Journal of Modern Education Review, ISSN 2155-7993, USA
February 2014, Volume 4, No. 2, pp. 97–104
Doi: 10.15341/jmer(2155-7993)/02.04.2014/003

http://www.academicstar.us

The Optimization of Interdisciplinary Competence Assessment on the Base of Multistage Adaptive Measurement

Victor Zvonnikov[1], Marina Chelyshkova[1], Malygin Alexey Aleksandrovich[2]
(1. State University of Management, Moscow, Russia; 2. Ivanovo State University of Chemistry & Technology, Ivanovo, Russia)

Abstract: In this paper the approach to increasing of measurements efficiency in competence assessment is offered. This approach is based in the mathematical models of the modern measurement theory (Item Response Theory) and the model of multistage measurements for interdisciplinary competence assessment. The IRT models are discussed in context of two problems: algorithms for adaptive testing and comparative analysis of their the possibilities for processing empirical results from the dichotomous and polytomous scored items. By this connection some models and some inequalities are resulted which allow to optimize competence assessment.

Key words: ability parameter, competence assessment, difficulty parameter, inter-disciplinary assessment

1. Introduction

The development of competence assessment in vocational training has the focus on the replacement of subject-based mono-disciplinary learning by the learning that is based on multi-disciplinary areas (modules). The new technologies of learning demands the development of innovative competence assessment which should be spent within the limits of High-Stakes Testing and be carry out the estimation of competence levels of development for each from the competencies set. This priority requirements at the development of competence assessment defines some necessary conditions for the models of measuring instrument as well for the technologies of measurement and interpretation of students results by threshold points. The competence approach has additional difficulties because we must provide the optimum efficiency of assessment in the conditions of a combination inter-disciplinary content of measuring instrument with high objectivity (reliability) and high validity of graduates scores.

For realization of these aims in inter-disciplinary competence assessment it is necessary to use multistage and adaptive measurement and special models which are offered in our paper. It allows to measure professional competencies during assessment with high objectivity (reliability) and high validity. The methodology of our approach includes some directions of researches: models of multistage measuring instruments, theory used for measuring instruments construction, algorithms of adaptive testing and further scaling and interpretation the

Victor Zvonnikov, Dr., Professor, Prorector, State University of Management; research areas: educational measurement and adaptive testing. E-mail: zvonnikov@mail.ru.
Marina Chelyshkova, Dr., Professor, Director of Quality Center, State University of Management; research areas: educational measurement and adaptive testing. E-mail: mchelyshkova@mail.ru.
Malygin Alexey Aleksandrovich, Candidate of Science (Pedagogical), Ivanovo State University of Chemistry and Technology; research areas: educational measurement, items response theory, computerized adaptive testing. E-mail: a_malygin@mail.ru.

graduate results for the quantitative assessment of professional competences. The choice of methodology was under the influence of main idea — to optimize competence assessment for scoring in assessment of professional competences.

The optimization requires measuring instrument including some subtests with different forms of inter-disciplinary items for multistage measurements (Chelyshkova M. B. & Zvonnikov V. I., 2012), the adaptation of difficulty items and their number in the conditions of high reliability and validity measurement results (Lord F. M., 1980; Zvonnikov V. & Chelyshkova M., 2013), using Item Response Theory to represent graduate's results at a uniform interval scale for developing large-scale competence assessment (Bond T. G., Fox C. M., 2007; Hambleton R. K., 1992; Lord F. M., 1980; Malygin A. A., 2012; Zvonnikov V. & Chelyshkova M., 2013) and the comparison of different mathematical IRT models on the base of Item Information Functions to draw the conclusion about optimal model combination for competence assessment (Baker F. B., 2004; Hambleton R. K., 1992; Lord F. M., 1980). All this topics are discuss in our paper.

