

---

ПСИХОЛОГИЯ ОБРАЗОВАНИЯ И ПЕДАГОГИЧЕСКАЯ ПСИХОЛОГИЯ  
EDUCATIONAL PSYCHOLOGY AND PEDAGOGICAL PSYCHOLOGY

---

**Exploring the Relationship between Performance and Response Process Data  
in Digital Literacy Assessment**

**Irina O. Tkachenko**

*National Research University Higher School of Economics, Moscow, Russia;  
Sirius University of Science and Technology, Sirius, Krasnodar region, Russia  
ORCID: <https://orcid.org/0000-0002-0406-7400>, e-mail: [iotkachenko@edu.hse.ru](mailto:iotkachenko@edu.hse.ru)*

**Ksenia V. Tarasova**

*National Research University Higher School of Economics, Moscow, Russia  
ORCID: <https://orcid.org/0000-0002-3915-3165>, e-mail: [ktarasova@hse.ru](mailto:ktarasova@hse.ru)*

**Daria A. Gracheva**

*National Research University Higher School of Economics, Moscow, Russia  
ORCID: <https://orcid.org/0000-0002-4646-7349>, e-mail: [dgracheva@hse.ru](mailto:dgracheva@hse.ru)*

Measuring complex latent constructs is challenging because of their multi-dimensionality. In this context, computer-based assessments have gained popularity due to its ability to handle large diverse data. The aim of the study is to investigate the interrelationship between performance, time, and actions in computer-based digital literacy assessment. The study involved more than 400 8<sup>th</sup>-grade schoolchildren (approximately 14—15 years old) from secondary schools in Russia. A subset was obtained from indicators capturing the demonstration of analysis of data, information, and digital content, which is a component of the information literacy in the digital literacy framework. The data was used to create latent models in the structural equation modeling framework. Confirmatory one-factor model for the Performance factor showed a good fit to the data (CFI=1; TLI=1; RMSEA=0). The model with dependencies among indicators demonstrated improved model fit ( $\chi^2_{(18)}=510,65$ ;  $p=0,05$ ) compared to the model without such dependencies. The results suggest that performance, time, and actions are interdependent. The findings underscore the need for a comprehensive approach to assessing digital literacy that accounts for these interdependencies, as well as investigating behavioral patterns of interaction with a large amount of information in the digital environment.

**Keywords:** computer-based assessment, digital literacy, evidence-centered design, structural equation modeling, confirmatory factor analysis, process data, response time, clicks.

**Funding.** This work is supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-10-2021-093; Project COG-RND-2104).

**For citation:** Tkachenko I.O., Tarasova K.V., Gracheva D.A. Exploring the Relationship between Performance and Response Process Data in Digital Literacy Assessment [Electronic resource]. *Sovremennaya zarubezhnaya psikhologiya* = Journal of Modern Foreign Psychology, 2024. Vol. 13, no. 1, pp. 58—68. DOI: <https://doi.org/10.17759/jmfp.2024130105> (In Russ.).

## Исследование взаимосвязи данных о результате и процессе выполнения заданий при оценке цифровой грамотности

**Ткаченко И.О.**

Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»),  
г. Москва; Научно-технологический университет «Сириус» (АНОО ВО «Университет “Сириус”»),  
пгт. Сириус, Российская Федерация  
ORCID: <https://orcid.org/0000-0002-0406-7400>, e-mail: [iotkachenko@edu.hse.ru](mailto:iotkachenko@edu.hse.ru)

**Тарасова К.В.**

Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»),  
г. Москва, Российская Федерация  
ORCID: <https://orcid.org/0000-0002-3915-3165>, e-mail: [ktarasova@hse.ru](mailto:ktarasova@hse.ru)

**Грачева Д.А.**

Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»),  
г. Москва, Российская Федерация  
ORCID: <https://orcid.org/0000-0002-4646-7349>, e-mail: [dgracheva@hse.ru](mailto:dgracheva@hse.ru)

Измерение комплексных латентных конструктов является непростой задачей. В этом контексте цифровые инструменты оценивания становятся все более популярными благодаря их способности обрабатывать большие и разнообразные данные. Целью исследования является изучение взаимосвязи между результатами тестирования, продолжительностью времени ответа и количеством предпринятых действий во время оценивания уровня цифровой грамотности. Участниками исследования стали более 400 восьмиклассников из московских школ. Для анализа были отобраны данные, отражающие навыки анализа данных, информации и цифрового контента, что является частью информационной грамотности в теоретической рамке цифровой грамотности. Применение моделирования структурными уравнениями позволило построить латентные модели, включая однофакторную модель «Результат тестирования» с высокими показателями соответствия и трехфакторную модель, демонстрирующую взаимосвязь между изучаемыми переменными. Результаты подчеркивают значимость интеграции этих взаимосвязей в цифровые инструменты оценивания цифровой грамотности и анализ поведенческих стратегий при работе с информацией в цифровом пространстве.

**Ключевые слова:** цифровые инструменты оценивания, цифровая грамотность, метод доказательной аргументации, моделирование структурными уравнениями, конфирматорный факторный анализ, данные о процессе ответа, время ответа, клики.

**Финансирование.** Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (Соглашение № 075-10-2021-093, Проект COG-RND-2104).

**Для цитаты:** Ткаченко И.О., Тарасова К.В., Грачева Д.А. Исследование взаимосвязи данных о результате и процессе выполнения заданий при оценке цифровой грамотности [Электронный ресурс] // Современная зарубежная психология. 2024. Том 13. № 1. С. 58—68. DOI: <https://doi.org/10.17759/jmfp.2024130105>

### Introduction

Accurate measuring of complex latent constructs presents a challenge due to its abstract and multifaceted nature. In this context, computer-based assessments (CBAs) have gained popularity due to its ability to handle large and diverse data [22]. CBAs leverage process data — from simple inputs, such as a button press, to complex patterns of actions — to analyze test-takers' activities that lead to specific outcomes [18]. These actions are typically captured in event log files as process data. Although there is no standardized method for selecting and generating this data, it provides new opportunities for researchers who work with data obtained by assessment tools with tasks in a digitally enriched environment.

