
ОТРАСЛЕВАЯ ПСИХОЛОГИЯ SPECIAL (BRANCH) PSYCHOLOGY

Методология «Evidence-Centered Design» для измерения комплексных психологических конструктов

Угланова И.Л.,

*стажер-исследователь лаборатории измерения новых конструктов и дизайна тестов,
Национальный исследовательский университет «Высшая школа экономики», Москва, Россия,
iuglanova@hse.ru*

Брун И.В.,

*заместитель руководителя лабораторией измерения новых конструктов и дизайна тестов,
Национальный исследовательский университет «Высшая школа экономики», Москва,
Россия, ibrun@hse.ru*

Васин Г.М.,

*младший научный сотрудник лаборатории измерения новых конструктов и дизайна тестов,
Национальный исследовательский университет «Высшая школа экономики», Москва, Россия,
gvasin@hse.ru*

В статье представлена новая методология создания инструментов измерения «Evidence-Centered Design» (ECD), в переводе — систематический подход к разработке тестов. ECD позволяет подойти к созданию инструмента с доказательной позиции: результаты тестирования рассматриваются как свидетельства выраженности определенной психологической черты, характеристики. Методология наиболее актуальна для измерения комплексных, многокомпонентных конструктов. Методологические положения иллюстрируются нашим опытом разработки инструмента измерения креативности и критического мышления в начальной школе.

Ключевые слова: Evidence-Centered Design, систематический подход к разработке тестов, критическое мышление, креативность.

Для цитаты:

Угланова И.Л., Брун И.В., Васин Г.М. Методология Evidence-Centered Design для измерения комплексных психологических конструктов [Электронный ресурс] // Современная зарубежная психология. 2018. Том 7. № 3. С. 18—27. doi: 10.17759/jmfp.2018070302

For citation:

Uglanova I.L., Brun I.V., Vasin G.M. Evidence-Centered Design method for measuring complex psychological constructs [Elektronnyy resurs]. *Journal of Modern Foreign Psychology*, 2018, vol. 7, no. 3, pp. 18—27. doi: 10.17759/jmfp.2018070302 (In Russ.; Abstr. in Engl.).

Введение

Тенденции XXI века в области измерений в психологии не просто делают процесс создания инструмента более объемным — они требуют новых методологических решений.

Динамичность современного мира ставит новые практические и исследовательские задачи в области психологического тестирования [17]. Все большее внимание уделяется многокомпонентным конструктам¹, таким как навыки XXI века (21st century skills) [11] и навыки высшего порядка (higher order thinking skills) [15]. В фокус исследований попадают не только резуль-

таты решения тестовых заданий, но и сам способ решения [18; 24].

Для измерения таких конструктов необходимы инструменты измерения, в которых активно используются компьютерные симуляции, аудио- и видео-элементы [24]. Это позволяет приблизить тестирование к реальной жизни, снизить тестовую тревожность у респондентов, автоматизировать и ускорить оценивание результатов [7; 25].

Цель статьи — представить современную методологию разработки инструментов измерения Evidence-Centered Design (ECD [19]), систематический подход к разработке тестов [2].

¹ Здесь и далее термин «конструкт» используется как обобщенное обозначение явно ненаблюдаемых, латентных психологических черт или характеристик, компетенций или навыков, которые могут быть оценены через релевантное наблюдаемое поведение.

Знакомство с новой методологией разработки тестов даст исследователям возможность более гибко выстраивать процесс разработки теста и подбирать более подходящие методологические приемы под свои цели и задачи.

В статье термины ECD и систематический подход используются как взаимозаменяемые. Общие методологические положения систематического подхода проиллюстрированы примерами из нашего опыта разработки инструмента измерения критического мышления и креативности в начальной школе.

Важным преимуществом методологии ECD выступает целостность и внутренняя согласованность всей процедуры разработки инструмента. Создание структуры и содержания заданий, определение правил начисления баллов, доставка инструмента респонденту, предоставление обратной связи — все элементы разработки инструмента в таком подходе согласованы между собой и заданы одним главным вектором — психологическим конструктом, который требуется измерить.

В логике ECD измерения идут от абстрактных, теоретических предположений о природе конструкта и о том, как он может быть измерен, к поиску конкретных, эмпирических свидетельств — наблюдаемых действий в процессе тестирования, позволяющих сделать обоснованные выводы о респондентах [20].

Методология ECD была разработана в компании «Educational Testing Service» (ETS; www.ets.org) в начале 2000-х гг. и используется в ряде тестологических компаний («Cisco», «GlassLab», «CRESST» «Labs», «CITO»).

В методологии ECD были разработаны как инструменты измерения академических достижений [13], так и инструменты оценки более общих когнитивных и личностных характеристик: креативности и коммуникативных навыков в решении задач (creative problem solving [17], collaborative problem solving [23]), систематического мышления (systems thinking [24]).

В России ECD применялся только однажды — при разработке инструмента измерения информационно-коммуникационной компетентности учеников средней школы [2].

Также мы обнаружили основные принципы ECD в работе [3], но без прямого упоминания методологии.

