New approaches in designing educational assessment instruments and its use in international studies: What's new in PISA 2018?

Dragica Pavlović Babić [1*]
[1] University of Belgrade, Faculty of Philosophy, Belgrade, Serbia
[*] dpavlovi@f.bg.ac.rs

Abstract: *The intention of this paper is to present the content and structure of the international assessment of student achievements OECD/Programme for International Student Assessment) and, in particular, innovative solutions in the testing techniques, competencies chosen to be assessed and contextual variables. All testing material is in the computer-based format which opens up new possibilities in assessing cognitive processes that are needed to solve tasks. Besides regularly assessed reading, mathematics and scientific literacy, new testing domains are financial literacy and global competence. A part of test material (in the domain of reading literacy is based on multistage adaptive testing design, which increase the informative value of achievement data. Among number of contextual factors, students are asked to perceive their well-being* exploring. Looking for students' feelings related to usual, everyday situations and activities, it is possible to conclude what they need to live happy and fulfilling life. Finally, the use of log file in order to define students' problem-solving strategies is discussed.*

Keywords: *assessment; education; adaptive testing; literacy; student achievement*

1. INTRODUCTION

OECD/PISA is currently the most comprehensive and the most referential international survey in the field of education. Its implementation has been organised by the OECD since 1997, in three-year cycles, with growing number of participating countries. In the 2018 study, over 600,000 students from over 80 countries, which make 87% of the world economy, have took part, among them over 8000 students from Serbia.

Based on PISA survey it is determined to what extent students, near the end of their general education period, have adopted knowledge and skills that are important to them in order to be successful in their professional and personal lives. Students are not only expected to reproduce acquired knowledge, but also to apply it in different, relevant situations outside the school. It is usual to use in PISA surveys the term literacy or competence instead of knowledge: be literate is about the knowledge considered to be the basic education capital that students need in order to continue with their education and to be successful

in their personal and professional roles they will find themselves in as adults. Be competent in this study does not mean only that someone has acquired appropriate knowledge, but also that he/she knows when and how to apply it. In other words, the emphasis is on functional knowledge, and all the tasks in the tests are related to real situations students may encounter (Pavlovic Babic & Baucal, 2013).

The basic research interest is not founded on the assessment of the extent to which knowledge foreseen by the curriculum has been adopted; it is rather based on whether students are able to use that knowledge and how. Adoption of literacy is a life-long process that does not take place only in school and exclusively through formal education, but also through interaction with parents and other adults, interaction with peers, media, direct and broader environment. One cannot expect fifteen-year-olds to have learned everything they would need as adults, but they should have a solid knowledge from key areas that will enable them further development and adoption of competencies. They also need to understand the fundamental processes and principles and to use them flexibly in various situations (OECD, 2013). PISA, therefore, measures how able the students are to respond to requests relevant for daily life and

to understand key concepts, and does not assess adoption of knowledge specific for a certain area.

Originally, tests were developed in paper-pencil format, but, lately, the format is changed into computer-based. Old items are transformed, and new ones developed to fit this format. Since the 2018 cycle, Serbia is participating in computer-based assessment, as well as the largest number of other participating countries.

But, apart from changing the test format, the 2018 cycle also brought some more innovations regarding test content, structure or format. The purpose of this paper is to present these innovations and new opportunities for researchers and practitioners.

2. STRUCTURE OF THE TEST: COGNITIVE DOMAINS

Since from the first study, conducted in 2000., achievements in three cognitive have been assessed: reading, mathematics and science literacy. In full range, test consists of 442 items in these three domains. As in every cycle, one of the testing domains is in the focus. This means that that one is examined in more detail, assessing student achievement in a number of subdomains. In the cycle 2018, this position belongs to the reading literacy, covered by 245 items, organized in 4 subdomains.

The structure and, consequently the number of test items, are defined by several dimensions, such as competencies (e.g. in Scientific literacy: Evaluate & design scientific enquiry, Explain phenomena scientifically, and Interpret data & evidence scientifically), situations (personal, public, occupational, educational), levels of achievement (usually 6 levels per domain are defined), text format (linear, non-linear, interactive, multiple).

