

Все, что вы хотели знать об инструменте «4К»,
но боялись спросить 😊



Как ECD работает на практике инструмента 4К

Уганова Ирина, научный сотрудник
Грачева Дарья, стажер-исследователь
Тарасов Сергей, стажер-исследователь

Цели доклада

Для слушателей:

1. Познакомиться с разработкой в методологии ECD и особенностями ее применения в заданиях сценарного типа для измерения сложных конструктов
2. Узнать о связке тест-девелопмента и психометрики

Для нас:

Обсудить важные для нас проблемы – рады комментариям и вопросам.

План доклада

- Как происходит создание заданий
- Что происходит после создания заданий
 - обработка данных
 - обратная связь
- Вопросы на обсуждение

Условные обозначения

КМ - критическое мышление

КР - креативность

КО - коммуникация

КП - кооперация

ОС - обратная связь

Индикатор - поведенческое проявление навыка в сценарном задании

Инструмент 4К – это...

Задания сценарного типа для измерения критического мышления, креативности, коммуникации и кооперации

- Игровой формат
- Автоматическая проверка
- Ведет историю с 2017 года

Сейчас в инструмент входит:

- 12 заданий для 4 класса
- 2 задания для 5 класса
- 12 заданий для 7 класса
- 4 задания в рамках проекта с Учи.ру (4-5 класс).

По некоторым заданиям было уже около 6 итераций разного масштаба правок (и вы поймете почему)

Дино	Ёж
Аквариум	Террариум
Монстр	Робот
Спектакль	Конкурс
Торт	Королевство
Путешествие	Лабиринт
Привидение (Old)	Инопланетянин
Привидение	
Пещера	Хижина
Вселенная	Подводный мир
Поход	Сад
Космический корабль	Племя
Мероприятие	Аттракционы
	Крушение корабля
	Экопарк

И еще большая команда 4К

Большой проект – большая команда

Угланова Ирина
руководитель проекта



Грачева Дарья
психометрист,
разработчик заданий,
аспирант, 2 курс



Тарасов Сергей
психометрист
разработчик заданий,
аспирант, 2 курс



Вырва Елена
разработчик
заданий, отв. за
ког.лаб.



Хмелевская Яна
разработчик заданий,
менеджер
магистрант, 2 курс



Сергеева Марина
разработчик заданий



Каган Софья
магистрант, 1 курс



Гранатурова
Александра
магистрант, 1 курс



+ программист, дизайнер, два иллюстратора

Как происходит создание заданий

Этапы

1. Определить цель тестирования

И это не всегда это просто!

Изначальная цель – мониторинг. Последовавшая – оценка прогресса в развития навыка.

Но такие задания (игровые, контекстные) отлично подходят для **формирующего** оценивания или для предоставления обратной связи более детализированной, чем необходимо в мониторинге.

Вопрос для обсуждения: для мониторинга допустима более низкая надежность. А для достижения достаточной надежности в более детальной ОС нужно увеличивать количество индикаторов. Но пользователям теста не нравятся длинные сценарии. Как выстроить методологически четкую процедуру контроля длины заданий?

Этапы

2. Построить теоретическую (концептуальную) рамку

Обычно мы работаем с различными теориями, которые, перемешиваясь между собой, создают концептуальную рамку для инструмента. Для ориентировки в теориях мы привлекаем экспертов по содержанию. Но возникает множество вопросов – сохраняется ли структура конструкта для разных возрастов, если конструкт состоит из фасетов, подразумевается ли их последовательность в поведении?

Пример проблемы - креативность. Существует ОГРОМНОЕ множество **разных** плохо структурированных теорий. Даже мета-анализы не помогают.

Этапы

3. Выбрать формат задания

Мы решили когда-то, что нам нужно нагородить огород:) И больше мы не сдаемся! Это из-за того, что мы уверены, что такая гибкая форма не ограничивает проявления конструкта.

Нам помогают другие исследования, где постулируется такая же позиция (Ku, 2009). Например, в тестах на collaborative problem solving активно используют имитации чатов, мы тоже их ввели.

Этапы

4. Техническая реализация заданий

Подготовка технических заданий внешним исполнителям: иллюстратору, дизайнеру и программисту

5. Реализация тестирования

Построение системы сборки теста (админка)

Подготовка форм обратной связи

Анализ данных

Форматы заданий – какие есть и зачем такие

Формальные операции

Способность к уравниванию традиционно проверялась в ходе интервью, но есть исследования, которые говорят о возможности введения стандартизированной диагностики.

Конкретно-операциональное мышление – задания с опорой на реальные объекты или их изображения, т.к. мышление преимущественно в образах.

Появление формальных операций позволяет мысленно устанавливать сложные логические отношения. Для диагностики формальных операций – текстовое предъявление задания.

Мы комбинируем текстовое и рисуночное предъявление, а также имитируем работу с физическими объектами (Уганова, Погожина, 2021).

Какое длинное письмо! Зачем тебя пригласила Валерия Ивановна? Выдели для меня предложение, где это сказано.

Дорогой эксперт!

Только на вас вся надежда! На мою семью наложено проклятие: многие её члены становятся привидением и живут в нашем доме. Они невидимые и даже беспредельные. Но я отчтёльно чувствую их присутствие. Мы не знаем, с чем это связано, и очень боимся стать привидениями сами.

Помогите разгадать секрет семейного проклятия. В комнатах дома хранятся вещи членов семьи. Они помогут в разгадке. Если думаться, привидений очень жалко - заперты в своих комнатах и выбираться не могут.

С уважением,
Валерия Ивановна.

Готово

Станет ли Клавдия привидением? Да Нет

Клавдия Ивановна	Любит чай	Носит платок с гербом	Ходит к занавеске
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Не привидение	Любит чай	Носит платок с гербом	Ходит к занавеске
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Богдана Васильевна сидит в кресле	Любит чай	Носит платок с гербом	Ходит к занавеске
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Не привидение	Любит чай	Носит платок с гербом	Ходит к занавеске
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Арина Сорокина	Любит чай	Носит платок с гербом	Ходит к занавеске
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Привидение	Любит чай	Носит платок с гербом	Ходит к занавеске
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Привидение	Любит чай	Носит платок с гербом	Ходит к занавеске
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Привидение	Любит чай	Носит платок с гербом	Ходит к занавеске
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Форматы заданий. Критическое мышление

Аквариум – последовательность действий для оценки КМ

Выдели предложения с важной информацией о том, что точно понадобится для аквариума, и скопируй их в блокнот. Когда закончишь, нажми кнопку «Готово».

Мы любим крабов! Обсуждение на Дети.ру

Обустройство аквариума начинается с укладки грунта. На дно аквариума для крабов нужно положить слой песка. А вот рыбкам подойдет более крупный грунт.

Кrabам нужно иногда залезать повыше, для этого в аквариум помещают камни. Рыбки тоже любят прятаться среди камней. Ты можешь принести камни с речки. Чтобы крабам было удобно, по [Добавить в блокнот](#) побольше и поменьше. Рыбкам нравятся более высокие камни. Так же каждому крабу понадобится домик из коряги или черепка, чтобы отдыхать днём. В магазине продаются домики в виде старинных замков. Рыбки

?

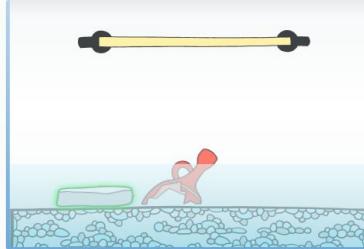
Готово

Блокнот

- Чтобы крабам было удобно, понадобятся камни побольше и поменьше.
- Рыбки тоже любят прятаться среди камней.
- А вот рыбкам подойдет более крупный грунт.

