Using Process Data from Completing a Task in Creative Thinking Assessment

Sergei V. Tarasov
HSE University, Moscow, Russia
ORCID: https://orcid.org/0000-0003-4151-115X, e-mail: svtarasov@hse.ru

Creative thinking is an important skill in the modern world, and assessing its efficiency using modern digital tools is an increasingly complex methodological task. Using the information about the process of solving the tasks into the assessment model using the digital testing mechanisms became a promising trend. The use of such data allows us to consider the processes of creative thinking in dynamics, which makes the assessment of the level of creativity more accurate and diverse. The paper presents an analysis of the work of 823 students in the 4th grade who were asked to create images in a closed simulation environment. In this way, we analyzed their creative and critical thinking. Then the sequences of actions of students at different levels of creative thinking were compared, and various strategies they used to complete creative thinking tasks were compared with strategies used for critical thinking tasks. The information on the process helps to understand how these tasks work and to perform validation study. It also enhance the scoring rules and the feedback that can be received after the test.

Keywords: creative thinking; process data; computer testing; N-grams.

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

Использование данных о процессе выполнения задания при оценке креативного мышления

Тарасов С.В.
ФГАОУ ВО «Национальный исследовательский университет
«Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»),
г. Москва, Российская Федерация
ORCID: https://orcid.org/0000-0003-4151-115X, e-mail: svtarasov@hse.ru

Креативное мышление является важным навыком современного мира, а его оценка с помощью современных цифровых инструментов становится все более сложной методологической задачей. Включение в модель оценки креативного мышления данных о процессе выполнения заданий является перспективным направлением, которое становится возможным в компьютерном тестировании. Применение таких данных позволяет учитывать процессы креативного мышления в динамике, что делает оценку уровня креативности учеников более точной и многогранной. В исследовании представлен анализ работ 823 учеников 4 класса, которые в ходе выполнения задания создавали изображения в закрытой симуляционной среде для оценки креативного и критического мышления. В эмпирической части с помощью критерия хи-квадрат были сравнены последовательности от одного до трех действий учеников с разным уровнем сформированности креативного мышления, а также при выполнении задания на креативное и критическое мышление. В результате было показано, что данные о процессе выполнения задания могут быть использованы при проверке качества заданий и валидизации инструментов измерения, а также расширяют систему подсчета баллов и обратную связь по результатам тестирования.

Ключевые слова: креативное мышление; процессные данные; компьютерное тестирование; N-grams.

Introduction

The education process identified some important abilities called either "key competences" or "21st century skills" [29], the mastery of which is necessary for successful realization in life. Creativity, or creative thinking, is one of these skills. While computers and various artificial intelligent systems are now replacing workers for many standard tasks [3], the ability to think creatively and solve problems, complex communications, and social skills are becoming increasingly significant skills in the labor market. This requirement forces teachers to consider how to develop these skills and psychometricians to consider how best to assess these skills.

Tarasov S.V.
Using Process Data from Completing a Task in Creative Thinking Assessment
Psychological Science and Education. 2023. Vol. 28, no. 4

In the field of creativity assessment in the educational sphere, there is a widespread approach representing four main categories of creativity definitions and research areas known as the 4Ps: product, process, person, and press of the environment [20]. In the following, we will review the instruments for measuring cognitive processes related to creativity and creative products or outcomes.

The process approach to measuring creativity focuses on specific cognitive processes that contribute to creativity. Traditional tests developed under this approach involve open-ended or poorly structured tasks that require the production of as many responses as possible, which are then assessed to determine various creativity factors, usually including fluency (number of responses), originality (statistical rarity), flexibility (number of different categories), and elaboration (number of details). The main idea of assessment is not only to consider the quantity of answers, but also their quality.

A major contribution to the assessment of creative abilities was made by Gilford [10], who developed the Structure of the Intelligent (SOI) model and divergent thinking tests, which is considered an important element of creativity. It is also worth noting that the most used creativity tests are the Torrance Tests of Creative Thinking (TTCT) [27], which have been translated into more than 40 languages. Torrance designed 12 tests for different ages, grouped into a verbal, visual, and audio battery. However, Torrens tests are labor-intensive in terms of training experts and administering testing. In addition, the tests do not measure all aspects of creativity, but they do not claim to do so either. In the field of creativity assessment, there is a position that no cognitive test is a predictor of creativity if it does not include affective and motivational factors [26].

The newer creativity tests for schoolchildren are most valuable as indicators of potential creativity assessment (The Evaluation of Potential Creativity, EPoC), which are developed as procedures that assess the "overlap area". The test consists of a series of subtests that were designed to measure both general and specific creative abilities in two areas — verbal/literary and graphic [4]. The EPoC subtests measure two key models of creative cognition — divergent and convergent thinking. Divergent thinking tasks involve creating as many drawings as possible using a simple abstract shape or familiar object. Verbal thinking tasks involve creating several simple story endings in response to a unique story beginning or, conversely, several plot twists in response to a unique denouement. In convergent thinking tasks in the graphic domain, test takers create a completed original drawing using at least four of the eight abstract shapes or familiar objects that serve as the basis for their composition. In the verbal/literary domain of the convergent thinking task, test takers are required to create a completed story based on either a given title or given fictional characters.

However, traditional approaches to creativity testing, such as paper-and-pencil testing, are already outdated and have several disadvantages. First, they require experts for assessment, which creates a high testing workload due to the complexity of administration and introduces the effect of subjectivity of experts into the final scores. In addition, traditional tests have psychometric problems, such as inconsistent evidence of reliability and validity of instruments in different samples, the influence of the test environment on test results, and outdated psychometric analyses [23].

In this regard, the assessment of complex skills requires using not only traditional testing formats, but also more modern formats. One such format is digital tasks containing interactive elements that often resemble games in their form of presentation [15]. Examples include a instrument for assessing convergent thinking — *BuzzWords* [13], divergent thinking — *Immune Defence*

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

[16], convergent and divergent thinking simultaneously — *Crea.blender* [19]. Often, instruments in digital environments not only have automatic scoring rules without expert involvement, but also collect data about the process of task performance.

Task process data refers to data collected from respondents interacting with the computerized assessment element. This is recorded in computer log files and is often represented as sequences of events (specific actions in the test environment) with time stamps [9]. Such data is collected and analyzed in the assessment of complex skills, such as collaborative problem solving [30], as well as in modern task formats that simulate video games [14].