In order to cover all these dimensions, a lot of items are needed. The PISA study uses a balanced incomplete block (BIB) design testing (Johnson, 1992, NAEP, 2001). BIB design has been developed for the needs of a large-scale testing in which the aim is to test a broader range of competencies or content, and to limit the time of examination subjects in a short time (2 to 3 hours). To obtain a reliable measure of the level of individual competence, a relatively large number of items are used. To reduce each respondent's test timing into the realistic framework, a set of items is divided into a number of blocks, and blocks are connected to a specific scheme in the brochure. Thus each booklet contains only a part of the items, which are chosen in a way that each brochure's content overlaps with several other brochures. The use of BIB designs has enabled the use of IRT techniques to analyze the data (Birnbaum, 1968; Lord, 1980, Bond & Fox, 2007).

Assessing literacy in reading, mathematics and science is comprehensive part of the PISA study. But besides these domains, the participating countries may choose to participate in the assessment of some other domains, also estimated as relevant for the educational status of young people. In this cycle, participating countries could choose to assess achievements in the field of financial literacy and global competence. Serbia chose to participate in both of these assessments.

2.1. Financial Literacy

In recent years, the awareness about the importance of financial literacy for the professional and everyday life of citizens has been increased, as well as the responsibility of the education system to cultivate this knowledge on a systematic way. Some prominent global trends underpin the rising interest in financial literacy as a key life skill, such as: transfer of risk from governments and employers to individuals, increased individual responsibility for financial decisions, growing numbers of consumers and increased number of available financial products and services.

PISA conceives of literacy as the capacity of students to apply knowledge and skills in key subject areas, and to analyze, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations. PISA is forward-looking, focusing on young people's ability to use their knowledge and skills to meet real-life challenges, rather than merely on the extent to which they have mastered specific curricular content (OECD, 2010).

Information about background, non-cognitive factors of achievement in the field of financial literacy is collected by a short student questionnaire. Items address aspects of the following areas identified as key areas: access to information and education; access to money and financial products; and spending and saving behavior. The questionnaire consists of a small number of questions that explore the range and types of students' interest in and their experience with financial matters.

Besides Serbia, 20 more countries are participating in students' assessment of financial literacy, such as Australia, Bulgaria, Brazil, Canada, Chile, Spain, Estonia, Finland, Georgia, Indonesia, Italy, Lithuania, Latvia, Netherland, Peru, Poland, Portugal, Russian Federation, Slovak Republic, and United States.

2.2. Global Competence

Can you deal with uncertainty? Can you deal with ambiguity? Are you ready to engage in different value systems? Are you competent to make sound judgements?

This type of competence is called Global competence and, first time from the begging of the PISA assessment, will be evaluated in this cycle.
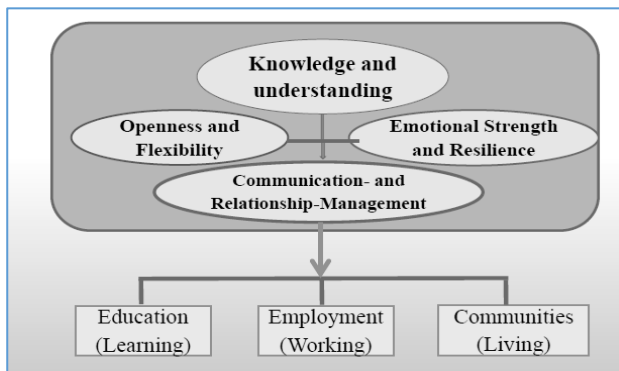

Figure 1. *Global competence: Dimensions and contexts*

Global Competence is defined as the capability and disposition to act and interact appropriately and effectively, both individually and collaboratively, when participating in an interconnected, interdependent and diverse world. This is a draft definition, and the construct will be more precisely defined and described in upcoming research report.