Готово

Помоги мне обустроить аквариум. Пользуйся подсказками в блокноте. Когда закончишь, нажми кнопку «Готово».



Готово ?

Домики Камни Грунт Разное Украшения Соль

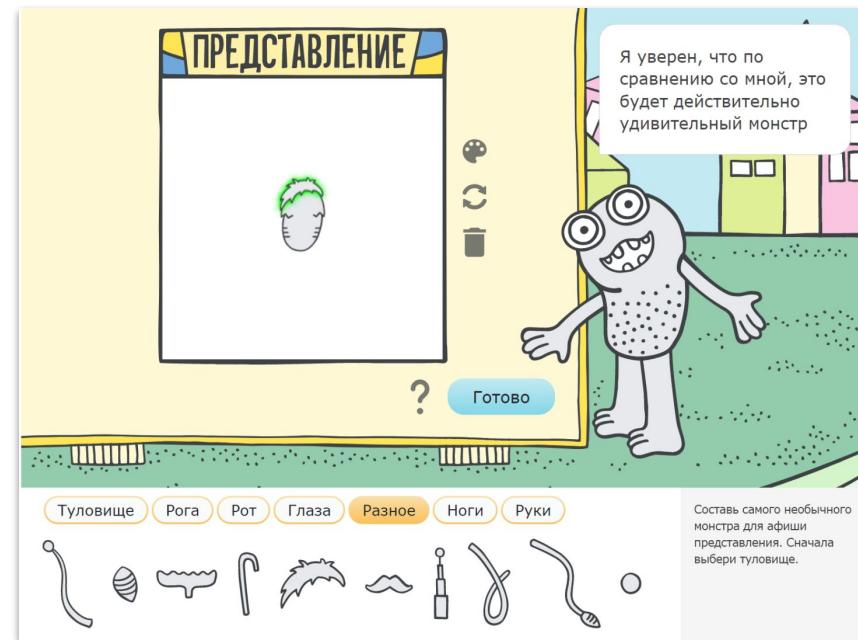
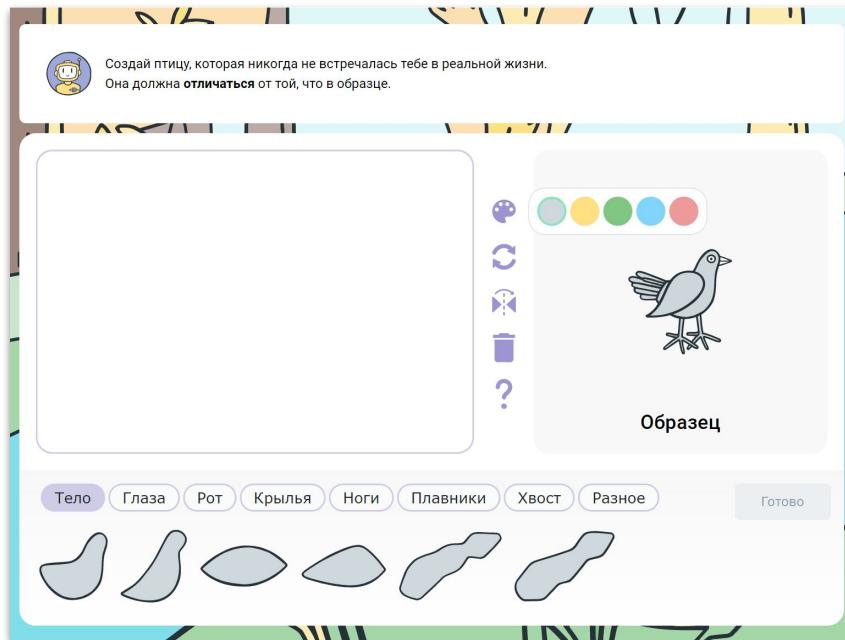
Коряга Коряга Коряга Черепок

Блокнот

- Чтобы крабам было удобно, понадобятся камни побольше и поменьше.
- Рыбки тоже любят прятаться среди камней.
- А вот рыбкам подойдет более крупный грунт.
- Также каждому крабу понадобится домик из коряги или черепка, чтобы отдыхать днём.

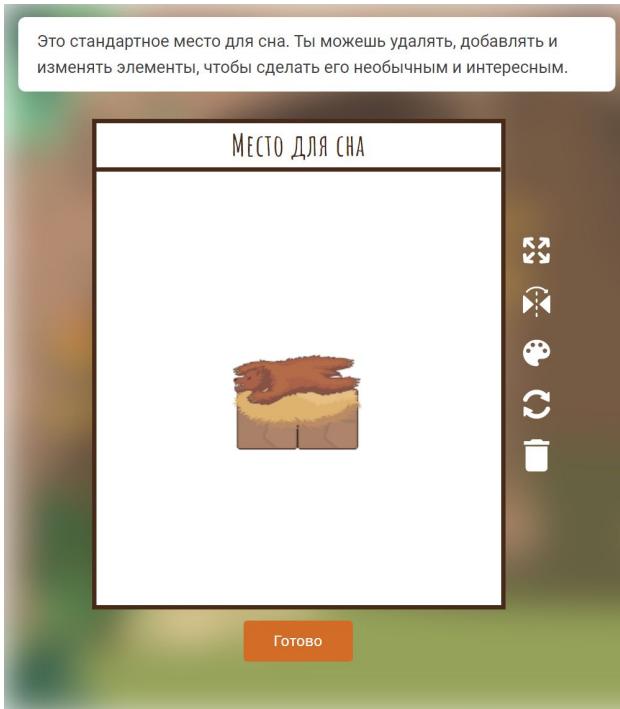
Форматы заданий. Креативность

Построенное изображение сравнивается с референсом, который может быть существующим или несуществующим



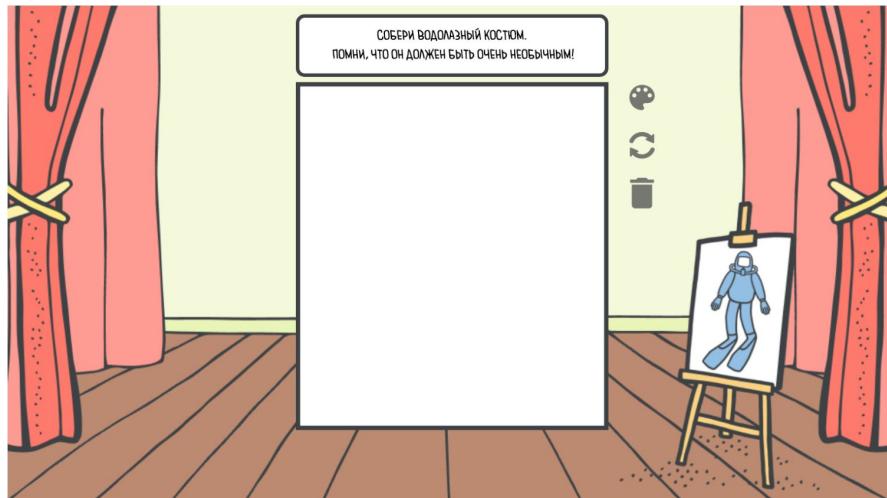
Форматы заданий. Креативность

Также референс также может быть помещен сразу в поле конструктора.