Для объяснения основных методологических положений систематического подхода мы приводим примеры из нашего опыта в разработке нового инструмента измерения критического мышления и креативности в конце начальной школы — 2К.

В рамках инструмента 2К критическое мышление и креативность рассматриваются с позиции компетентностного подхода [1]. Инструмент состоит из заданий сценарного типа, которые предъявляются респондентам на компьютере в онлайн режиме. В заданиях сценарного типа респондент становится героем истории, подготовленной разработчиками инструмента. В таких историях заложены несколько проблем, которые тестируемый решает, последовательно справляясь с отдельными элементами заданий. В инструменте 2К

для определения уровня креативности и критического мышления используется три сценарных задания, которые включают 31 индикатор, на которых строится оценивание. При этом каждый индикатор встречается в сценариях несколько раз.

Общие принципы систематического подхода к разработке тестов

Основная цель ECD — обосновать аспекты дизайна теста, его разработки, шкалирования и использования с помощью принципов формальной логики [20]. Измерение рассматривается как процесс аргументации какого-либо вывода о респондентах на основе ограниченного количества свидетельств того, что респонденты делают/говорят в ситуации тестирования. Центральное понятие здесь — суждение (вывод; claim) о респонденте; о том, что он знает или умеет делать в реальной жизни. Весь процесс разработки теста направлен на то, чтобы собрать как можно больше доказательств того, что вывод, сделанный о респонденте на основе его тестового балла, отражает действительность.

В рамках ECD разработка и применение инструмента измерения условно делятся на 5 этапов (layers):

- 1) Анализ области компетенций (Domain Analysis).
- 2) Моделирование области компетенций (Domain Modeling).
- 3) Разработка концептуальной рамки инструмента (Conceptual Assessment Framework; CAF).
- 4) Сборка инструмента (Assembly Model).
- 5) Доставка инструмента (Delivery Model).

Этап 1. Анализ области компетенций

На первом этапе создания инструмента разработчики собирают всю важную информацию об измеряемом конструкте на основании обзора литературы, полуструктурированных интервью с представителями целевой аудитории и обсуждений с экспертами.

Основная цель этого этапа — понять, какие именно психологические черты или характеристики, компетенции или знания составляют измеряемый конструкт, как проявляется измеряемый конструкт в поведении, чем различаются люди, у которых он сформирован на разном уровне. Анализ области компетенции посвящен выявлению таких элементов реального мира и аспектов поведения людей, которые могут быть релевантны исследуемому конструкту. Помимо этого, на данном этапе разработчики отвечают на вопрос о том, с какими другими конструктами связан измеряемый конструкт и каков характер этой связи.

2К: пример реализации этапа

В ходе реализации первого этапа ECD в инструменте 2К были проанализированы документы:

1. Международные рамки и подходы к определению навыков, компетенций и грамотностей 21-го века

(например, World Economic Forum, Partnership for 21st Century Skills, University of Melbourne, UNESCO и др.).

2. Текст ФГОС НОО в части метапредметных образовательных результатов.

3. Существующие подходы к определению критического мышления и креативности и инструменты их измерения.

В результате мы установили, что важность критического мышления и креативности подчеркивается во всех международных рамках навыков XXI века. С эмпирической точки зрения, конструкт «креативность» существенно лучше проработан исследователями, по сравнению с «критическим мышлением». В России нет ни одного инструмента измерения критического мышления для целевой аудитории 9—11 лет с установленными психометрическими характеристиками и опубликованными данными о валидности и надежности, в то время как для измерения креативности — есть.

Этап 2. Моделирование области компетенций

На этом этапе разработчики структурируют информацию, полученную в процессе анализа области компетенций, в форме аргументов, базируясь на принципах логики, наглядно представленных в диаграмме Тулмина [26] (рис. 1).

На рис. 1 А — наблюдаемое поведение респондента, которое он демонстрирует в процессе прохождения теста (data), Б — то суждение (вывод), который разработчики хотят сделать о респонденте по результатам выполнения теста (claim), и В — аргумент, который показывает, почему наблюдаемое поведение А является свидетельством сформированности измеряемого конструкта Б (warrant). Альтернативное объяснение (Г, alternative explanation) показывает другие возможные причины наблюдаемого поведения А, не связанные с аргументом В.

2К: пример реализации этапа

На этапе моделирования области компетенций мы систематизировали информацию, полученную на предыдущем этапе, определили структуру измеряемых конструктов и потенциальное поведение, которое может свидетельствовать об их сформированности. Также мы сформулировали основные выводы,

которые будут сделаны об учениках по результатам тестирования:

а) ученик обладает определенным уровнем критического мышления, которое заключается в умении анализировать информацию, делать вывод на ее основе и аргументировать собственный вывод;

б) ученик обладает определенным уровнем креативности, которая заключается в любознательности относительно объекта работы и воображении.

В основу определения критического мышления легли работы: И.Р. Лаи [14], Р. Пола [22], П. Фасионе [10], Р.Х. Энниса [9], О.Л. Лиу и коллег [16]. О.Л. Лиу и коллеги пишут, что ученик, обладающий высоким уровнем критического мышления, «критически оценивает аргументы, допущения, абстрактные понятия и данные (которые могут быть неполными), чтобы вынести суждение и сформулировать соответствующие вопросы для достижения решения или набора решений» [16, с. 18].

