

Тарасова Ксения
к.п.н., зам. заведующего
Лаборатории измерения новых конструкторов
и дизайна тестов ЦПИО НИУ ВШЭ





01

**Парадигма
обоснованного
дизайна
Principled Assessment
Design**

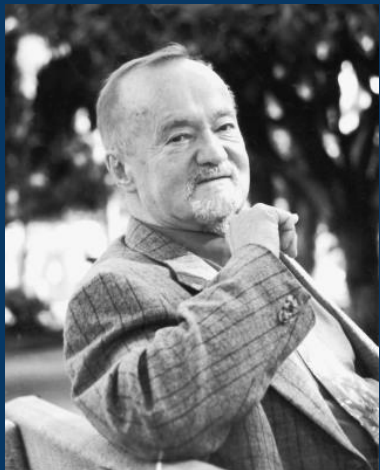
Что это такое?

В рамках обоснованного дизайна оценивание понимается как построение аргумента, включающего в себя некое **заявление**, которое мы хотим сделать о респонденте, **данные**, которыми мы располагаем, и **инструментарий**, позволяющий связать данные с предполагаемым заявлением.

Данные, связанные с заявлением посредством инструментария служат доказательством, подтверждающим это заявление.

1. Что мы оцениваем, почему и зачем? Какой вывод мы хотим сделать о оцениваемом.
2. Какие доказательства для этого необходимы.
3. Где и как возможно получить эти доказательства.
4. Как мы интерпретируем доказательства, как и на основании чего связываем их с предметом оценки.
5. Возможно ли распространить вывод, сделанный на основе этих доказательств, на ситуации, выходящие за рамки тех, что были предложены в процессе оценивания?

(Messick, S. 1992; Mislevy, R. J. 1994)



Структура аргумента Тулмина

**Данные, полученные в
результате
деятельности
(data)**

Наблюдаемое поведение респондента, которое он демонстрирует в процессе прохождения теста

Структура аргумента Тулмина

Суждение (вывод), который
разработчики хотят сделать о
респонденте по результатам
выполнения теста

**Заявление о
респонденте
(claim)**

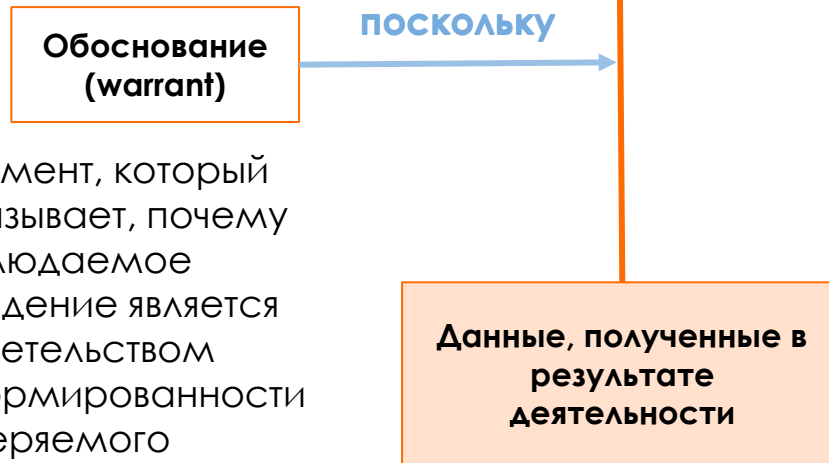
**Данные, полученные в
результате
деятельности
(data)**



