values and taking values in the interval (0,5; 2,5), and a_i – is a parameter characterizing the structure of testee's knowledge structure.

Three-parameter A. Birnbaum model(1968):

$$P_j(\theta) = c_j + (1-c_j)\frac{e^{1,7a_j(\theta-\beta_j)}}{1+e^{a_j 1,7(\theta-\beta_j)}}, \quad (7)$$

where all the symbols remain the same, and $c_j$is a parameter characterizing the probability of giving a correct answer to j item in case the answer is guessed.

For implementation of any strategy described above it is necessary to fulfil the following conditions:

–availability of assessment tools and measurement instruments base with stable characteristics and obtained by model IRT selected;

–availability of computer programs or software-tools, where one or more of models IRT selected are used, contributing to the obtaining of the maximum accuracy when assessing student level of training;

–having specification, which provides content validity of the simulated test.

From a didactic point of view, the latter condition is of special importance. Valuable elements of learning outcomes and competencies demonstration, the assessment of which is planned in specification, are to be considered in order to get unbiased and comparable results when measuring.

The authors of the article in experimental work made their choice in favor of flexible strategy (Figure 7 shows the implementation of the algorithm). This decision was made due to the comparative theoretical analysis of literature on adaptive testing results (van der Linden& Glas, 2010; Wainer, 2000; Weiss, 1983), as well as accumulated practical experience on beta-testing of educational measurements and types of measuring instruments different models, including paper-and-pencil and computer testing with different options of presenting the items (determinate or random) and time restrictions (Chelyshkova, 2001;Malygin, 2012) .

In this study for the practical implementation two-parameter model A. Birnbaum (1968) was selected (5) due to the following reasons. In case of a large array of on-line statistical data processing when multistage adaptive measurement takes place, the choice of excessively complex model (i.e. three-parameter model, calculating the possibility of guessing the correct answer) results in less reliable and valid results of measurement due to the poor convergence of iterative methods of latent variable accuracy determining - θlevel of ability and a long duration test.

