

Отечественная и зарубежная педагогика. 2022. Т. 1, № 3. С. 83–97.  
Domestic and foreign pedagogy. 2022. Vol. 1, no. 3. P. 83–97.

Научная статья  
УДК 37.02  
doi: 10.24412/2224–0772–2022–84–83–97.

## ПОДХОДЫ К РАЗРАБОТКЕ ВАРИАНТОВ ЗАДАНИЙ СЦЕНАРНОГО ТИПА В РАМКАХ МЕТОДА ДОКАЗАТЕЛЬНОЙ АРГУМЕНТАЦИИ



Д. А. Грачева

Дарья Александровна Грачева<sup>1</sup>, Ксения Вадимовна  
Тарасова<sup>2</sup>

<sup>1, 2</sup> Национальный исследовательский университет  
«Высшая школа экономики», Москва, Россия

<sup>1</sup> dgracheva@hse.ru

<sup>2</sup> ktarasova@hse.ru

**Аннотация.** *Введение.* В статье рассматриваются подходы к созданию заданий сценарного типа, разработанных в логике доказательной аргументации (Evidence-Centered Design (ECD)), которые предлагают цифровую тестовую среду для оценки критического мышления. Создание нескольких вариантов таких заданий является трудоемким процессом. Одно из решений этой проблемы предполагает определение подходов разработки, которые позволят упростить процесс создания вариантов. Цель статьи — описать и сравнить подходы к разработке вариантов заданий сценарного типа в рамках метода доказательной аргументации.

*Методология и методы исследования.* Создание вариантов происходило по разным направлениям подхода клонирования, которое заключалось в выявлении обязательных и вариативных элементов заданий. Первое направление фокусировалось на проработке контента. Второе направление дополнительно предполагало соблюдение синтаксических



К. В. Тарасова

структур предложений в стимульном материале. Связи между вариантами сценариев оценивались в рамках методологии конфирматорного факторного анализа. Выборка исследования составила 389 учащихся начальной школы, которые проходили сценарии во всех вариантах.

*Результаты исследования.* Анализ показал, что сценарии, реализованные в разных проблемных ситуациях, не оценивают способности к критическому мышлению одинаковым образом. Связи между вариантами сценариев, разработанных по разным направлениям подхода клонирования, значимо не различаются.

*Заключение.* Оптимальным подходом к разработке вариантов сценарных заданий является подход клонирования, который учитывает комплексную структуру конструкта и воспроизводит насыщенную цифровую среду. Результаты исследования позволяют утверждать, что при создании вариантов заданий сценарного типа, разработанных в рамках ECD, могут использоваться оба обозначенных направления подхода клонирования.

**Ключевые слова:** задания сценарного типа, варианты теста, метод доказательной аргументации, Evidence-Centered Design (ECD), разработка заданий, комплексные конструкты, навыки XXI века

**Для цитирования:** Грачева Д. А., Тарасова К. В. Подходы к разработке вариантов заданий сценарного типа в рамках метода доказательной аргументации // Отечественная и зарубежная педагогика. 2022. Т. 1, № 3 (84). С. 83–97. doi: 10.24412/2224-0772-2022-84-83-97.

Original article

### Approaches to the Development of Scenario-Based Task Forms Within the Framework of Evidence-Centered Design

Daria A. Gracheva<sup>1</sup>, Ksenia V. Tarasova<sup>2</sup>

<sup>1,2</sup> National Research University Higher School of Economics, Moscow, Russia

<sup>1</sup> dgracheva@hse.ru

<sup>2</sup> ktarasova@hse.ru

**Abstract. Introduction.** The article discusses approaches to the development of scenario-based tasks aimed at measuring critical thinking within the Evidence-Centered Design (ECD). Creating several forms of the tasks is a time-consuming process. One of the solutions to this problem involves defining development approaches that will simplify the process of task development. The purpose of the article is to describe and compare approaches to the development of scenario-based tasks within the framework of Evidence-Centered Design.

**Research Methods.** To create task forms we use the cloning approach which is based on the determination of radicals and incidentals. Creating the form

of the first task we mostly pay attention to task content which adequately represents main features of the scenario. For the form of the second task grammatical structure of sentences is considered as radicals. To compare the task forms, we used the Confirmatory Factor Analysis approach. In total, data from 389 primary school students were used in the analysis.

*Results.* The analysis showed that scenarios with different contexts do not measure critical thinking in the same way. The correlations between the scenario forms developed under different test development approaches do not significantly differ.

