SAM – Toolkit to assess primary school students' academic achievements

Piotr Nezhnov[*]

Introduction

The assessment of student achievement is one of the crucial issues of general education. Obviously, adequate assessment is an important precondition for effective interaction between teachers, school administrations and education management officials. Such a feedback system, including the use of testing materials, can present the community with currently recognized education objectives and, by establishing the level of students' achievement, provide a basis for making coordinated pedagogic and management decisions.

Clearly, to implement such an important function the assessment should meet the challenges which are faced at different levels of the education system. Thus, on one hand there are tasks of education system governance as a whole. And here an accurate measurement of learning achievements is very important because it provides the possibility of objectively ranging the success of regional and municipal systems of education, educational approaches, schools, students and making adequate policy and managerial decisions, e.g. providing assistance for certain activities and institutions, distributing limited resources (grants, scholarships) etc. On the other hand, there are problems related to the educational process itself. Program designers, training specialists, school administrators, teachers – those who are directly responsible for education results and improvement of school efficiency – participate in solving these issues. And in this context the qualitative diagnostics aspect of assessment becomes especially significant.

[*] Center for International Cooperation in Education Development (CICED) of the Academy of National Economy under the Government of the Russian Federation, nejnoff@gmail.com.

So the toolkit that would assure the adjustment of activity to different education system participants should ideally provide both aspects of assessment. However, the main efforts in toolkit development are, at present, related to the methodology of measurement, which has an obvious priority and also has a tendency to replace the methodology of qualitative analysis. This can be illustrated by international monitoring studies, which demonstrate the advanced examples of toolkit development for assessing students' achievement. Here one can refer to PIRLS (Progress in International Reading Literacy Study), TIMSS *(*Trends in International Mathematics and Science Study), PISA (Programme for International Student Assessment), the results of which cause great interest in specialists, policy makers and global society (Mullis, Kennedy et al., 2006; Mullis, Martin et al., 2003; OECD, 2003).

To begin with, the test materials of these studies are developed on the basis of one or another cognitive classification, which determines kinds of activity, typical for different aspects or levels of learning content attainment. In other words, the starting point in test development is a hypothetical cognitive structural picture of evaluated features (degree of training, attainment). E.g. in TIMSS test package for monitoring mathematics' mastering, a differentiation of four types of activity is introduced which make a sort of taxonomy: Knowing Facts and Procedures; Using Concepts; Solving Routine Problems; Reasoning. This qualitative typology serves as a basis for developing a set of test items.

However, later on the package is evaluated by the statistical processing of received data to create a scale, which no longer indicates the initial classification of the four types of activity Instead of this, four levels of mastering mathematics appear (advanced, high, middle, low), thus meeting the statistical (i.e. quantitative) criteria.

It should be noted that, finally, a statistical analysis is performed to determine test items completed by every group of students. So every level, with a certain probability, is referred to a set of items and then to a relevant list of skills, i.e. levels of competence acquire probability content parameters. Thereby, test results, being scaled, give the ground not only for ranking the examinees, but also for grouping them into four categories, with regard to mastering the certain set of skills. However, the qualitative difference between the four groups of skills turns out to be rather vague and requires additional conceptualization, which seems to be problematic in this case.

Such an approach to developing toolkits for the assessment of school students' achievements prevails nowadays. The main reason for that is a complicated situation about the use of taxonomies of educational goals – hierarchical classifications, which model an education process and are necessary prerequisites for quality assurance of education results.

The first peculiarity of the situation is the substantial number of different taxonomies. Following the classical version of B. Bloom (Bloom et al., 1956), a variety of level-related schemes have been developed, which differ from their original one way or another. In particular, the versions of I.J. Lerner, M. V. Simonov, and other authors appeared in Russian education (Lerner, 1980; Simonov et al., 1999). In the international education environment the cognitive classifications developed in TIMSS, PIRLS and PISA became widely known. The second peculiarity is a lack of a common theoretical background, which would provide the possibility of rational comparison and mutual criticism of various versions with the further generalization. The implicit compliance of pedagogy with the mechanistic approaches (association psychology, behaviorism) not adequate for the interpretation of education process (Davydov, 1972), makes the majority of existing taxonomies of equal worth from the philosophical point of view. As a result, the subject of discussion moves to the area of pedagogical expertise, thus partly explaining why this multiplicity is so stable.