The draft definition shows that the construct is highly saturated with personal attitudes and values. Four dimensions are relevant for understanding the General competence, those are: Knowledge and understanding, Emotional strength and resilience, Communication and relationship-management, and Openness and Flexibility. Items are situated in three contexts: Education (Learning), Employment (Working), and Communities (Living).

Besides Serbia, 31 more countries are participating in students' assessment of global competence, which clearly show the interest for this competence as an educational outcome.

Despite the expressed interest, it is questionable to what extent the data collected will be suitable for the interpretation and formulation of educational policies. Without the pilot study we do not know is the construct well conceptualized (the question of construct validity), is it well differentiated in relation to related concepts and is it well operationalized for measurement needs.

3. BACKGROUND VARIABLES

Along with knowledge tests, questionnaires for students and schools are also used. They are used to collect information on different factors that may be relevant for performance, e.g. material and educational resources families have; students' view on learning, motivation for learning, strategies and habits in relation to learning; students' ability to apply modern information technologies, school's contribution to computer literacy; different aspects of school operation such as: characteristics of

teaching staff (level of education, professional motivation, styles of work), size of a class, composition (homogeneity or heterogeneity), atmosphere in the classroom and the school, attitude of the teaching staff towards students, sense of belonging to the school, school anxiety; material resources of the school, mode of financing (public or private), management and decision-making process, involvement of parents in processes and decision-making in the school, etc.

Based on the information from the questionnaires, it is possible to link student performance to students' and schools' characteristics and identify the most relevant factors of achievement.

Besides the student questionnaire, five more questionnaires are administrated: 1. Use of informative technologies (in school and at home), 2. Plans for educational career, 3. Questions about experiences with financial means and institutions, 4. Questionnaire about multicultural and global topics' experiences (which follows assessment of Global competence), and 5. Students' well-being questionnaire. Two last ones are newly developed and first time administrated. These questionnaires will be described in more detail.

Questionnaire for parents is also developed, but Serbia does not participate in this option.

Schools (principals) are asked to fulfill the school questionnaire which covers topics such as: Resources, human and materials, Decision making, including financial decision, Participation of students and parents, Assessment, including reporting to parents.

3.1. Global competence questionnaire

Some global developments, challenges and trends are recognized as important and students are asked to demonstrate how they understand their causes, implications, and / or possibility for change. They are asked about their personal experience with topics like: the student's engagement with others about global events/issues (e.g. chat online), extent of exposure to/awareness of global developments/challenges/trends (e.g. via news, other media). The second topic is migration/movement of student, for example: the student's own background; the student's experience living abroad; the possibility of the student working/studying abroad in the future; demographics in his/her community from other countries/cultures. The third topic is student interaction with or exposure to people from other countries/cultures; student's travel experience; student's degree of curiosity/ motivation to travel.

3.2. Students' Well-being questionnaire

Students' well-being refers to psychological, cognitive, social and physical functioning that students need to live happy and fulfilling life. Numerous recent research shows that the sense of

well-being correlates significantly with his/her achievement in the academic sphere, but also with other indicators of success and good adaptation, e.g. in an emotional or social sphere. So, the evaluation of the students' well-being must be sensitive to both their actual of present states and achievements.

The questionnaire consists of a small number of questions exploring students feeling related to usual, everyday situations and activities, for example, feelings during math classes, on the break, while doing homework or physical activities.

4. ADAPTIVE TESTING

Computerized adaptive tests (CATs) are a sophisticated method of test delivery based on item response theory (IRT). They operate by adapting both the difficulty and quantity of items seen by each examinee (Way, Twing, Camara, Sweeney & Mazzeo, 2010).