На основе указанных работ были выделены два компонента компетенции «критическое мышление»: анализ информации (далее в тексте этот компонент будет называться «анализ» для краткости); вывод и аргументация.

Под анализом информации понимается работа с имеющейся информацией в соответствии с целями и условиями поставленной задачи. Результатом такой работы становится категоризация информации на факты, мнения и др.

В процессе анализа человек определяет взаимоотношения между элементами текста: связи, противоречия, выявляет избыточную и недостающую для решения проблемы информацию, а также оценивает ее — выносит суждения о достоверности фактов, сильных и слабых сторонах аргументов, важности отдельных элементов для решения поставленной задачи и т. д.

Аналогично были даны определения для компонента «вывод и аргументация» и для компетенции «креативность» и ее составляющих. Определения прошли экспертные обсуждения с педагогами начальной школы и возрастными психологами.

В рамках этой статьи мы описываем только компонент «анализ» компетенции «критическое мышление».

Для каждого компонента критического мышления и креативности был сформирован набор свидетельств проявления наблюдаемого поведения, а для каждого свидетельства была построена диаграмма Тулмина.



Рис. 1. Общий вид диаграммы Тулмина

Например, анализ подразумевает определение респондентом информации, нужной для решения поставленной задачи (рис. 2). В поведении учеников это проявляется, когда они выделяют информацию, необходимую для решения поставленной задачи. Если респондент копирует выделенную информацию в специальное поле (А), это свидетельствует о том, что он определяет выделенную информацию как важную для решения задачи (Б), на основании того, что у него была инструкция, в которой говорится о том, что всю полезную информацию следует перенести в специальное поле (В).

Помимо аргумента, нужно рассмотреть и альтернативные объяснения наблюдаемого поведения (Г).

Альтернативные объяснения в нашем примере призваны ответить на вопрос: «Чем может объясняться перенос некоторой информации в специальное поле, если не анализом?»

Например, ученик может перенести информацию в специальное поле, потому что пробовал разные варианты действий, которые позволяет сделать интерфейс, или ученик забыл о том, что ему нужно сделать.

На основе информации об измеряемых конструктах, которую мы получили в ходе обзора исследований и работы с экспертами, мы выделили следующие ожидаемые результаты (work products) респондента при тестировании. Например, определение категорий информации в тексте (факта и мнения, причины и следствия), оценка достоверности источника информации.

Для того чтобы разработать инструмент, измеряющий критическое мышление, необходимо четко обозначить границы этого конструкта и указать на связь с другими связанными конструктами.

Например, в работе Р. Фасионе [10] как один из компонентов критического мышления встречается саморегуляция, которая важна для преодоления неопределенности (новизны) и условно разделяется на мотивацию и толерантность к неопределенности. Однако спорным остается вопрос, являются ли эти конструкты частью критического мышления или всего лишь «пересекаются», т. е. связаны с ним. С учетом

накопленных исследований [например: 12] мы предполагаем, что сопротивление неопределенности является важным сопутствующим фактором для проявления критического мышления, однако не является одним из его компонентов.

Этап 3. Концептуальная рамка инструмента

Третий этап — разработка концептуальной рамки, он касается описания всех элементов инструмента измерения, т. е. спецификации теста. Концептуальная рамка связывает результаты двух предыдущих этапов с заданиями, заложенными в инструмент, и действиями респондентов при их решении.

Концептуальная рамка инструмента состоит из трех моделей (рис. 3).

1. Модель конструкта (Proficiency Model, Student model).
2. Модель заданий (Task Model).
3. Модель свидетельств (Evidence Model) [5].

Модель 1. Модель конструкта

Модель конструкта отражает то, какие психологические характеристики, компетенции или знания респондента являются целью оценивания. В отличие от предыдущих этапов создания инструмента, в этой модели измеряемый конструкт описывается не нарративно, а в более технической манере, с указанием конкретных переменных, отражающих измеряемый конструкт и различные уровни его сформированности.

Модель конструкта может включать и другие характеристики респондентов, которые могут повлиять на выполнение заданий теста и, таким образом, сказаться на интерпретации результатов измерения. Например: что в целом знают и умеют респонденты (в определенном возрасте), насколько хорошо они знакомы с компьютером, насколько велика необходимость специальных изменений предъявления инструмента для тестируемых с ОВЗ. Эта информация помогает разработать наиболее подходящие для целевой аудитории задания [19; 20].