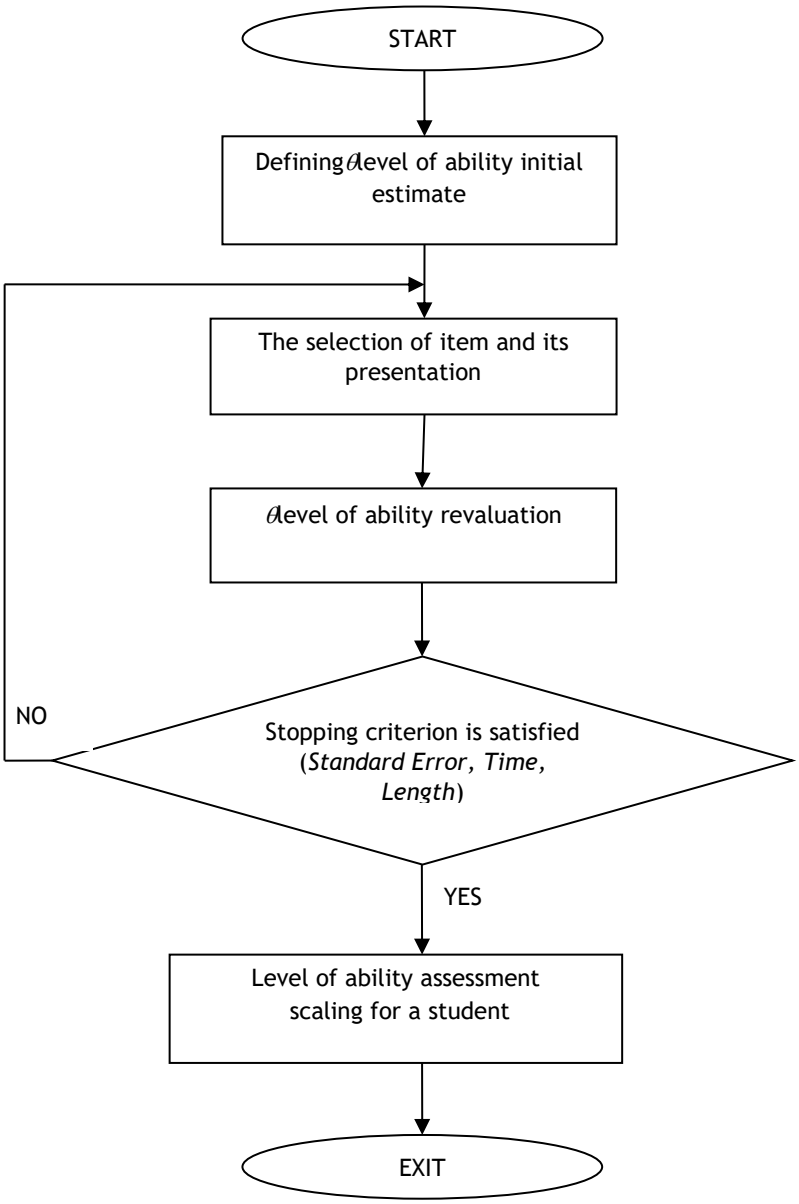

Figure 7. Flexible strategy algorithm of adaptive measurement

The advantages of the chosen strategy are:

–values of measurement error as a stopping criterion of adaptive measurement algorithm. It is obtained by evaluating the variable measurement, which allows to predict the reliability of the test;

–individual adaptive trajectory measurements for each student;

–the possibility of difficulty variation and differentiation ability in selection of the items optimized for each student.

The results of university students' questionnaire survey (series of 218 results) and personal communication on learning outcomes assessment led to the conclusion of a positive attitude to such assessment methods. They are based, first, on the principles of equality and objectivity, and, second, are organized through computer systems, information and communication technology.

Along with the above mentioned advantages, still remains the question of how to enter the adaptive mode of measurement, i.e. from task of what difficulty level one should start. Several approaches for the start of measurement procedure can be suggested. The first approach is with no information on student initial level of ability, it is necessary to focus on the average rating in the group. The second approach relates to the two-stage measurement and is based on the use of the pretest implementation results. The third approach suggests to start with a relatively easy item, giving time to the student for adaptation.

The fragment of the adaptive measurement algorithm is below.

The adaptive measurement algorithm process.

Initial conditions: availability of measurement instruments base with stable values of $\beta_j$item difficulty and its a_j differentiation ability.

There are 10 items from the base:

j item	β_j	a_j
1	-2.237	0.397
2	-1.116	0.537
3	-0.469	1.261
4	-0.103	0.857
5	0.067	1.471
6	0.241	0.92
7	0.495	1.382
8	0.801	0.94
9	1.17	1.29
10	1.496	1.44

Step 1. The student is given the easiest item from the base (according to this table, it is item 1).

Step 2. The student's answer is correct, i.e. his pattern of responses is {1}.

Step 3. In accordance with pattern of responses, likelihood function for the entire range of θvalues with known β_1=-2.237 and a_1=0.397 is calculated. θvalue, for which the likelihood function takes the maximum value, is chosen. The general view of the likelihood function is $L(x_j \mid \theta_i)=P_j^{x_j}Q_j^{1-x_j}$, where $P_j = \frac{e^{1,7a_j(\theta-\beta_j)}}{1+e^{1,7a_j(\theta-\beta_j)}}$is the probability of a correct answer to j item, and $Q_j = 1 - P_j$ isthe probability of a wrong answer.

After calculation, we find that likelihood function L_1receives the maximum value at $\theta = 4$.

*Step 4. T*he values of the information function at θ=4 for the remaining items with β_j difficulty are calculated and the item with the maximum value of this function is presented. In our case, it is item# 8$I(\theta = 4)$=0.013, so it will be given to a student.

Step 5. Depending on the answer of the student (true / false), the likelihood function is being recalculated and its maximum value is selected. Let us assume that the student has given the wrong answer. Then his pattern of responses becomes {1; 0}, taking into concideration his previous answer. The likelihood function in this case is as follows: $L(1; 0)=P_1(\theta)\cdot Q_2(\theta)$. The maximum value of this function will be achieved at θ=0.5.

Step 6. For θ = -0.5 level of ability the item with the highest value of the information function is being selected. In our case, it is item3.