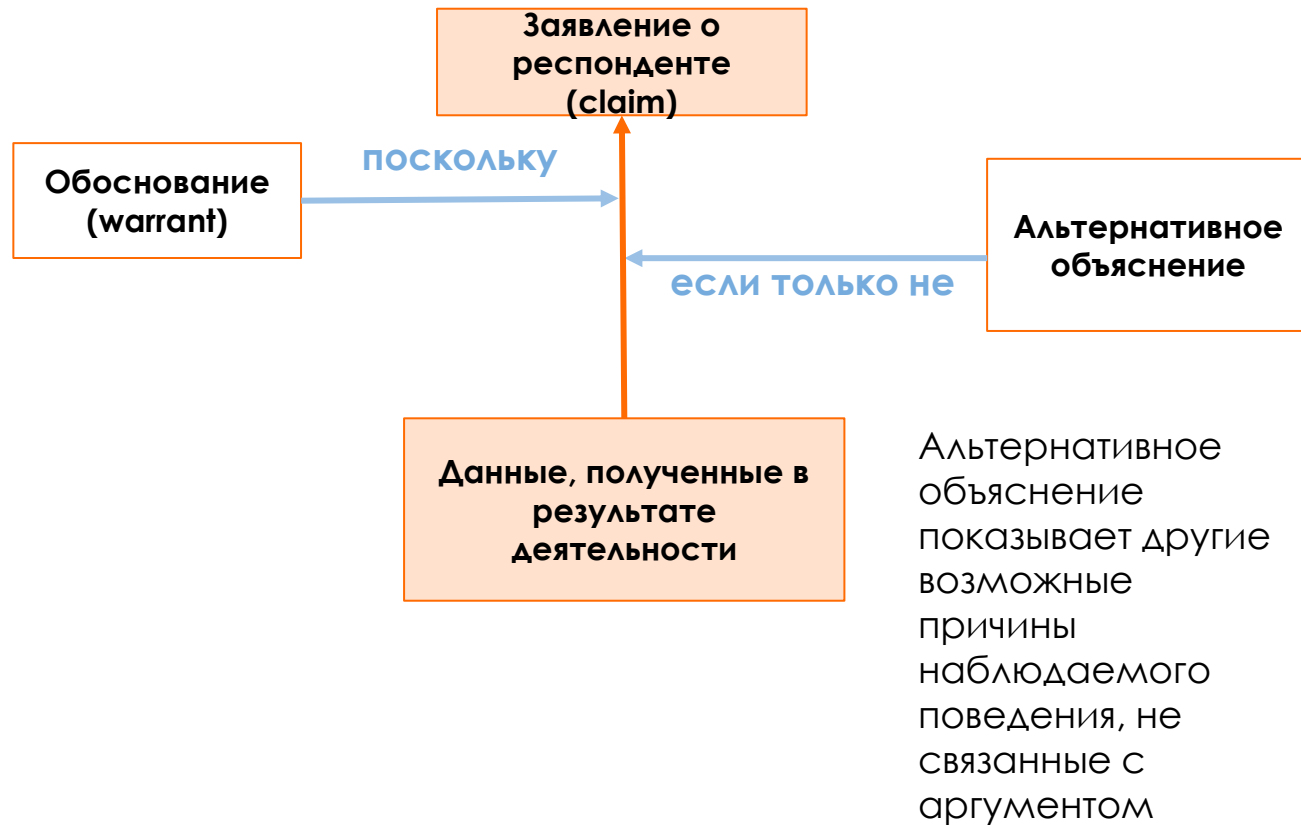
Структура аргумента Тулмина

Аргумент, который показывает, почему наблюдаемое поведение является свидетельством сформированности измеряемого конструкта.

Теории говорят нам...



Структура аргумента Тулмина



Структура аргумента Тулмина



Процесс доказательного мышления при создании инструментов оценивания формализован в форме простой схемы (Pellegrino, 2001)



1. Результаты познавательной деятельности - то, что предполагается оценить. Это научные, эмпирически подтвержденные теории, сообщающие, как оцениваемый приобретает, осваивает и демонстрирует компетенции, которые мы собираемся оценивать. Это могут быть теории обучения, основания, полученные из области когнитивной или поведенческой психологии или нейробиологии, теории, описывающие поведение человека при взаимодействии в цифровых средах и многое другое, в зависимости от имеющихся возможностей и поставленных задач. Этот компонент позволяет нам корректно и точно сформулировать заявление, которое мы хотим сделать относительно респондента.



2. Наблюдение - совокупность представлений о том, какие ситуации дадут доказательства, необходимые, чтобы сделать заявление. Это набор представлений о том, при выполнении каких действий и заданий, могут проявиться интересующие компетенции, которые нужно измерить. Эти ситуации, действия и задания не случайны, а обусловлены предыдущим компонентом. Теории, входящие в компонент «результаты познавательной деятельности», дают основания для того, чтобы сказать, что, выполняя конкретное задание, оцениваемый демонстрирует именно те навыки, которые мы хотим оценить. Этот компонент отвечает за данные, на основе которых мы сможем сделать заявление, и отчасти обуславливает связь между данными и заявлением.



3. Интерпретация – процесс, в ходе которого мы придаем собранным доказательствам смысл. Это методы и инструменты анализа, объясняющие, почему и как данные, полученные из компонента «наблюдение» связаны с оцениваемыми компетенциями, с заявлением, которое мы собираемся сделать. Методы, входящие в этот компонент, также обусловлены и находятся в связи с предыдущими двумя.



02

Методология Evidence-Centered Design

Метод доказательной аргументации (Evidence-Centered Design)



Robert Mislevy

Основывается на принципах формальной логики, выстраивая систему аргументации от тех действий, которые совершает респондент в процессе решения заданий к выводу о сформированности компетенции

Разработан в начале 21 века и используется для измерения современных сложных латентных конструктов

Что такое ECD?

Это набор взаимосвязанных процедур призванных прояснить два вопроса:

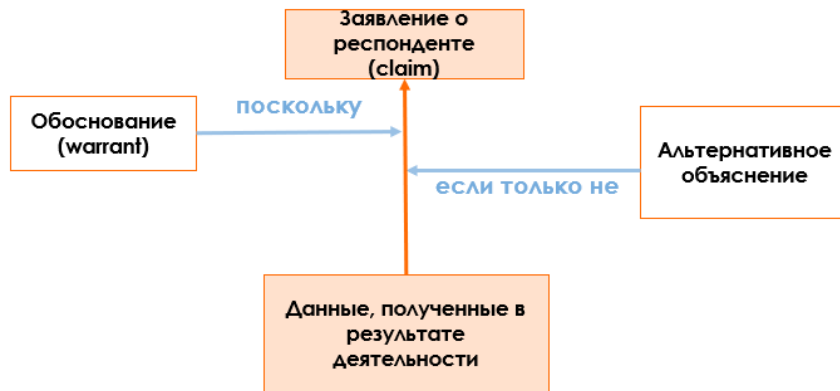
- Какие свидетельства компетентности (или знаний или выраженности конструкта) тестируемого мы можем непосредственно наблюдать?
- Как структурировать ситуацию, чтобы увидеть эти свидетельства?

Цель: собрать как можно больше свидетельств того, что выводы по тесту валидны в отношении тестируемых

Подходит для измерения сложных конструктов и конструирования технологически усовершенствованных заданий (СРВА)



- Claim – заявление (вывод) о тестируемом, которое планируется сделать по результатам тестирования
- Evidence – свидетельства, которые нужно собрать в ходе выполнения теста, для того, чтобы сделать вывод о тестируемом
- Assessment argument – если A, то B, потому что B.



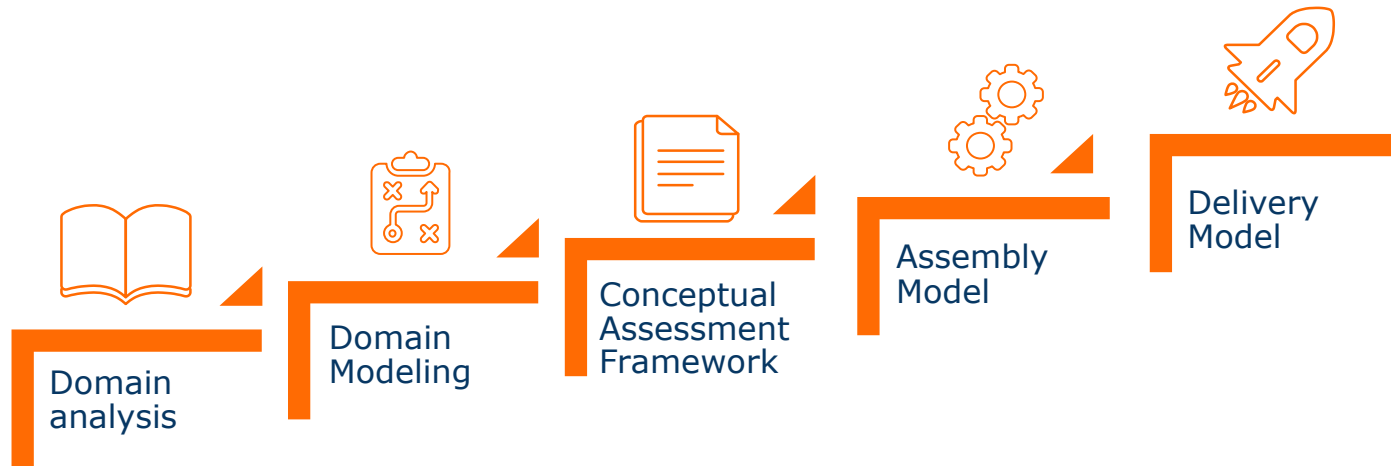
Анализ области компетенций (Domain Analysis)