Форматы заданий. Креативность

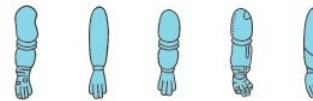
В некоторых заданиях оценивается выбор более детализированных элементов внутри категории



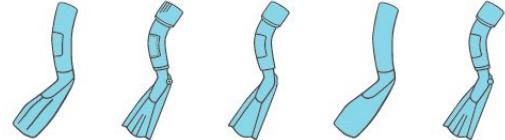
Готово



Корпус Рукава Штанины Шлем Разное



Корпус Рукава Штанины Шлем Разное

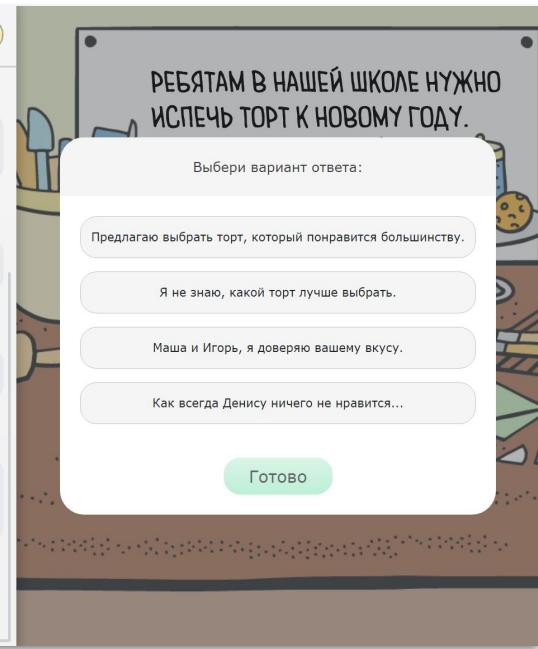
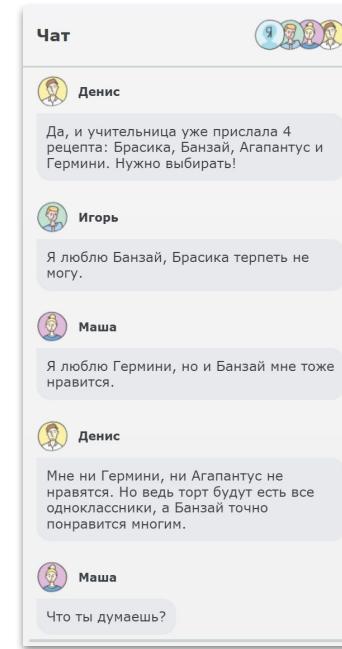
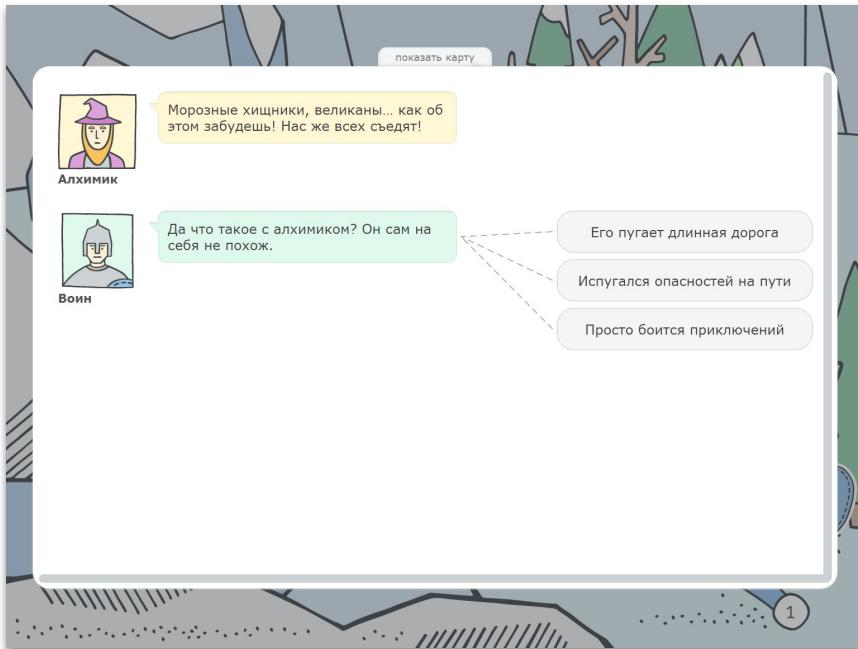


Корпус Рукава Штанины Шлем Разное



Форматы заданий. Коммуникация и коопeração

Задания построены на общении с аватарами, которое происходит с помощью
ответа на реплику аватара или выбора реплики в чате



Форматы заданий. Коммуникация и кооперация

В заданиях нет совершенной свободы для тестируемого.

Для снятия всех индикаторов мы должны приводить всех к одним и тем же сюжетным точкам. Для этого используются следующие стратегии:

- Пересечение
- Спасение

Источник: "Convergence" (да, также, как в статистике) and "Rescue"
(He, von Davier, Greiff, & Steinhauer, Borysewicz, 2017)

Пересечение и спасение – методы разработки заданий

Пересечение

Пересечение - ситуация, когда независимо от того, какой выбор делает тестируемый, сюжет задания приводит его к одной и той же стадии решения проблемы (точки конвергенции).

 **Оля**
Любит читать и сочинять истории

Вот, какие у меня есть книги:
- История изучения океана
- Кто живёт на дне моря
- Как делают одежду для подводного плавания

Я помогу тебе выбрать книгу

Перейти к своему заданию

 **Оля**
Любит читать и сочинять истории

И всё таки я не могу определиться.
Какую мне взять книжку?



Оля

Любит читать и сочинять истории

Нажми на название книжки, которую мне нужно прочитать.

История изучения океана

Кто живёт на дне моря

Как делают одежду для подводного плавания

! Даже если Т. решил не помочь, ему все равно надо сделать выбор

Спасение

Спасение - предоставление необходимой информации через агентов (аватаров) или какой-либо другой механизм в ситуации, когда тестируемый делает неоптимальный выбор.

 **Петя**
Любит рисовать и раскрашивать

Теперь нам нужно распределить задачи и решить, кто что будет делать. Рядом с каждым заданием нужно отметить одного участника, который будет делать это задание.

Прочитать, как выглядит водолазный костюм

- Петя Оля Я

Раскрасить детали костюма

- Петя Оля Я

Собрать костюм (самое важное)