*Conclusions.* The optimal approach to the development of scenario-based tasks is the cloning approach, which considers the complex structure of the construct and digital environment. The results of the study show that both variations of cloning approaches could be used to develop scenario-based tasks within the framework of Evidence-Centered Design.

**Keywords:** scenario-based tasks, task forms, Evidence-Centered Design, task development, complex constructs, 21st century skills

**For citation:** Gracheva D. A., Tarasova K. V. Approaches to the development of scenario-based task forms within the framework of evidence-centered design *Domestic and Foreign Pedagogy*. 2022;1(3):83–97. (In Russ.). <https://doi.org/10.24412/2224-0772-2022-84-83-97>.

## Введение

В настоящий момент компьютерное тестирование представляет собой развитую индустрию, охватывающую большое разнообразие тестов в различных областях. При этом измерения отходят от использования традиционных типов заданий с выбором варианта ответа. Большое внимание уделяется разработкам с использованием современных технологий. Наряду с тем, что одни представляют собой только переход из бумажной формы в компьютерную, другие объединены общей чертой — тестируемые могут продемонстрировать уровень владения навыком через наблюдаемое поведение в заранее установленной ситуации [19]. Ввиду развития технологий, позволяющих моделировать разнообразную цифровую тестовую среду, появился термин Virtual performance-based assessment (VPBA) [3].

В широком смысле суть VPBA заключается в создании среды для испытуемых, в которой они могут взаимодействовать с моделируемыми ситуациями таким образом, чтобы разработчики теста могли получить доказательства способности тестируемых действовать в таких ситуациях. К VPBA относятся задания сценарного типа, которые фокусируются как на самом решении проблемы, так и на процессе решения, и представляют

собой среду, в которой тестируемые демонстрируют уровень владения навыками в ответ на проблему, часто приближенную к реальной жизни. Это позволяет говорить о заданиях сценарного типа как о заданиях со сложной структурой, предполагающих нелинейное продвижение по тесту и взаимодействие с интерактивной средой. Сегодня сценарные задания успешно реализуются в цифровой среде с использованием игровых элементов и симуляций, например для оценки критического мышления [5; 24], навыка решения проблем в рамках международного исследования PISA [13].

В сочетании с методологией разработки инструментов с фокусом на наблюдаемое поведение тестируемого — методом доказательной аргументации (Evidence-Centered Design, ECD) — сценарные задания создают возможности для оценивания навыков более высокого порядка для диагностики текущего состояния или отслеживания изменений [20; 23]. Однако при повторном тестировании одними заданиями участники могут показать более высокие результаты за счет не только прироста способностей, но и эффекта практики. Кроме того, многократное использование заданий ведет к снижению мотивации к выполнению: тестируемые слабо вовлечены в предлагаемую проблемную ситуацию, потому что уже с ней знакомы. Таким образом, возникает необходимость создания вариантов заданий.

По сравнению с традиционными тестами разработка заданий сценарного типа, обладающих сложной структурой и направленных на измерение комплексных конструкторов, является трудоемким процессом. В таком случае возможность разработки нескольких вариантов сценариев в сжатые сроки ограничена. Одно из решений этой проблемы предполагает определение подходов разработки, которые позволят ускорить и упростить создание вариантов сценарных заданий. Отмечается, что существующие руководства для написания заданий закрытого типа в этом случае часто оказываются бесполезными [9], а подход ECD недостаточно изучен в приложении к созданию вариантов таких заданий с разным контекстом и содержанием.

**Цель статьи** — описать и сравнить подходы к разработке вариантов заданий сценарного типа в рамках метода доказательной аргументации.

В первой части статьи описываются общие подходы к разработке вариантов теста. Во второй части статьи на данных тестирования представлены результаты разработки и сравнения вариантов сценарных заданий для измерения критического мышления в начальной школе.

### Общие подходы к разработке вариантов теста

Один из первых подходов к разработке заданий для вариантов теста основывался на практическом опыте экспертов и тестовой спецификации. Критика подхода указывала на низкую эффективность работы экспертов — созданные задания часто не проходят психометрическую проверку качества [11].

Другой подход к разработке вариантов заключался в отборе заданий из банка заданий с использованием статистических методов. Этот подход основывается на случайном распределении заданий по вариантам теста, который в числе прочего может осуществляться при контроле психометрических характеристик [15].

Как альтернатива субъективному и статистическому подходу широкое развитие получила идея автоматической генерации заданий на основе форм заданий [4; 21]. Для этого разрабатывалось задание («задание-родитель»), которое затем представлялось в виде формы с пропущенными элементами. Заменяя пропущенные элементы (например, числовые данные, содержательные понятия и др.), возможно быстро создать варианты задания с одной синтаксической структурой [21].