For example, in the framework of the Programme for International Student Assessment (PISA) in 2012, students' problem-solving skills were assessed using a set of interactive test items [17]. The test is designed to measure individuals' ability to cognitively process information to understand and solve problem situations where the method of solution is not immediately obvious. By exploring the problem situation and interacting with the computer environment, students found pieces of information that would be useful for solving the problem at hand. In addition to the answers to the test questions, students' behavioral data is also recorded to have an insight into their problem-solving strategies. For PISA problem-solving tasks, the learner's explicit answers to each question are product data, and the series of clicks and inputs made during his or her interaction with the questions, as well as the timestamps associated with each action, are process data of the task. Different methods for analyzing process data in relation to one of the 2012 PISA tasks were shown in [18], whose findings suggest that the choice of data analysis methods for analyzing the process of task performance in a saturated digital environment depends on the purpose of the analysis and the structure of the data.

The aim of the study is to determine the possibilities of using data on the task performance process in the framework of evaluating creative thinking using an instrument in a digital environment. Based on the indicators in the instrument, we get information about the level of creativity skill development, so we can compare the processes of solving a task among students with different levels of creativity skill development. In this instrument, described in detail in the next section, not only creativity but also critical thinking is assessed using the same format tasks.

Thus, in the study, we want to answer the following research questions regarding the data on the task process:

1. What actions distinguish the process of performing a creative thinking task for students with a high level of creative thinking from performing it for students with a low level?

2. What actions distinguish the process of performing a creative thinking task from the process of performing a critical thinking task?

Methods

The instrument

This study analyzes one of the tasks of the instrument "4K" for 21st-century skills assessment among fourth-grade schoolchildren, developed by the Centre for Psychometrics and Measurements in Education (Laboratory for New Constructs and Test Design) of Institute of Education of HSE University. The instrument consists of several scenario-type tasks that evaluate four skills: creative and critical thinking, communication, and cooperation. Moreover, most tasks allow to measure several skills.

The instrument is presented to students in computerized form. The task screen is interactive: the test taker clicks on an area of his/her choice and then sees a

Tarasov S.V.
Using Process Data from Completing a Task in Creative Thinking Assessment
Psychological Science and Education. 2023. Vol. 28, no. 4

pre-prepared specific response from the testing system. This format allows for the demonstration of complex skills, as well as keeps test takers motivated and reduces test anxiety. At the same time, the simulation of a real environment allows for more accurate recording of observed behavior, i.e., evidence that the test taker possesses a specific skill.

This paper considers the task «Monster», which is aimed at assessing creative and critical thinking. The general context of the task is imaginary and consists in the fact that the test taker found himself at a holiday performance in the City of Monsters. The task consists of description screens and a constructor (Figures 1—2) with which the test taker creates images.

In the first part of this task, aimed at as-sessing creative thinking, each test taker creates three monsters for the performance poster, which should be surprising and un-usual and different from the locals. The interface of this part of the task is shown in Fig. 1. The constructor consists of a canvas in the center on which elements from the bottom panel can be placed. All elements are divided into categories (torso, arms, various, etc.) and can be used an unlim-ited number of times (except for the torso, which must necessarily be one and is fixed in the center of the canvas). On the right side of the screen was a local resident, which was considered a reference for the purpose of originality scoring. Also, on the screen above the reference and on the right side of the bottom panel are instructions for completing the task.
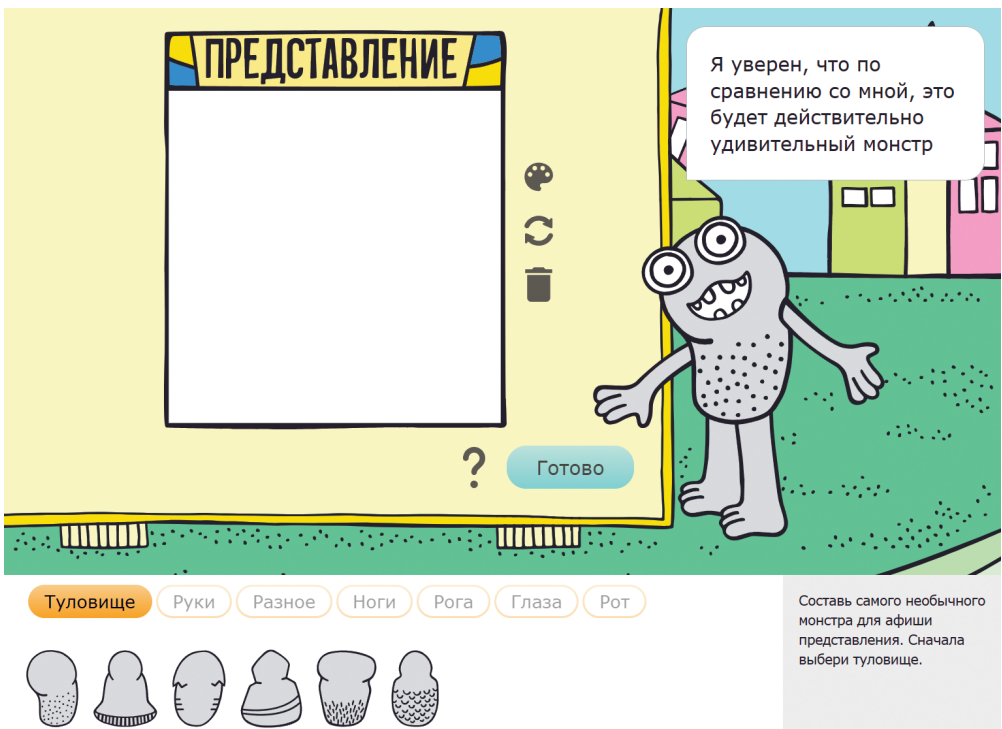

Fig. 1. Interface of the task «Monster» for evaluating creative thinking

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

In the second part of the task, it turns out that an emergency incident occurred during the performance and the test taker must draw up a sketch of the criminal based on a short description. The interface of this part of the task is not much different from the interface of the previous part and is shown in Fig. 2. The main difference lies in the instructions, according to which the test taker must create a sketch of a criminal monster, and not an unusual monster.

Construct operationalization

To create the measurement instrument, we defined the theoretical framework of the creativity construct [1], which is based on both the cognitive approach [10, 27] and the concept of structured imagination [28].

Creativity within this instrument includes two sub-constructs:

1. *Originality* — the ability to generate new ideas and problems solutions, which can be expressed as new ideas and construct unobvious relations between existing ideas.