Моделирование области компетенций (Domain Modeling)

Разработка концептуальной рамки инструмента (Conceptual Assessment Framework; CAF)

Воплощение теста (Assembly Model, Assessment Implementation)

Применение (Delivery Model)



Анализ области компетенций

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.

Основа для построения аргумента.

- Что мы знаем об этой области (содержание, понятия, терминология, инструменты и репрезентативные формы, концепции и т.д.)?
- Какой подход к рассмотрению выбираем (когнитивный, бихевиористский, интеракционистский и т.п.)
- В каких ситуациях она важна (выделяются повторяющиеся аспекты ситуаций)?
- Как они проявляются (виды знаний и умений, которые считаются важными в данной области)?
- Ключевые компетенции? Связь с другими компетенциями (конструктами) и характер этой связи
- Job analysis, User experience (выявление таких элементов реального мира и аспектов поведения людей, которые могут быть релевантны исследуемому конструкту)
- (Внешние) причины и следствия компетенций
- Нормативные акты, рамочные документы (стандарты содержания)



Domain
analysis

Анализ области компетенций: результат

Проанализировано:

- 126 определений
- 7 глобальных рамок
- Более 200 источников просмотрено
- Около 90 отсеяно
- Около 120-130 использовано (около 40 книг)
- Более 100 инструментов

Дано определение конструкта

Определены компоненты



Domain
analysis

Уровень моделирования состоит из систематических структур для организации содержания, выявленного в ходе анализа области компетенций, в терминах оценочного аргумента (базируясь на принципах логики).

- Как мы концептуализируем ключевые аспекты области компетенций?
- Как внутри выбранного подхода рассматриваются и определяются компетенции?
- С какими компетенциями связаны измеряемые конструкторы?
- Сколько субкомпетенций? Как они связаны между собой?
- Как связаны элементы ситуаций, в которых компетенция может проявиться?

оригинальное определение термина(ов), использующихся в названии субконструкта	библиографическая ссылка	возможное операционное определение субконструкта	библиографическая ссылка	как субконструкт проявляется в поведении (на уровне стратегий/тактик)	библиографическая ссылка
conceptual representation - the preverbal message that is a semantic representation that refers to some state of affairs. That state of affairs can be in any mode of thought, but the message must be in propositional form	1-Levelt J.W. (1989) Speaking From Intention to Articulation A Bradford Book	процесс в результате которого мысль превращается в мысленное сообщение, имеющее не лингвистический, нелинейный и довербальный характер	1-Berger, Charles R..(2010) The Handbook of Communication Science . SAGE Publications. Kindle Edition. 2-Konopka A.E , Brown-Schmidt S. (2014) Message Encoding in The Oxford Handbook of Language Production Edited by Matthew Goldrick Victor Ferreira Michele Miozzo Oxford University Press 3-Kormos J. (2006) Speech Production	1-способность указывать наиболее значимые свойства (объекты) в начале сообщения (в начале описания готового сообщения) Стратегия- при описании предложенной ситуации трестируемый начинает описание с наиболее значимого объекта Стратегия- при описании	1-Konopka A.E , Brown-Schmidt S. (2014) Message Encoding in The Oxford Handbook of Language Production Edited by Matthew Goldrick Victor Ferreira Michele Miozzo Oxford University Press 2-Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In F. Ferreira & J. Henderson (Eds.), The integration of language, vision



1. Определение структуры измеряемых конструктов и потенциальное поведение, которое может свидетельствовать об их сформированности (поведенческая онтология)

- Knowledge map, Concept map - инструмент, применяемый для преподавания и оценивания. Служит для в для иллюстрации наших декларативных знаний об области компетенций (факты, определения, утверждения и тд) и процедуральных знаний об области компетенций(как что-то работает, процессы, планы и тд).

В измерениях позволяет:

- установить общую терминологию и определения
- ясно определить и проиллюстрировать связи между концепциями и навыками (знаниями, способностями)
- сделать определения и связи между концепциями понятными и согласованными для людей, имеющих разный образовательный бэкграунд



Семантические связи

Causal - X создает или изменяет эффект на Y – потому что, вследствие

Hierarchical - X находится внутри или является частью Y, X-часть, категория, класс Y - часть, целое, тип

Temporal - описывает временные отношения. Предпосылка к, шаг после, в процессе и тд