Examining a test-taker's interaction with the elements of the test digital environment allows researchers to delve into the cognitive processes, emotions, motives, decision-making strategies, and problem-solving behaviors of test-takers, offering insights into their performance and construct understanding [21]. These processes are responsible for generating observed variations in test scores. Thus, Yamamoto and Lennon suggested applying process data such as time spent during CBAs to identify response patterns that reflect data fabrication [26]. Yu and colleagues investigated the utility and fairness of different test-takers' data sources for predicting college performance success [19]. They found that combining institutional data and process data (total number of clicks and total time) led to the most significant accuracy increase in predicting both

short-term and long-term test-takers' success. Teig and colleagues expanded the amount of process data used by the colleagues and identified three distinct exploration behaviors — strategic, emergent, and disengaged — that also varied in terms of item accuracy [17]. Moreover, process data can reveal patterns of response, such as speed-accuracy trade-offs [16], or suggest potential sources of measurement error, such as difficulties with certain types of items or response biases [2]. This information can be used to refine the scoring algorithms used to interpret test results and improve the accuracy and reliability of test scores [20; 27].

Unlike conventional information about the correctness of the answer, the study of response processes aims to capture the unique testing situation. Such possibilities provide a foundation for exploring broader aspects of digital literacy (DL) proficiency, a crucial component in today's digital age. DL is a complex latent construct that is defined as the ability to safely use digital technologies for searching, analyzing, creating, managing information, communication, and collaboration to solve problems in a digital environment to meet personal, educational, and professional needs [1]. DL may be reflected in behavior and reaction patterns that can take on a wide range of forms based on the diverse social practices that enable people to comprehend, communicate, and use information in digital contexts [9]. Researchers and global organizations in the field of education note the importance but challenge of its measurement [12; 13; 14; 15; 23]. To address response process validity challenges during measuring DL, the researchers use CBAs, where detailed process data may be gathered and reveal additional validity evidence [6]. For instance, Bartolom and colleagues found that the participants' response time and eye movement patterns varied significantly based on the scores they achieved during students' DL assessment using CBA [3]. In another study by Li and colleagues, the authors collected object-, time-, and click-related process data from students' interaction with the CBA tool and extracted feature variables related to DL in order to construct the measurement model [9]. However, to the best of our knowledge, none of the existing studies were focused at revealing the relationship between DL and process data collected during its assessment in the Russian sample. Thus, this study raises a research question: how performance induced by complex latent construct of DL correlates with invested time and the number of actions in the context of DL assessment?

## Methods

### Participants

The study involved 440 8<sup>th</sup>-grade schoolchildren (approximately 14—15 years old) from 15 secondary schools in Moscow, Russia. School staff made their own decisions about the number of test-takers, which resulted in a lack of additional descriptive characteristics for the sample. The

schoolchildren completed the test individually and online, using school computers and sitting in small isolated groups. The test sessions were overseen by specially trained administrators, ensuring standardized administration.

### Instrument

The computer-based DL assessment tool was developed based on an evidence-centered design methodology [8] in a format of ten interlinked interactive scenario-based tasks. Such format offers more authenticity compared to classical test forms [25], because it replicates real-world situations that test-takers could face and creates a rich and immersive digital environment that captures behavior that matches the measured construct [24]. Such types of tasks help to resolve the internal motivation issue of performance and increase the reliability of the results obtained, which is especially important for tests with low stakes [4; 5; 11]. Throughout the development process, the equal coverage concept was adhered to, meaning that every task was designed to assess DL components in a way that would fairly cover every subcomponent included in the theoretical frame: information literacy, computational literacy, technical literacy, digital communication, digital security [1]. Thus, the assessment in the digital environment is based on observable actions of test-takers that reflect the construct of DL being measured.

The test follows a story where a test-taker and their virtual classmate create a digital project for a contest. Each task is a self-contained scenario with some units of activity for a test-taker, which provokes observed behavior and outcome. The narratives of the tasks cover every stage involved: from finding the contest using a simulated search engine to submitting the developed project on the contest's website in a specific format. During testing, each test-taker's performance trace is recorded in a log file.

Fig. 1 shows a fragment of the task, where a test-taker is instructed to choose a program that meets certain criteria. To do this, a test-taker needs to demonstrate the ability to analyze, interpret, and critically evaluate data, information, and digital content, considering available information about the authority, reliability, and credibility of the source (analysis of data, information, and digital content is a component of the information literacy in DL framework). Here, as a part of the broader program search task, a test-taker has to evaluate feedback from the users based on their reputation and credibility. After drawing a conclusion, a test-taker needs to reply to a virtual classmate in the messenger by sending the name of the chosen program selected from the drop-down list. The following process data is saved in a log file: 1) response time 2) and the number of clicks, including scrolling produced and clicks in the messenger field to select a response. Moreover, the polytomous score (0, 1, or 2) that reflects the test-taker's proficiency in evaluating the authority of the information source is assigned based on the chosen answer option (program) and is recorded in the same log file.