To evaluate the originality of the image constructed by the test taker, we compare it with the reference image. A reference is a proxy image that reflects the image that is most frequently found in the sample. When creating a reference for a task, we identify the most typical elements and their number through cognitive interviews and quantitative research at the pilot stage. The focus of the task on creativity is determined by the fact that as part of the task, test takers are asked to create a new image that should differ from the original one.
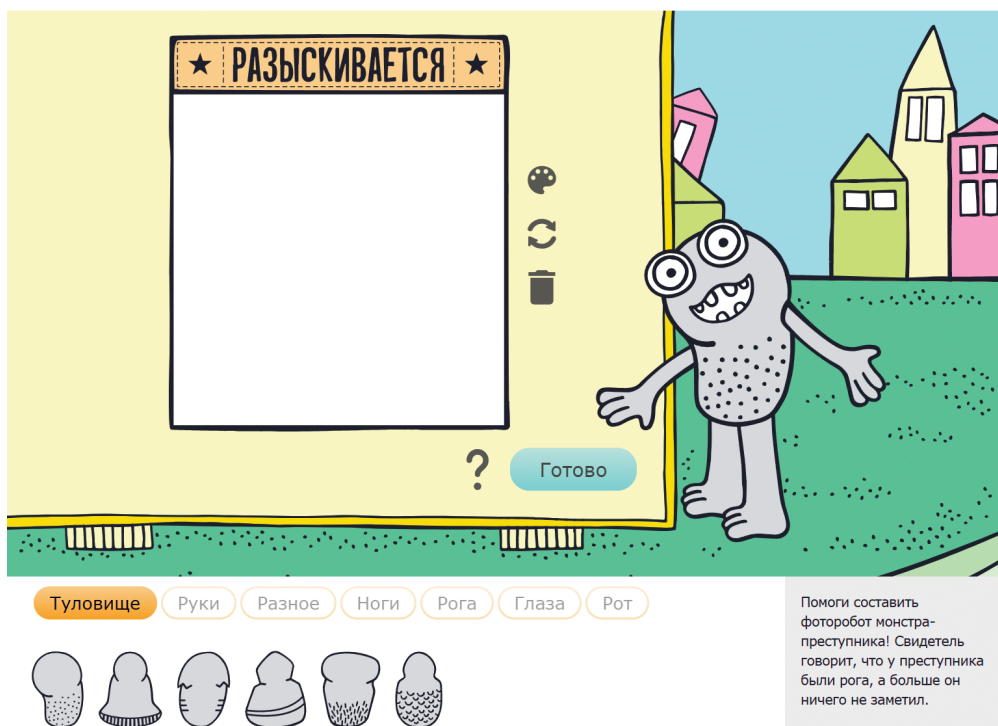

Fig. 2. Interface of the task «Monster» for evaluating critical thinking

Tarasov S.V.
Using Process Data from Completing a Task in Creative Thinking Assessment
Psychological Science and Education. 2023. Vol. 28, no. 4

2. *Elaboration* — the ability to work out the proposed idea in depth, with a high degree of detail.

Similar to the measurement in the cognitive approach, in our study, elaboration is measured as the number of elements that are used in the solution, and their functional diversity.

Sample and procedure

Testing was conducted in the spring of 2022 in several cities of Russia. From all the samples, students were selected who fully completed the task «Monster» (each image contains at least 2 elements). Thus, the base for the analysis consisted of 823 fourth grade students.

The average task completion time is 15 minutes. Testing took place in schools under the supervision of a teacher with the consent of parents in accordance with research ethics. The results of the students were anonymized for analysis purposes.

Indicators

In scenario tasks of the instrument "4K", skills are evaluated using indicators. Indicators are defined as the behavioral manifestations of a skill that were included in the scenario set by the developer, for example, test taker chosen or did not choose the correct behavior in the situation that arose. Behavioral display of creativity and critical thinking in the task «Monster», we consider the features of images created by test takers. These image features (indicators) have predefined scoring rules, shown in Table 1. The scoring rules were applied to each image separately. Below are the scoring rules based on the indicators of originality and elaboration.

In addition to the indicator, the image creating process was recorded for each

Table 1

Scoring rules for creativity assessment in the task "Monster"

Indicator	Scoring rules
Originality	
origin1	1 — number of elements from the category Legs is different from 2 0 — number of elements from the category Legs is equal to 2
origin2	1 — number of elements from the category Hands is different from 2 0 — number of elements from the category Hands is equal to 2
origin3	1 — number of elements from the category Mouth is different from 1 0 — number of elements from the category Mouth is equal to 1
origin4	1 — number of elements from the category Eyes is different from 2 0 — number of elements from the category Eyes is equal to 2
symmetLegs1	1 — lack of symmetry for elements from the category Legs 0 — there is symmetry for elements from the category Legs
symmetHands1	1 — lack of symmetry for elements from the category Hands 0 — there is symmetry for elements from the category Hands
position1	1 — at least one element from the category Legs is not in the leg's slots 0 — all elements from the category Legs are in the leg's slots NA — elements from the category Legs are not used at all
position2	1 — at least one element from the category Hands is not in the hand's slots 0 — all elements from the category Hands are in the hand's slots NA — elements from the category Hands are not used at all
position3	1 — at least one element from the category Mouth is in the upper part of the torso 0 — all elements from the category Mouth are at the top part of the torso NA — elements from the category Mouth are not used at all

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

position4	1 — at least one element from the category Eyes is in the upper part of the torso 0 — all elements from the category Eyes are at the top part of the torso NA — elements from the category Eyes are not used at all
Elaboration	
nElements	2 = 9 and more elements 1 = 5 to 8 elements 0 = 0 to 4 elements except elements from the category Various
nVarious	2 = 2 or more elements from the category Various 1 = 1 element from the category Various 0 = elements from the category Eyes are not used at all
color	1 — changed the color of the monster at least once 0 — did not change color
turn	1 — turned any element at least 1 time 0 — did not turn elements

test taker in the form of a comma-separated record of all actions, for example, "Start, Add_Torso, Add_Hands, Add_Hands, Add_Legs, Add_Legs, Add_Mouth, Color, Add_Eyes, Finish". All possible actions are described in Table 2.

Data analysis methodology

To assess the level of creativity, the methodology of confirmatory factor analysis (CFA) on ordinal variables was used [6]. We use the weighted least squares method using the polychoric correlation matrix (WLSMV) for reliable estimation due to the categorical order of the data [7]. The model's fit to data was determined based on the comparative fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). The first two indexes show the distance of the constructed model from the

Table 2

Actions in the Action Log

Action log entry	Action
Start	Beginning of the task
Finish	End of the task
Color	Changing the monster color
Add_X, where X can be: Torso Hands Legs Horns Eyes Mouth Various	Adding an element from the corresponding category
Tur_X, where X is same as the case with Add, except Torso, because the torso is always located in the center of the canvas and does not rotate	Rotate an element from the corresponding category
Del_X, where X is same as the case with Add	Delete an element from the corresponding category
Que	Using help by clicking on the question mark

Tarasov S.V.
Using Process Data from Completing a Task in Creative Thinking Assessment
Psychological Science and Education. 2023. Vol. 28, no. 4

null model, in which no variable is associated with another [5], and the last one is based on the analysis of model residuals [25]. We used the following critical values of the co-efficients, following the generally accepted cut-offs (Yu, 2002): CLI>0.95; TLI>0.95; RMSEA≤0.06. To improve the quality of the model, modification indices based on the analysis of model residues were used.