One can hope to overcome this difficult situation by developing integral theoretical approaches which bring into the system concepts from the main constituents of education process (such as “learning”, “development”, “education result”, “understanding”, “competence” etc.) and by giving footholds to pedagogical taxonomies suitable for making an assessment instrument. I think that some prerequisites for providing a solution to this problem have been created by the Russian psychological school of L.S. Vygotsky.

Theoretical background

According to L. Vygotsky the growth (maturation) of a child is a specific process which should be interpreted psychologically as “cultural development” (Vygotsky, vol. 3, pp. 85, 225, 291-316). Education process in a wide sense of its meaning is an organized version of cultural development. The main propositions of this model are as follows:

1) the cultural development has genetic and functional aspects. The genetic aspect relates to the age cycle (and respective educational stages), and the functional aspect relates to the process of learning (Vygotsky, vol. 2, p. 305; vol. 3, pp. 128, 147-151);
2) learning is a necessary condition for cultural development, because it stimulates and gives direction to this process. Through learning a student is appropriating sign structures (systems of concepts, schemes, rules, samples), which crystallize culturally normal patterns of acting. These

structures function as psychological tools, i.e. they provide orientation (guidelines) for actions (Vygotsky, vol. 3, p. 78-86);

3) transmitting the sign structures to the child is only the beginning of the educational process. 'Seeds of knowledge' create a zone of proximal development (ZPD), and a process of functional development starts - during which a child fulfils the ZPD. That means he or she actively reconstructs and adopts the substance of accepted knowledge and the ability to act is growing. This pattern reconstruction and adoption is exactly the inner spontaneous process the results of which (intermediate and final) are really important for teachers, school administrators and other members of the educational community (Vygotsky, vol. 2, pp. 188-202, 244-267);
4) functional development is not a monotonous movement from an imperfect skill or competence to a perfect one, but is a successive attainment of three key points (or levels) connected with three possible options of retaining a sign as a cultural pattern holder:
   – retaining "external structure of sign"; "external, associative", "magic" usage of signs; usage "without understanding how they work" (Vygotsky, vol. 2, pp. 115; v.3, p. 157, 159, 161);
   – retaining a sign with understanding how it works; reconstruction and usage of the "connection" crystallized in a sign (Vygotsky, vol. 3, p. 159);
   – "imbedding" a sign "in the structures of mind"; "the last phase of internalization" (Vygotsky, vol. 2, pp. 108-117; vol. 3, pp. 157-163).

Later, due to the research of V.V. Davydov, D.B.Elkonin, B.D.Elkonin, P.J.Galperin, P.G.Nezhnov, O.V. Savelieva, A.V.Zaporozhets and other authors (Davydov, 1996, pp. 228-237; 1988; B.Elkonin, 1994; D.Elkonin, 1989, pp. 60-77, 494-495; Galperin, 1998 ; Nezhnov, 2007; Savelieva, 1989; Zaporozhets, 2000 et al.) the key points referred to were specified by associating them with the main types of pattern guidelines crystallized in a sign:

– level 1 – reproductive or formal – reconstruction of external characteristics of cultural pattern of action (algorithms, rules, forms of actions);
– level 2 – reflexive or essential – reconstruction of fundamentals (substantial relationship) of a pattern of action;
– level 3 – functional – reconstruction of possibilities of a pattern of action.

These three successive points of adoption of the cultural pattern create a basic taxonomy of educational objectives; this taxonomy has some psychological background, i.e. it hypothesises psychological structures which are crucial in passing from an immature stage to a mature one in terms of skill. In this taxonomy a level designates a type of cultural sample retention by a child, with the resulting possibilities of thinking and acting, as a qualitative characteristic of the educational result.