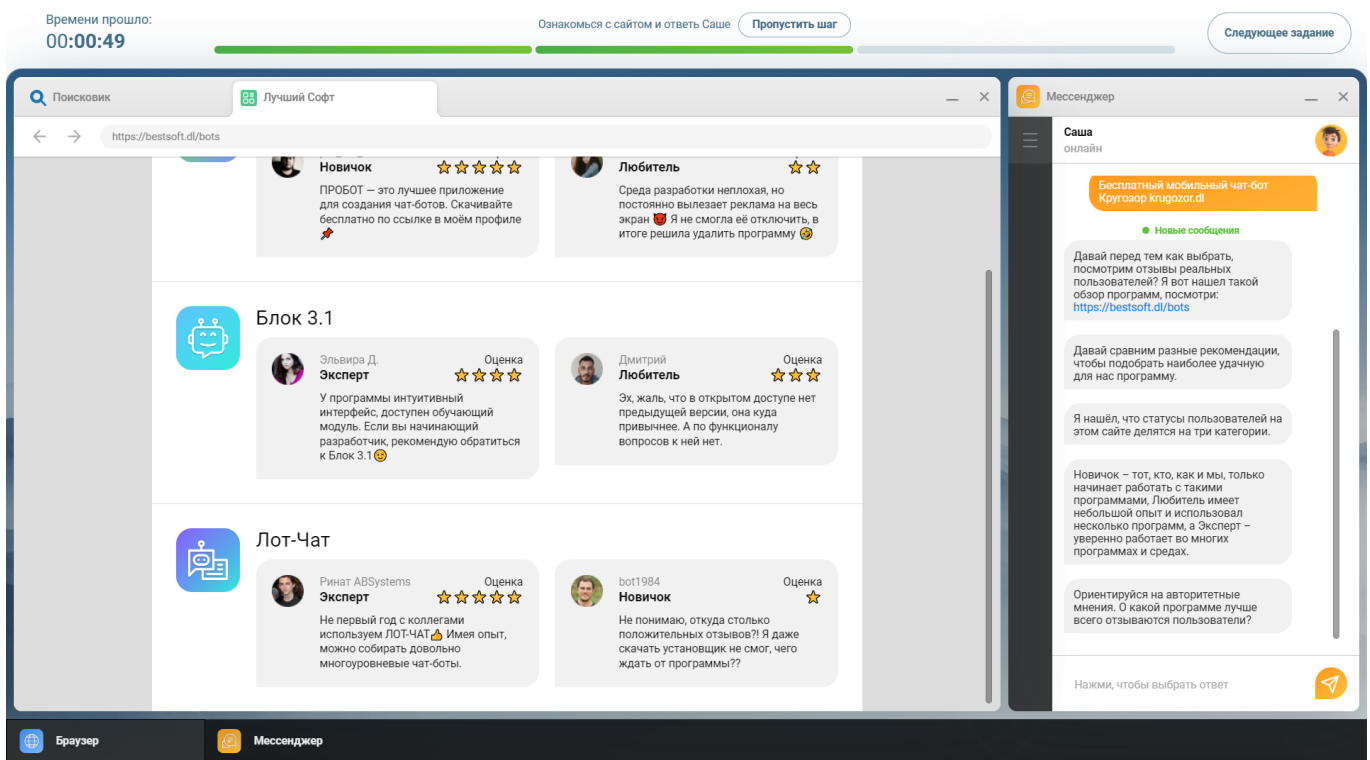


Fig. 1. Example of the unit of activity in DL assessment tool

### Data Analysis Strategy

Analysis was conducted in R version 4.2.1 in package “lavaan”. During data preparation, all test-takers’ profiles with zero values for the response time and the number of clicks were deleted, as well as profiles with missing values for the obtained scores. As a result, the sample consisted of 426 students.

To address the aim, we applied the model proposed by De Boeck and Scalise within the structural equation modeling framework [7]. For the purposes of this study, we decided to analyze data collected only from units of activity that call for the demonstration of cognitive skills linked to analysis of data, information, and digital content. For each unit of activity, we considered the following indicators: 1) the score that was obtained based on predefined evidence rules related to the digital literacy behavior, 2) the time spent from the moment the stem was presented till the moment the answer was sent, 3) and the total number of clicks made. The aggregated data from all the included into analysis units of activity was then categorized into three distinct factors: Performance, Time, and Action. The Performance factor includes scores obtained in the six units of activity; the Time factor includes the log-transformed numbers of seconds spent for completing each unit of activity; the Action factor includes log-transformed numbers of clicks taken by test-takers during each unit of activity. More parameters were added to the model to account for additional dependencies: residual covariances (correlations) for the Action indicators and the Performance indicators, direct effect of the Action indicators on the Time indicators, direct effect of the Performance indicators on the Time indicators. The diagram of the model is shown in Fig. 2.

The construction of the mentioned model was carried out in three distinct steps. Firstly, a separate confirmatory

factor analysis model was constructed for the Performance factor to ensure the internal structure of the scale. Secondly, the three-factor model without dependencies was specified. Lastly, additional dependencies described above were added to the model.

To estimate the models’ parameters, the Robust Maximum Likelihood (MLR) estimator was used to account for the non-normality of the data [23]. The model fit was assessed according to the following goodness of fit indices: Comparative Fit Index (CFI)>0,95; Tucker-Lewis Index (TLI) > 0,95; Root Mean Square Error of Approximation (RMSEA)<0,06 [10].

### Results

Confirmatory one-factor model for the Performance factor showed a good fit to the data (CFI=1; TLI=1; RMSEA=0). Standardized factor loadings ranged from 0,196 to 0,595 and were significant ( $p<0,05$ ). Thereby, we can use it in the further analysis to investigate the relationship between the Performance, Time, and Action factors.

Then, the three-factor model without dependencies was estimated and showed a poor fit with the data (CFI=0,548; TLI=0,476; RMSEA=0,096). Three-factor model with dependencies demonstrated improved model fit (CFI=0,966; TLI=0,954; RMSEA=0,028). Model comparison using Scaled Chi-Squared Difference Test confirmed that the model with dependencies has significantly better fit ( $\chi^2_{(18)}=510,65; p<0,05$ ). For the latter model, standardized loadings of all the Performance, Time, and Action indicators were significant ( $p<0,05$ ; Tab. 1).