The unit of analysis in models is the indicator. The relationship between the indicator and the factor can be described based on the factor loading of the indicator in the CFA model. A statistically significant and positive factor loading indicates the presence of this relationship, and a higher value of the factor loading indicates a higher relationship of the indicator with the studied factor.

Factor scores from the CFA model calculated using the regression method are used to assess the level of ability formation [for more information, see 8]. Factor scores are a standardized continuous scale with a mean of 0 and a standard deviation equal to the root of the square of the multiple correlation between all indicators and the factor. Thus, the obtained factor scores can be used for further ranking of test takers.

To work with data from the task performance process, the recorded variables were divided into N-grams, small sequences of certain actions, where N is the number of actions in this sequence. The shortest sequences are unigrams, i.e. sequences of one action. Bigrams and trigrams are also used in the work — sequences of two and three actions, respectively. If the test taker performs 16 actions while completing the task, then this sequence contains 16 unigrams, 15 bigrams, and 14 trigrams. N-grams can be repeated both between different test takers and within the process of a task performed by one test taker. Based on the Table 2 we can assume that there will be 24 unique unigrams in the task, and the "Start" and "Finish" unigrams must occur in all tasks. Due to the different frequency of occurrence, certain sequences make a different contribution to distinguishing groups of test takers, so the accepted practice when working with N-grams is to weigh them [12] according to the formula

$$w(i,j) = \begin{cases} \left(1 + \log(n_{i,j})\right) * \log\left(\frac{N}{n_i}\right), & \text{если } n_{i,j} > 0 \\ 0, & \text{если } n_{i,j} = 0 \end{cases}$$

where i,j is a certain action i (N-gram) in a certain sequence j,

N is the total number of sequences,

$n_{i,j}$ — frequency of action i in the sequence j,

n_i — frequency of action i in all sequences.

To compare the task performance process, we calculate the weighted frequency of N-grams in different subgroups to determine how much one group differs from another in terms of action frequencies. According to the null hypothesis, two sets are randomly equivalent, so the distribution of their actions is proportional to each other. To assess the deviation from this null hypothesis, the chi-square criterion ($\chi2$) is calculated [for more information, see 2]. N-grams with higher values of $\chi2$ are the sequences of actions that distinguish the task performance process in the selected subgroups. At the significance level of 0.05, the critical value of $\chi2$ is 3.84. That is, if the observed value exceeds it, then we can be 95% sure that the action occurred more often in one of the two subgroups. The higher the value of $\chi2$, the greater the differences between the subgroups.

All calculations were performed using the programming language for statistical data processing R version 4.3.1 using the lavaan package [21] for evaluating CFA models and the ngram package [24] for calculating N-grams.

Results

Assessment the level of creativity
First, models of confirmatory factor analysis were constructed to determine the level of creativity according to the indicators included in the task. Two separate models

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

were constructed for the Originality and Elaboration. The fit indexes of both models are presented in Table 3 and indicate that both models fit well to the data.

Table 3

Indicators of compliance данным with the CFA model dataей КФА

Index	Originality	Elaboration
CFI	0.980	0.973
TLI	0.979	0.961
RMSEA	0.039	0.045

Figure 3 shows the factor structure of the model for the Originality. For all three images of monsters, correlations between indicators of symmetry and hand position, symmetry and the original number of hands, symmetry and the original number of legs were added to the model to better match the data.

All standardized factor loadings are significant (p < 0.05) and range from 0.22 to 0.67 with an average value of 0.45, which is an acceptable value.
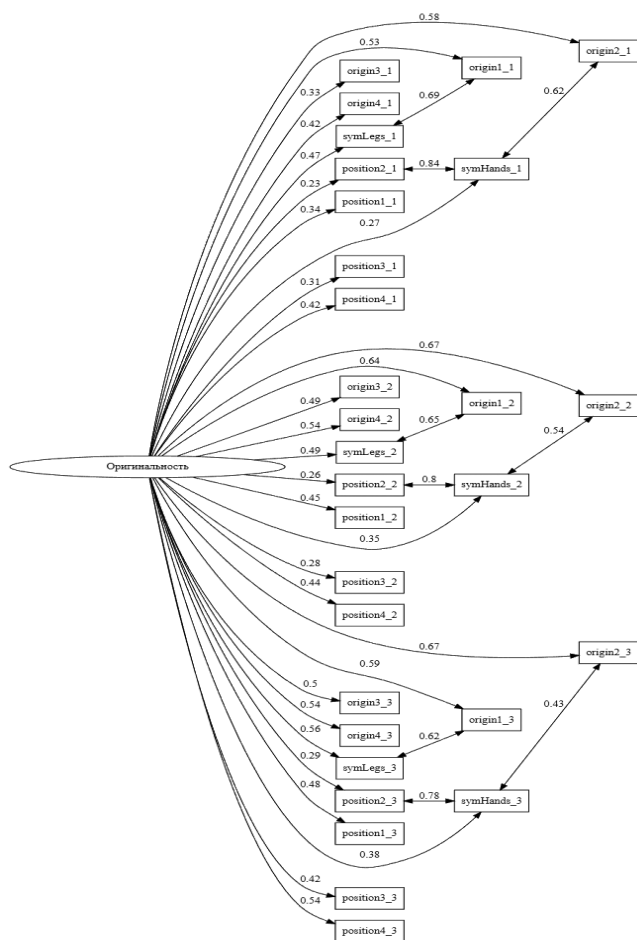

Fig. 3. Factor structure of the Originality

Tarasov S.V.
Using Process Data from Completing a Task in Creative Thinking Assessment
Psychological Science and Education. 2023. Vol. 28, no. 4

Figure 4 shows the factor structure of the model for the Elaboration. In this model, correlations were also added between related indicators, namely rotation, the number of elements except the category Various, and the number of elements from the category Various.

All standardized factor loadings are significant ($p<0.05$) and range from 0.16 to 0.83 with an average value of 0.41.

Comparison of the process of performing a creative thinking task by students with high and low levels of creative thinking

To answer the first research question, we looked at the first images of a monster that were created by test takers when solving a task aimed at assessment creative thinking. Unigrams, bigrams, and trigrams were constructed, since longer sequences do not have a high frequency in the data and are difficult to interpret. Table 4-5 shows the chi-square value when comparing two groups of test takers by the level of formation of two creativity sub-constructs (originality and elaboration), and 200 students with the highest factor score for the corresponding substructure were selected in the high — level group, and 200 students with the lowest factor score were selected in the low-level group.