Позже было предложено классифицировать элементы на обязательные (radicals) и вариативные (incidentals) [14]. Для создания взаимозаменяемых вариантов только вариативные элементы подлежат изменению, а обязательные элементы, определяющие трудность заданий, остаются неизменными [4]. Такие задания называют изоморфными или клонами, а процесс разработки известен как подход клонирования.

Применение субъективного и статистического подходов ограничено в случаях заданий со сложной структурой [7]. Создание банка сценарных заданий нецелесообразно, а тестовая спецификация не может с точностью воспроизвести особенности свободной тестовой среды или природу комплексного конструкта, что неизбежно приводит к потере качества теста и появлению различий в вариантах.

Для заданий со сложной структурой рекомендуют использовать подход клонирования, при этом процесс разработки может быть индивидуализирован в зависимости от инструмента [7; 8]. В частности, разные направления подхода клонирования определяются элементами теста, которые являются обязательными или вариативными. Кроме того, технология автоматической генерации для таких заданий затруднена, поэтому процесс разработки может включать привлечение экспертов.

Клонирование использовалось при создании симуляций в области сетевых технологий [10], медицинских ситуационных тестов [18], компьютерных тестов для менеджеров [17].

Принимая во внимание предыдущий опыт создания вариантов заданий со сложной структурой, далее мы представим результаты разработки вариантов заданий сценарного типа и сравнение подходов к их клонированию.

### Методология и методы исследования

#### *Описание инструмента*

В исследовании используются сценарные задания («Аквариум» и «Динозавр») инструмента «4К», созданного в рамках метода доказательной аргументации для измерения анализа информации как части критического мышления [1].

Каждый элемент сценарного задания проектируется как свидетельство (индикатор) компетенций. При выполнении заданий учащиеся проявляют способность выделять релевантную информацию, выбирать надежные источники. Оба сценария демонстрируют хорошее психометрическое качество и были валидизированы для учащихся младшей школы [2].

Одна из особенностей таких заданий состоит в том, что индикаторы связаны общим контекстом и контентом. **Контекст** определяет основную проблемную ситуацию задания и развитие ситуации (сюжета) — последовательность действий, отношения между этапами задания, персонажами и пр. **Контент** определяет содержательное наполнение сценария.

В основе сценария «Аквариум» лежит ситуация, где тестируемому необходимо собрать информацию, которая поможет обустроить среду обитания для питомца, а в сценарии «Динозавр» тестируемому необходимо собрать информацию для доклада о редком животном.

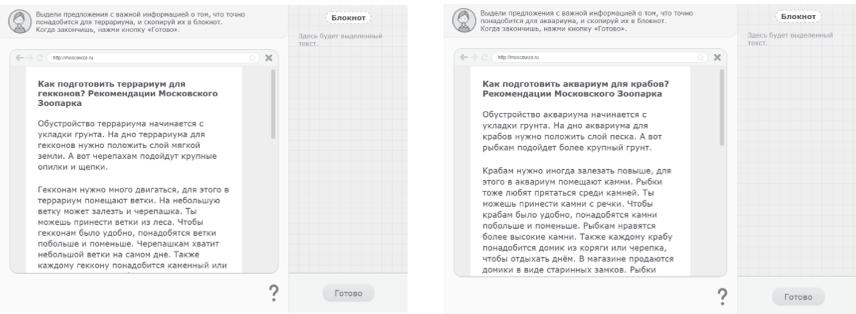
Другая особенность — использование компьютерных симуляций. Это делает задание более реалистичным, позволяет использовать технические средства для оценки комплексных навыков. В сценариях используется симулятор поисковой системы, где тестируемый выбирает запрос в поисковой строке браузера, источник информации и изучает информацию по заданной теме.

*Разработка вариантов заданий сценарного типа*

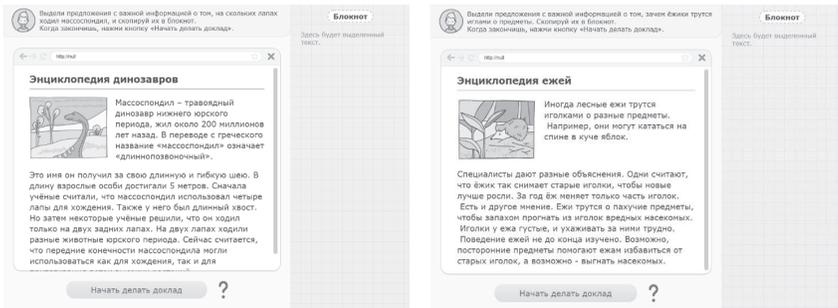
При создании вариантов для сценариев «Аквариум» и «Динозавр» мы ориентировались на подход клонирования. Для этого были определены наборы обязательных и вариативных элементов.