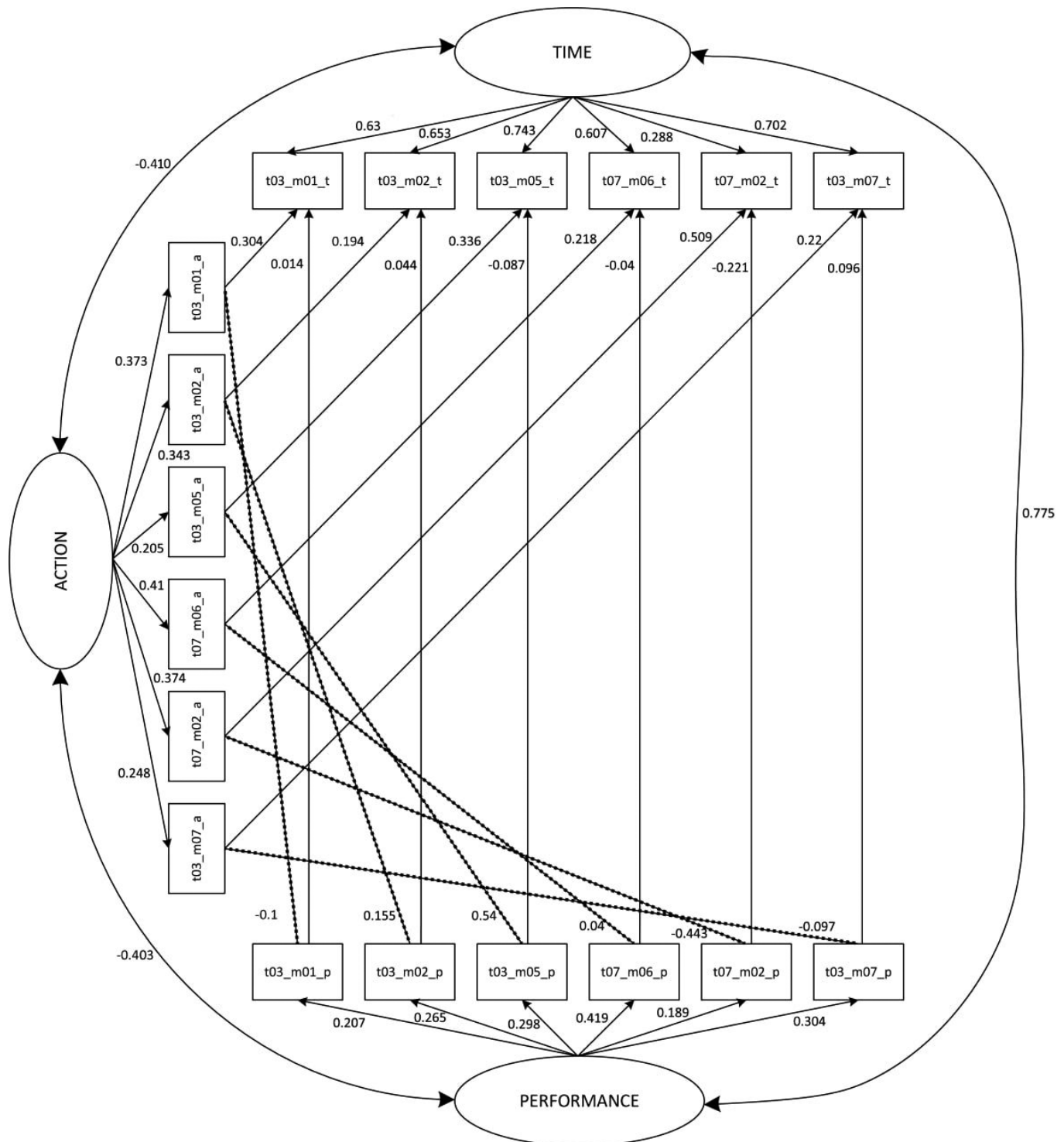


Fig. 2. Diagram for three-factor CFA model with additional dependencies

Note: each variable name follows the format *IndicatorName\_Factor*. For example, the time spent on completing a unit of activity containing indicator t03\_m02 is denoted as *t03\_m02\_t*; “→” — direct effect; “- - -” — correlated residuals.

Following the aim of the paper, we explore correlations between latent factors as well as patterns of additional dependencies (Tab. 2). Correlations between latent factors were as follows: for the Performance and Time factors:  $r=0,775$ ,  $p<0,05$ ; for the Performance and Action factors:  $r=-0,403$ ,  $p<0,05$ ; for the Action and Time factors:  $r=-0,410$ ,  $p<0,05$ . Thus, higher performance requires more time and fewer actions. The Time factor is negatively related to the Action factor, which means

that students who performed more actions were faster in providing answers. However, the additional direct effects of the Action indicators on the Time indicators were positive ( $ps<0,05$ ). These additional dependencies are not explained by latent factors and can, therefore, be observed because they belong to the same indicators (students' behavior within the unit of activity has similarities). All the other additional dependencies did not show a stable pattern and were only significant for certain indicators.

Table 1

Standardized factor loadings in the three-factor mode with dependencies

Indicator	Standardized factor loading
<i>the Performance factor</i>	
t03_m01_p	0,207*
t03_m02p	0,265*
t03_m05_p	0,298*
t03_m07_p	0,304*
t07_m06_p	0,419*
t07_m02_p	0,189*
<i>the Time factor</i>	
t03_m01_t	0,63*
t03_m02_t	0,653*
t03_m05_t	0,743*
t03_m07_t	0,702*
t07_m06_t	0,607*
t07_m02_t	0,288*
<i>the Action factor</i>	
t03_m01_a	0,373*
t03_m02_a	0,343*
t03_m05_a	0,205*
t03_m07_a	0,248*
t07_m06_a	0,41*
t07_m02_a	0,374*

Note: “\*” — Standardized factor loading was found to be significant at  $p < 0,05$ .

## Discussion

In this study, we examined the interrelationship between process data and performance on a digital assessment tool developed in the format of CBA. For a subset of indicators capturing proficiency in the analysis of data, information, and digital content, data for obtained score, time, and number of actions was collected and used to create latent models in the structural equation modeling framework. The initial one-factor model for the Performance factor showed a good fit to the data and optimal and statistically significant standardized factor loadings. The model containing three factors (Performance, Time, and Action) showed a poor fit to the data but improved significantly upon the inclusion of additional dependencies among behavior within the same indicators as was suggested by Boeck and Scalise [7]. This underscores the complexity of factors influencing accurate DL assessment.

Following the aim of the study, special attention should be paid to the correlations between performance and process data. The significant positive correlation between the Performance factor and the Time factor is in line with the previous studies [7]. This could imply that more time invested in processing leads to higher performance. However, this relationship might be attributed to the assessment instrument’s nature in our case. The tasks include extensive information to analyze (in the form of websites). Showing proficiency in such conditions is supposed to be time-consuming. Moreover, the variability in dependencies shows certain

Table 2

Standardized effects for dependences in the three-factor model with dependencies

Indicator	Dependencies	Standardized effect
t03_m01	t03_m01_p → t03_m01_t	0,014
	t03_m01_p ~ t03_m01_a	-0,1
	t03_m01_a → t03_m01_t	0,304*
t03_m02	t03_m02_p → t03_m02_t	0,044
	t03_m02_p ~ t03_m02_a	0,155*
	t03_m02_a → t03_m02_t	0,194*
t03_m05	t03_m05_p → t03_m05_t	-0,087*
	t03_m05_p ~ t03_m05_a	0,54
	t03_m05_a → t03_m05_t	0,336*
t03_m07	t03_m07_p → t03_m07_t	0,096*
	t03_m07_p ~ t03_m07_a	-0,097
	t03_m07_a → t03_m07_t	0,22*
t07_m06	t07_m06_p → t03_m06_t	-0,04
	t07_m06_p ~ t03_m06_a	0,04
	t07_m06_a → t03_m06_t	0,218*
t07_m02	t07_m02_p → t03_m02_t	-0,221*
	t07_m02_p ~ t03_m02_a	-0,443*
	t07_m02_a → t03_m02_t	0,509*

Note: “\*” — Standardized factor loading was found to be significant at  $p < 0,05$ ; “~” — residual correlation between the Performance and Action factors; “→” — direct effect (the Action indicators on the Time indicators, the Performance indicators on the Time indicators).

patterns only being significant for specific indicators that may reveal context dependency.