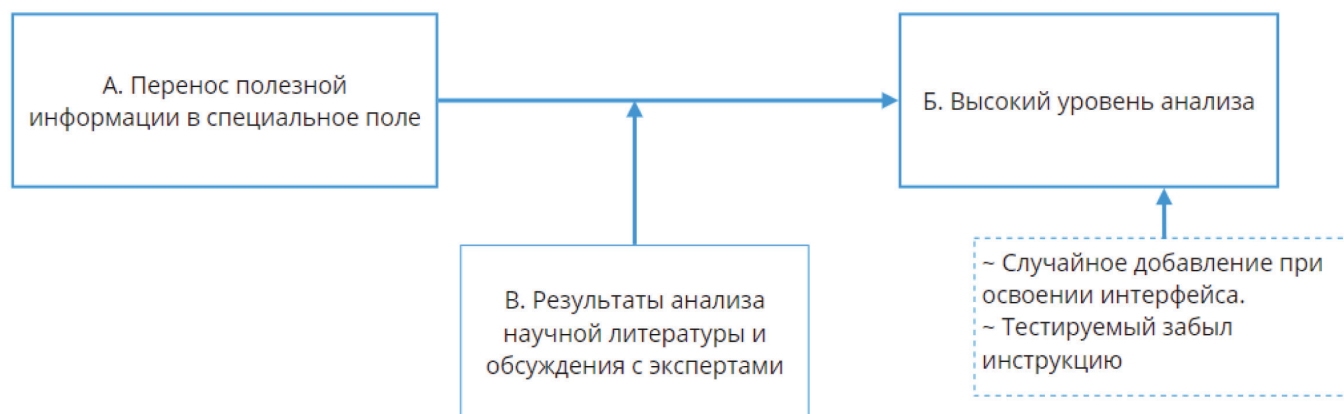


Рис. 2. Диаграмма Тулмина на примере компонента критического мышления «анализ»



Рис. 3. Связь моделей, составляющих концептуальную рамку инструмента

2К: пример реализации этапа

В инструменте 2К модель конструкта представляет собой упорядоченное описание структуры исследуемых конструктов и ожидаемого поведения респондента.

Мы выделили три уровня сформированности критического мышления: продвинутый, базовый и развивающийся. Респонденты на продвинутом уровне выбирают только надежные и релевантные источники информации, верно идентифицируют факт и мнение, понимают, какая информация нужна для решения проблемы и какой не хватает. Базовый уровень критического мышления присваивается, если респондент корректно работает с информацией, но допускает ошибки в оценке того, какие источники дают надежную, правдивую информацию. Развивающийся уровень означает, что респондент не только не понимает, какие из источников дают достоверную информацию, но и не может идентифицировать, какая информация будет полезна при решении проблемы и т. д.

В модели конструкта обосновывается, почему «анализ» является компонентом критического мышления и какой вес разные свидетельства имеют в итоговом уровне критического мышления. Согласно нашим предположениям, «анализ» является частью критического мышления, поскольку, чтобы успешно использовать информацию в жизни и делать с ее помощью обоснованные выводы, необходимо уметь декомпозировать информацию и оценивать ее свойства (релевантность, надежность). Все больше информации мы получаем в электронном виде, через интернет, поэтому в инструменте 2К мы опираемся на такое поведение, которое проявляется при анализе информации в цифровой среде.

В решение заданий на «анализ» могут внести вклад и дополнительные факторы: опыт работы на компьютере; навыки чтения; внимательность и заинтересованность,

поскольку прохождение теста является добровольным. Отметим, что для выполнения заданий не важна успеваемость учеников по школьным предметам, так как задания не опираются на учебный контекст.

Модель 2. Модель заданий

Модели заданий помогают систематизировать информацию, которую разработчики собирают о поведении респондента в ситуации тестирования. Благодаря модели конструкта, уже известно, что конкретно нужно измерить и какие допущения принимаются о дополнительных компетенциях и мотивации респондента.

Модель задания помогает проконтролировать, что представленный респонденту стимульный материал действительно подходит для измерения характеристик респондента, обозначенных в модели конструкта.

Модель задания включает в себя описание всего, с чем взаимодействует респондент при решении задания: тип задания (например, сценарного типа, эссе, с множественным выбором и др.), стимульный материал, инструкции, степень интерактивности и т. д., обычно в форме таблицы или структурированного списка. Она также включает и описание ожидаемых результатов — той части наблюдаемого поведения, которая релевантна измеряемому конструкту. Из хорошо разработанной модели задания очевидно, почему определенное задание служит для измерения определенного конструкта и каким образом ожидаемые результаты выполнения этого задания свидетельствуют о навыках респондента.

Модель задания может описывать целое семейство заданий, подчиненных одной логике и измеряющих одни и те же индикаторы исследуемого конструкта, в неизменной последовательности. Кроме обязательных элементов заданий, эта модель также включает набор

вариативных элементов, за счет которых можно менять трудность задания или создавать задания-клоны.

2К: пример реализации этапа

Рассмотрим модель задания, которая предназначена для измерения компонента «анализ». Согласно модели конструкта, умение анализировать информацию, заключается, в частности, в определении необходимой и/или недостающей информации для решения проблемы и в оценивании достоверности различных источников информации. «Анализ» предполагает сложное поведение, поэтому для его наблюдения наиболее подходящим будет сценарный тип задания. Важно, что в модели конструкта знание школьной программы не должно быть связано с результатами теста, поэтому следует максимально избегать в задании школьного контекста.

В итоге было разработано задание о содержании крабов в качестве домашних животных и постройке аквариума для них. Обязательные элементы сценария: поиск надежного источника информации о жизни крабов в домашних условиях и выбор полезной информации о том, каким должен быть их аквариум. Вариативный компонент: количество источников информации (чем их больше, тем задание труднее).