*Step 7. N*ew pattern of responses is arranged by adding a value of 0 or 1, depending on his answer to the current item. Let us assume that the answer was correct, then the pattern of responses is {1, 0, 1}, for which the likelihood function is being re-calculated again: $L(1; 0; 1)=P_1(\theta)\cdot Q_2(\theta)\cdot P_3(\theta)$. It is again θvalue, corresponding to the maximum value. In the example studied, the maximum value of the likelihood function is achieved at θ = 0.0

Step 8. Step 6 repeats. Item 5 is given.

Step 9. Step 7 repeats.

And so until either the condition $|\theta_k - \theta_{k-1}| \leq \varepsilon$if fulfilled or stopping rule does not work, for example, it happens when standard measurement error is $(\theta) = \frac{1}{\sqrt{I(\theta)}} \leq 0.6$.

Let us assume that in our example, after successful fulfillment of item 5 a new value θ = 0,5 is obtained. The student gives a wrong answer to the next question. The pattern of responses is {1; 0; 1; 1; 0}. The likehood function is respectively $L(1; 0; 1; 1; 0)=P_1(\theta)\cdot Q_2(\theta)\cdot P_3(\theta)\cdot P_4(\theta)\cdot Q_5(\theta)$. The maximum value of a new likelihood function is now achieved at θ = 0,0. In this case, we obtain the following sequence of level of ability values {4.0; -0.5; 0.0; 0.5; 0.0}.

Step 10 . Calculating the standard measurement error using the above formula gives a value of 0.58, which is less than the established criterion. Then student's level of ability in logits is equal to θ = 0.0 with a standard measurement error of 0.58.

For the result notification to a student, his individual score of level of ability is to be scaled.

Thus, each student passes the test in his own individual mode with different pace and number of items. It is clear that first to complete the test are students with very high and low level of ability, since only these two groups are given the items corresponding to their actual level of knowledge and easy to be fulfilled. The latter increases the motivation to fulfil the items given.

The condition of achieving the planned standard measurement appears as a stopping rule of adaptive measurement. Then, the final assessment of the level of ability is expressed as $\theta \pm Se(\theta)$.

By now, in education knowledge field the aspects, related to technologic and software implementation of adaptive learning and testing algorithms under conditions of engineering student training (Teryuha, 2006); increase the

effectiveness of adaptive testing of training quality of students of the Humanities (Gorbachev , 2006); maintenance of computer adaptive testing in a secondary vocational education (Minko , 2010) were studied. It can be assumed that the examined issue of learning outcomes assessment in the frames of increasing the effectiveness of assessment procedure has its unique solution in a rapidly changing technological and information professional education components.

The results of authors' experimental work provide reasonable data to claim that observance of educational (individualization, differentiation, interactivity, systematicity) and technological (availability of measurement instruments base with stable evaluations of their difficulty parameters and differentiation ability, software-tools environment, readiness of teaching staff) terms, as well as the use of the described algorithm of multistage adaptive measurement implementation, do not underestimate the measurements accuracy, and thereby the effectiveness of student assessment procedures is ensured.

Among the important advantages of multistage adaptive measurements implementation are:

1) high efficiency, achieved by minimizing the number of items and test duration, when the measurement error is not greater (equal to or lower than) than making one in traditional tests, similar in content;

2) the high level of sensitivity, substantially excluding the possibility of cheating, tips and other undesirable actions during tests;

3) individualization of the test pace provided by adaptive algorithms and related software, with the help of which the selection of the next item on another difficulty layer (measuring instrument) takes place only after the previous item has been fulfilled;

4) increasing the level of motivation for learning outcomes assessment for weaker students by eliminating excessively difficult items, causing the growth of anxiety factor and fear;

5) immediate result notification on an interval scale of test scores for each testee right after finishing his test with individual sets of item;

6) elimination of time, organizational and financial expenses on standardization for establishing test standards due to the lack of traditional measurements of fixed length.

Conclusion

In the context of educational programs spectrum expansion and the development of new learning forms, the student and learning outcomes assessment procedure should be conducted in immediate and highly-efficient mode. It is stated that the multistage adaptive measurements lead to the increase of student assessment procedures effectiveness through the use of IRTmodels (in particular, two-parameter A. Birnbaum model), creating a situation of success for each student by selecting the tasks, appropriate for his level of training and the implementation of individualization principle.