Conversely, the negative correlation between the Performance factor and the Action factor was observed. The digital environment in the assessment tool lacked obligatory interactive elements, such as the need for plenty of clicks. That potentially identifies a specific clicking pattern of individuals with low-ability. Such behavior may also reflect attempts to find additional information, hints; understanding of the digital environment functionality; or merely curiosity.

Another negative correlation between the Time factor and the Action factor implies that high levels of interaction with the digital environment do not necessarily lead to increased time spent. It may be explained by the diversity of the sample, including test-takers who prefer trial-and-error methods in situations when they are not sure about the correctness of the answer. However, these results raise questions about the interpretation drawn. The same result was observed with a different instrument measuring a distinct complex latent construct, leading us to believe that this is not related to the construct per se but rather to behavioral patterns in the digital environment [7]. It suggests the need for further exploration of behavioral strategies and latent profiles of test-takers. This also may shed light on how modern children interact with large volumes of information in digital environments, hinting that such behavior might be a characteristic of this cohort. In general, the finding that higher performance needs more time and fewer actions

enhances the understanding of DL assessment, implying that quicker response may not always be an indicator of efficacy, particularly if a lot of actions are taken.

Despite these insights, the study has limitations. The sample was restricted to 8<sup>th</sup>-grade students from Moscow, which may limit the generalizability of our findings. Moreover, due to the high level of anonymity of test-takers associated with a child sample recruited through private and public schools, it is impossible to check whether the findings are consistent in specific subsamples. Additionally, the instrument used has specific characteristics that do not universally apply to all DL assessment tools. Future research should aim to replicate these findings in diverse assessment settings and with varied age groups. Another limitation is that in the formed factors, the data used was not from all the units of activity. Moreover, adding finer-grained process data could reveal more about how test-takers manage and distribute their time and actions throughout the unit of activity. Further research might concentrate on breaking down time into different phases of answering or distinguishing click areas. Moreover, incorporating other process data,

such as duration and frequency of pauses or changes in the answer option, may reveal the complex interplay between them and the performance. Additionally, the data analysis was restricted to linear relationships among the factors. Exploring non-linear dynamics could offer a deeper understanding of the complexities involved in the assessment.

## Conclusion

This research demonstrates that the obtained score, response time, number of clicks are interdependent, as evidenced by the confirmatory factor analysis outcomes. The findings underscore the need for a comprehensive approach to assessing DL that accounts for these interdependencies. The relationships identified can be used to refine the DL assessment tool tasks, and potentially considered in the development of other instruments for measuring complex latent constructs using CBA. Moreover, educational environments that provide real-time feedback may be designed not only to consider the correctness but also the efficiency of time spent and actions.

## Краткое изложение содержания статьи на русском языке

### Введение

Измерение комплексных латентных конструктов, таких как цифровая грамотность, является непростой задачей для исследователей. В этом контексте цифровые инструменты оценивания выглядят как многообещающее решение, поскольку предоставляют возможности для сбора и анализа обширных мультимодальных данных, собранных в процессе тестирования. Изучение взаимодействия тестируемых с цифровыми инструментами оценивания способствует более глубокому пониманию когнитивных процессов и стратегий, влияющих на получаемые результаты. Анализ затраченного времени и количества кликов для прохождения тестирования может указывать как на потенциальные несоответствия в данных [26], так и на скорость и точность ответов тестируемых [17]. Интеграция дополнительной информации о тестируемых, такой как пол, уровень дохода или среднегодовая оценка, с данными о взаимодействии с цифровыми инструментами оценки улучшает прогнозирование учебных достижений [16; 19]. Однако в России такой подход к анализу данных не применялся. Целью представленного исследования является анализ взаимосвязи между результатом тестирования, затраченным временем и объемом совершенных действий во время оценки цифровой грамотности школьников.

### Методы

В исследовании приняли участие 440 учащихся 8-го класса (14—15 лет) из 15 московских школ.

Тестирование проводилось в форме индивидуального онлайн-теста на школьных компьютерах. Участники работали в малых группах под наблюдением сотрудников, прошедших специальное обучение. Используемый цифровой инструмент для оценки цифровой грамотности состоит из десяти интерактивных сценарных заданий, имитирующих ситуации из реальной жизни, что повышает мотивацию к выполнению. Каждое задание направлено на оценку конкретных аспектов цифровой грамотности. Для дальнейшего анализа были отобраны данные, отражающие навыки анализа данных, информации и цифрового контента, что является частью информационной грамотности в теоретической рамке цифровой грамотности.

### Результаты

Для проведения статистического анализа были использованы программы R и RStudio и пакет «lavaan». На этапе предварительной обработки данных были исключены записи тестируемых, содержащие нулевые или отсутствующие значения по параметрам времени и количества кликов, а также отсутствующие значения полученного балла. Итоговый анализ проводился на основе данных 426 тестируемых. Была использована модель структурного уравнения (structural equation model), основанная на подходе П. Де Бока (P. De Boek) и К. Скэлиса (K. Scalise) [7]. Она включает в себя три основных фактора: «Результат тестирования» (основан на шести политомических показателях правильности ответов), «Время» (логарифмически преобразованное время выполнения модулей) и «Действие» (логарифмически преобразованное количество кликов по модулям).

Дополнительно модель включала остаточные ковариации между факторами «Действие» и «Результат тестирования», а также прямое влияние факторов «Действие» на «Время» и «Результат тестирования» на «Время».

Сначала была построена однофакторная модель «Результат тестирования». После этого создана трехфакторная модель, для которой первоначально не предполагались взаимосвязи между факторами. Однако далее такие зависимости были добавлены. Для оценки параметров моделей использовался робастный алгоритм максимального правдоподобия (robust maximum likelihood estimation, MLR), что позволило учесть ненормальность распределения данных. Для оценки качества моделей использовались сравнительный индекс соответствия (comparative fit index, CFI) и индекс Такера-Льюиса (Tucker Lewis index, TLI).