Модель 3. Модель свидетельств

Модель свидетельств связывает информацию из модели конструкта и моделей заданий. Она разбивает ожидаемые результаты респондентов на отдельные уровни и подробно описывает, каким образом каждый из них связан с уровнем выраженности конструкта. Для этого требуется структурировать ожидаемые результаты. Поскольку ожидаемые результаты связаны с латентным конструктом через модели заданий, это позволяет нам проинтерпретировать появление определенных индикаторов как свидетельство определенного уровня выраженности конструкта.

Первая часть модели свидетельств — правила оценки ожидаемых результатов (evidence rules): как каждый уровень индикаторов отличается друг от друга в контексте их связи с измеряемым конструктом.

Вторая часть модели свидетельств — модель измерений (measurement model). Это математическая модель, которая связывает уровни индикаторов и уровни латентных конструктов респондентов. Всякому практическому измерению присуща погрешность, и для качественного измерения ее необходимо учесть. Обычно по этой причине для модели измерений используется одна из двух популярных психометрических парадигм: классическая теория тестов (КТТ) или современная теория тестов (IRT). Выбор метода статистического анализа данных — один из ключевых этапов разработки теста и для каждого конкретного инструмента разработчики сами принимают обоснованное решение.

2К: пример реализации этапа

В одном из сценариев, направленных на измерение «анализа», согласно модели заданий, респонденту

предъявляются три источника информации, различные по достоверности. Максимальный балл начисляется, если респондент выбирает наиболее достоверный источник в первую очередь. Например, респонденту требуется определить, что необходимо купить для содержания краба дома, используя симуляцию поисковика. Один из ожидаемых результатов здесь связан с тем, какую из трех возможных ссылок, предоставляемых поисковиком, респондент выбирает в первую очередь. Если первой выбрана ссылка «Рекомендации сотрудников зоопарка» (достоверный источник), это рассматривается как индикатор высокого уровня компонента «анализ». У этого индикатора всего два уровня, и если в первую очередь респондент открывает другие ссылки, это считается индикатором низкого уровня «анализа».

Таким же образом расписаны все остальные индикаторы этого задания, и благодаря готовым моделям задания и конструкта можно проследить цепочку логики от того, что мы хотим измерить, к тому, как оно проявляется в поведении.

Теперь рассмотрим модель измерений для инструмента 2К. Модель конструкта указывает на то, что латентные конструкты связаны между собой определенным образом, и это необходимо учесть в модели измерений. Для этого мы построили байесовскую сеть — направленный ациклический граф, узлы которого отражают измеряемые конструкты и их индикаторы, а дуги — связи между ними. Связи между индикаторами и соответствующими конструктами представлены через одну из моделей современной теории тестов — Graded Response Model [6; 21].

Этап 4. Сборка инструмента

Этап сборки инструмента посвящен разработке и подготовке к использованию конкретных элементов, описанных в концептуальной рамке инструмента. Этот этап включает в себя создание конкретных заданий, полное описание рубрик оценивания, инструкций по сборке варианта теста, а также апробацию заданий.

2К: пример реализации этапа

В соответствии с концептуальной рамкой мы разработали три интерактивных задания сценарного типа, которые основаны на повседневном и фантастическом контексте. Одно из них заключается в том, чтобы найти и проанализировать информацию о содержании крабов в домашних условиях и построить на ее основе аквариум из предложенных элементов.

В ранней версии задания респондентам предлагалось собрать аквариум для черепахи. Однако при проведении первой апробации выяснилось, что дети, у которых дома есть черепахи, легче справляются с этим заданием, чем респонденты с близким уровнем критического мышления, но не имеющим дома черепах. Поэтому в итоговом задании было выбрано более редкое для содержания в домашних условиях животное — краб.

Этап 5. Доставка инструмента

Наконец, на последнем, пятом, этапе разработчики прорабатывают вопросы, связанные с применением, практической реализацией инструмента. Этот этап описывает правила проведения тестирования (администрирования) и не касается содержания теста [4].

2К: пример реализации этапа

Инструмент измерения 2К создавался для применения в компьютерной форме. Для реализации тестирования был создан сайт с запрограммированными заданиями инструмента. Апробации позволили выявить и учесть ряд проблем (например, проблему недостаточной мощности школьных компьютеров, на которых проводилось тестирование).

Заключение

Целью статьи было познакомить читателя с новой методологией создания инструментов измерения сложных, многокомпонентных конструкций.

Мы охарактеризовали систематический подход к разработке тестов с иллюстративными примерами из нашего опыта создания инструмента измерения критического мышления и креативности — 2К.

Однако ни одна методология не свободна от ограничений.

Основные ограничения систематического подхода сводятся к ресурсозатратности: каждый этап требует детальной проработки, которая занимает большое количество усилий, времени и денежных ресурсов. Создание инструмента для измерения сложного конструкта в этом подходе предполагает наличие большой команды: психометриков и разработчиков заданий, экспертов содержательной области, представителей целевой группы для апробации, возможно, иллюстраторов и менеджеров, а в случае компьютерного тестирования — проектировщиков интерфейсов и программистов.