2. The Models for Multistage Measurements

For the choice of stages number and kinds of subtests in measuring instrument we took into consideration the following requirements which are the set of didactic conditions:

- The interdisciplinary competencies are regarded as a key dimension of vocational competence assessment.
- The choice of measuring instrument components should provide the maximum "transparency" of assessment results, ease and simplicity of result interpretations for students and teachers and high correlation with requirements of Educational Standards.
- The measuring instrument structure should promote high validity of threshold points for classification of graduates.
- During carrying out of multistage measurements the growth of number of stages does not provide automatically quality of measurement results, therefore the quality of each component in measuring instrument should be in the center of developer's attention.
- The constructing of multistage measurement models should be based on the system approach and be conformable for age group of trainees.
- For constructing of multistage measurement instrument it is necessary to use the adaptive testing. On the base of adaptive testing methods we can minimize the measurement errors and duration of testing and maximize the validity of graduates scores.

It is possible to offer the model of multistage measurements for the purpose of competence assessment on the base of requirements of Educational Standards. It has three-level of multistage measurements and interdisciplinary competence in accordance of competence assessment (Figure 1).

This model provides higher differentiating effect in graduate's scores but demands more expenses for carrying out High Stakes Testing in competence assessment. In this model special attention is given to mini-cases. A key quality to consider in using mini-cases is authenticity, the degree to which the assessment reflects the competence development. So mini-cases are the most authentic tools for competence assessment.

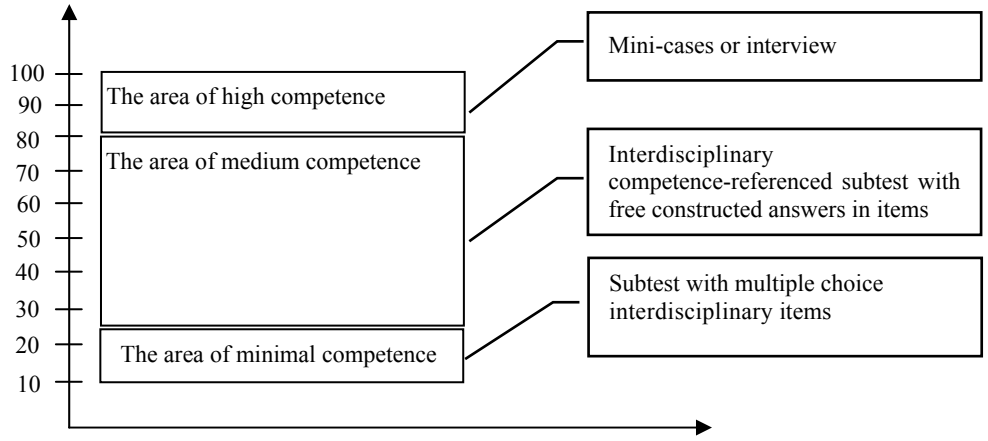

Figure 1 The Model of Measuring Instrument for Three Stage Competence Assessment

3. The Optimization of Measurement Methods for Various Stages

The development of Item Response Theory (IRT) models have allowed to compare the learner's level of ability and level of item difficulty. The idea of comparison has been realized in adaptive testing on the basis of item selection by the equation $\theta = \beta$, where the scores of ability parameter θ and the scores of difficulty parameter β are in logits (Baker F. B., 2004; Lord F. M., 1980; Van der Linden W. J., 2010). It is the main advantage of IRT where item location (β) and the person trait level (θ) are indexed on the same metric. The equation $\theta = \beta$ helps to optimize item selection for assessment. It provides the minimum time of testing for each learner in measurements with high estimations reliability for every learner's score.

In adaptive testing the value of probability of correct item performances $P_i(\theta_i - \beta)$ for assessment is set by the inequality $|P_i(\theta_i - \beta) - 0.5| < 0.1$, where θ_i- level of i^{th} learner ability, β- difficulty of items and all item are locally independent (Malygin A. A., 2012; Zvonnikov V. & Chelyshkova M., 2013). Within the one-parametrical dichotomous model G. Rash (Baker F. B., 2004) it is possible to write probabilities P_{ij} in the form of

$$P_{ij} = \frac{e^{1,7(\theta_i - \beta_j)}}{1 + e^{1,7(\theta_i - \beta_j)}}, \quad Q_{ij} = \frac{1}{1 + e^{\theta_i - \beta_j}}$$ and $P_{ij} = 1 - Q_{ij}$ and β_j - difficulty of j^{th} item.