Among unigrams, there are no significant differences between students with a high and low level of originality. Based on bigrams, we can conclude that high-level students are more likely to use elements from the category Various, which is logical, since this category contains unique elements that responsible for originality of
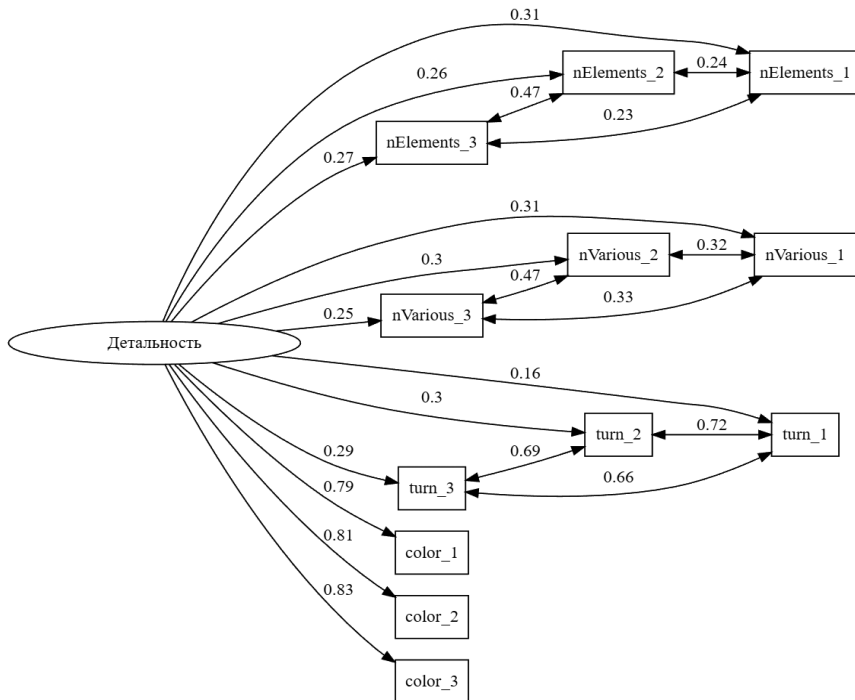

Fig. 4. Factor structure of the Elaboration

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

Table 4

Difference in the frequency of N-grams in the process of performing tasks for creative thinking in students with high and low levels of originality

High level		Low level	
Action	χ^2	Action	χ^2
unigrams ($\chi^2>5$)			
Tur_Various	5,3	Del_Mouth	8
		Tur_Legs	6,7
		Tur_Mouth	6,5
bigrams ($\chi^2>10$)			
Add_Various Add_Eyes	26,1	Tur_Hands Color	17
Add_Various Add_Various	17,1	Add_Eyes Add_Eyes	14,6
Add_Mouth Add_Mouth	15,1	Add_Mouth Tur_Mouth	12,9
Del_Legs Add_Eyes	15,1	Tur_Mouth Tur_Mouth	12,9
Add_Horns Add_Legs	14,8	Color Add_Horns	11,3
Add_Legs Finish	14,2	Tur_Horns Add_Various	10,8
Add_Various Del_Hands	10,3	Tur_Hands Add_Mouth	10,2
Color Del_Legs	10,3		
Del_Horns Add_Various	10,3		
Del_Horns Finish	10,3		
Del_Legs Color	10,3		
Tur_Horns Add_Legs	10,3		
trigrams ($\chi^2>20$)			
Add_Hands Add_Hands Add_Hands	33	Add_Mouth Add_Eyes Add_Eyes	24,5
Add_Eyes Add_Eyes Add_Eyes	27,1	Add_Eyes Add_Eyes Finish	20
Add_Hands Add_Legs Finish	27,1	Del_Torso Add_Torso Add_Legs	20
Add_Legs Add_Various Add_Hands	27,1		
Add_Mouth Add_Legs Finish	23,3		

the final image. Students with a high level of originality also have bigrams that include the action of deleting an element, which can indicate a creative process when the student tries different elements to create an original image. Among the trigrams, the greatest difference between students with different levels of originality is observed in the sequences of adding three elements from the category (hands and eyes). This confirms the existing scoring rules for originality, since for these categories the number of elements in the reference is two, and if the test taker adds three elements at once, this is associated with a high score for originality.

Students with a high level of elaboration are significantly more likely to change the color of the monster and rotate elements from different categories. This is consistent with the scoring rules for elaboration, as these actions are associated with higher scores for elaboration indicators. At the same time, students with a low level of elaboration remove items from different categories, which reduces the total number of monster items and indicates a low level of elaboration of the test taker. In addition, it is observed that unigram addressed to help button is more common in students with a low level of elaboration. This can be explained by the fact that such students

Tarasov S.V.
Using Process Data from Completing a Task in Creative Thinking Assessment
Psychological Science and Education. 2023. Vol. 28, no. 4

Table 5

Difference in the frequency of N-grams in the process of performing tasks for creative thinking in students with high and low levels of elaboration

High level		Low level	
Action	χ2	Action	χ2
unigrams (χ2>4)			
Color	64,4	Del_Torso	18,1
Tur_Horns	27	Que	18
Tur_Eyes	9,4	Add_Mouth	5,9
Tur_Various	8,6	Del_Eyes	4,8
Tur_Legs	7,2	Del_Hands	4,6
bigrams (χ2>35)			
Add_Torso Color	75,8	Add_Torso Add_Mouth	75,5
Color Color	59,6	Add_Torso Add_Eyes	73
Color Finish	47,1	Add_Torso Add_Hands	70,7
Color Add_Eyes	44,9	Add_Legs Finish	65
Color Add_Legs	43,4	Add_Hands Finish	62,2
Color Add_Hands	40,9	Add_Hands Add_Legs	58,4
Add_Horns Tur_Horns	39,5	Add_Eyes Finish	57
Color Add_Various	39,2	Add_Torso Add_Legs	47,9
Color Add_Mouth	35,6		
trigrams (χ2 > 25 for high level and χ2 > 45 for low level)			
Start Add_Torso Color	72,8	Start Add_Torso Add_Mouth	88,9
Add_Torso Color Color	42,6	Start Add_Torso Add_Hands	83,6
Add_Horns Tur_Horns Tur_Horns	37,2	Add_Hands Add_Hands Finish	74,3
Color Add_Legs Add_Legs	35,9	Add_Torso Add_Hands Add_Hands	72,4
Color Add_Hands Add_Hands	34	Start Add_Torso Add_Legs	66,7
Tur_Horns Tur_Horns Tur_Horns	32,5	Start Add_Torso Add_Eyes	61,9
Add_Horns Add_Horns Tur_Horns	30	Add_Torso Add_Legs Add_Legs	59
Add_Torso Color Add_Legs	30	Add_Hands Add_Hands Add_Legs	55,8
Color Add_Eyes Add_Eyes	30	Add_Mouth Add_Eyes Finish	51,2
Add_Torso Color Add_Eyes	26,7	Add_Torso Add_Mouth Add_Hands	51,2
Color Color Color	26,7	Add_Legs Add_Legs Finish	50,6
		Add_Torso Add_Eyes Add_Eyes	48,9

may not have understood how to work in the constructor and therefore could not create a monster and show their creativity.