Результаты анализа однофакторной модели «Результат тестирования» показали отличное соответствие с данными (CFI=1, TLI=1, RMSEA=0), при этом факторные нагрузки варьировались от 0,196 до 0,595. Дальнейший анализ показал, что трехфакторная модель без учета взаимосвязей между факторами демонстрирует недостаточное соответствие данным (CFI=0,548; TLI=0,476; RMSEA=0,096). При этом модель, включающая зависимости между факторами, показывает значительно лучшие результаты (CFI=0,966; TLI=0,954; RMSEA=0,028). Сравнение соответствия двух последних моделей данным подтвердило статистически значимое улучшение в модели, учитывающей взаимосвязи ( $\chi^2_{(18)}=510,65$ ;  $p<0,05$ ). Факторные нагрузки модели, учитывающей взаимосвязи варьировались от 0,189 до 0,743 и были статистически значимыми ( $p<0,05$ ).

Выявлена значительная положительная корреляция между факторами «Результат тестирования» и «Время» ( $r=0,775$ ;  $p<0,05$ ). Подобные результаты могут быть связаны с особенностями самого инструмента, содержащего большое количество текста для анализа. Кроме того, обнаружены отрицательные корреляции между факторами «Результат тестирования» и «Действие» ( $r=-0,403$ ,  $p<0,05$ ), а также между факторами «Время» и «Действие» ( $r=-0,410$ ;  $p<0,05$ ), что приводит к интересной закономерности: увеличение числа кликов не способствует повышению результативности или увеличению времени, необходимого для выполнения задач. Напротив, большое количество кликов может отражать более хаотичное или менее целенаправленное взаимодействие с цифровыми ресурсами, что в итоге не приводит к повышению качества выполнения заданий.

## Обсуждение результатов

Результаты подчеркивают важность баланса между качеством и количеством взаимодействий с цифровыми ресурсами в контексте оценки цифровой грамотности. Более высокие достижения коррелируют с умением эффективно распоряжаться временем и минимизировать количество ненужных действий, что может отражать более глубокое понимание проанализированной информации и лучшую стратегию решения задач. Результаты исследования позволяют отметить значимость интегрированного подхода к оценке цифровой грамотности, который учитывает не только точность ответов, но и эффективность использования времени и количества действий.

## References

1. Avdeeva S., Tarasova K. Ob otsenke tsifrovoy gramotnosti: metodologiya, kontseptual'naya model' i instrument izmereniya [Digital Literacy Assessment: Methodology, Conceptual Model and Measurement Tool]. *Voprosy obrazovaniya = Educational Studies (Moscow)*, 2023, no. 2, pp. 8—32. DOI:10.17323/1814-9545-2023-2-8-32\_(In Russ.).
2. Zhang S., Wang Z., Qi J., Liu J., Ying Z. Accurate assessment via process data. *Psychometrika*, 2023. Vol. 88, no. 1, pp. 76—97. DOI:10.1007/s11336-022-09880-8
3. Bartolomé, J., Garaizar, P., & Bastida, L. (2020, October). Validating item response processes in digital competence assessment through eye-tracking techniques. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 738-746). DOI:10.1145/3434780.3436641
4. Bergner Y., von Davier A.A. Process data in NAEP: Past, present, and future. *Journal of Educational and Behavioral Statistics*, 2019. Vol. 44, no. 6, pp. 706—732. DOI:10.3102/1076998618784700
5. Hamari J., Shernoff D.J., Rowe E., Coller B., Asbell-Clarke J., Edwards T. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in human behavior*, 2016. Vol. 54, pp. 170—179. DOI:10.1016/j.chb.2015.07.045
6. Cui Y., Chen F., Lutsyk A., Leighton J.P., Cutumisu M. Data literacy assessments: A systematic literature review. *Assessment in Education: Principles, Policy & Practice*, 2023. Vol. 30, no. 1, pp. 76—96. DOI: 10.1080/0969594x.2023.2182737
7. De Boeck P., Scalise K. Collaborative problem solving: Processing actions, time, and performance. *Frontiers in psychology*, 2019. Vol. 10, article ID 1280, 9 p. DOI:10.3389/fpsyg.2019.01280
8. Mislevy R.J., Behrens J.T., Dicerbo K.E., Levy R. Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of educational data mining*, 2012. Vol. 4, no. 1, pp. 11—48. DOI:10.5281/zenodo.3554641