Considered to be on the leading edge of assessment technology, computer-adaptive tests represent an attempt to measure the abilities of individual students more precisely and to get more information about student' ability. Computer-adaptive tests try to match the knowledge and ability of a test taker. Basically, there's a simple strategy in test design: if a student gives a wrong answer, the computer follows up with an easier question; if the student answers correctly, the next question will be more difficult, as it is showed in Figure 2.
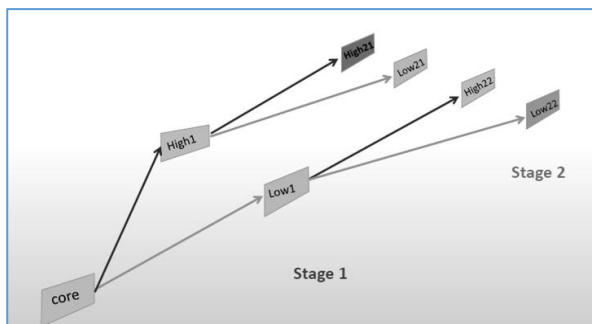

Figure 2. *Paths in multistage adaptive testing, basic model*

There are a lot of advantages if the adaptive testing is used. From the student perspective, this is more efficient way since it is time-saving. It will take less time since only those questions considered appropriate for the student are offered. The most current forms of computer-adaptive testing are typically administered online, and because the scoring is computerized, teachers and students can get test results more quickly than with paper-pencil tests. That way, the pool of items needed for test is much smaller, which means less time for test development and/or smaller number of item writers.

From the perspective of data use, a wide range of benefits is provided using adaptive testing: shorter tests and reduction of cost and examinee fatigue; more precise scores; more control of score precision (accuracy) which makes the test much more fair while traditional tests measure the middle students well but not the top or bottom students; greater test security because everyone is not seeing the same form; a better experience for examinees, as they only see items relevant for them, providing an appropriate challenge which can increase students' motivation; immediate score reporting; more frequent retesting is possible; individual pacing of tests; examinees move at their own speed; storing results in a database immediately makes data management easier.

5. USE OF LOG FILES

In PISA 2018 cycle is possible to analyze log files for computer tested reading items.

Basically, log files could have only limited number of information: 1. Variable „time" corresponds to the time when the event was written to the log file. This is measured in seconds, and 2. Variables „event", „event_value" and „event_detail", together, can be used to describe the activity of students or the system during the test. Those activities are: start item, end item, click, double click.

Although it is possible to extract only a limited number of information, their significance for understanding the work strategies and their effects on the achievements, is potentially high. Some researcher confirmed that a general indicator of the motivational level could be predicted from very basic data commonly recorded in log files (Cocea, & Weibelzahl, 2007). Within PISA assessment some analysis of relationship between time spent on the task and success in task solving has been done (OECD, 2015).

REFERENCES

[1] Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In: F. M. Lord & M. R. Novick (Eds.) *Statistical theories of mental test scores. Reading*. MA: Addison-Wesley.

[2] Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences*. New Jersey: LEA.

[3] Cocea, M. & Weibelzahl, S. (2007). Cross-system validation of engagement prediction from log files. *Second European Conference on Technology Enhanced Learning (EC-TEL 2007).* Crete, Greece.

[4] Johnson, E. G. (1992).The design of the national assessment of educational progress. Journal of Educational Measurement, 29, 95-110.

[5] Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

[6] NAEP. (2001). *The NAEP 1998: Technical reports.* Washington: UN department of Education.

[7] OECD. (2010). *PISA 2009 Framework: Key Competencies in Reading, Mathematics and Science.* Paris: OECD Publishing.

[8] OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy.* Paris: OECD Publishing. *http://dx.doi.org/10.1787/9789264190511-en*

[9] OECD. (2015). *Students, Computers and Learning: Making the Connection*, Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264239555-en

[10] Pavlović Babić, D. & Baucal, A. (2013). *Podrži me, inspiriši me: PISA 2012 u Srbiji.* Beograd: Institut za psihologiju i Centar za primenjenu psihologiju.

[11] Way, W., Twing, J., Camara, W., Sweeney, K. & Mazzeo, J. (2010). *Some consideration related to the use of adaptive testing for the common core assessment.* London: ETS, Pearson, College Board.