Implications and Recommendations

This article can be of interest for experts in learning outcomes assessment quality and for teaching staff, aimed at practical implementation of new

assessment forms and technologies into student assessment process, and at the use of mathematic methodic of construction and IRT educational tests.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Evgenij M. Dorozhkinis Doctor of Education, FullProfessor, Rector of Russian State Vocational Pedagogical University, Ekaterinburg, Russia.

Marina B. Chelyshkovais Doctor of Education,Full Professor, Depute Director of the Department of Management Education Quality in State University of Management, Moscow, Russia

Alexey A. Malyginis Candidate of Education, Head of the Department of Educational Programs Office in Ivanovo State University, Ivanovo, Russia.

Irina A. Toymentsevais Doctor of Education, Professor of Samara State University of Economics, Samara, Russia.

Tatiana Y. Anopchenko is Dean of Management Faculty, Doctor of Economics, Professor, Southern Federal University, Rostov-on-Don, Russia.

References

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability / In: F.M. Lord and M.R. Novick. Statistical Theories of Mental Test Scores. Reading, Mass: Addison - Wesly, 568 p.

Boyatzis, R. (2008). *The Competent manager. A model for effective performance.* Moscow: HIPPO publ., 352 p.

Chelyshkova, M. B. (2001). *Adaptive testing in education (theory, methodology, technology)*. Moscow: Research Center of Training Quality Problems, 165 p.

Chelyshkova, M. B.. (2002). *Theory and practice of designing of pedagogical tests*. Moscow: Logos, 432 p.

Crocker, L. &Algina, J. (2006). *Introduction to Classical and Modern Test Theory*. Pacific Grove, CA: Wadsworth, 527 p.

Dorozhkin, E.M., Zaitseva, E.V. & Tatarskikh, B.Y. (2016). Impact of Student Government Bodies on Students' Professional Development . *IEJME-Mathematics Education, 11*(7), 2666-2677.

Gorbachev, V. T. (2006).*Improving the efficiency of adaptive testing the quality of student learning in higher education of a humanitarian profile*. PhD Thesis. Moscow: Military University of the Russian Federation Ministry of Defense, 258 p.

Hambleton, R. K., Swaminathan, H.& Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. N.-Y. :Sage Publications, 174 p.

Lord, F.M. (1980). *Application of Item Response Theory to practical testing problems*. Hillsdale N.-J. : Lawrence Erlbaum Ass.,Publ, 266 p.

Malygin, A. A. (2011). Computer adaptive testing to quality assurance of distance learning. *Vestnikuniversiteta*, *4*, 166–169.

Malygin, A. A. (2012). *Adaptive testing in distance learning*. Ivanovo: ISUCT, 138 p.

Minko, N. T. (2010).*Pedagogical support of computerized adaptive testing in the context of personal education*. PhD Thesis. Ulan-Ude: Buryat State University, 209 p.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. With a foreword and afterword by B.D. Wright. The Univ. of Chicago Press. Chicago & London, 199 p.

Spencer, L. M.&Spencer, S. M. (2005). *Competence at Work. Models for Superior Performance*. Moscow: HIPPO, 384 p.

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science. New Series*, Vol. 103, 2684, 677-680.

Teryuha, R. V. (2006). *The technology of computerized adaptive testing in vocational training of engineers*. PhD Thesis. Krasnodar: Kuban State University, 261 p.

Van der Linden, W. J.&Glas, C. A. W. (2010). *Elements of adaptive testing, Statistical for social and behavioral sciences*. Springer Science + Business Media, LLC, 437 p.

Wainer, H.(2000). *Computerized adaptive testing: A Primer*. – 2nd edition. – Mahwah, NJ :ELawrence Erlbaum Associates, 278 p.

Weiss, D. J. (1983). *New horizons in testing: Latent trait theory and computerized adaptive testing*. N. Y. : Academic Press, 380 p.

Zvonnikov, V. I. (2006). *Measurement and the quality of education*. Moscow: Logos. 312 p.