- Петя Оля Я

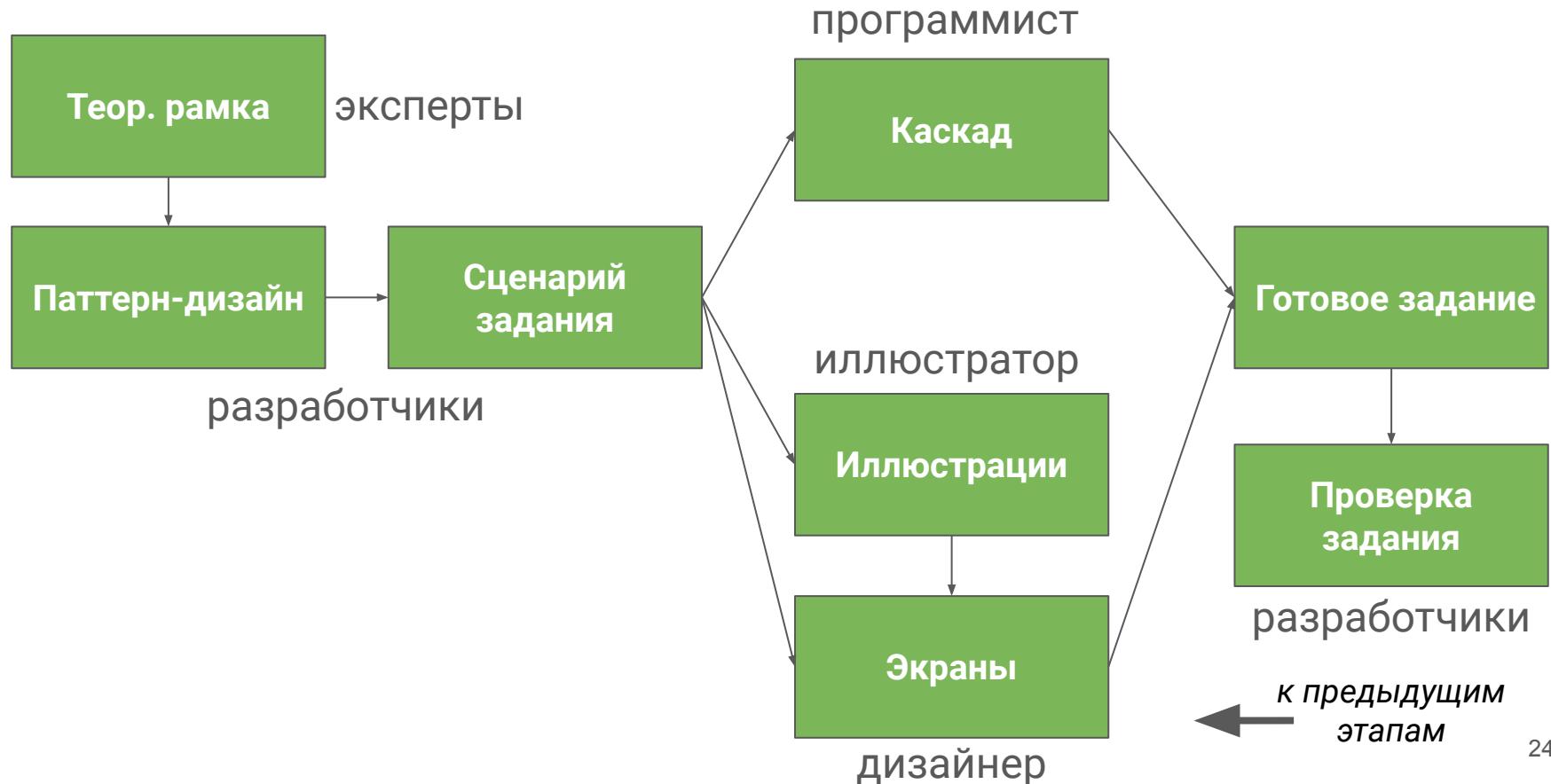
 **Оля**
Любит читать и сочинять истории

Режиссёр сказал мне прочесть описание, потому что у меня болит рука. А тебе сделать самое важное, собрать костюм. Петя будет красить

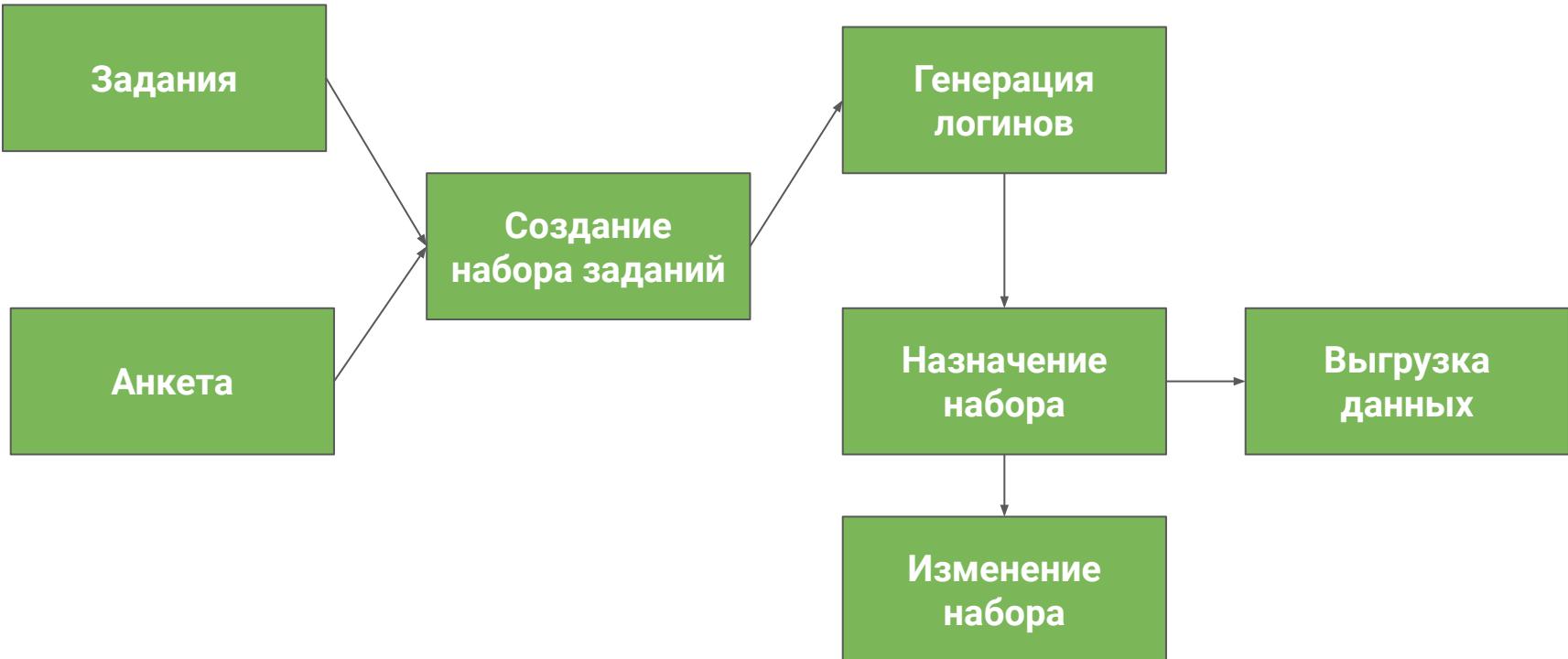
! При правильном распределении сообщение Оли не появляется