Первый подход клонирования (задания «Динозавр» — «Еж») предполагает наличие следующих обязательных элементов: контекст, свидетельства проявления критического мышления, механики, симуляторы, последовательность действий внутри сценария.

При втором подходе клонирования (задания «Аквариум» — «Террариум») использовались все перечисленные выше обязательные элементы и дополнительно соблюдалась эквивалентность синтаксических структур предложений в стимульном материале. Второй подход клонирования за счет дополнительных ограничений мы считаем более строгим направлением в подходе клонирования. Примеры экранов заданий приведены на Рисунках 1–4.



*Рисунок 1–2. Пример экрана задания «Аквариум» (слева) и «Террариум» (справа)*



*Рисунок 3–4. Пример экрана задания «Динозавр» (слева) и «Еж» (справа)*

В качестве основного вариативного элемента для всех сценариев используется содержание сценария (контент). Иными словами, мы меняем объекты ситуации, которые предложены в сценарии. Кроме того, вариативность вариантов сценариев достигается заменой иллюстративных материалов, чтобы задания не казались тестируемым одинаковыми (например, Рисунок 3–4).

Таким образом, первый подход клонирования включает следующие этапы:

1. Подбор нового контента сценария при соблюдении проблемной ситуации и сюжета, симуляций, механик, свидетельств конструкта.

*Пример: задания «Аквариум» — «Террариум»*

В оригинальном варианте сценария «Аквариум» перед тестируемым ставится задача обустроить **аквариум для крабов**. Для второго варианта сценария требовалось подобрать новый контент для той же ситуации, который будет максимально похож на оригинал и понятен целевой аудитории (учащиеся четвертых классов). Так, во втором варианте («Террариум») тестируемому необходимо собрать информацию, которая поможет обустроить **террариум для гекконов**.

*Пример: задания «Динозавр» — «Еж»*

В оригинальном варианте сценария «Динозавр» тестируемому предлагается проанализировать текст о **динозавре**, где представлены три точки зрения относительно проблемы (на скольких лапах ходил динозавр). Дополнительная сложность в подборе нового контента заключалась в том, что информация о животном должна быть реальной, а не выдуманной разработчиками. С учетом характеристик оригинального контента для второго варианта был подобран новый контент: редкое животное (**еж**) и проблема, которая рассматривается с трех позиций (зачем ежи трутся иголками о предметы).

2. Изменение дизайна экранов (замена иллюстративных материалов, цветов) — по возможности.

Второй, более строгий, подход клонирования дополнительно включает третий этап:

3. Создание форм предложений для обеспечения эквивалентных структур текстов в стимульном материале.

*Пример: задания «Аквариум» — «Террариум»*

В ходе разработки на основе оригинального сценария «Аквариум» для

каждого предложения создавались формы предложений, где пропущенные элементы заполнялись с учетом нового контента. Например, форма предложения из сценария «Аквариум» имеет вид: *«на дно (аквариума для крабов) нужно положить (слой песка)»*. Предложения для второго варианта сценария имеют следующий вид: *«на дно (террариума для гекконов) нужно положить (слой мягкой земли)»*.

**Гипотеза исследования** заключается в том, что варианты, созданные по более строгому направлению подхода клонирования, будут связаны сильнее, чем другая пара вариантов, где использовался подход с меньшими ограничениями. Также мы ожидаем, что варианты сценариев с единым контекстом демонстрируют большую согласованность, чем сценарии, реализованные в разных контекстах (например, «Аквариум» и «Динозавр»).

Для проверки гипотез было проведено исследование, в котором приняли участие 463 учащихся четвертых классов школ российских регионов. Тестирование проходило в школах в присутствии учителя и состояло из двух частей. В первой части учащиеся проходили задания «Аквариум» и «Террариум», во второй части — «Динозавр» и «Еж». Порядок предъявления одного из вариантов в каждой паре сценариев определялся случайным образом. Прохождение четырех сценариев занимало в среднем 35–40 минут.

Ввиду того, что тестируемые проходили варианты сценариев подряд с небольшим перерывом, мы исключили из анализа профили, где время тестирования хотя бы одного из сценариев было экстремально низким (не позволяло прочитать текст задания и дать вдумчивый ответ). Итоговая выборка исследования составила 381 учащийся.