Comparison of the process of completing tasks for creative and critical thinking

To answer the second research question, the first image of the monster was taken each, which were created by test takers while solving two parts of the task: to assess creative and critical thinking. As in the previous section, unigrams, bigrams, and trigrams were constructed, but among the trigrams, we selected those that occur at least 10 times in all solutions to remove very rare sequences, even if they allow us to distinguish solutions. Table 6 shows

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

the chi-square value when comparing two groups of monsters (N-grams with the highest chi-square value in each group were selected).

Among unigrams there are clear differences between the solutions. In the creative thinking task, test takers are significantly more likely to delete elements in all 7 categories. In the critical thinking task, test takers significantly more often add elements from the category Horns, which is explained by the task, because when compiling a sketch of a criminal, who according to eyewitnesses had horns, the addition of elements from this category shows that the test taker has correctly understood the task and has shown critical thinking. But at the same time, in the critical thinking task, test takers are more likely to add elements from other categories as well and turn hands and horns. Based on bigrams and trigrams, we can say that in the creative thinking task, students more often change the color of the monster, while in the critical thinking task, students leave its original color (gray). They also remove and rotate

Table 6

Difference in the frequency of N-grams in the process of performing tasks for creative and critical thinking

Creative thinking		Critical thinking	
м	χ2	Action	χ2
unigrams ($\chi 2>15$)			
Del_Legs	101,8	Add_Horns	73,8
Del_Hands	52,5	Tur_Hands	50,6
Del_Mouth	38,5	Tur_Horns	23,1
Del_Eyes	34,2	Add_Mouth	21,9
Del_Torso	30,3	Add_Eyes	20
Del_Horns	27,6	Add_Hands	18,3
Del_Various	24	Add_Legs	17,2
bigrams ($\chi 2>60$)			
Del_Hands Add_Hands	111,1	Add_Torso Add_Horns	238,4
Del_Legs Add_Legs	96,6	Add_Torso Add_Eyes	88,7
Add_Legs Del_Legs	88,2	Add_Eyes Finish	83,1
Color Color	79,4	Add_Horns Finish	72,1
Del_Legs Del_Legs	78,2	Add_Legs Finish	70,6
Add_Hands Del_Hands	64,7		
trigrams ($\chi 2>60$)			
Add_Torso Color Color	106,1	Start Add_Torso Add_Horns	187,8
Del_Legs Add_Legs Add_Legs	103,9	Add_Torso Add_Horns Add_Horns	109,9
Add_Hands Del_Hands Add_Hands	99,4	Add_Torso Add_Horns Add_Mouth	100
Add_Legs Add_Legs Del_Legs	85,5	Start Add_Torso Add_Eyes	79
Del_Hands Add_Hands Add_Hands	78,6	Add_Torso Add_Horns Add_Legs	74,6
Add_Legs Del_Legs Add_Legs	77	Add_Legs Add_Eyes Finish	68,4
Del_Hands Add_Hands Tur_Hands	66,2		
Add_Torso Color Add_Legs	63		
Tur_Legs Tur_Legs Tur_Legs	60,9		
Color Add_Legs Add_Legs	60,8		

Tarasov S.V.
Using Process Data from Completing a Task in Creative Thinking Assessment
Psychological Science and Education. 2023. Vol. 28, no. 4

elements from different categories, which indicates a creative process, while in the critical thinking task test takers try to simply assemble the monster from parts, so after weighing, actions related to adding elements prevail in the process of completing the critical thinking task.

Discussion of the results and conclusion

The purpose of this work was to determine the possibilities of using process data of the task performance in the framework of assessment creative thinking. Using instrument is implemented in a digital environment and consisting of two parts: for evaluating creative and critical thinking.

At the first stage of the analysis, CFA models were constructed to obtain test takers' scores for creative thinking. Further, these scores were used to identify groups of test takers with high and low levels of originality and elaboration.

At the next stage, the method of N-grams, small sequences of actions extracted from the test takers' action log, was chosen to analyze data about the task performance process. Sequences from one to three actions were used.

We used the $\chi 2$ criterion to compare the frequency of occurrence of different N-grams in students with high and low levels of creative thinking. As a result, a few N-grams were identified that are associated with different levels of creative thinking. For example, using an item from the category Various indicates a high level of originality, and deleting an item from the category Torso indicates a low level of elaboration.

In the second research question, we compared the performance of creative and critical thinking tasks implemented in the same constructor. Based on the frequency of occurrence of certain N-grams, we found out that different tasks work accordingly. So, in the creative thinking task, students are more likely to delete

and rotate elements, change the color of the image, and in the critical thinking task, students are more likely to add those elements that are necessary for the correct solution of the task.

As a conclusion, it can be noted that the chosen method of analyzing data on the task performance process turned out to be useful and can be used to understand the behavior of test takers with different levels of skill development (in this study, creative thinking), it can also be usedin order to check the quality of tasks.

Using data about the task performance process can also enrich the scoring system and feedback on test results. For example, when analyzing students with a high and low level of originality, we found a correlation between the presence of bigrams in the test taker's profile(deleting and then adding elements), and a high level of originality. This relationship is not considered in the current scoring system. However, special attention should be paid in order to better understand the intended use of such data. This is important because expanding towards the accumulation of more complex data can challenge traditional approaches to scaling assessment results in educational testing and can be handled inadequately by test users [22].

Based on the analysis of the task performance process, we found that other characteristics of the test takers, such as their level of digital literacy, can also influence the result of the creative thinking task. This can be supported by the fact that students with a low level of elaboration more likely refered to the task reference, which contained on working with the constructor, than students with a high level. It can be assumed that due to the low level of digital literacy, they did not understand how to perform the task of evaluating creative thinking in a digital environment. This observation describes one of the limitations of modern assessment

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

instruments, namely the possible interference of other factors.

The limitations of this study include a small sample for analysis, which is why some N-grams were very rare in the aggregate of solutions. For this reason, during the analysis stage, we combined actions by category (for example, adding or turning a hand), and did not consider specific elements within the category (adding hand #1 or turning hand #2). Analysis at the level of individual elements is part of our plans to refine the constructor test takers use to create images. There is a hypothesis that certain elements within categories may be associated with the level of creativity, for example, test takers with a high level of originality will tend to add hand #3 than other hands.