After some transformations of inequality for probability of correct item concerning parameter difficulty we have values β in the range $-0.20 < \theta_i - \beta < 0.24$ or $\theta_i + 1.96Se(\theta) - 0.24 < \beta < \theta_i - 1.96\ Se(\theta) + 0.20$, taking into account borders of a confidential interval for parameter estimations at a significance value $\alpha = 0.05$. Such items are optimum for assessment on the base of adaptive testing.

Within two-parametrical dichotomous model of IRT (Baker F. B., 2004) it is possible to write probabilities P_i in the form of formula

$$P_i(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_j(\theta - \beta_j)} e^{-\frac{z^2}{2}} dz$$, where

$$a_j = \frac{(r_{bis})_j}{\sqrt{1 - (r_{bis})_j^2}}$$ and r_{bis} - biserial correlation coefficient

As well as earlier for assessment the probability of correct item performance is defined by inequality $|P_i [a_i \cdot (\theta_i$

- β)] – 0.5| < 0.1 and product $a_i \cdot (\theta_i - \beta)$ will change in interval (0.20; 0.24). Unlike the situation considered above within one-parametrical model, now the range of estimations of item difficulty parameter values will be defined not only by value θ_i, but also by value a_i. So for i^{th} learner in the assumption of positive values a_i the inequality for β looks like $\theta_i - 1/a_i \cdot 0.24 < \beta < \theta_i + 1/a_i \cdot 0.20$ (Zvonnikov V. & Chelyshkova M., 2013). Inequality $\theta_i - 1/a_i \cdot 0.24 < \beta < \theta_i + 1/a_i \cdot 0.20$ allows to draw an interesting conclusion about length of interval on an axis of item difficulties. In case of i^{th} learner structure of knowledge is high quality, corresponding to great values a_i, the borders of interval $\theta_i - 1/a_i \cdot 0.24 < \beta < \theta_i + 1/a_i \cdot 0.20$ for the organization effective adaptive assessment decrease. If values of parameter a_i begin to decrease, the width of interval defined by inequality $\theta_i - 1/a_i \cdot 0.24 < \beta < \theta_i + 1/a_i \cdot 0.20$ increases. Noted effect shows the influence of structure parameter values on borders of item difficulty ranges in assessment.

The algorithm of adaptive testing is shown in Figure 2.