At the first level of formation the generalization of a skill is minimal and covers a narrow spectrum of typical situations. The second level provides the possibility, in principle, to fulfill all kinds of tasks which are in accordance with the cultural pattern. On the third level of formation where the psychological nature is the least studied (Galperin, 1998, pp. 364, 388; Nezhnov, 2007; Zaporozhets, 2000, pp. 518-659) the skill is characterized by functionality, i.e. by the possibility of being used freely in different contexts. This level of skill mastery corresponds to some extent with the definition of such a category of key competences as “using tools interactively” (http://www.oecd.org/dataoecd/47/61/35070367.pdf, p. 5).

Method

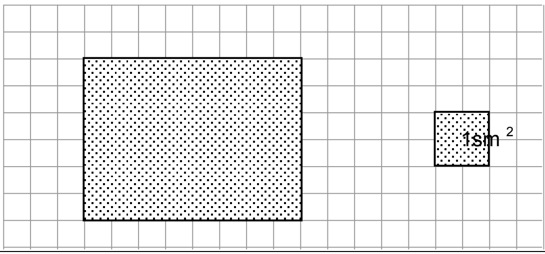
SAM is the attempt of the group of Russian specialists headed by P. Nezhnov, E.Kardanova and B. Elconin to develop a measurement toolkit on the basis of their described model (Nezhnov et al., 2009). The aim of the toolkit is to assess the subject competences of primary school students (age 10-11) on three basic levels: formal, essential and functional.

The particular characteristic of this instrument is: for each subject content area test items have been developed which correspond to the 1st, 2nd and 3rd level. Three items relating to the same content at different levels, and making a natural hierarchy based on their difficulty, form a block. Each block works as a detector that characterizes the level (quality) of mastering the relevant part of the learning program. A set of such blocks forms a test notebook.

The blocks are being developed according to the task typology made on the basis of the level model. Briefly, a typical task, which can be solved using a standard algorithm, corresponds to the first level of competence. Where the standard procedure cannot be performed, and it is necessary to show substantial understanding of the situation in order to work out an appropriate scheme of acting, the task will correspond to the second level. But if such a task requires figuring out several different solution and choosing the one that meets certain contextual requirements, then it will correspond to the third level. The task typology is still awaiting final specification.

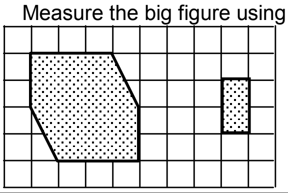
Examples of item blocks for assessing math competence are given below (Fig. 1 and 2).

Measure the big shaded figure in square centimeters. Write down your answer.



The answer:__________

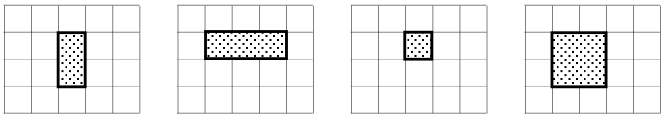
Measure the big figure using the unit. Write down your answer.



The answer: _______________

Children were measuring the area of the same figure using different unit areas.

Vasya Nick Sasha Tanya



The results of three measurements are presented in the right column of the table. Identify which result belongs to which child. Write down the children names in the left column.

Children names	Measurement results
	12
	24
	48

Figure 1 – *Block of test items (value measurement)*

Thus, all three items are related to area measurement. The first item requires applying a rule - it is a direct application of a standard measure to the object being measured (correct answers – 63%). The second item requires the children to realize the inadequacy of the standard way of acting, to identify the essential relationship (between a measure and a number) and to fulfill the task, e.g. by modifying the object to make the situation typical (correct answers – 32%). The third item requires fluent ability to operate with relations between a value, a measure and a number in order to establish several hypothesis and test them (correct answers – 15%).