Методология ECD позволяет объединить наблюдаемое поведение и латентные конструкты в единый логический аргумент. Это полезно как на этапе разработки измерительных инструментов, так и на этапе их отладки.

Поскольку в сумме все этапы разработки дают исчерпывающее описание готового инструмента, легко вернуться к любому из них для работы над ошибками.

Например, обсуждение с экспертами вносит уточнение в теоретическую рамку, а апробации позволяют выявить сильные и слабые стороны конкретных заданий и сложности в администрировании.

Вся полученная информация необходима разработчикам, чтобы сделать инструмент лучше — сделать его таким, чтобы он позволил валидно и надежно измерять нужную черту или компетенцию [8].

ЛИТЕРАТУРА

1. Бояцис Р. Компетентный менеджер. Модель эффективной работы. М.: ХИППО, 2008. 340 с.
2. Оценка информационно-коммуникационной компетентности учащихся: подходы, инструмент, валидность и надежность результатов [Электронный ресурс] / С.М. Авдеева [и др.] // Вопросы образования. 2017. № 4. С. 104—132. URL: <https://cyberleninka.ru/article/n/otsenka-informatsionno-kommunikatsionnoy-kompetentnosti-uchaschihsya-podhody-instrument-validnost-i-nadezhnost-rezultatov> (дата обращения: 04.07.2018).
3. Хлопотов М.В. Применение байесовской сети при построении моделей для оценки уровня сформированности компетенций [Электронный ресурс] // Интернет-журнал «Науковедение». 2014. № 5(24). С. 1—28. URL: <https://cyberleninka.ru/article/n/primenenie-bayesovskoy-seti-pri-postroenii-modeley-dlya-otsenki-urovnya-sformirovannosti-kompetentsiy> (дата обращения: 26.08.2018).
4. Almond R., Steinberg L., Mislevy R. Enhancing the design and delivery of assessment systems: A four-process architecture [Электронный ресурс] // The Journal of Technology, Learning and Assessment. 2002. Vol. 1. № 5. P. 1—64. URL: <https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1671/1509> (дата обращения: 26.08.2018).
5. An evidence centered design for learning and assessment in the digital world [Электронный ресурс]: CRESST Report 778 / J.T. Behrens [et al.]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), 2010. 48 p. URL: <https://files.eric.ed.gov/fulltext/ED520431.pdf> (дата обращения: 26.08.2018).
6. Bayesian networks in educational assessment / R.G. Almond [et al.]. New York: Springer, 2015. 662 p. doi:10.1007/978-1-4939-2125-6
7. de Klerk S., Veldkamp B.P., Eggen T.J.H.M. Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example // Computers & education. 2015. Vol. 85. P. 23—34. doi:10.1016/j.compedu.2014.12.020
8. Debugging the evidence chain [Электронный ресурс] / R.G. Almond [et al.] // 2013 Uai Application Workshops: Big data meet complex models and models for spatial, temporal and network data (uai2013aw / Eds. R.G. Almond, O. Mengshoel. Seattle, WA: Association For Uncertainty In Artificial Intelligence, 2013. P. 1—10. URL: <http://ceur-ws.org/Vol-1024/paper-01.pdf> (дата обращения: 26.08.2018).
9. Ennis R.H. Critical thinking assessment // Theory into practice. 1993. Vol. 32. № 3. P. 179—186. doi:10.1080/00405849309543594
10. Facione P. Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction [Электронный ресурс]. Newark: California Academic Press, 1990. 112. p. URL: <https://philarchive.org/archive/FACSTA> (дата обращения: 26.08.2018).