Техническая реализация заданий

Краткая схема создания заданий



Сборка теста



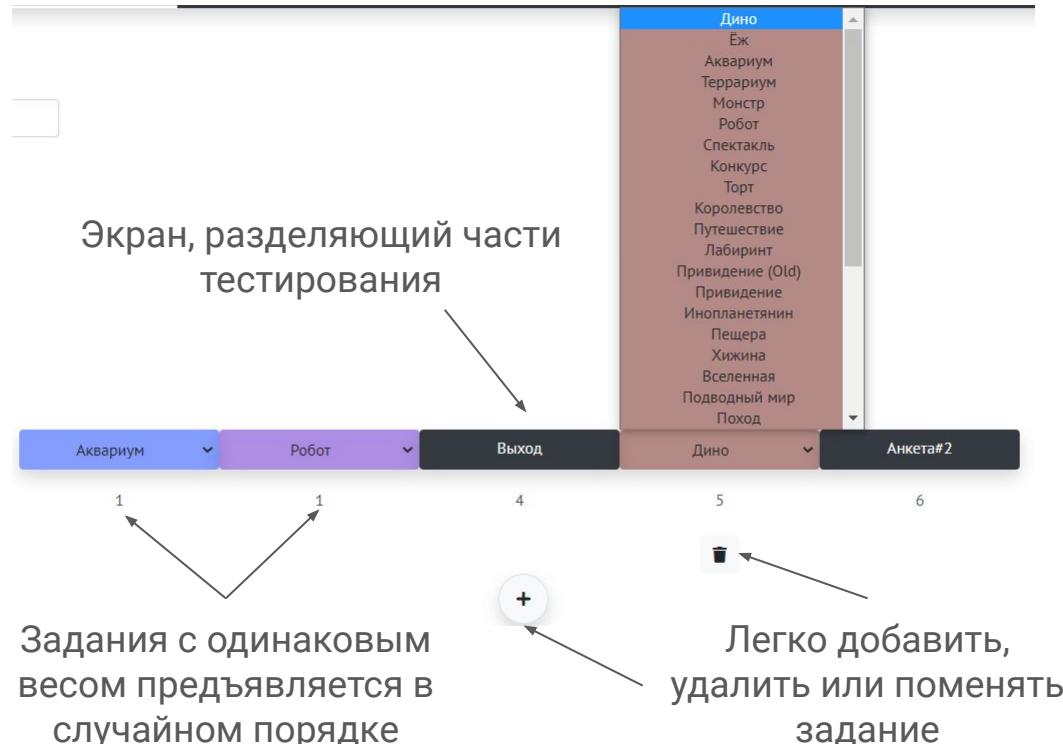
Создание набора заданий

Конструктор наборов

Система весов:

случайный порядок заданий

Разделительные экраны



Генерация логинов

Уникальный логин с номером региона, школы и класса - возможность фильтрации

Создание логинов для одной школы за раз (трудоемко, когда много школ)

Назначение созданных наборов в процентном соотношении (50% первый набор, 50% второй набор или 100% один набор или другое разбиение)

Создание новых логинов

Префикс ▾

Номер региона

1

Номер школы

2

▲ Класс №1 +

Номер класса

1

Количество учеников (количество логинов)

25

Класс

4

Поиск по названию набора

ID#2 9/09_Ког.лаба_Экопарк 15 %

ID#1 9/09_Ког.лаба 40 %

Суммарное значение процентов не может быть меньше 100

Создать

Выгрузка данных

Возможности выгрузки:

- картинок (напр., монстров)
- индикаторов
- индикаторов с параметрами

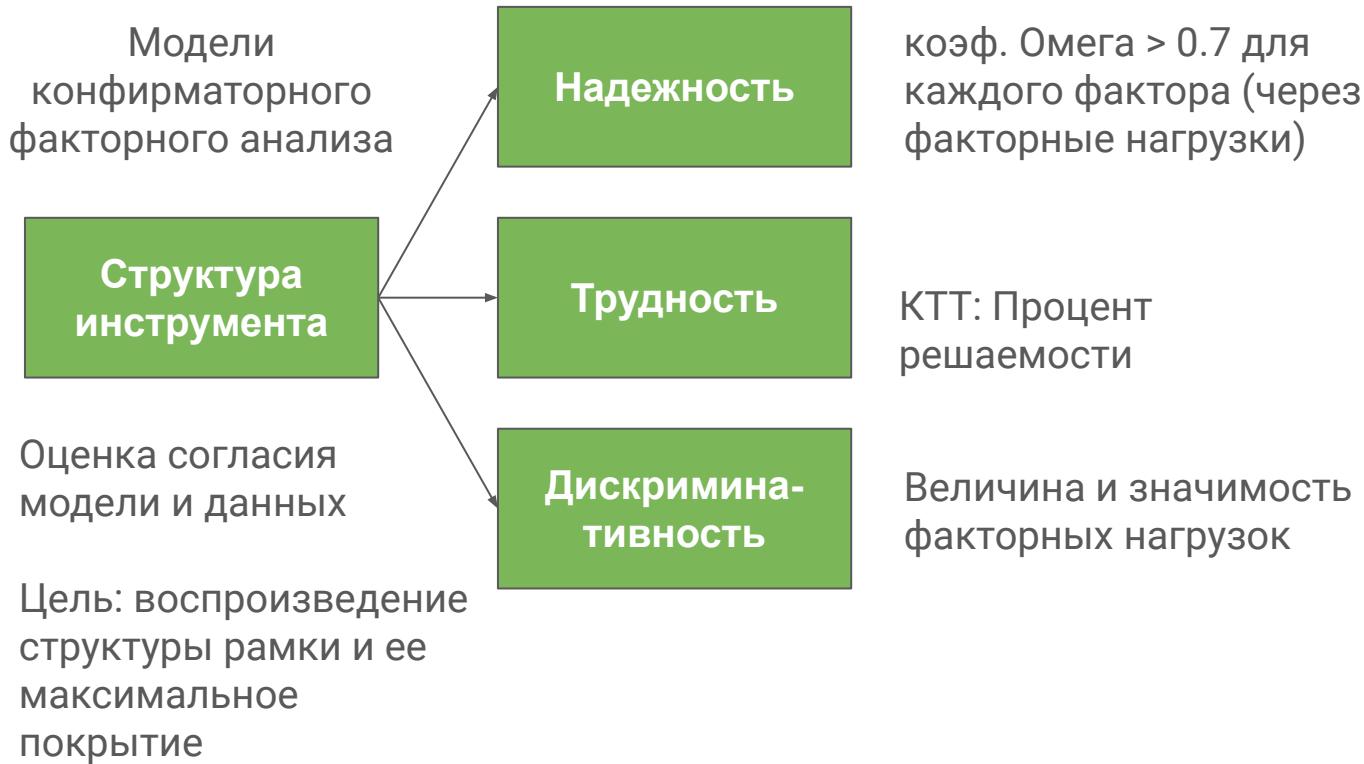
С применением фильтров по:

- логину
- дате
- заданиям
- и т.д.

The screenshot shows a user interface for data export. At the top right are buttons for 'Экспорт' (Export) and 'Создать Логины' (Create Logins). On the left, there's a dropdown menu 'Архивы' (Archives). A sidebar on the right lists categories: 'Результаты' (Results), 'Результаты (только индикаторы)' (Results (only indicators)), 'Нетика' (Netaika), 'Логины' (Logins), and 'Картинки' (Pictures). The main area contains several input fields and dropdowns for filtering: 'Логин (поиск)', 'Префикс', 'Номер региона', 'Номер школы', 'Номер класса', 'Класс', 'Дата начала тестирования', 'Дата создания', 'Дата изменения', and 'Дата последней авторизации'. Below these are two rows of checkboxes for filters: 'Дино', 'Ёж', 'Аквариум', 'Террариум', 'Монстр', 'Робот', 'Спектакль', 'Конкурс', 'Торт', 'Королевство', 'Путешествие', 'Лабиринт', 'Привидение (Old)', 'Привидение', 'Инопланетянин', 'Пещера', 'Хижина', 'Вселенная', 'Подводный мир', 'Поход', 'Сад', 'Космический корабль', 'Племя', 'Мероприятие', 'Аттракционы', 'Крушение корабля', 'Эколарк', 'Геном людей', 'Атомная энергетика', 'Происхождение сумчатых', 'Персональные данные', and 'Персональные данные (New)'. At the bottom, there are two sliders: 'Количество выполненных заданий (больше или равно)' (Number of completed tasks (greater than or equal to)) with a value of 0, and 'Количество заданий в teste ("-1" - фильтр отключен)' (Number of tasks in teste ("-1" - filter disabled)) with a value of -1.

Что происходит после создания заданий

Обработка данных



Локальная зависимость

Стандартное допущение любой вероятностной психометрической модели – условная независимость (conditional independence) всех наблюдаемых переменных

$$P(X_1 = x_1 \text{ and } X_2 = x_2 | \theta) = P(X_1 = x_1 | \theta) * P(X_2 = x_2 | \theta)$$

Но наблюдаемые переменные могут быть связаны не только за счет общего целевого конструкта. Есть разные источники (Yen, 1993) возникновения локальной зависимости.