What is the result of dividing 10472 by 34? Answer:___________
Petya copied a multiplication task from a textbook. He copied the first of the two multipliers correctly: 7, but he accidentally transposed the figures in the other multiplier. Because of this mistake, his result of multiplication was 147. What result would Petya get if he copied the task correctly? Answer: ___________
What is the highest result one can get by replacing the letters with digits in the following expression: AB5 + BC2 (different letters must have different numerical values)? Answer: ___________

Figure 2 – *Block of test items (number and computing)*

The first task implies direct application of a calculating rule, i.e. algorithm (correct answers – 70%). The second one requires analyzing the erroneous arithmetic operation (in the context of the positional principle) and finding a way to correct it (34%). And finally, the third task involves using the positional principle of testing several versions and finding the one that satisfies the maximum condition (15%).

On the basis of test results (mathematic processing of data for each level) a student profile can be composed, as well as a class profile and a profile of the student cohort for each tested subject (Fig. 3).

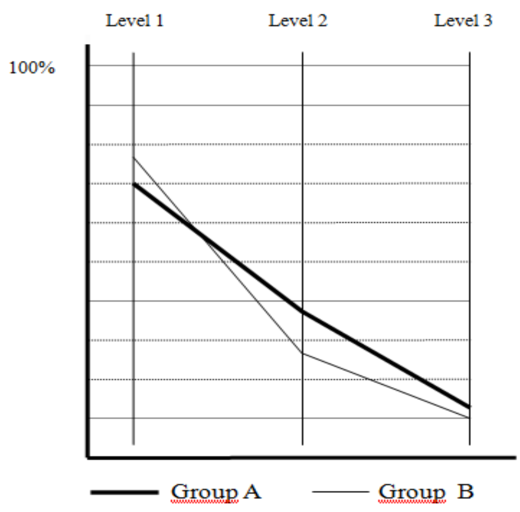


Figure 3 – *Math profiles of two groups of students*

Commento [MP1]: Inviare una nuova immagine senza sottolineature o un file di Excel

The test results can as well be referred to an integral scale, making it possible to range the success of test fulfillment by students, classes, etc.

Such presentation of test data which gives structural vision of the competence being formed, opens a way to deeper interpretation of learning outputs. Thus, the diagram (Fig. 3) presenting profiles of two groups of primary school students, which have practically the same scores on the integral scale, allows us to make the statement that Group A demonstrates better understanding of learnt material than Group B (indicators of Group A exceed those of Group B at level 2). Using the primary data one can also identify those parts of the content which each group failed to master at the second level. As for the scale of the third level, it demonstrates the line of progress, which at primary school, according to expectations related to age, should only begin to show.

Design

Participants

Data for this study were collected during SAM pilot testing in the Krasnoyarsk region of the Russian Federation. The sampling procedure includes two variables: type of school and school location. All examinees were 11-year-old students of the last (fourth) grade of primary school. The total number of participants for this pilot test was 418.

Instrument

Currently, SAM includes tests to assess math, natural sciences and Russian language competencies. In this article we will only review the materials of the math test developed by the group of specialists headed by of S. Gorbov.

Five content areas were included in the test on mathematical competence: numbers and calculations; value measurement; mathematical regularities; dependence between values; geometry elements. Test items for each content area were developed in accordance with the three levels of mastery described above. The test contains 15 three-level units (blocks of items) and the total number of items is equal to 45.

The test is assumed to be multidimensional. Items of each competence level form a subscale, so there are three subscales with 15 items for each one. Each item belongs to only one subscale (dimension). All items were scored dichotomously.

Measurement model

All subscales of the SAM tests measure related (but supposedly different) latent characteristics of the examinees. There are three approaches in the item response modeling to such kind of tests. Firstly, we can ignore multidimensionality in the test and apply a unidimensional model. Secondly, we can recognize multidimensionality and apply a unidimensional model to each dimension consecutively. And thirdly, we can apply multidimensional models.

At the stage of SAM validation all three approaches were applied. Members of the Rasch family of item response models were employed (The Unidimensional Rasch model and its extension – The Multidimensional Random coefficients multinomial logit model (MRCMLM)). Unidimensional and multidimensional analyses were conducted with ConQuest (http://assess.com).

Additionally classical test analysis was conducted at the first stage for items and test analysis.