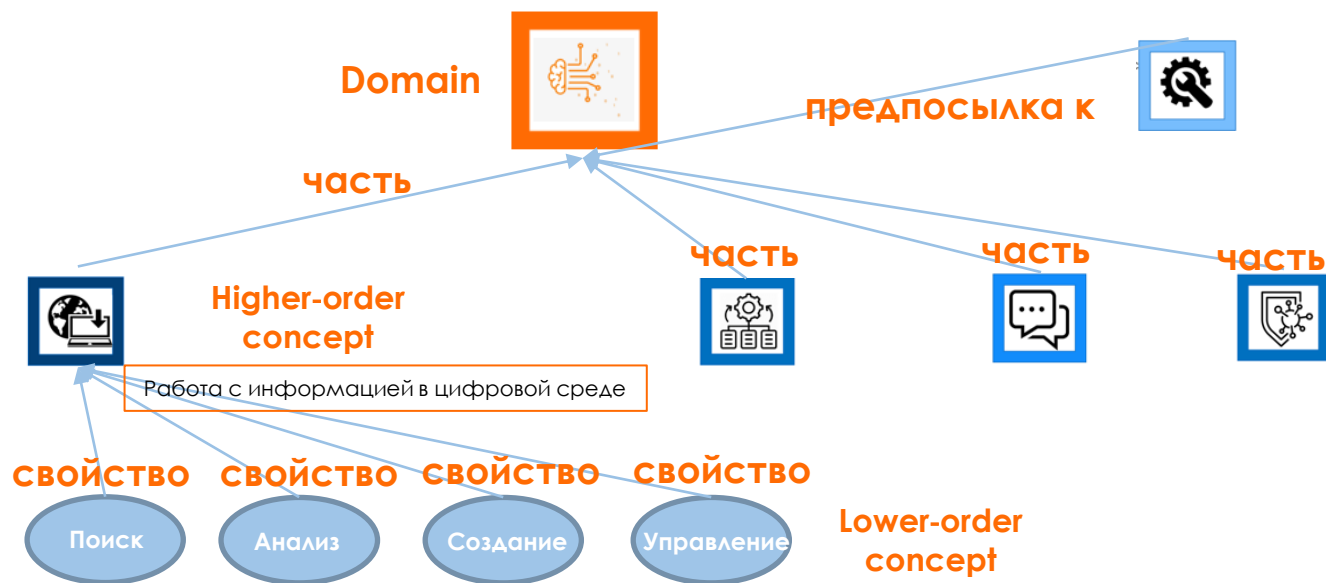
Functional - X создано для или может выполнять функцию в отношении Y - проявление для, контролирует, использует, обеспечивает и тд

Property - X свойство, характеристика, отличительная черта Y - свойство

Location - определяет пространственные отношения - работает на (уровне), под, над и тд



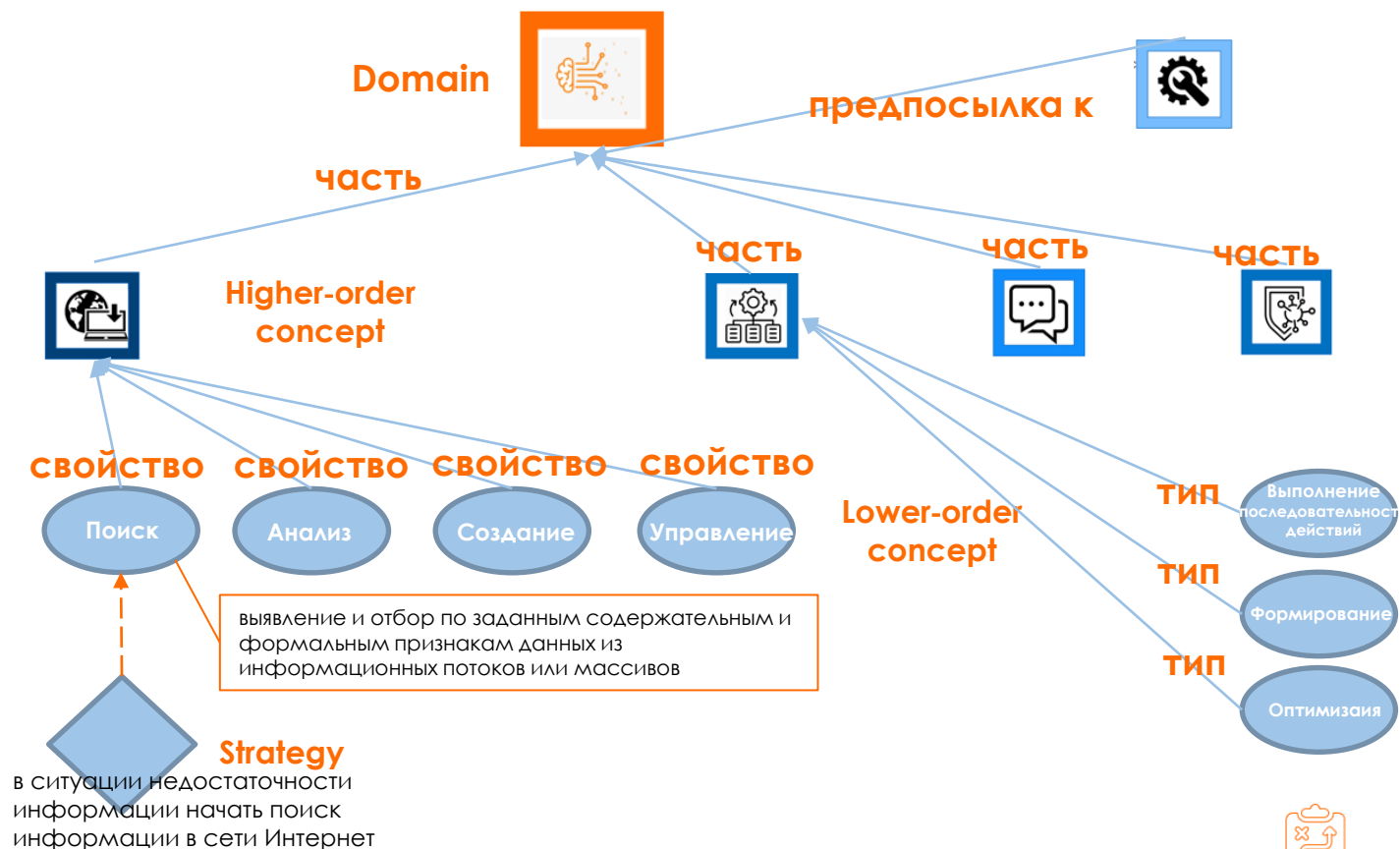
Моделирование области компетенций: результат онтология



Добавление стратегий в онтологию операционализирует конструкт на уровне поведения (демонстрирует овладение каждой концепцией). Это позволяет делать выводы о навыках, понимании или знаниях непосредственно из поведенческих данных в ситуации тестирования, явно связывая стратегии низкого уровня с концепциями, для которых они предоставляют доказательства, таким образом явно определяя, как будет выглядеть правильная реакция или конкретное неправильное представление на поведенческом уровне.