### *Стратегия анализа*

В данной статье мы рассматриваем четыре задания сценарного типа как четыре способа измерения двух способностей к анализу (способность выбирать надежный источник информации и выделять релевантную информацию). В сценариях надежность оценивается через выбор наиболее надежной ссылки из предложенных поисковой системой, а релевантность оценивается через выделение предложений, отражающих полезную информацию относительно задачи. Таким образом, четыре сценария содержат всего 24 дихотомических индикатора релевантности и 6 дихотомических индикаторов надежности.

## Подходы к разработке вариантов заданий сценарного типа ... I

Исследование связей между разными способами измерения позволит оценить степень согласованности сценариев между собой и проверить гипотезы исследования. Для этой цели была построена бифакторная модель в рамках конфирматорного факторного анализа (КФА), где факторы способностей и способов измерения (сценариев) ортогональны. Каждый индикатор сценария относится и к фактору способности, и к фактору способа измерения (обозначено как  $Y_{ij}$ , где  $i$  — фактор способа измерения,  $j$  — фактор способности). Модель представлена на Рисунке 5.

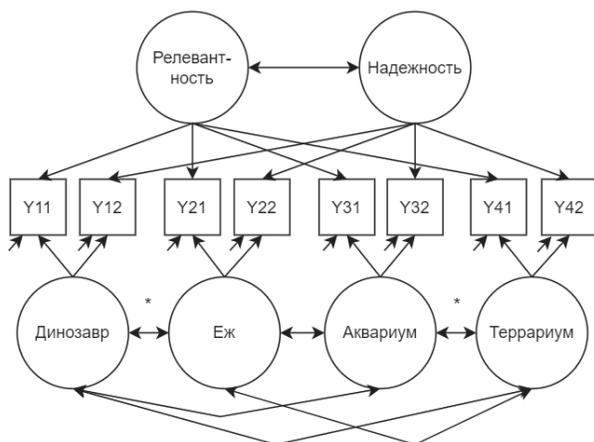


Рисунок 5. Бифакторная модель КФА

Примечание. В верхней части: Релевантность (1) и Надежность (2) — факторы способностей. В нижней части: «Динозавр» (1), «Еж» (2), «Аквариум» (3), «Террариум» (4) — факторы сценариев. \* — корреляции между вариантами сценариев.

Качество модели определялось по следующим критериям соответствия модели и данных: CFI (TLI) > 0.95, RMSEA < 0.05 [22]. Для оценки параметров использовался метод взвешенных квадратов (WLSMV). Для сравнения двух подходов создания вариантов заданий будет проверена гипотеза о равенстве корреляций между факторами вариантов одного сценария (отмечены \* на Рисунке 5). Анализ проводился в R.

### Результаты исследования и обсуждение

Построенная модель хорошо согласуется с данными ( $\chi^2(df) = 290.05$

(372), CFI = 0.979, TLI = 0.976, RMSEA = 0.027). Чтобы оценить степень согласованности сценариев, рассмотрим связи между факторами сценариев. Незначимые корреляции между сценариями с разным контекстом (от 0.1 до 0.2 при  $p > 0.05$ ) подтверждают второе положение гипотезы исследования — сценарии с разным контекстом не оценивают способности к анализу информации одинаковым образом. Их совместное использование в тестировании позволяет более полно и всесторонне оценить навык анализа информации.

Варианты сценариев демонстрируют статистически значимую связь на уровне 0.05 со средним размером эффекта: «Аквариум» и «Террариум» (0.513), «Динозавр» и «Еж» (0.575). Между корреляциями не обнаружено статистически значимых различий ( $z = -1.2$ ,  $p = 0.11$ ). Гипотеза, предполагающая более сильную связь между вариантами сценариев с единичными синтаксическими структурами предложений, не подтвердилась.

Исследование разных направлений клонирования заданий проводилось на примере ситуационного теста в бланковой форме для другого конструкта [18], поэтому сравнение полученных результатов с предыдущим опытом не является корректным. Нам неизвестно об исследованиях, которые сравнивали разные подходы разработки на примере заданий сценарного типа в компьютерной форме. Варианты сценарных заданий используются для оценки критического мышления и креативности, однако в статьях опускаются подробности, описывающие подход к разработке [16; 24].