9. Li J., Bai J., Zhu S., Yang H.H. Game-Based Assessment of Students' Digital Literacy Using Evidence-Centered Game Design. *Electronics*, 2024. Vol. 13(2), Article ID 385, 19 p. DOI:10.3390/electronics13020385
10. Hu L., Bentler P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 1999. Vol. 6, no. 1, pp. 1—55. DOI:10.1080/10705519909540118
11. Nichols S.L., Dawson H.S. Assessment as a context for student engagement. In Christenson S.L., Reschly A.L., Wylie C. (eds.), *Handbook of research on student engagement*. Boston: Springer Science+Business Media, 2012, pp. 457—477. DOI:10.1007/978-1-4614-2018-7\_22
12. Oliveri M.E., Mislevy R.J. Introduction to “Challenges and opportunities in the design of ‘next-generation assessments of 21st century skills’” special issue. *International Journal of Testing*, 2019. Vol. 19, no. 2, pp. 97—102. DOI:10.1080/15305058.2019.1608551
13. Peng D., Yu Z. A literature review of digital literacy over two decades. *Education Research International*, 2022. Vol. 2022, article ID 2533413, 8 p. DOI:10.1155/2022/2533413
14. OECD. Recommendation of the Council on Children in the Digital Environment [Electronic resource]. Paris: OECD, 2022. 14 p. URL: <https://legalinstruments.oecd.org/public/doc/272/272.en.pdf> (Accessed 26.02.2024).
15. Laanpere M., UNESCO, UNESCO Institute for Statistics. Recommendations on assessment tools for monitoring digital literacy within unesco's digital literacy global framework. Montreal: UNESCO Institute for Statistics, 2019. 23 p. DOI:10.15220/2019-56-en
16. Domingue B.W., Kanopka K., Stenhaus B. et al. Speed—Accuracy Trade-Off? Not So Fast: Marginal Changes in Speed Have Inconsistent Relationships with Accuracy in Real-World Settings. *Journal of Educational and Behavioral Statistics*, 2022. Vol. 47, no. 5, pp. 576—602. DOI:10.3102/10769986221099906
17. Teig N., Scherer R., Kjærnsli M. Identifying patterns of students' performance on simulated inquiry tasks using PISA 2015 log-file data. *Journal of Research in Science Teaching*, 2020. Vol. 57, no. 9, pp. 1400—1429. DOI:10.1002/tea.21657
18. Heinonen J., Aro T., Ahonen T., Poikkeus A.-M. Test-taking behaviors in a neurocognitive assessment: Associations with school-age outcomes in a Finnish longitudinal follow-up. *Psychological assessment*, 2011. Vol. 23, no. 1, pp. 184—192. DOI:10.1037/a0021291
19. Yu R., Li Q., Fischer C., Doroudi S., Xu D. Towards Accurate and Fair Prediction of College Success: Evaluating Different Sources of Student Data [Electronic resource]. In Rafferty A.N., Whitehill J., Romero C., Cavalli-Sforza V. (eds.), *Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference (July 10—13, 2020)*. Montreal: International educational data mining society, 2020, pp. 292—301. URL: <https://files.eric.ed.gov/fulltext/ED608066.pdf> (Accessed 26.02.2024).
20. Zumbo B.D., Hubley A.M. (eds.). Understanding and investigating response processes in validation research. Cham: Springer International Publishing, 2017. 383 p. DOI:10.1007/978-3-319-56129-5
21. Ercikan K., Pellegrino J.W. (eds.). Validation of score meaning for the next generation of assessments: The use of response processes. N.Y.; London: Taylor & Francis, 2017. 165 p.
22. Andrews-Todd J., Mislevy R.J., LaMar M., de Klerk S. Virtual performance-based assessments. In von Davier A.A., Mislevy R.J., Hao J. (eds.), *Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment: With Examples in R and Python*. Berlin: Springer, 2021, pp. 45—60.
23. Vuorikari R., Stefano K., Yves P. DigComp 2.2: The Digital Competence Framework for Citizens-With new examples of knowledge, skills and attitudes. Luxembourg: Publications Office of the European Union, 2022. 133 p. DOI:10.2760/115376
24. Wang J., Wang X. Structural equation modeling: Applications using Mplus. New Jersey: John Wiley & Sons, 2019. 536 p. DOI:10.1002/9781119422730
25. Wirth J. Computer-based tests: Alternatives for test and item design. In Hartig J., Klieme E., Leutner D. (eds.), *Assessment of competencies in educational contexts*. G ttingen: Hogrefe & Huber Publishers, 2008, pp. 235—252.
26. Yamamoto K., Lennon M.L. Understanding and detecting data fabrication in large-scale assessments. *Quality Assurance in Education*, 2018. Vol. 26, no. 2, pp. 196—212. DOI:10.1108/QAE-07-2017-0038
27. Zumbo B.D., Maddox B., Care N.M. Process and product in computer-based assessments: Clearing the ground for a holistic validity framework. *European Journal of Psychological Assessment*, 2023. Vol. 39, no. 4, pp. 252—262. DOI:10.1027/1015-5759/a000748

## Литература

1. Авдеева С., Тарасова К. Об оценке цифровой грамотности: методология, концептуальная модель и инструмент измерения // Вопросы образования. 2023. № 2. С. 8—32. DOI:10.17323/1814-9545-2023-2-8-32
2. Accurate assessment via process data / S. Zhang, Z. Wang, J. Qi, J. Liu, Z. Ying // Psychometrika. 2023. Vol. 88. № 1. P. 76—97. DOI:10.1007/s11336-022-09880-8
3. Bartolomé J., Garaizar P., Bastida L. Validating item response processes in digital competence assessment through eye-tracking techniques // In Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality. 2020. P. 738—746. DOI:10.1145/3434780.3436641