Further areas of research include analyzing data on the task performance time, namely, the time between certain actions in the task execution sequence, as well as using other approaches to analyze data on the task solving process in a digital environment. One of these approaches is the Longest Common Sequence method [11], a sequence analysis method used in natural language processing and biostatistics in order to understand the strategy of test takers when solving digital problems. This approach can be applied to the task «Monster» considered in the article, since among the solutions you can select non-original solutions (repeating the reference) and calculate how much the students' solutions will differ from such a solution. It is also possible to calculate the proximity of decisions made by different students, which will allow us to identify certain clusters of decisions that can be used to draw conclusions about the creative thinking of test takers.

References

1. Uglanova I.L., Orel E.A., Brun I.V. Izmerenie kreativnosti i kriticheskogo myshlenija v nachal'noj shkole [Measuring creativity and critical thinking in primary school]. *Psihologicheskij Zhurnal [Psychological Journal]*. 2020, no. 6(41), pp. 96—107. (In Russ.).
2. Agresti A. Categorical data analysis. New York, NY: John Wiley & Sons, Inc, 1990.
3. Autor D.H., Levy F., Murnane R.J. The skill content of recent technological change: an empirical exploration. *Quarterly Journal of Economics*, 2003. Vol. 118, pp. 1279—1333. DOI:10.1162/003355303322552801
4. Barbot B., Besançon M., Lubart T. The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences*, 2016. Vol. 52, pp. 178—187.
5. Bentler P.M. Comparative fit indexes in structural models. *Psychological Bulletin*, 1990. Vol. 107(2), pp. 238—246. DOI:10.1037/0033-2909.107.2.238
6. Bock R.D., Gibbons R., Muraki E. Full-information item factor analysis. *Applied psychological measurement*, 1988. Vol. 12(3), pp. 261—280.
7. Brown T.A., Moore M.T. Confirmatory factor analysis. *Handbook of structural equation modeling*, 2012, pp. 361—379.
8. DiStefano C., Zhu M., Mîndrilă D. Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research, and Evaluation*, 2009. Vol. 14, no. 20. DOI:10.7275/da8t-4g52
9. Griffin P., Care E. Assessment and teaching of 21st century skills: Methods and approach. Springer, 2014.
10. Guilford J.P. The nature of human intelligence. McGraw-Hill, 1967.
11. He Q., Borgonovi F., Paccagnella M. Leveraging process data to assess adults' problem-solving skills: Identifying generalized behavioral patterns with sequence mining. *Computers and Education*, 2021. Vol. 166, 104170. DOI:10.1016/j.compedu.2021.104170
12. He Q., von Davier M. Analyzing process data from problem-solving items with N-grams: Insights from a computer-based large-scale assessment. In R. Yigal, F. Steve, M. Maryam (Eds.). Handbook of research on technology tools for real-world skill development. Hershey, PA: Information Science Reference, 2016, pp. 749—776.
13. Howell S., Veale T. Designing serious games with linguistic resources. *Proceedings of the 4th International Conference on Foundations of Digital Games*, 2009, pp. 291—298. DOI:10.1145/1536513.1536564
14. Kerr D., Chung G.K.W.K., Iseli M.R. The feasibility of using cluster analysis to examine log data from educational video games. *CRESST Report 790*, 2011.
15. Kim Y.J., Shute V.J. Opportunities and challenges in assessing and supporting creativity in video games. In G.P. Green, & J.C. Kaufman (Eds.). Video games and creativity. Elsevier Academic Press,

Tarasov S.V.
Using Process Data from Completing a Task in Creative Thinking Assessment
Psychological Science and Education. 2023. Vol. 28, no. 4

2015, pp. 99—117. DOI:10.1016/B978-0-12-801462-2.00005-9

16. Krebs E., Jaschek C., von Thienen J., Borchart K.-P., Meinel C., Kolodny O. Designing a Video Game to Measure Creativity. *2020 IEEE Conference on Games (CoG)*, 2020, pp. 407—414. DOI:10.1109/CoG47356.2020.9231672

17. OECD. "Problem-Solving Framework". In PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishing, Paris, 2013. DOI:10.1787/9789264190511-6-en

18. Qiao X., Jiao H. Data Mining Techniques in Analyzing Process Data: A Didactic. *Frontiers in Psychology*, 2018, no. 9:2231. DOI:10.3389/fpsyg.2018.02231

19. Rafner J., Hjorth A., Risi S., Philipsen L., Dumas C., Biskjær M.M., Noy L., Tylén K., Bergenholtz C., Lynch J., Zana B., Sherson J. crea.blender: A Neural Network-Based Image Generation Game to Assess Creativity. In Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play. Association for Computing Machinery, 2020, pp. 340—344. DOI:10.1145/3383668.3419907

20. Rhodes M. An Analysis of Creativity. *The Phi Delta Kappan*, 1961. Vol. 42(7), pp. 305—310.

21. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 2012. Vol. 48(2), pp. 1—36. DOI:10.18637/jss.v048.i02

22. Rupp A. Feature selection for choosing and assembling measurement models: A building-block-based organisation. *International Journal of Testing*, 2002. Vol. 2(3/4), pp. 311—360.

23. Said-Metwaly S., Van den Noortgate W., Kyndt E. Methodological issues in measuring creativity: A systematic literature review. *Creativity. Theories-Research-Applications*, 2017. Vol. 4(2), pp. 276—301.

24. Schmidt D., Heckendorf C. "ngram: Fast n-Gram Tokenization." R package version 3.2.2, 2022.

25. Shi D., Maydeu-Olivares A., Rosseel Y. Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, 2020. Vol. 27(1), pp. 1—15.

26. Sternberg R.J., Lubart T.I. Defying the crowd: Cultivating creativity in a culture of conformity. New York: Free Press, 1995

27. Torrance E.P. Guiding creative talent. Prentice-Hall, Inc., 1962. DOI:10.1037/13134-000

28. Ward T.B. Structured Imagination: the Role of Category Structure in Exemplar Generation. *Cognitive Psychology*, 1994. Vol. 27(1), pp. 1—40. DOI:10.1006/cogp.1994.1010

29. Schwab K. The Global Competitiveness Report. World Economic Forum, 2018.

30. Yuan J., Xiao Y., Liu H. Assessment of Collaborative Problem Solving Based on Process Stream Data: A New Paradigm for Extracting Indicators and Modeling Dyad Data. *Frontiers in Psychology*, 2019, no. 10:369. DOI:10.3389/fpsyg.2019.00369

Литература

1. *Угланова И.Л., Орел Е.А., Брун И.В.* Измерение креативности и критического мышления в начальной школе // Психологический Журнал. 2020. № 6(41). С. 96—107.