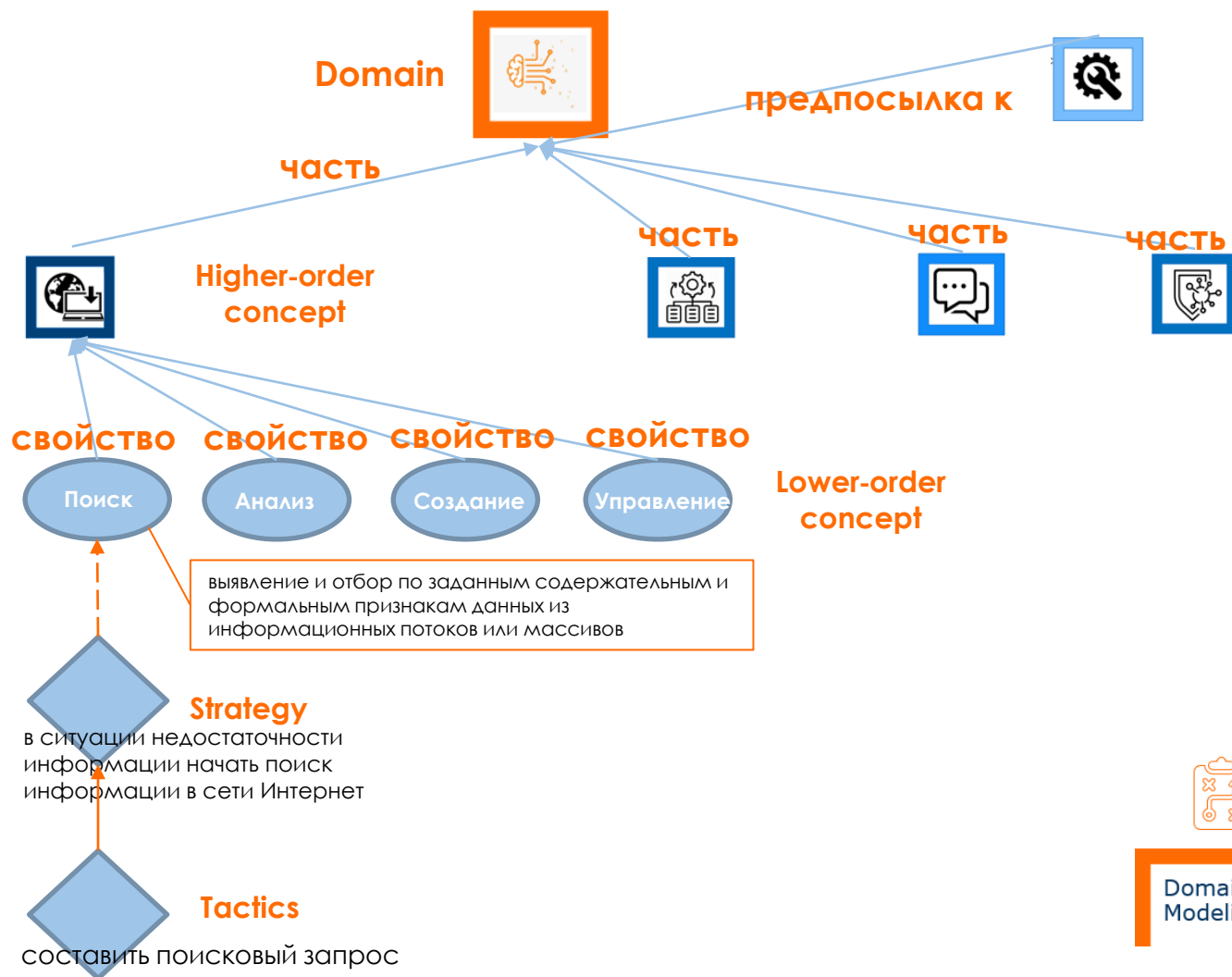


Моделирование области компетенций: результат онтология



Domain
Modeling

Моделирование области компетенций: результат ОНТОЛОГИЯ



2. принятие решения о том, как результаты оценивания будут интерпретироваться (**разработка показателей уровней достижений (Achievement Level Descriptors/Performance Level Descriptors)**).

Правильная разработка показателей уровней достижений - важное условие качества разрабатываемого инструмента инструмента оценивания (валидности и надежности его результатов).

3. **Установление стандартов** (standard setting) - это процесс определения одной или нескольких пороговых оценок теста. Установление стандартов является одним из важнейших шагов в процессе дизайна и разработки теста. На каком уровне респондент должен выполнить тест, для того чтобы соответствовать тому или иному уровню, принятому при разработке этого теста.



Моделирование **области** **компетенций:** **результат** **Pattern Design**

Помогает заполнить пробелы аргумента. Поскольку структура шаблона проектирования неявно содержит структуру аргумента, заполнение шаблона делает явными отношения между частями информации с точки зрения ролей, которые они выполняют.

Название	Описание	Как это в дальнейшем ляжет в САФ
Обоснование	Объяснение почему этот шаблон является важным, позволяет собрать необходимые свидетельства о ключевых компетенциях	
Ключевые компетенции	Описание КК, являющиеся целью этого паттерна	Student Model
Дополнительные компетенции	Другие знания / навыки/умения, которые могут потребоваться	Student Model
Потенциальное наблюдаемое поведение	То наблюдаемое поведение респондентов при взаимодействии с заданием, которое будет свидетельствовать о проявлении КК Касаются конкретных аспектов результатов деятельности. Описывают качества, сильные стороны или степени характеристик продуктов деятельности - свидетельства, которые продукты труда передают о КК (например, " количество... ", " качество... ", " уровень... ", "вид...").	Evidence Model
Потенциальный результат деятельности (work products)	Результат (например, устный ответ, письменное описание), в котором мы можем обнаружить проявление (свидетельства) КК	Task Model
Характерные особенности	Обязательные элементы ситуаций, в которых можно получить свидетельства	Task Model
Вариативные элементы	Аспекты ситуаций, которые можно варьировать для того, чтобы сместить сложность или фокус	Task Model

Разработка концептуальной рамки инструмента



Модель конструкта
(Student Model)

Модель задания
(Task Model)

Модель свидетельств
(Evidence Model)

1) ЧТО ИЗМЕРЯЕМ?