4. Bergner Y., von Davier A.A. Process data in NAEP: Past, present, and future // *Journal of Educational and Behavioral Statistics*. 2019. Vol. 44. № 6. P. 706—732. DOI:10.3102/1076998618784700
5. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning / J. Hamari, D.J. Shernoff, E. Rowe, B. Collier, J. Asbell-Clarke, T. Edwards // *Computers in human behavior*. 2016. Vol. 54. P. 170—179. DOI:10.1016/j.chb.2015.07.045
6. Data literacy assessments: A systematic literature review / Y. Cui, F. Chen, A. Lutsyk, J.P. Leighton, M. Cutumisu // *Assessment in Education: Principles, Policy & Practice*. 2023. Vol. 30. № 1. P. 76—96. DOI: 10.1080/0969594x.2023.2182737
7. De Boeck P., Scalise K. Collaborative problem solving: Processing actions, time, and performance // *Frontiers in psychology*. 2019. Vol. 10. Article ID 1280. 9 p. DOI:10.3389/fpsyg.2019.01280
8. Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining / R.J. Mislevy, J.T. Behrens, K.E. Dicerbo, R. Levy // *Journal of educational data mining*. 2012. Vol. 4. № 1. P. 11—48. DOI:10.5281/zenodo.3554641
9. Game-Based Assessment of Students' Digital Literacy Using Evidence-Centered Game Design / J. Li, J. Bai, S. Zhu, H. H. Yang // *Electronics*. 2024. Vol. 13(2). Article ID 385. 19 p. DOI: 10.3390/electronics13020385
10. Hu L., Bentler P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives // *Structural equation modeling: a multidisciplinary journal*. 1999. Vol. 6. № 1. P. 1—55. DOI:10.1080/10705519909540118
11. Nichols S.L., Dawson H.S. Assessment as a context for student engagement // *Handbook of research on student engagement* / Eds. S.L. Christenson, A.L. Reschly, C. Wylie. Boston: Springer Science+Business Media, 2012. P. 457—477. DOI:10.1007/978-1-4614-2018-7\_22
12. Oliveri M.E., Mislevy R.J. Introduction to “Challenges and opportunities in the design of ‘next-generation assessments of 21st century skills’” special issue // *International Journal of Testing*. 2019. Vol. 19. № 2. P. 97—102. DOI:10.1080/15305058.2019.1608551
13. Peng D., Yu Z. A literature review of digital literacy over two decades // *Education Research International*. 2022. Vol. 2022. Article ID 2533413. 8 p. DOI:10.1155/2022/2533413
14. Recommendation of the Council on Children in the Digital Environment [Электронный ресурс] / OECD. Paris: OECD, 2022. 14 p. URL: <https://legalinstruments.oecd.org/public/doc/272/272.en.pdf> (дата обращения: 26.02.2024).
15. Recommendations on assessment tools for monitoring digital literacy within unesco's digital literacy global framework / M. Laanpere, UNESCO, UNESCO Institute for Statistics. Montreal: UNESCO Institute for Statistics, 2019. 23 p. DOI:10.15220/2019-56-en
16. Speed—Accuracy Trade-Off? Not So Fast: Marginal Changes in Speed Have Inconsistent Relationships with Accuracy in Real-World Settings / B.W. Domingue, K. Kanopka, B. Stenhaug [et al.] // *Journal of Educational and Behavioral Statistics*. 2022. Vol. 47. № 5. P. 576—602. DOI:10.3102/10769986221099906
17. Teig N., Scherer R., Kjærnsli M. Identifying patterns of students' performance on simulated inquiry tasks using PISA 2015 log-file data // *Journal of Research in Science Teaching*. 2020. Vol. 57. № 9. P. 1400—1429. DOI:10.1002/tea.21657
18. Test-taking behaviors in a neurocognitive assessment: Associations with school-age outcomes in a Finnish longitudinal follow-up / J. Heinonen, T. Aro, T. Ahonen, A.-M. Poikkeus // *Psychological assessment*. 2011. Vol. 23. № 1. P. 184—192. DOI:10.1037/a0021291
19. Towards Accurate and Fair Prediction of College Success: Evaluating Different Sources of Student Data [Электронный ресурс] / R. Yu, Q. Li, C. Fischer, S. Doroudi, D. Xu // *Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference (July 10—13, 2020)* / Eds. A.N. Rafferty, J. Whitehill, C. Romero, V. Cavalli-Sforza. Montreal: International educational data mining society, 2020. P. 292—301. URL: <https://files.eric.ed.gov/fulltext/ED608066.pdf> (дата обращения: 26.02.2024).
20. Understanding and investigating response processes in validation research / Eds. B.D. Zumbo, A.M. Hubley. Cham: Springer International Publishing, 2017. 383 p. DOI:10.1007/978-3-319-56129-5
21. Validation of score meaning for the next generation of assessments: The use of response processes / Eds. K. Ercikan, J.W. Pellegrino. N.Y.; London: Taylor & Francis, 2017. 165 p.
22. Virtual performance-based assessments / J. Andrews-Todd, R.J. Mislevy, M. LaMar, S. de Klerk // *Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment: With Examples in R and Python* // Eds. A.A. von Davier, R.J. Mislevy, J. Hao. Berlin: Springer, 2021. P. 45—60.
23. Vuorikari R., Stefano K., Yves P. DigComp 2.2: The Digital Competence Framework for Citizens-With new examples of knowledge, skills and attitudes. Luxembourg: Publications Office of the European Union, 2022. 133 p. DOI:10.2760/115376
24. Wang J., Wang X. Structural equation modeling: Applications using Mplus. New Jersey: John Wiley & Sons, 2019. 536 p. DOI:10.1002/9781119422730
25. Wirth J. Computer-based tests: Alternatives for test and item design // *Assessment of competencies in educational contexts* / Eds. J. Hartig, E. Klieme, D. Leutner. G ttingen: Hogrefe & Huber Publishers, 2008. P. 235—252.
26. Yamamoto K., Lennon M.L. Understanding and detecting data fabrication in large-scale assessments // *Quality Assurance in Education*. 2018. Vol. 26. № 2. P. 196—212. DOI:10.1108/QAE-07-2017-0038

27. Zumbo B.D., Maddox B., Care N.M. Process and product in computer-based assessments: Clearing the ground for a holistic validity framework // *European Journal of Psychological Assessment*. 2023. Vol. 39. № 4. P. 252—262. DOI: 10.1027/1015-5759/a000748

### **Information about the authors**

*Irina O. Tkachenko*, Master student, National Research University Higher School of Economics, Moscow, Russia; Junior Researcher, Sirius University of Science and Technology, Sirius, Krasnodar region, Russia, ORCID: <https://orcid.org/0000-0002-0406-7400>, e-mail: [iotkachenko@edu.hse.ru](mailto:iotkachenko@edu.hse.ru)

*Ksenia V. Tarasova*, PhD in Education, Director, Centre for Psychometrics and Measurement in Education, National Research University Higher School of Economics, Moscow, Russia, ORCID: <https://orcid.org/0000-0002-3915-3165>, e-mail: [ktarasova@hse.ru](mailto:ktarasova@hse.ru)

*Daria A. Gracheva*, Junior Research Fellow, Laboratory for New Construct Measurement and Test Design, Centre for Psychometrics and Measurement in Education, National Research University Higher School of Economics, Moscow, Russia, ORCID: <https://orcid.org/0000-0002-4646-7349>, e-mail: [dgracheva@hse.ru](mailto:dgracheva@hse.ru)

### **Информация об авторах**

*Ткаченко Ирина Олеговна*, студент магистратуры, Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»), Москва; младший научный сотрудник, Научный центр когнитивных исследований, Научно-технологический университет «Сириус» (АНОО ВО «Университет “Сириус”»), пгт. Сириус, Российская Федерация, ORCID: <https://orcid.org/0000-0002-0406-7400>, e-mail: [iotkachenko@edu.hse.ru](mailto:iotkachenko@edu.hse.ru)

*Тарасова Ксения Вадимовна*, кандидат педагогических наук, директор Центра психометрики и измерений в образовании, Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0002-3915-3165>, e-mail: [ktarasova@hse.ru](mailto:ktarasova@hse.ru)

*Грачева Дарья Александровна*, младший научный сотрудник лаборатории измерения новых конструктов и дизайна тестов Центра психометрики и измерений в образовании, Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0002-4646-7349>, e-mail: [dgracheva@hse.ru](mailto:dgracheva@hse.ru)

Получена 31.01.2024

Принята в печать 11.03.2024

Received 31.01.2024

Accepted 11.03.2024