Для простых форматов заданий подход клонирования ассоциируется с упрощенным подходом к разработке, провоцирует «натаскивание». Несмотря на то, что сложная структура сценарных заданий снижает вероятность научения, воспроизведение большинства элементов сценария приведет к созданию похожих заданий, которые могут восприниматься тестируемыми как идентичные. При этом отмечается, что важным аспектом тестирования навыков высокого порядка является применение навыков в разных проблемных ситуациях [24]. С другой стороны, создание вариантов на основе уже валидизированного оригинального задания позволит получить задания с высокой степенью согласованности результатов и в целом обеспечит качество разрабатываемого инструмента в рамках методологии ECD.

Создание взаимозаменяемых вариантов заданий сценарного типа остается нетривиальной задачей. Понимание того, какие элементы за-

дания относятся к обязательным и вариативным, позволит не только упростить процесс разработки, но и избежать некорректных результатов. Исследования влияния элементов заданий на результаты тестирования проводились для разных конструкторов и форм заданий: письменные задания по английскому языку [12], тест читательской грамотности [6], что делает возможным в перспективе перенести это направление исследования и на задания сценарного типа.

Тем не менее результаты тестирования такими заданиями подвержены влиянию контекста или контента задания, которые сложно оценить [3], поэтому исследования в области разработки вариантов сценариев не должны ограничиваться только разбиением задания на элементы. Поиск подхода к разработке заданий сценарного типа призван найти баланс между сопоставимостью результатов, согласованностью заданий и их единообразностью.

Результаты исследования следует принимать с учетом ограничений. В работе подходы клонирования были опробованы на разных сценариях (по количеству индикаторов, длине, контексту и пр.), поэтому результаты могут быть ревалидизированы на нескольких вариантах одного задания сценарного типа. Кроме того, варианты заданий предъявлялись тестируемому сразу, без перерыва, поэтому мы не исключаем влияние научения на результаты.

### *Заключение*

Целью статьи было представление читателю подходов к разработке вариантов заданий сценарного типа в рамках метода доказательной аргументации (ECD). Важной особенностью сценарных заданий является возможность оценить сформированность навыков через наблюдаемое поведение в заданной ситуации. Однако процесс разработки таких заданий трудоемкий. Создание вариантов позволяет не только разнообразить содержание проблемных ситуаций, но и снизить затраты по сравнению с разработкой новых заданий, что особенно актуально для заданий с насыщенной тестовой средой: симуляциями, игровыми элементами.

В данном исследовании мы фокусировались на анализе существующих подходов к созданию вариантов тестовых заданий и последующем обосновании выбора оптимального — подхода клонирования — на примере создания вариантов двух заданий сценарного типа для оценки критического мышления в начальной школе. Реализация подхода кло-

нирования опиралась на идею выявления обязательных и вариативных элементов задания. В более строгом направлении клонирования заданий реализация подхода предполагала изменение содержания сценария (контента) при соблюдении синтаксической структуры предложений в обоих вариантах. Другое направление не накладывало требований о синтаксической эквивалентности стимульных материалов задания и опиралось на проработку контента. Сравнительный анализ показал, что варианты сценариев в рамках одной проблемной ситуации более согласованы между собой, чем сценарии, реализованные в разных контекстах. При этом гипотеза, предполагающая, что более строгое направление клонирования позволит создать максимально похожие варианты сценариев, не подтвердилась.

Результаты исследования позволяют утверждать, что при создании вариантов заданий сценарного типа, разработанных в рамках подхода ECD, могут использоваться оба обозначенных направления клонирования. Кроме того, строгое воспроизведение элементов сценария не приводит к большей согласованности вариантов. Это дает возможность разработчикам инструментов измерения комплексных навыков использовать различное содержательное наполнение сценариев, тем самым снижая их узнаваемость и вероятность научения.

#### Список источников

1. Уланова И. Л., Брун И. В., Васин Г. М. Методология Evidence-Centered Design для измерения комплексных психологических конструктов // Современная зарубежная психология. 2018. Т. 7, № 3. С. 18–27.
2. Уланова И. Л., Орел Е. А., Брун И. В. Измерение креативности и критического мышления в начальной школе // Психологический журнал. 2020. Т. 41, № 6. С. 96–107.
3. Andrews-Todd J., Mislevy R. J., LaMar M., et al. Virtual performance-based assessments // Computational psychometrics: New methodologies for a new generation of digital learning and assessment. Springer, Cham, 2021. P. 45–60.
4. Bejar I. I. A generative approach to psychological and educational measurement // ETS Research Report Series. 1991. No. 1. P. 1–54.
5. Braun H. I., Shavelson R. J., Zlatkin-Troitschanskaia O., et al. Performance assessment of critical thinking: Conceptualization, design, and implementation // Frontiers in Education. 2020. No. 5. P. 1–10.
6. Buerger S., Kroehne U., Koehler C., et al. What makes the difference? The impact of item properties on mode effects in reading assessments // Studies in Educational Evaluation. 2019. No. 62. P. 1–9.
7. Clause C. S., Mullins M. E., Nee M. T., et al. Parallel test form development: A procedure for alternate predictors and an example // Personnel Psychology. 1998. Vol. 51, No. 1. P. 193–208.
8. Davey T., Ferrara, S., Shavelson R., et al. Psychometric considerations for the next generation of performance assessment. Washington, DC: Center for K-12 Assessment & Performance Management, Educational Testing Service, 2015. P. 1–100.
9. Draaijer S., Hartog R. J. M. Design patterns for digital item types in higher education // E-Journal of instructional science and technology. 2007. Vol. 10, No. 1. P. 1–32.
10. Fay D. M., Levy R., Mehta V. Investigating psychometric isomorphism for traditional and perfor-