Analysis

Firstly we will give a summary of the basic results of item analysis according to Classical Test Theory. Means of difficulty levels of test items for the 1st and 2nd levels equal 59% (rank from 43 to 75) and 20% (rank from 4 to 39). The 3rd level items turned out to be much more difficult in comparison with other levels' items for the given sample of examinees: the mean of difficulty levels is 6% (rank from 1 to 15). Thus, on the whole, the 1-st level items are easier than the 2nd level items, and these in their turn are easier than the 3-d level items.

A coefficient of point-biserial correlation was used to assess discrimination. The average values for the items of the 1st, 2nd and 3rd levels are 0.38, 0.33 and 0.18 respectively. So items of the 3rd level have smaller discrimination in comparison with items of the 1st and 2nd levels. This can be explained partly by the high difficulty of the 3-d level items for the given sample of examinees.

On the whole, all items have satisfactory quality characteristics. Furthermore, reliability of the analyzable test is quite high – 0.85 (the Coefficient Alpha was used as a reliability coefficient).

From the theoretical model from which the toolkit is being developed, it is evident that there must be a hierarchy of difficulty inside each block of three items. All items blocks of the analyzable test satisfy this requirement: success of item solving decreases from the 1^{st} level to the 3^{rd} one in each bloc. It serves as an additional argument in favor of validity of the toolkit.

Analysis of the test under the Modern test theory using different approaches described above allowed us to make the following conclusions:

The test can be considered as an essential unidimensional one, i.e. aimed to measure one latent variable which is the level of mathematical competence of the students. Therefore the test results can provide an integral estimation of the test takers' mathematical competence. The reliability of this assessment is 0.85, which is quite high.

The consecutive approach is unacceptable for the SAM math data: there is a substantial reduction in reliability for separate subscales in comparison with other approaches and the standard errors of students' measurement by each subscale are extremely high. This means that separate subscales cannot be considered as independent measurements. This result is expected due to the small number of items in each subscale and in light of the fact that all items were scored dichotomously.

The results of scaling the MASS mathematical data using unidimensional and multidimensional models are shown in Table 1.

Table 1 – *Summary of unidimensional and multidimensional model scaling*

			Model Summary	
Number of parameters		*Deviance*	*Reliability*	*Standard error mean of students' estimation (logits)*
Unidimensional	46	14,919.96	.85	.45
	Multidimensional	51	14,891.1	
Dimension 1			.83	.48
Dimension 2			.80	.53
Dimension 3			.61	.54

An analysis of the table shows that the reliability of students' estimation is quite high under both the unidimensional and the multidimensional models. What's more, under the multidimensional approach the reliability for each dimension comes closer to the unidimensional reliability estimate.

It is known that goodness of fit of the model can be evaluated using the deviance index. For two models, one a special case of the other, the difference in deviances has approximately a chi-squared distribution with degrees of freedom equal to the difference between the numbers of parameters in the two models. As Table 1 indicates, the difference in deviance between the two models is 28.86. This difference is approximately distributed as a chi-square with 5 degrees of freedom. This suggests that the multidimensional model fits the data much better than the unidimensional model. Additionally, the comparison of these two models was implemented by means of Akaike's Information Criterion (AIC), which is a transformation of the Deviance index. For the unidimensional model AIC index is equal to 15011.96 and for the multidimensional model it is 14993.1. Thus the multidimensional model provides the best explanation of the data. This provides statistical support for the use of the three-dimensional model where different dimensions are based on competence levels.

A multidimensional (3-level) approach, even with a small number of tasks for each level, allows us to make quite a reliable assessment of test takers on Levels 1 and 2 (measurement reliability is 0.83 and 0.8 respectively). Measurement reliability on Level 3 is a little lower (0.65), which can be explained by the fact that the majority of Level 3 tasks have turned out to be very difficult for the given group of tested students. In order to get more stable and accurate characteristics of the Level 3 tasks it is intended to try them on a different sample group of test takers (i.e. secondary school students).