Например, локальная зависимость из-за принадлежности к одному стимульному материалу (source of evidence) – из-за эффекта контекста задания (DiCerbo, Xu, Levy, & Lai, Holland, 2017).

В психометрической литературе также известен термин Testlet Effect.

Проблемы со структурой. Пример 1

Предполагается, что в КР есть две составляющие: Оригинальность и Детальность.

В теориях: противоречивые данные о силе их связи (Hee Kim, 2006; Bart et al., 2017).

У нас:

- в 4 классе – ортогональные факторы.
- в 7 классе – один фактор.

В разных классах разные модели заданий: в системе начисления баллов выбор элементов для оценки оригинальности по-разному влияет на оценку детальности.

Оценка оригинальности и детальности строится сразу на одном созданном учеником объекте. Кажется, эта проблема преодолима только если делать разные объекты для оригинальности и детальности.

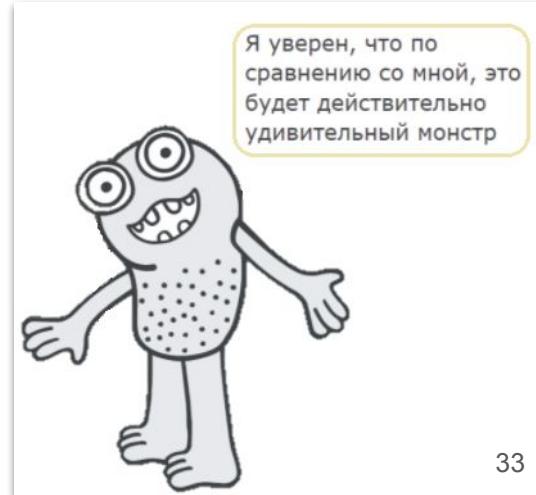
Проблемы со структурой. Пример 2

Индикаторы из разных заданий (Экопарк и Монстр) не вкладываются в фактор Оригинальность.

Мы объясняем, что это из-за особенностей контекста: в одном - реалистичный, в другом - фантастический. То есть у элементов разные формы, а реалистичный референс - более очевидный.



Образец



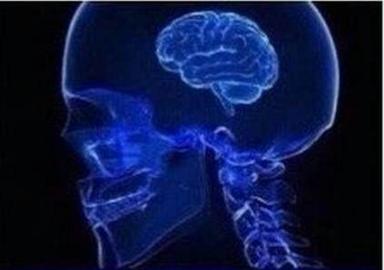
Проблемы со структурой. Пример 3

По теоретической рамке,
КО состоит из трех факторов, КП – из четырех.

По результатам анализа данных –
это один конструкт. Связи между факторами
были порядка 0,9. И это результат двух
итераций, на второй из которых принимались
отдельные усилия по разграничению
факторов.

Проблема – предоставление подробной
обратной связи.

Фактор на
КОКП



Отдельные
факторы на
КО и КП



Факторы на
составляющие



Факторы на
способности



Проблемы с дискриминативностью

Неправильный скоринг

Проверить выгрузку данных (правильный ответ закодирован как 0)

Проблемное задание

Найти причины (плохие дистракторы, проблемы с интерфейсом, непонятная инструкция или бог знает что)

Старая версия: необходимо поставить камень выше уровня воды → отрицательная дискр.

Ник:

Важно, чтобы у каждого краба был отдельный островок, чтобы иногда он мог вылезать из воды, и свой домик. Ты можешь сделать их самостоятельно.

The screenshot shows a mobile application interface. At the top, there is a search icon and the text "Задание" (Task). A yellow callout bubble says: "С помощью этой нужной информации, которая была в тексте, помоги мне обустроить аквариум для двух наших крабов!" (Using this useful information from the text, help me to arrange an aquarium for our two crabs!). Below this is a large rectangular area containing a blue water layer. Inside the water, there is a grey rock and some green aquatic plants. At the bottom of this area is a green button labeled "Готово" (Done). Below the water layer, there is a horizontal menu bar with several tabs: "Грунт" (Soil), "Островок" (Island), "Украшения" (Decorations), and "Разное" (Miscellaneous), with "Разное" being the active tab. Below the menu, there are three items with arrows pointing to them: a small green plant, a larger green plant, and a blue stick-like object.

Итоговое решение: убрать уровень воды из задания → исправление индикатора с отр. дискр.

Проблемы с трудностью

Индикатор креативности: преобразование предмета

Очень трудные индикаторы: учащиеся ничего не меняли

Старая версия:



Причины: Непредсказуемость и комплексность интерактивной тестовой среды, сложные схемы

скоринга, непонятная инструкция
Решение: перенести инструкцию на экран

После изменений:



Проблемы из-за локальной зависимости

Разнообразие контекстов заданий решает две проблемы: повышение экологической валидности и разнообразие факторов контекста.

Когда мы стоим факторную модель, мы **не знаем**, что это за выделенный фактор.

Это может быть как целевой навык, так и эффект контекста **целиком**.

Мы делаем разные задания в том числе, чтобы
НЕ ВСЯ ДИСПЕРСИЯ БЫЛА ИЗ-ЗА КОНТЕКСТА.

Проблемы из-за локальной зависимости

То есть мы сами создаем дополнительные источники дисперсии, но без них не можем.

Важно, чтобы разработчики знали, что принадлежность к одному стимульному материалу (source of evidence) приводит к локальной зависимости.

Локальная зависимость иногда называется unmodeled correlations. Чтобы учесть эту доп.связь индикаторов, можно вводить доп.факторы – “факторы контекстов”

Дополнительные факторы “отщипывают” ранее необъясненную дисперсию. На практике это часто уменьшает размер факторных нагрузок на целевой фактор.

Иногда (если индикаторы принадлежат одному фактору) решить локальную зависимость можно путем объединения индикаторов в полигомомический. А это ведет к формальному уменьшению количества индикаторов, что уменьшит надежность.



Проблемы из-за локальной зависимости. Пример 1

Традиционный пример – common passage

Внутри осмысленного абзаца текста предложения связаны.

Поэтому если задание предполагает выделение предложений из текста – и разработчик, и психометрист должны быть готовы к тому, что каждый выбор будет зависеть не только от целевой способности, но и от самого факта принадлежности к тексту.

Выдели предложения с важной информацией о том, что точно понадобится для аквариума, и скопируй их в блокнот. Когда закончишь, нажми кнопку «Готово».

Как подготовить аквариум для крабов? Рекомендации Московского Зоопарка

Обустройство аквариума начинается с укладки грунта. На дно аквариума для крабов нужно положить слой песка. А вот рыбкам подойдет более [Добавить в блокнот](#)

Крабам нужно иногда залезать повыше, для этого в аквариум помещают камни. Рыбки тоже любят прятаться среди камней. Ты можешь принести камни с речки. Чтобы крабам было удобно, понадобятся камни побольше и поменьше. Рыбкам нравятся более высокие камни. Также каждому крабу понадобится домик из коряги или черепка, чтобы отдыхать днём. В магазине продаются домики в виде старинных замков. Рыбки

Блокнот

1. Обустройство аквариума начинается с укладки грунта.

2. На дно аквариума для крабов нужно положить слой песка.

3. А вот рыбкам подойдет более крупный грунт.