mance-based assessment // *Journal of Educational Measurement*. 2018. Vol. 55, No. 1. P. 52–77.

11. *Gierl M.J., Haladyna T.M.* Automatic item generation: Theory and practice. Routledge, 2012. P. 256
12. *Homayounzadeh M., Saadat M., Ahmadi A.* Investigating the effect of source characteristics on task comparability in integrated writing tasks // *Assessing Writing*. 2019. No. 41. P. 25–46.
13. *Stadler M., Herborn K., Mustafić M., et al.* The assessment of collaborative problem solving in PISA 2015: An investigation of the validity of the PISA 2015 CPS tasks // *Computers & Education*. 2020. No. 157. P. 1–11.
14. *Irvine S.H., Dann P.L., Anderson J.D.* Towards a theory of algorithm-determined cognitive test construction // *British Journal of Psychology*. 1990. Vol. 81, No. 2. P. 173–195.
15. *Irvine S.H., Kyllonen P.C.* Item generation for test development. Routledge, 2013. 444 p.
16. *Jankowska D.M., Karwowski M.* Measuring creative imagery abilities // *Frontiers in psychology*. 2015. No. 6. P. 1–17.
17. *Lievens F., Anseel F.* Creating alternate in-basket forms through cloning: Some preliminary results // *International Journal of Selection and Assessment*. 2007. Vol. 15, No. 4. P. 428–433.
18. *Lievens F., Sackett P.R.* Situational judgment tests in high-stakes settings: Issues and strategies with generating alternate forms // *Journal of Applied Psychology*. 2007. Vol. 92, No. 4. P. 1043–1055.
19. *Messick S.* The interplay of evidence and consequences in the validation of performance assessments // *Educational researcher*. 1994. Vol. 23, No. 2. P. 13–23.
20. *Mislevy R.J., Almond R.G., Lukas J.F.* A brief introduction to evidence-centered design // *ETS Research Report Series*. 2003. No. 1. P. 1–29.
21. *Osburn H.G.* Item sampling for achievement testing // *Educational and Psychological Measurement*. 1968. Vol. 28, No. 1. P. 95–104.
22. *Roos J.M., Bauldry S.* Confirmatory factor analysis. SAGE Publications, 2021. P. 144.
23. *Seelow D.* The art of assessment: Using game-based assessments to disrupt, innovate, reform and transform testing // *Journal of Applied Testing Technology*. 2019. Vol. 20, No. 1. P. 1–16.
24. *Wang D., Liu H., Hau K.T.* Automated and interactive game-based assessment of critical thinking // *Education and Information Technologies*. 2021. P. 1–23. DOI: <https://doi.org/10.1007/s10639-021-10777-9>.