11. Griffin P., Care E., McGaw B. The changing role of education and schools // Assessment and teaching of 21st century skills / Eds. P. Griffin, E. Care. Dordrecht: Springer, 2012. P. 1—15. doi:10.1007/978-94-007-2324-5_1
12. Grit: perseverance and passion for long-term goals / A.L. Duckworth [et al.] // Journal of personality and social psychology. 2007. Vol. 92. № 6. P. 1087—1101. doi:10.1037/0022-3514.92.6.1087
13. Introduction to the biomass project: An illustration of evidence-centered assessment design and delivery capability [Электронный ресурс]: CSE Report 609 / L.S. Steinberg [et al.]; National Center for Research on Evaluation, Standards, and Student Testing; California University; Center for the Study of Evaluation. Los Angeles, CA, 2003. 57 p. URL: <https://files.eric.ed.gov/fulltext/ED483027.pdf> (дата обращения: 26.08.2018).
14. Lai E.R. Critical thinking: A literature review [Электронный ресурс]: Research Report. Vol. 6. Pearson, 2011. 49 p. URL: <http://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf> (дата обращения: 26.08.2018).
15. Lewis A., Smith D. Defining higher order thinking // Theory into practice. 1993. Vol. 32. № 3. P. 131—137. doi:10.1080/00405849309543588
16. Liu O.L., Frankel L., Roohr K.C. Assessing Critical Thinking in Higher Education: Current State and Directions for Next Generation Assessment // ETS Research Report Series. 2014. Vol. 2014. № 1. P. 1—23. doi:10.1002/ets2.12009
17. Melding the power of serious games and embedded assessment to monitor and foster learning / V.J. Shute [et al.] // Serious games: Mechanisms and effects / Eds. U. Ritterfeld, M. Cody, P. Vorderer. New York; London: Routledge; Taylor and Francis, 2009. P. 295—321.
18. Messick S. Alternative modes of assessment, uniform standards of validity: Research Report. Princeton, N.J.: Educational Testing Service, 1994. 25 p. (ETS Research Report Series. № 2). doi:10.1002/j.2333-8504.1994.tb01634.x
19. Mislevy R.J., Almond R.G., Lukas J.F. A brief introduction to evidence-centered design: Research Report. Princeton, N.J.: Educational Testing Service, 2003. 37 p. (ETS Research Report Series. Vol. 2003. № 1.). doi:10.1002/j.2333-8504.2003.tb01908.x
20. Mislevy R.J., Haertel G.D. Implications of evidence-centered design for educational testing // Educational Measurement: Issues and Practice. 2006. Vol. 25. № 4. P. 6—20. doi:10.1111/j.1745-3992.2006.00075.x
21. Modeling conditional probabilities in complex educational assessments [Электронный ресурс] / R.J. Mislevy [et al.]; Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California. Los Angeles, 2001. 56 p. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.322.4516&rep=rep1&type=pdf> (дата обращения: 26.08.2018).
22. Paul R. Critical thinking: What, why, and how // New directions for community colleges. 1992. Vol. 1992. № 77. P. 3—24. doi:10.1002/cc.36819927703
23. Polyak S.T., von Davier A.A., Peterschmidt K. Computational Psychometrics for the Measurement of Collaborative Problem Solving Skills // Frontiers in psychology. 2017. Vol. 8. P. 20—29. doi:10.3389/fpsyg.2017.02029
24. Psychometric considerations in game-based assessment / R.J. Mislevy [et al.]. [Redwood City, CA]: GlassLab Report, 2014. 158 p.
25. Shute V.J., Rahimi S. Review of computer-based assessment for learning in elementary and secondary education // Journal of Computer Assisted Learning. 2017. Vol. 33. № 1. P. 1—19. doi:10.1111/jcal.12172
26. Toulmin S.E. The uses of argument. Cambridge: Cambridge University Press, 1958. 262 p.

Evidence-Centered Design method for measuring complex psychological constructs

Uglanova I.L.,

research intern at Laboratory of new constructs measurement and tests designing, National Research University
«Higher School of Economics», Moscow, Russia,
iuglanova@hse.ru

Brun I.V.,

deputy head of Laboratory of new constructs measurement and tests designing, National Research University
«Higher School of Economics», Moscow, Russia,
ibrun@hse.ru

Vasin G.M.,

junior researcher at Laboratory of new constructs measurement and tests designing, National Research University
«Higher School of Economics», Moscow, Russia,
gvasin@hse.ru

The article introduces a new method in creating measuring tests — Evidence-Centered Design (ECD) or a systematical approach in test-creating. ECD allows using evidentiary framework while creating the test: the testing results are considered to be the evidence of certain psychological trait or characteristic expression. This method is the most relevant for measuring complex, multicomponent constructs. Methodological positions are presented through our experience in developing a new test for measuring creativity and critical thinking in primary school.

Keywords: Evidence-Centered Design, systematical approach in test-creating, critical thinking, creativity.

REFERENCES

1. Boyatsis R. Kompetentnyi menedzher. Model' effektivnoi raboty [Competent manager. Model of effective work]. Moscow: KhIPPO, 2008. 340 p. (In Russ.).
2. Avdeeva S.M. et al. Otsenka informatsionno-kommunikatsionnoi kompetentnosti uchashchikhsya: podkhody, instrument, validnost' i nadezhnost' rezul'tatov [Assessment of information and communication competence of students: approaches, tools, validity and reliability of results] [Elektronnyi resurs]. *Voprosy obrazovaniya [Education issues]*, 2017, no. 4, pp. 104—132. Available at: <https://cyberleninka.ru/article/n/otsenka-informatsionno-kommunikatsionnoy-kompetentnosti-uchashchihsya-podhody-instrument-validnost-i-nadezhnost-rezultatov> (Accessed 04.07.2018). (In Russ.; Abstr. in Engl.).
3. Khlopotov M.V. Primenenie baiesovskoi seti pri postroenii modelei dlya otsenki urovnya sformirovannosti kompetentsii [Bayesian network application in building models for assessing the level of competence formation] [Elektronnyi resurs]. *Internet-zhurnal Naukovedenie [Internet Journal of Science]*, 2014, no. 5(24), pp. 1—28. Available at: <https://cyberleninka.ru/article/n/primenenie-bayesovskoy-seti-pri-postroenii-modeley-dlya-otsenki-urovnya-sformirovannosti-kompetentsiy> (Accessed 26.08.2018). (In Russ.).
4. Almond R., Steinberg L., Mislevy R. Enhancing the design and delivery of assessment systems: A four-process architecture. *The Journal of Technology, Learning and Assessment*, 2002, vol. 1, no. 5, pp. 1—64. Available at: <https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1671/1509> (Accessed 26.08.2018).
5. Behrens J.T. et al. An evidence centered design for learning and assessment in the digital world [Elektronnyi resurs]: CRESST Report 778. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), 2010. 48 p. Available at: <https://files.eric.ed.gov/fulltext/ED520431.pdf> (Accessed 26.08.2018).
6. Almond R.G. et al. Bayesian networks in educational assessment. New York: Springer, 2015. 662 p. doi:10.1007/978-1-4939-2125-6
7. de Klerk S., Veldkamp B.P., Eggen T.J.H.M., Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & education*, 2015, vol. 85, pp. 23—34. doi:10.1016/j.compedu.2014.12.020
8. Almond R.G. et al. Debugging the evidence chain [Elektronnyi resurs]. In Almond R.G., Mengshoel O. (eds.) Proceedings of the 2013 uai application workshops: Big data meet complex models and models for spatial, temporal and network data (uai2013aw). Seattle, WA: Association For Uncertainty In Artificial Intelligence, 2013. pp. 1—10. Available at: <http://ceur-ws.org/Vol-1024/paper-01.pdf> (Accessed 26.08.2018).
9. Ennis R.H. Critical thinking assessment. *Theory into practice*, 1993, vol. 32, no. 3, pp. 179—186. doi:10.1080/00405849309543594