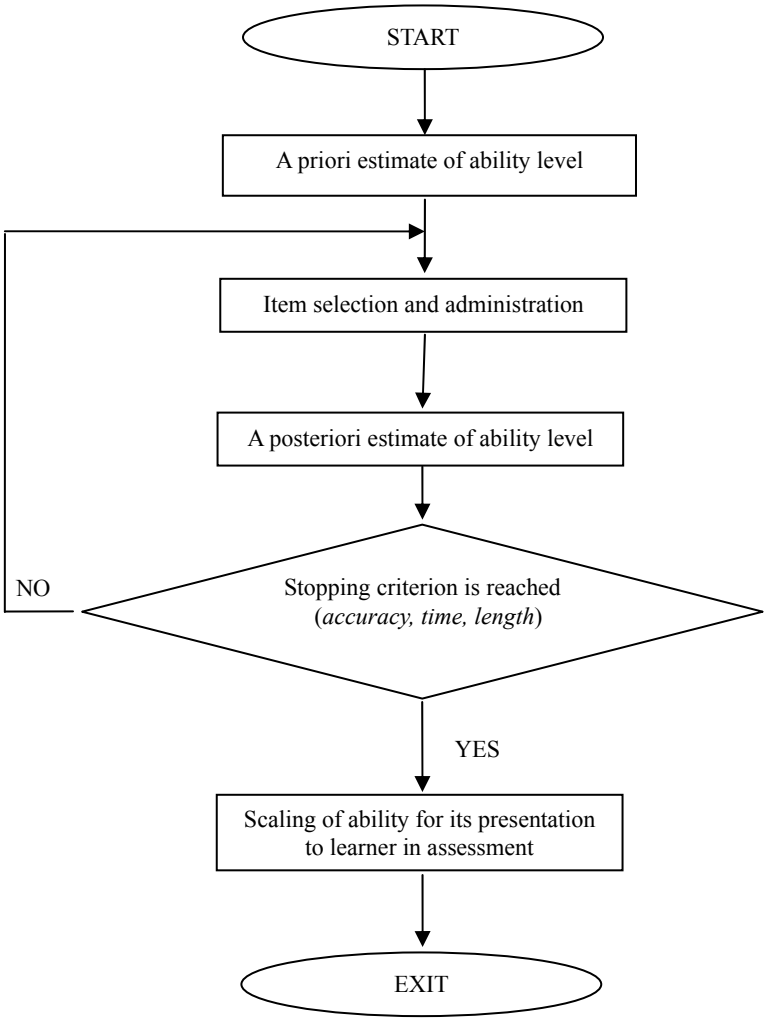

Figure 2 Algorithm of Flexi-Level Adaptive Testing

The application of adaptive testing in the modern variant demands the automated check of item performances, therefore it is possible to use adaptive testing during competence assessment only for first subtest with multiple choice interdisciplinary items. There are items with multiple response options (second and third components of model in Figure 1) in the model of measuring instrument which are typical for competence assessment. Such

items provide partial credit for partially correct answers when each response option is scored separately. So, the second and third of measurement stages suppose the experts participation for checking. Polytomous IRT models for such items look like quite differently from dichotomous models. In this cases, it is necessary to use Generalized Partial Credit Model (Baker F. B., 2004) which has the equation

$$P_{jg} = \frac{e^{aj}[\varphi_g(\theta - b_i) + \sum_{g=1}^{l} \tau_g}{\sum_{h=1}^{m} e^{aj}[\varphi_h(\theta - b_i) + \sum_{g=1}^{h} \tau_g]}$$

where

a_j is the slope parameter representing item discrimination;

b_j is the item location parameter;

ϕ_g is the usual scoring function that equals the category count for the specific category (g) being modeled

τ_g is the threshold parameter representing the category boundary locations relative to the item location parameter.

4. The Comparative Analysis the Possibilities of IRT Models.

The comparative analysis the possibilities of IRT models for measurement error minimization in multistage competence assessment was spent by the Item Information Functions for dichotomous and polytomous IRT models. As the whole, the equation for Item Information Function can be written in the form

$$I(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

where I(θ) denotes test information, conditional on θ, and L represents the likelihood function of measurement instrument. Despite of sign a minus the results is s positive value because the second derivate of the log itself is always negative.

This equation of Item Information Function for IRT models has enough simple form

$$I_j(\theta) = \frac{[P_j^{'}(\theta)]^2}{[P_j(\theta)][Q_j(\theta)]}$$

These formulas were used for experimental data received by three-level model of measuring instrument model by classical and adaptive technology of testing administration. The data processing was spent on the basis of polytomous and dichotomous models of IRT by the ConQuest (Baker F. B., 2004). We received some conclusions from total Item Information Curves which are represented in Figures 3, 4 and 5.

Figure 3 shows two curves of Item Information Functions. Continuous line corresponds to the case when the items of the bottom minimum range of competence were estimated dichotomous without adaptive testing technology and for processing empirical results the dichotomous model G. Rasch was used. The empirical results from the middle and top part of competence have been processed by Generalized Partial Credit Model. In the same Figure the shaped line shows the case when all empirical results from three stages have been presented in dichotomous form and have been processed by dichotomous model G. Rasch.

Figure 4 shows two curves of Item Information Functions too. But in this case without adaptive testing technology for processing of empirical results received by items of the bottom minimum range of competence two-parametrical dichotomous IRT model was used. Another empirical result has been processed by Generalized Partial Credit Model. The shaped line shows the same case as the continuous line in Figure 3.

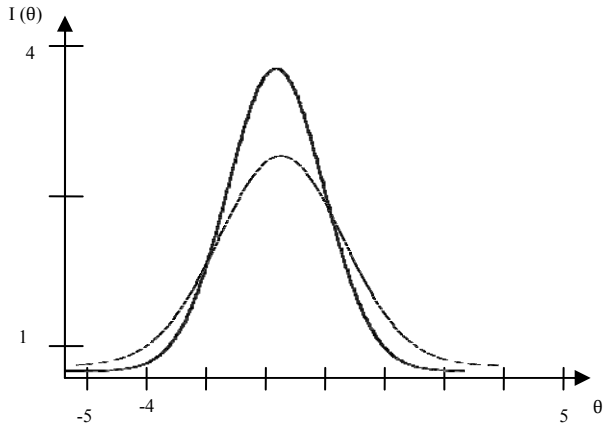

Figure 3 Two Curves of Item Information Functions (Dichotomous Model G. Rasch and Generalized Partial Credit Model)