### References

1. *Uglanova I.L., Brun I.V., Vasin G.M.* Metodologiya Evidence-Centered Design dlya izmereniya kompleksnyh psihologicheskikh konstruktov // *Sovremennaya zarubezhnaya psihologiya*. 2018. T. 7, № 3. S. 18–27. [In Rus].
2. *Uglanova I.L., Orel E.A., Brun I.V.* Izmerenie kreativnosti i kriticheskogo myshleniya v nachal'noj shkole // *Psihologicheskij zhurnal*. 2020. T. 41, № 6. S. 96–107. [In Rus].
3. *Andrews-Todd J., Mislevy R.J., LaMar M., et al.* Virtual performance-based assessments // *Computational psychometrics: New methodologies for a new generation of digital learning and assessment*. Springer, Cham, 2021. P. 45–60.
4. *Bejar I.I.* A generative approach to psychological and educational measurement // *ETS Research Report Series*. 1991. No. 1. P. 1–54.
5. *Braun H.I., Shavelson R.J., Zlatkin-Troitschanskaia O., et al.* Performance assessment of critical thinking: Conceptualization, design, and implementation // *Frontiers in Education*. 2020. No. 5. P. 1–10.
6. *Buerger S., Kroehne U., Koehler C., et al.* What makes the difference? The impact of item properties on mode effects in reading assessments // *Studies in Educational Evaluation*. 2019. No. 62. P. 1–9.
7. *Clause C.S., Mullins M.E., Nee M.T., et al.* Parallel test form development: A procedure for alternate predictors and an example // *Personnel Psychology*. 1998. Vol. 51, No. 1. P. 193–208.
8. *Davey T., Ferrara, S., Shavelson R., et al.* Psychometric considerations for the next generation of performance assessment. Washington, DC: Center for K-12 Assessment & Performance Management, Educational Testing Service, 2015. P. 1–100.
9. *Draaijer S., Hartog R.J.M.* Design patterns for digital item types in higher education // *E-Journal of instructional science and technology*. 2007. Vol. 10, No. 1. P. 1–32.
10. *Fay D.M., Levy R., Mehta V.* Investigating psychometric isomorphism for traditional and

performance-based assessment // *Journal of Educational Measurement*. 2018. Vol. 55, No. 1. P. 52–77.

11. *Gierl M. J., Haladyna T. M.* Automatic item generation: Theory and practice. Routledge, 2012. P. 256

12. *Homayounzadeh M., Saadat M., Ahmadi A.* Investigating the effect of source characteristics on task comparability in integrated writing tasks // *Assessing Writing*. 2019. No. 41. P. 25–46.

13. *Stadler M., Herborn K., Mustafić M., et al.* The assessment of collaborative problem solving in PISA 2015: An investigation of the validity of the PISA 2015 CPS tasks // *Computers & Education*. 2020. No. 157. P. 1–11.

14. *Irvine S. H., Dann P. L., Anderson J. D.* Towards a theory of algorithm-determined cognitive test construction // *British Journal of Psychology*. 1990. Vol. 81, No. 2. P. 173–195.

15. *Irvine S. H., Kyllonen P. C.* Item generation for test development. Routledge, 2013. 444 p.

16. *Jankowska D. M., Karwowski M.* Measuring creative imagery abilities // *Frontiers in psychology*. 2015. No. 6. P. 1–17.

17. *Lievens F., Anseel F.* Creating alternate in-basket forms through cloning: Some preliminary results // *International Journal of Selection and Assessment*. 2007. Vol. 15, No. 4. P. 428–433.

18. *Lievens F., Sackett P. R.* Situational judgment tests in high-stakes settings: Issues and strategies with generating alternate forms // *Journal of Applied Psychology*. 2007. Vol. 92, No. 4. P. 1043–1055.

19. *Messick S.* The interplay of evidence and consequences in the validation of performance assessments // *Educational researcher*. 1994. Vol. 23, No. 2. P. 13–23.

20. *Mislevy R. J., Almond R. G., Lukas J. F.* A brief introduction to evidence-centered design // *ETS Research Report Series*. 2003. No. 1. P. 1–29.

21. *Osburn H. G.* Item sampling for achievement testing // *Educational and Psychological Measurement*. 1968. Vol. 28, No. 1. P. 95–104.

22. *Roos J. M., Bauldry S.* Confirmatory factor analysis. SAGE Publications, 2021. P. 144.

23. *Seelow D.* The art of assessment: Using game-based assessments to disrupt, innovate, reform and transform testing // *Journal of Applied Testing Technology*. 2019. Vol. 20, No. 1. P. 1–16.

24. *Wang D., Liu H., Hau K. T.* Automated and interactive game-based assessment of critical thinking // *Education and Information Technologies*. 2021. P. 1–23. DOI: <https://doi.org/10.1007/s10639-021-10777-9>.

### Информация об авторах

Д. А. Грачева — аспирант, стажер-исследователь Центра психометрики и измерений в образовании

К. В. Тарасова — кандидат педагогических наук, заместитель заведующего Лабораторией измерения новых конструкторов и дизайна тестов

### Information about the authors

D. A. Gracheva — PhD-student, Research Assistant at the Center for Psychometrics and Measurement in Education

K. V. Tarasova — PhD (Education), Deputy Head of the Laboratory for Measuring New Constructs and Test Design

Статья поступила в редакцию 13.04.2022; одобрена после рецензирования 28.04.2022; принята к публикации 04.05.2022.

The article was submitted 13.04.2022; approved after reviewing 28.04.2022; accepted for publication 04.05.2022.