Discussion

The results of the analyses reveal that the reliability of students' estimation is quite high under both the unidimensional and the multidimensional models. The comparison of the models reveals that the multidimensional model fits the data significantly better than the unidimensional model. This provides statistical support for the use of the three-dimensional model where different dimensions are based on the competence levels.

The statistical data show that it is appropriate to use the unidimensional model to obtain an integral index of the test takers' mathematical competence and to use the multidimensional 3-level model to assess their mathematical competence on each separate level.

The analysis of the test items allowed us to outline the ways to improve the instrument in terms of reliability and accuracy of measurement. Particularly, we have found out that a part of 2nd and 3rd Level tasks could be made a little simpler, while still meeting the level criteria. For instance, the task to find the maximum value of mathematical expression AB5 + BC2 in the initial version contained the formula ABC + BDA, where six rather than four letters needed to be replaced by numbers. According to our criteria, this task corresponds to the 3rd level but has greater complexity. That means that using this task we underestimate those children who meet the 3rd level by the type of orientation, but are not mature enough to manage the additional difficulty. Therefore, developing tasks in accordance with the lower boundary of each level (e.g. with four letters rather than six in the task discussed here) can increase the accuracy with which the blocks determine the qualitative structure of students' math competence.

Other important aspects of the SAM test investigation remain beyond this paper. They include detailed analysis of test items and subscales; tests forms equating; scaling students' estimates on different dimensions; construction of student's profile, etc. We hope to cover these issues in future publications.

Conclusion

This research has proven that the suggested approach is feasible. It is not absolutely new in the field of educational achievements' assessment. Its distinctive feature is in that it advances the methodology of qualitative analysis to the foreground, whereas in other approaches this methodology plays a more subordinate role.

Thus the framework of the tool is formed by the taxonomy of quality levels of the subject competencies, which is based explicitly on an integrated psychological model of interest to educators and psychologists. Furthermore, the model which is outlined only briefly in this paper can be easily expanded in a more detailed version, for which there is a considerable body of relevant theoretical and experimental research on age and functional development. Here we can refer not only to the works of Vygotsky and his followers, but also to Piaget, Wallon, Levin and many other famous authors who contributed to the understanding of cultural development. All this enhances prerequisites for the interpretation of test results and the extension of psycho-pedagogical hypotheses. For example, according to the periodization of the cultural development (Elkonin, 1986) in elementary school, mastering the subject material at the functional level should not be expected as an age norm. And this gives the basis for theoretical interpretation of the test results and for determining the teaching strategy to be used.

Approbation of this instrument, in order to determine the principles of interpretation and usage of the test results, is planned to take place in the context of pedagogical research as well as school practice.

Acknowledgment

The author would like to thank E. Kardanova for the help in preparing the psychometric part of this article, and B. Elkonin for essential advice with regard to the psychological issues. I am grateful to V .Bolotov, A. Volkov, E. Lenskaya and I. Frumin for multiple discussions related to the instrument and for promoting its development. I'd like to thank B.Soloviov the head of CICED for the careful organizational support of our work. And I am extremely thankful to my colleagues for their devoted work within this project.

References

Bloom, B., Englehart, M. Furst, E., Hill, W., Krathwohl, D. (1956), *Taxonomy of Educational Objectives: The Classification of Educational goals. Handbook I: Cognitive Domain*, New York-Toronto: Longmans, Green.

Davydov, V.V. (1988), *Learning Activity: The Main Problems needing Further Research*, multidisciplinary newsletter for activity theory, 1/2, pp. 29-36.

Davydov, V.V. (1996), *The Theory of Developing Learning*, Moscow: Intor, Russian ed.

Davydov, V.V. (1972), *Types of Generalization in Learning*, Moscow: Pedagogy, Russian ed.

Elkonin, B.D. (1994), *Introduction to the Psychology of Development*, Moscow: Trivola, Russian ed.

Elkonin, D.B. (1989), *Selected Psychological Works*, Moscow: Pedagogy, Russian ed.