2. *Agresti A.* Categorical data analysis. New York, NY: John Wiley & Sons, Inc, 1990.

3. *Autor D.H., Levy F., Murnane R.J.* The skill content of recent technological change: an empirical exploration // Quarterly Journal of Economics. 2003. Vol. 118. P. 1279—1333. DOI:10.1162/003355303322552801

4. *Barbot B., Besançon M., Lubart T.* The generality-specificity of creativity: Exploring the structure of creative potential with EPoC // Learning and Individual Differences. 2016. Vol. 52. P. 178—187.

5. *Bentler P.M.* Comparative fit indexes in structural models // Psychological Bulletin. 1990. Vol. 107(2). P. 238—246. DOI:10.1037/0033-2909.107.2.238

6. *Bock R.D., Gibbons R., Muraki E.* Full-information item factor analysis // Applied psychological measurement. 1988. Vol. 12(3). P. 261—280.

7. *Brown T.A., Moore M.T.* Confirmatory factor analysis // Handbook of structural equation modeling. 2012. P. 361—379.

8. *DiStefano C., Zhu M., Mĩndrilă D.* Understanding and Using Factor Scores: Considerations for the Applied Researcher // Practical Assessment, Research, and Evaluation. 2009. Vol. 14, Article 20. DOI:10.7275/da8t-4g52

9. *Griffin P., Care E.* Assessment and teaching of 21st century skills: Methods and approach. Springer, 2014.

10. *Guilford J.P.* The nature of human intelligence. McGraw-Hill, 1967.

11. *He Q., Borgonovi F., Paccagnella M.* Leveraging process data to assess adults' problem-solving skills: Identifying generalized behavioral patterns with sequence mining // Computers and Education. 2021. Vol. 166, 104170. DOI:10.1016/j.compedu.2021.104170

12. *He Q., von Davier M.* Analyzing process data from problem-solving items with N-grams: Insights from a computer-based large-scale assessment / In R. Yigal, F. Steve, M. Maryam (Eds.). // Handbook of research on technology tools for real-world skill development. Hershey, PA: Information Science Reference, 2016. P. 749—776.

13. *Howell S., Veale T.* Designing serious games with linguistic resources // Proceedings of the 4th International

Тарасов С.В.
Использование данных о процессе выполнения задания при оценке креативного мышления
Психологическая наука и образование. 2023. Т. 28. № 4

Conference on Foundations of Digital Games. 2009. P. 291—298. DOI:10.1145/1536513.1536564

14. *Kerr D., Chung G.K.W.K., Iseli M.R.* The feasibility of using cluster analysis to examine log data from educational video games // CRESST Report 790. 2011.

15. *Kim Y.J., Shute V.J.* Opportunities and challenges in assessing and supporting creativity in video games / In G. P. Green, & J. C. Kaufman (Eds.) // Video games and creativity. Elsevier Academic Press. 2015. P. 99—117. DOI:10.1016/B978-0-12-801462-2.00005-9

16. *Krebs E., Jaschek C., von Thienen J., Borchart K.-P., Meinel C., Kolodny O.* Designing a Video Game to Measure Creativity // 2020 IEEE Conference on Games (CoG). 2020. P. 407—414. DOI:10.1109/CoG47356.2020.9231672

17. OECD. "Problem-Solving Framework" // In PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishing, Paris. 2013. DOI:10.1787/9789264190511-6-en

18. *Qiao X., Jiao H.* Data Mining Techniques in Analyzing Process Data: A Didactic // Frontiers in Psychology. 2018. № 9:2231. DOI:10.3389/fpsyg.2018.02231

19. *Rafner J., Hjorth A., Risi S., Philipsen L., Dumas C., Biskjær M.M., Noy L., Tylén K., Bergenholtz C., Lynch J., Zana B., Sherson J.* crea.blender: A Neural Network-Based Image Generation Game to Assess Creativity // In Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play. Association for Computing Machinery. 2020. P. 340—344. DOI:10.1145/3383668.3419907

20. *Rhodes M.* An Analysis of Creativity // The Phi Delta Kappan. 1961. Vol. 42(7). P. 305—310.

21. *Rosseel Y.* lavaan: An R Package for Structural Equation Modeling // Journal of Statistical Software. 2012. Vol. 48(2). P. 1—36. DOI:10.18637/jss.v048.i02

22. *Rupp A.* Feature selection for choosing and assembling measurement models: A building-block-based organisation // International Journal of Testing. 2002. Vol. 2(3/4). P. 311—360.

23. *Said-Metwaly S., Van den Noortgate W., Kyndt E.* Methodological issues in measuring creativity: A systematic literature review // Creativity. Theories-Research-Applications. 2017. Vol. 4(2). P. 276—301.

24. *Schmidt D., Heckendorf C.* "ngram: Fast n-Gram Tokenization." R package version 3.2.2. 2022.

25. *Shi D., Maydeu-Olivares A., Rosseel Y.* Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA // Structural Equation Modeling: A Multidisciplinary Journal. 2020. Vol. 27(1). P. 1—15.

26. *Sternberg R.J., Lubart T.I.* Defying the crowd: Cultivating creativity in a culture of conformity // New York: Free Press, 1995

27. *Torrance E.P.* Guiding creative talent // Prentice-Hall, Inc. 1962. DOI:10.1037/13134-000

28. *Ward T.B.* Structured Imagination: the Role of Category Structure in Exemplar Generation // Cognitive Psychology. 1994. Vol. 27(1). P. 1—40. DOI:10.1006/cogp.1994.1010

29. Schwab K. The Global Competitiveness Report // World Economic Forum. 2018.

30. *Yuan J., Xiao Y., Liu H.* Assessment of Collaborative Problem Solving Based on Process Stream Data: A New Paradigm for Extracting Indicators and Modeling Dyad Data // Frontiers in Psychology. 2019. № 10:369. DOI:10.3389/fpsyg.2019.00369

Information about the authors
Sergei V. Tarasov, PhD Student, Research Assistant at Center for Psychometrics and Measurement in Education, Institute of Education, HSE University, Moscow, Russia, ORCID: https://orcid.org/0000-0003-4151-115X, e-mail: svtarasov@hse.ru

Информация об авторах
Тарасов Сергей Владимирович, аспирант, стажер-исследователь Центра психометрики и измерений в образовании, Институт образования, ФГАОУ ВО «Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»), г. Москва, Российская Федерация, ORCID: https://orcid.org/0000-0003-4151-115X, e-mail: svtarasov@hse.ru