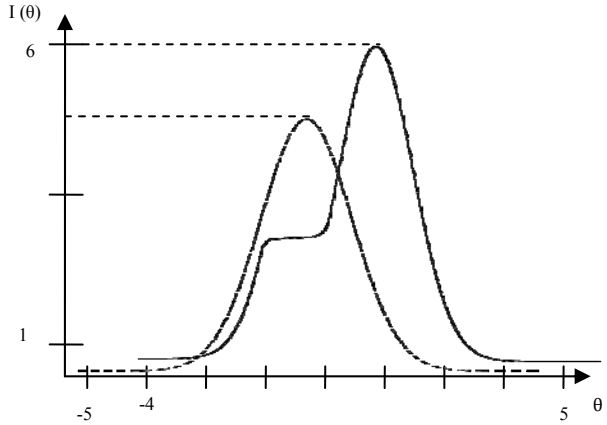

Figure 4 Two Curves of Item Information Functions (Two-Parametrical Dichotomous Model of IRT and Generalized Partial Credit Model)

In Figure 4 the top of a continuous curve reaches points with ordinate 6. But it has flat part that allows to draw the conclusion about its wrong form.

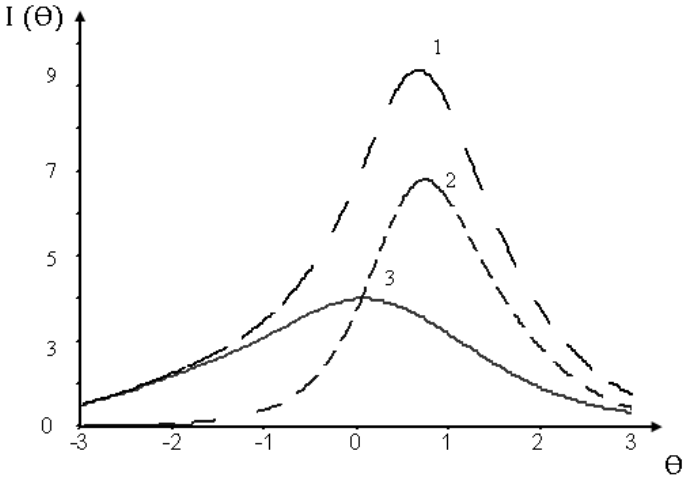

Figure 5 Three Curves of Item Information Functions (with Using Adaptive Testing Technology and Generalized Partial Credit Model)

Figure 5 shows three curves of Item Information Functions too. Curve 1 corresponds to the case when adaptive testing technology was applied during the first subtest administration and for processing of empirical results two-parametrical dichotomous IRT model was used. Another subtests for stages 2 and 3 have been administrated by classical way and for performance checking some experts involved. All empirical results for stages 2 and 3 have been processed by Generalized Partial Credit Model. The curve with number 2 is the one that applies without adaptive testing technology. For processing of empirical results received by items of the bottom minimum range of competence one-parametrical dichotomous model of IRT was used. The empirical results from stages 2 and 3 have been processed by Generalized Partial Credit Model. The third curve of information function shows the case when all empirical results from three stages have been presented in dichotomous form and have been processed by dichotomous model G. Rasch.

5. Algorithms for First Stage of Competence Assessment by Adaptive Testing

Algorithms of assessment demand rescoring learner's ability after performance every item of the adaptive test. If we use new symbol $T_j(\theta)$ instead of accepted earlier probability of the right answer $P_j(\theta)$ and designate observable dichotomizing results of examinee answers of the adaptive test by symbols $\{x_1, x_2, \ldots, x_j, \ldots x_k\}$ (j = 1, 2, …, k) we can enter likelihood function for Rasch model scores on "k" step of adaptive testing

$$L_k(\theta) = \prod_{j=1}^{k} [T_j(\theta)]^{x_j} \cdot [1 - T_j(\theta)]^{1-x_j},$$

where $L_k(\theta)$ - likelihood function.