Galperin, P.J. (1998), *Psychology as an Objective Science*, Moscow-Voronezh: Institute of Practical Psychology, Russian ed.

Lerner I.J. (1980), *The Learning Process and its logic*, M., Russian ed.

Majorov, A.N. (2002), *Theory and Practice of Developing Tests for a System of Education*, Moscow: Intellect-Center, Russian ed.

Mullis, I.V.S., Kennedy, A.M., Martin, M.O., Sainsbury, M. (2006), *PIRLS 2006. Assessment Framework and Specifications*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, 2nd ed.

Mullis, I.V.S., Martin, M.O., Smith, A. et al. (2003), *TIMSS Assessment Frameworks and Specifications 2003*, International Study Center, Lynch School of Education, Boston College, 2nd ed.

Nezhnov, P. (2007), "Mediating and spontaneity in the 'cultural development' model", *Moscow State University Bulletin*, Series 14: Psychology, 1, pp. 133-146, Russian ed.

Commento [MP2]: Luogo di edizione: editore

Nezhnov, P. (2009), *Toolkit for Assessment of Subject Competences of Primary School Students*, Innovation in Assessment to Meet Changing Needs, The 10th Annual AEA-Europe Conference, Malta, p. 22.

Nezhnov, P.G., Froumin, I.D., Khasan, B.I., Elkonin, B.D. (eds.) (2009), *Diagnostics of Learning Achievements in Primary School*, Moscow: Open Institute "Developing learning", Russian ed.

Savelieva, O.V. (1989), *Psychological criteria of the Quality of Knowledge of Primary School Students*, Author's abstract of dissertation, Moscow: The Institute of Psychology – The Academy of Pedagogical Sciences of the USSR, Russian ed.

Simonov, V.P. (1999), *Diagnostics of Students' Training Level*, Moscow: МПА, Russian ed.

The Definition and Selection of Competencies. Executive Summary (2005), retrieved from http://www.oecd.org/dataoecd/47/61/35070367.pdf.

The PISA 2003 Assessment Framework. Mathematics, Reading, Science and Problem Solving Knowledge and Skills (2003), Paris: OECD.

Vygotsky, L.S. (1982, 1983), *Collected Works*, Voll. 2-3, Moscow: Pedagogy, Russian ed.

Zaporozhets, A.V. (2000), *Psychology of Action*, Moscow-Voronezh: Psychological-Sociological Institute, Russian ed.

Russian editions

Выготский Л.С. Собр. соч, тт. 2,3. М.: Педагогика, 1982, 1983.

Гальперин П.Я. Психология как объективная наука. Москва-Воронеж: Институт практической психологии, 1998.

Давыдов В.В. Теория развивающего обучения. М.:ИНТОР, 1996.

Давыдов В.В. Виды обобщения в обучении. М.: Педагогика, 1972.

Запорожец А.В. Психология действия. Москва-Воронеж: НПО «МОДЭК», 2000.

Ковалева Г.С. Зарубежный опыт построения и актуальные проблемы развития образовательного тестирования. М., 2000. http://centeroko.ru/public.htm#other_pub.

Лернер И.Я. Процесс обучения и его закономерности. М.: Знание, 1980.

Майоров А.Н. Теория и практика создания тестов для системы образования. М.:Интеллект-центр, 2002.

Нежнов П.Г. Опосредствование и спонтанность в модели «культурного развития». Вестн.Моск.Ун-та. Серия 14. Психология , 2007, 1, с. 133-146.

Нежнов П.Г., Фрумин И.Д., Хасан Б.И., Эльконин Б.Д. (Ред.). Диагностика учебной успешности в начальной школе. М.: ОИРО, 2009.

Савельева О.В. Психологические критерии качества знаний младших школьников. Автореф… канд. дисс. М.: ИП РАО. 1989.

Симонов В.П. Диагностика степени обученности учащихся: Учебно-справочное пособие. М.: МПА, 1999.
Эльконин Б.Д. Введение в психологию развития. М.: Тривола, 1994.
Эльконин Д.Б. Избранные психологические труды. М.: Педагогика, 1989.