10. Facione P. Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction [Elektronnyi resurs]. Newark: California Academic Press, 1990. 112 p.. Available at: <https://philarchive.org/archive/FACCTA> (Accessed 26.08.2018).
11. Griffin P., Care E., McGaw B. The changing role of education and schools // Assessment and teaching of 21st century skills. Dordrecht: Springer, 2012. pp. 1—15. doi:10.1007/978-94-007-2324-5_1
12. Duckworth A.L. et al. Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, 2007, vol. 92, no. 6, pp. 1087—1101. doi:10.1037/0022-3514.92.6.1087
13. Steinberg L.S. et al. Introduction to the biomass project: An illustration of evidence-centered assessment design and delivery capability [Elektronnyi resurs]: CSE Report 609. Center for the Study of Evaluation National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles, CA, 2003. 57 p. Available at: <https://files.eric.ed.gov/fulltext/ED483027.pdf> (Accessed 26.08.2018).
14. Lai E.R. Critical thinking: A literature review [Elektronnyi resurs]: Research Report. Vol. 6.. Pearson, 2011. 49 p. Available at: <http://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf> (Accessed 26.08.2018).
15. Lewis A., Smith D. Defining higher order thinking. *Theory into practice*, 1993, vol. 32, no. 3, pp. 131—137. doi:10.1080/00405849309543588
16. Liu O.L., Frankel L., Roohr K.C. Assessing Critical Thinking in Higher Education: Current State and Directions for Next Generation Assessment. *ETS Research Report Series*, 2014, vol. 2014, no. 1, pp. 1—23. doi:10.1002/ets2.12009
17. Shute V.J. et al. Melding the power of serious games and embedded assessment to monitor and foster learning. In Ritterfeld U., Cody M., Vorderer P. (Eds.). *Serious games: Mechanisms and effects*. New York; London: Routledge; Taylor and Francis, 2009, pp. 295—321.
18. Messick S. Alternative modes of assessment, uniform standards of validity. Princeton, N.J.: Educational Testing Service, 1994. 25 p. (ETS Research Report Series. № 2). doi:10.1002/j.2333-8504.1994.tb01634.x
19. Mislavy R.J., Almond R.G., Lukas J.F. A brief introduction to evidence-centered design: Research Report. Princeton, N.J.: Educational Testing Service, 2003. 37 p. (ETS Research Report Series. Vol. 2003. № 1.). doi:10.1002/j.2333-8504.2003.tb01908.x
20. Mislavy R.J., Haertel G.D. Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 2006, vol. 25, no. 4, pp. 6—20. doi:10.1111/j.1745-3992.2006.00075.x
21. Mislavy R.J. et al. Modeling conditional probabilities in complex educational assessments [Elektronnyi resurs]. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California. Los Angeles, 2001. 56 p. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.322.4516&rep=rep1&type=pdf> (Accessed 26.08.2018).
22. Paul R. Critical thinking: What, why, and how. *New directions for community colleges*, 1992, vol. 1992, no. 77, pp. 3—24. doi:10.1002/cc.36819927703
23. Polyak S.T., von Davier A.A., Peterschmidt K. Computational Psychometrics for the Measurement of Collaborative Problem Solving Skills. *Frontiers in psychology*, 2017, vol. 8, pp. 20—29. doi:10.3389/fpsyg.2017.02029
24. Mislavy R.J. et al. Psychometric considerations in game-based assessment. GlassLab Report, 2014. 158 p.
25. Shute V.J., Rahimi S. Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, 2017, vol. 33, no. 1, pp. 1—19. doi:10.1111/jcal.12172
26. Toulmin S.E. The uses of argument. Cambridge: Cambridge University Press, 1958. 262 p.