Определяются психологические характеристики, компетенции или знания респондента, которые являются целью оценивания

2) ГДЕ ИЗМЕРЯЕМ?

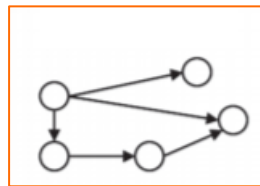
Тип задания, проблемная ситуация, стимульный материал, ожидаемые результаты (наблюдаемое поведение, которое релевантно измеряемому конструкту)

3) КАК ИЗМЕРЯЕМ?

Описывает, каким образом каждый из ожидаемых результатов связан с уровнем выраженности конструкта, шкала

Что мы хотим измерить?

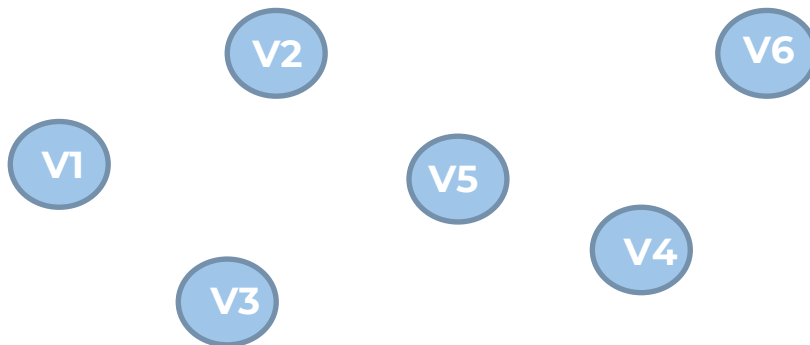
СМ - это набор переменных, которые соответствуют атрибутам учащихся, таким как навыки, знания и способности, о которых вы хотите заявить.



Предположим, мы хотим заявить о способности человека играть в теннис.

Переменные (полный набор переменных, связанных с навыками игры в теннис)?

- Какими навыками должен обладать хороший теннисист?
- Что делают новички?

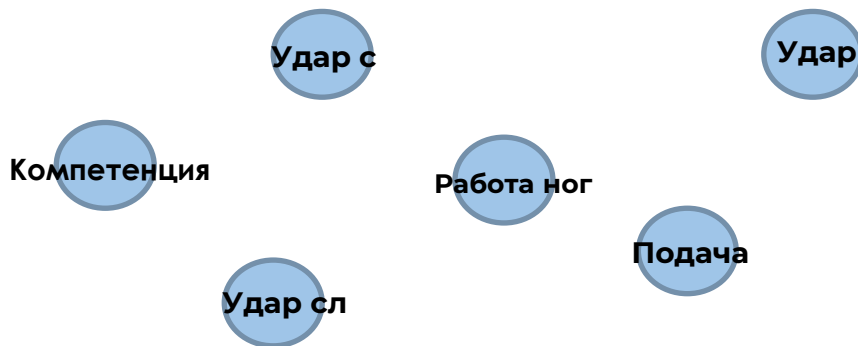


Conceptual Assessment Framework

Модель конструкта

Предположим, мы хотим заявить о способности человека играть в теннис.

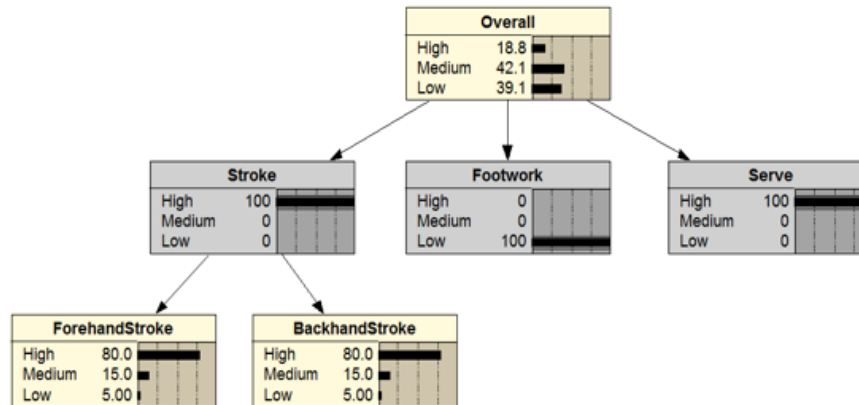
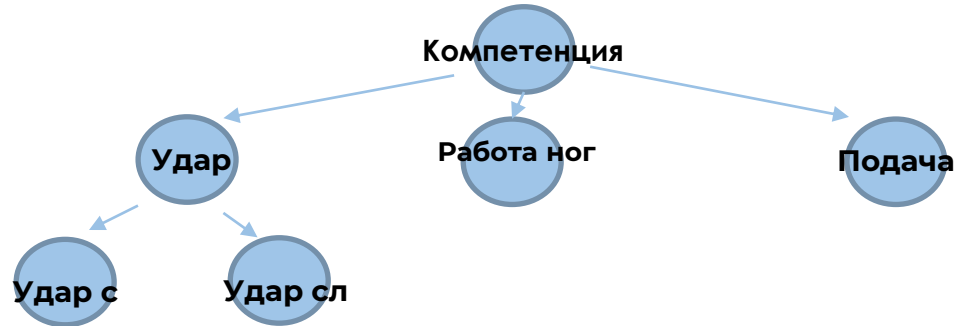
Выбор переменных: удар (как справа, так и слева), работа ног и подача.



Conceptual Assessment Framework

Модель конструкта

- Структура
- «Веса» - теннисные эксперты говорят, что самый важный навык для теннисного выступления-это работа ног, потому что хорошая работа ног является предварительным условием хорошего удара.