Готово

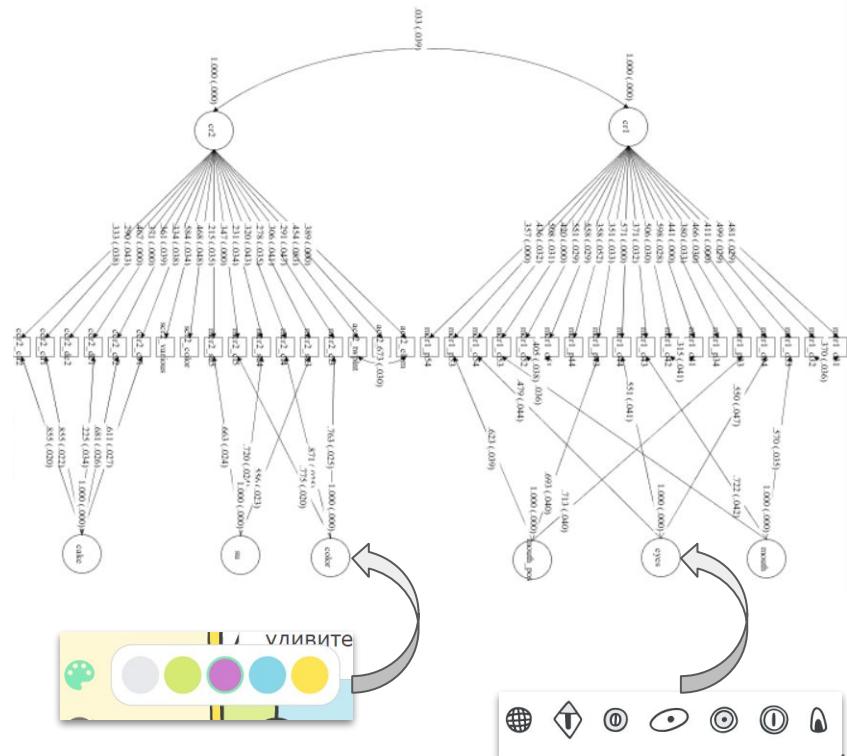
Проблемы из-за локальной зависимости. Пример 2

Монстр:

Если необходимо в интерактивной среде построить несколько похожих объектов, то добавление разных предметов может быть связано.

Фактор цвета, поворота и т.д.

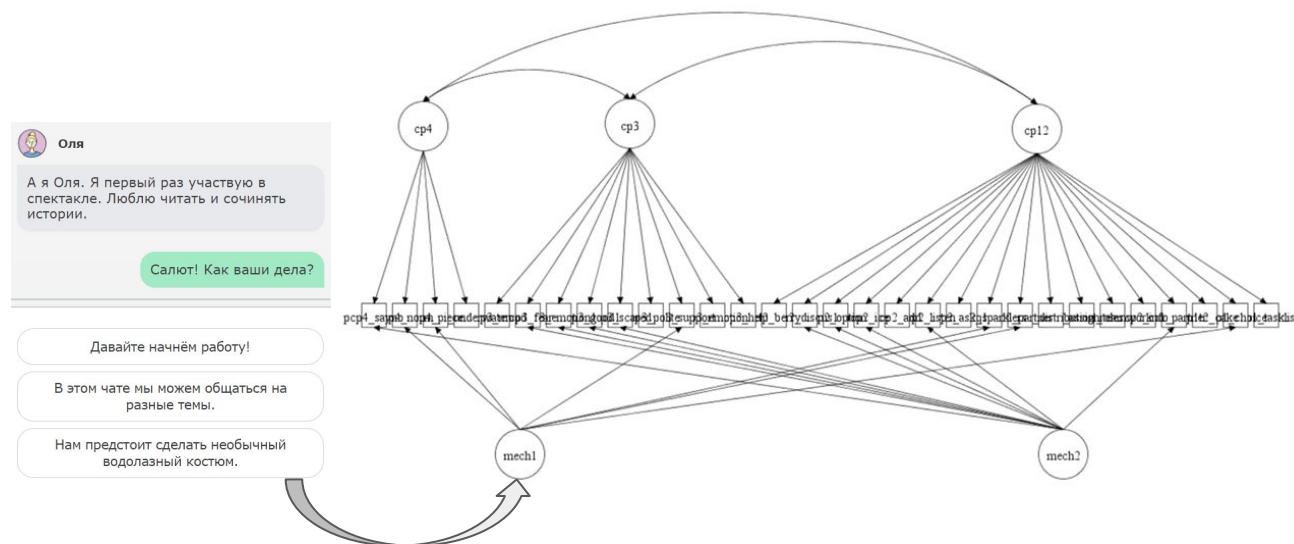
Если выбирает определенную ногу, то к ней определенную руку?



Проблемы из-за локальной зависимости. Пример 3

В заданиях для КО и КП есть своя механика – ответ на реплику в чате.

Она приводит к появлению одного возможного фактора контекста, который встречается в двух сценариях – Спектакль и Торт.



Скоринг в гибкой среде

Разработчик может задавать не только линейную структуру, но и сложные взаимосвязи

Но моделировать это непросто - теоретически можно в BN, IRTrees...

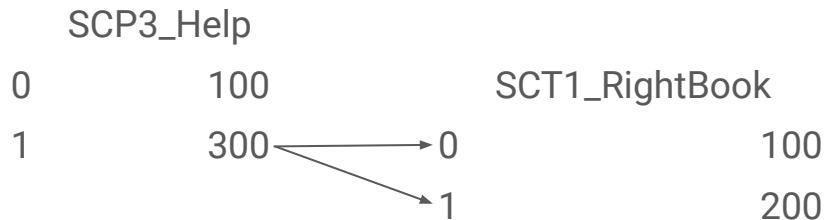
Необходимо это учитывать в скоринге и не допускать взаимоисключающих действий.

Скоринг в гибкой среде

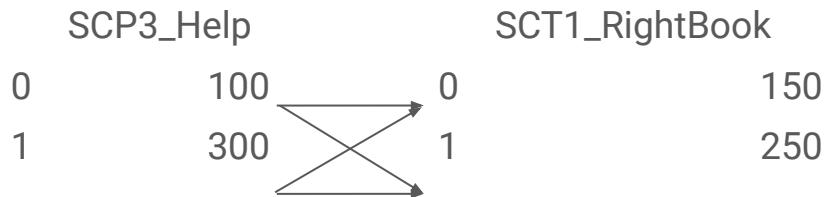
Возможные способы исправления:

- Стратегии **пересечения** и спасения

Без пересечения:



С пересечением:



SCP3_Help



Я помогу тебе выбрать книгу



Перейти к своему заданию



Оля

Любит читать и сочинять истории

И всё таки я не могу определиться.
Какую мне взять книжку?



Оля

Любит читать и сочинять истории

Нажми на название книжки, которую
мне нужно прочитать.



История изучения океана



Кто живёт на дне моря



Как делают одежду для подводного
плавания

Скоринг в гибкой среде

- Объединение индикаторов, если они относятся к одной способности:

Петя: Хочешь работать самостоятельно или будем работать все вместе?

Варианты ответа:

- Мне удобнее работать самостоятельно
- Давайте работать все вместе



SCP2_TeamWork (0 1)

Только если выбрал "Мне удобнее работать самостоятельно", то:

Оля: Но ведь все задания связаны друг с другом!

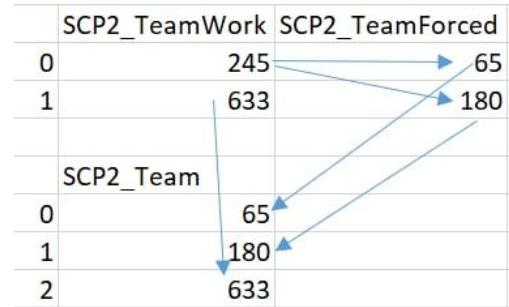
Петя: Поэтому нам нужно втроём их обсуждать.