The a posterior estimations of learner's parameter $\bar{\theta}$ after performance item "k" looks like

$$\bar{\theta}_k = \sum_{q=1}^{Q} t_q \cdot L_k(t_q) \cdot W(t_q) / \sum_{q=1}^{Q} L_k(t_q) \cdot W(t_q),$$

where t_q – quadrature points dividing the interval of possible distribution of measured variable θ from - 4 to +4 logits on equal parts and q = 1, 2, … Q. For the chosen number of quadrature points $t_{q+1} - t_q = 0.1$ and q = 1, 2, … Q; w (t_q) - weights in quadrature points, recalculated after performance of each next item of the adaptive test and

$$(\sum_{q=1}^{Q} W(t_q) = 1)$$

$L_k(t_q)$ - values of likelihood function in quadrature points.

The a posterior estimation of standard deviation for θ looks like

$$S_{ap}(\theta) = [\sum_{q=1}^{Q} (t_q - \bar{\theta}_k)^2 \cdot L_k(t_q) \cdot W(tq) / \sum_{q=1}^{Q} L_k(t_q) \cdot W(t_q)]^{\frac{1}{2}},$$

Where S_{ap} - a posterior estimation of standard deviation.

6. Conclusions

It is possible to formulate some conclusions creating the necessary preconditions for optimization competence assessment:

(1) For carrying out High Stakes Testing in competence assessment we must use multistage measuring instrument model. Three stages are optimum. This model with three stages provides higher differentiating effect in competence assessment. First stage is subtest with multiple choice interdisciplinary items, second stage is subtest with interdisciplinary competence-referenced items with free constructed answers and for third stage at area of high competence we must use mini-cases or interview. The mini-cases are the most authentic measuring instrument for competence assessment.

(2) The following parameters may serve as the indicators of competence assessment efficiency:

- time for tools administration and conduction of assessment,
- number of items needed for assessment,
- value of measurement error of learner's scores in assessment.

Our issue shows that all indicators can be realized if adaptive testing technology was applied during the first subtest administration and subtests for stages 2 and 3 must been administrated by classical way. The comparison of different mathematical IRT models for processing of empirical results on the base of Item Information Functions has allowed to draw the conclusion about the advantages of combination dichotomous model G. Rasch and Generalized Partial Credit Model for competence assessment.

(3) Item selection for adaptive testing with difficulty $\theta_i - 0.24 < \beta < \theta_i + 0.20$ allows to optimize first stage of competence assessment. If the learner's structure of knowledge is high quality, corresponding to great values a_i, the borders of interval $\theta_i - 1/a_i \cdot 0.24 < \beta < \theta_i + 1/a_i \cdot 0.20$ for the organization effective adaptive assessment decrease. Otherwise they increase.

References

Bond T. G. and Fox C. M. (2007). "Applying the Rasch model: Fundamental measurement in the human sciences", Lawrence Erlbaum Associates.

Baker F. B. (2004). *Item Response Theory: Parameter Estimation Techniques*, ASC. Univ. Ave.

Chelyshkova M. B. and Zvonnikov V. I. (2012). *Assessment of Educational Quality in Certification: Competency Approach*, M.: Logos.

Hambleton R. K. (1992). "Measurement advances to address educational policy questions", in: Plomp T., Pieters J. & Feteris A. (Eds.), *European Conference on Educational Research*, Enshede: Department of Education, University of Twente, pp. 681–684.

Lord F. M. (1980). *Application of Item Response Theory to Practical Testing Problems*, Hillsdale, NJ: Lawrence Erlbaum.

Malygin A. A. (2012). *Adaptive Testing in Distance Learning*, Ivanovo: ISUCT.

Zvonnikov V. and Chelyshkova M. (2013). "The optimization of formative and summative assessment by adaptive testing and zones of students development", *Journal of Psychosocial Research*, No. 1, Publications Pvt Ltd, New Delhi.

Van der Linden W. J. (2010). *Elements of Adaptive Testing, Statistical for Social and Behavioral Sciences*, Springer Science, Business Media, LLC.