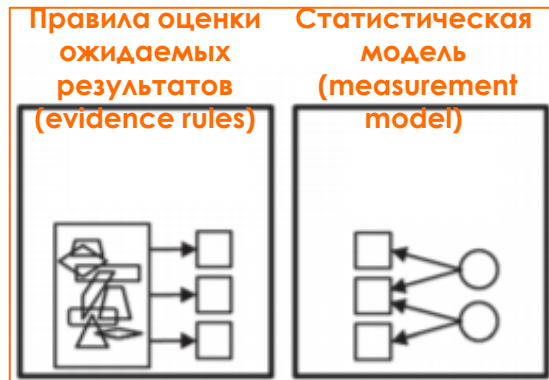


- как наблюдаемые действия могут быть использованы в качестве доказательств для переменных модели компетенций.

Evidence Model

Модель

свидетельств



как каждый уровень индикаторов отличается друг от друга в контексте их связи с измеряемым конструктом

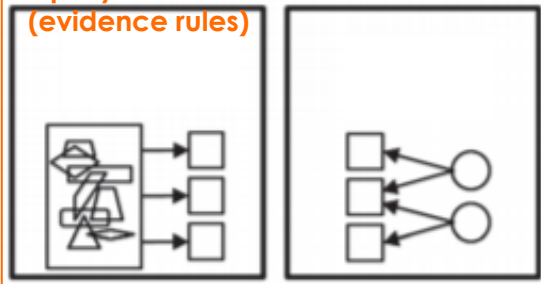
математическая модель, которая связывает уровни индикаторов и уровни латентных конструктов респондентов. Сети Байеса позволяют моделировать любой уровень сложности структуры, поэтому они особенно хорошо подходят для обсчета результатов теста, основанного на ECD. Выбор метода статистического анализа данных — один из ключевых этапов разработки теста и зависит от целей оценивания и типа создаваемого инструмента.

Conceptual Assessment Framework

Модель свидетельств

- как наблюдаемые действия могут быть использованы в качестве доказательств для переменных модели компетенций.

Правила оценки
ожидаемых
результатов
(evidence rules)



Удар с

Индикатор/ Балл

0

1

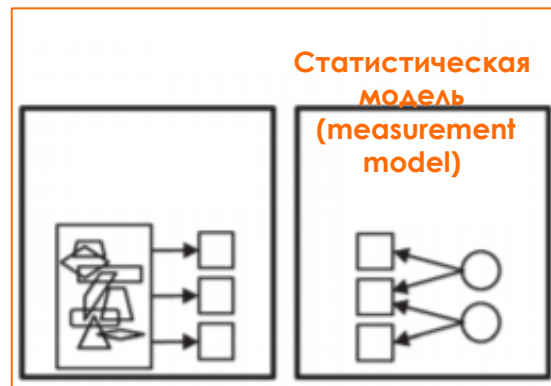
2

Индикатор/ Балл	0	1	2
Форма удара	Неподходящая форма	Подходящая форма, но не соблюден тайминг	Подходящая форма, и соблюден тайминг
Контроль направления мяча с помощью удара справа	Мяч приземлился за пределами линии.	Мяч приземлился внутри линии, но противнику легко отбить его.	Мяч приземлился внутри линии, и противнику трудно отбить его.
Сила удара	Мяч не пересек сетку	Мяч пересек сетку, но противнику легко отбить его.	Мяч пересек сетку, и противнику не удалось отбить его.

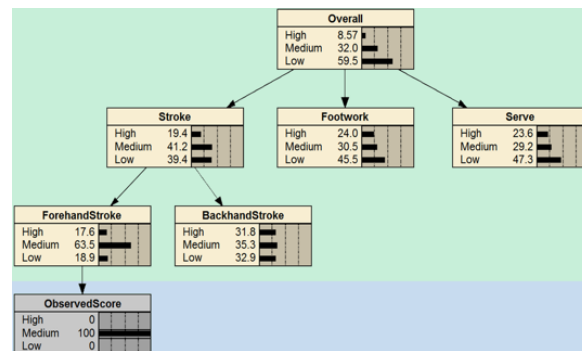
Conceptual Assessment Framework

Модель свидетельств

- как наблюдаемые действия могут быть использованы в качестве доказательств для переменных модели компетенций.



- пропорция (3 балла из 6 – 0.5)
- cut-scores (0.68 – 1.00 High; **0.34 – 0.67 Medium**; 0.00 – 0.33 Low)



Conceptual Assessment Framework

Модель
свидетельств

Conceptual Assessment Framework

Модель задания

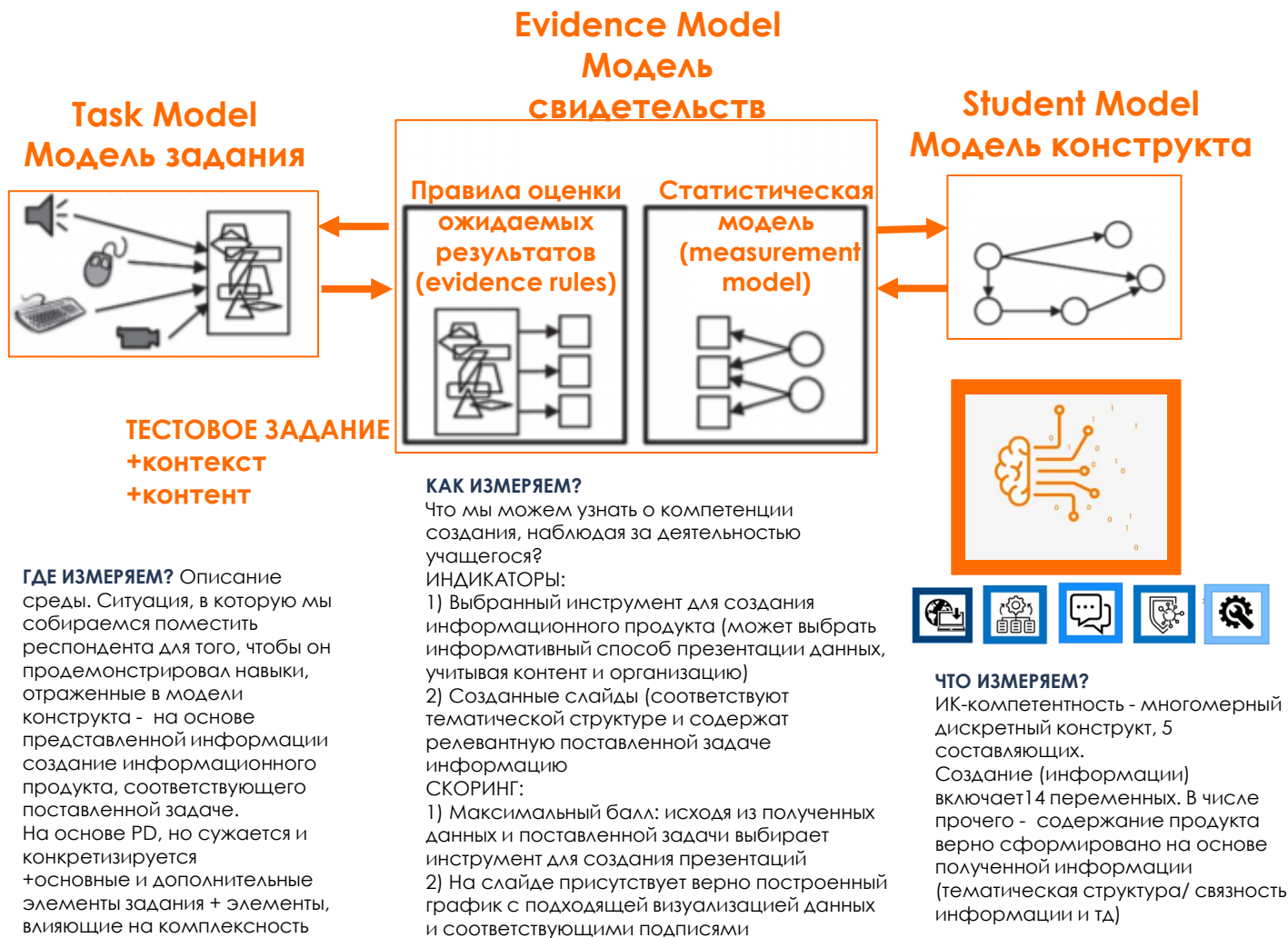
- Ситуация, в которой могут быть воспроизведено то, что мы хотим наблюдать (пример с теннисом: Multiple choice, эссе, демонстрация на корте)
- Вариации ситуации (виды кортов: грунтовые, травяные, хард, ковровые, паркетные, резиновые и асфальтовые)

Выбор средств презентации заданий (presentation model)

В каком виде представлять задания?

Решение о формате теста: бумага/компьютер, множественный выбор/виньетки/сценарий (сценарное задание на экране компьютера - веб-интерфейс, симулирующий различные программные продукты)

Conceptual Assessment Framework



- **Инструкции по сборке варианта теста (assembly model)** –

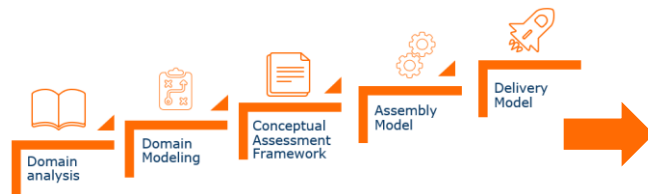
Сколько и каких заданий требуется?

Критерии отбора заданий для конкретного тестирования (количество и список характеристик тестовых заданий, которые должны присутствовать в каждом варианте теста)

- **Способы администрирования (delivery model)**

Какие средства и технологии использовать для представления заданий и сбора ответов?

Варианты обеспечения администрирования: как доставить задание, как записать ответ, как обработать результат (правила скоринга), вернуть ли результат тестируемому, как записать результат и мета-информацию для анализа заданий (ICLTest локально-сетевой способ администрирования, в присутствии администратора, самостоятельное, результаты деятельности, включая контекстуальные, автоматически записываются на сервер через интернет-соединение)



Набор конкретных ответов по этим пунктам дает **Спецификацию теста** (Behrens, Mislevy et al. 2010)

Литература

- Mesick, S. (1992) The Interplay of Evidence and Consequences in the Validation of Performance Assessment ETS Research Report Series, 1992: i-42. doi:10.1002/j.2333-8504.1992.tb01470.x
- Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59(4), 439–483. doi:10.1007/bf02294388
- Toulmin, S. (1958 (2003)) The Uses of Argument Updated Edition Cambridge University Press
- Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6-20.
- Mislevy, R. J., & Riconscente, M. (2005). Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9). Menlo Park, CA: SRI International.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). A brief introduction to evidence-centered design (CSE Technical Report 632). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, UCLA. Retrieved from <http://www.cse.ucla.edu/products/reports/r632.pdf>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the role of task model variables in assessment design. In S. H. Irvine & P. C. Kyllonen (Eds.) *Item generation for test development* (pp. 97-128). Mahwah, New Jersey: Lawrence Erlbaum.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-62.

Литература

- Mislevy R.J. (2018) Sociocognitive Foundations of Educational Measurement p.73 Routledge
- Ercikan, K; Oliveri, M. E. (2016). In Search of Validity Evidence in Support of the Interpretation and Use of Assessments of Complex Constructs: Discussion of Research on Assessing 21st Century Skills. Applied Measurement in Education, 29(4), 310–318. doi 10.1080/08957347.2016.1209210
- Kerr, D; Andrews-Todd, J; Mislevy, R. (2016). The In-Task Assessment Framework for Behavioral Data. in The Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications (pp.472-507) John Wiley 10.1002/9781118956588.ch20.
- Marion, S; Landl E (2017) Principled Assessment Design for the Performance Assessment of Competency Education
https://www.nciea.org/sites/default/files/publications/PACE%20Principled%20assessment%20design_092417.pdf
- Pellegrino J.W; Chudowsky N; Glaser, R. (2001) Knowing what Students Know The Science and Design of Educational Assessment Board on Testing and Assessment Center for Education Division of Behavioral and Social Sciences and Education National Research Council National Academy Press.
- Ferrara, S; Lai, E; Nichols, P (2016). Principled Approaches to Assessment Design, Development, and Implementation in The Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications (p.42) John Wiley. 10.1002/9781118956588.ch3