Варианты ответа:

- Раз вы оба так думаете, давайте вместе.
- Я всё-таки хочу работать отдельно.



SCP2_TeamForced (0 1)



SCP2_Team (0 1 2)

Скоринг в гибкой среде

У нас есть опыт в автоматическом скоринге:

- текстов
- конструкторов

Но в этой презентации мы это не разбираем подробности.

Если у вас есть интерес в этом, давайте обсудим вместе

Обратная связь



Как показывать результаты?

Традиционные способы: относительно других тестируемых или по достижению уровня мастерства

Относительно других – по баллам (процентилям и т.д.) и по уровням. Сейчас наша обратная связь – по уровням, чтобы можно было дать содержательное описание в удобном для учителя формате.

Выделение уровней по распределениям: фактор-скоры модели КФА разделяются на 3 уровня по правилу среднее $\pm 1SD$.

Модель КФА – потому что необходимо моделировать множество факторов: для учета сложной структуры конструкта и локальных зависимостей. Другие модели с дискретными латентными переменными (напр. DCM) не подходят.

Можно использовать Байесовскую сеть, но методология имеет мало практики для оценки качества модели. Однако для предоставления результатов – очень удобный подход.

По достижению уровня мастерства – есть конструкты, которые выражаются именно в такой логике. Например, формальные операции – они или сформированы, или не-сформированы. А КР? А КО и КП?

Как обять необъятное? Покрытие рамки индикаторами

Цель тестирования:
предоставить валидную и
надежную обратную связь.

Мы вводим такой уровень
подробности ОС, который
обеспечен достаточным
количеством и
разнообразием
индикаторов

Время		ИТОГ	6	10	12	17	13	7
Креативность			Дино	Монстр	Аквариум	Путешествие	Торт	Спектакль
Составляющая	Оригинальность (CR1)	64	0	30	0	0	0	4
	Детальность (CR2)	37	0	12	3	0	6	4
ИТОГ		101	0	42	3	0	6	8
Критическое мышление								
Составляющая	Способности		Дино	Монстр	Аквариум	Зелье	Торт	Спектакль
	1 - Способность выделять <u>надежные</u> , заслуживающие доверия источники информации и использовать информацию для решения проблемы только из таких источников		2	1	2	-	2	1
СТ1 Анализ	2 - Способность выделять <u>релевантную</u> , полезную информацию для решения проблемы		12	4	9	-	2	1
	3 - Способность собирать наиболее <u>полную</u> информацию для решения проблемы, искать дополнительные источники информации, оценивать полноту информации		3		1	1	1	
	4 - Способность выделять основную мысль в тексте		17	5	12	0	3	1
	1 - Способность формулировать собственное решение проблемы (вывод), соответствующее целям и условиям задания. Вывод формулируется на основании проведенного анализа (текста)		24	1	3	9	7	5
СТ2 Вывод	2 - Способность изменять способы решения и выводы при изменении условий		2		2	-	2	
	3 - Способность аргументировать выбранную позицию с использованием достоверных элементов текста		2			-	2	
	4 - Способность признавать существование других, отличных от его, точек зрения		0			-	11	2
	ИТОГ		61	44	4	35	9	2

Чем подробнее обратная связь, тем полезнее тест.

НО

Чем подробнее обратная связь, тем длиннее тест.

Что видит пользователь теста?

На сегодняшний день в ОС для учителя/родителя включены:

- описание навыков в рамках инструмента
- описание уровня: по общему навыку и по составляющим (если есть)
- сводный результат по классу и таблица с результатами каждого ученика (по ID).

Форма ОС для ученика включает краткое оптимистичное описание уровня.

Сейчас ведем работу по улучшению формы ОС:

Должны ли быть примеры заданий в тексте ОС? Если да, то насколько подробно?

Нужно ли давать учителю и/или ребенку советы по улучшению прохождения теста?

Что нужно включать в ОС?

Уровни развития

Развивающийся уровень

Общий балл Критическое мышление: Ученики не обращаются к надежным источникам информации или выбирают сомнительные источники, часто в тексте выбирают не релевантную цели задания информацию, не обращаются к дополнительным источникам информации. Вывод относительно проблемы часто оказывается неоптимальным, не чувствительным к изменениям условий. Ученики редко представляют качественную аргументацию принятого решения.

По составляющим:

Анализ. Надежность источников информации

Ученики редко отдают предпочтение надёжным источникам информации, выбирают вместо них недостоверные информационные ресурсы.

Результаты тестирования

Пришло время узнать
о твоих сверхспособностях
чутоку больше

КРИТИЧЕСКОЕ МЫШЛЕНИЕ
Продвинутый уровень

Ты отлично справляешься! Ты знаешь, что не всем источникам информации можно доверять одинаково и выбираешь только самые надежные. При работе с текстом ты выбираешь только самые важные идеи, которые помогут решить задачу. Если ты оказываешься в ситуации, что информации не хватает и не получается принять взвешенное решение, ты стараешься найти дополнительные источники. При выборе решения ты внимательно относишься к условиям задачи и учитывашь всю информацию из надежных источников.

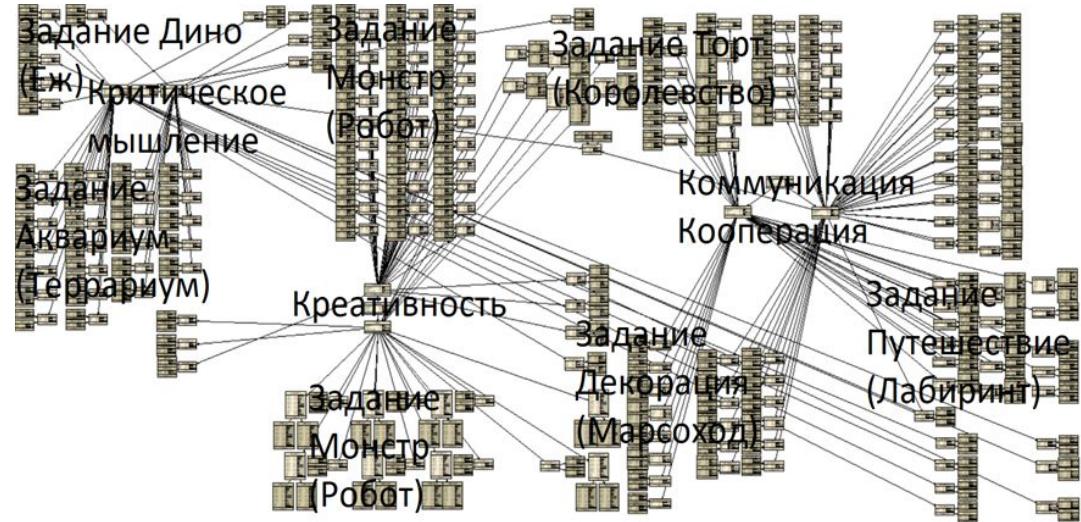
Автоматизация ОС

Обработка результатов тестирования может осуществляться с помощью Байесовской сети.

Можно в программе Netica.

Netica имеет интуитивный и приятный интерфейс пользователя для ввода топологии сети. Но у нее есть минусы:

- требует развертывания отдельного модуля на сервере
- много времени на построение сети
- платная



Автоматизация ОС

Поэтому мы взяли пакет **bnlearn** в R:

- Обучение - на основе уровней из КФА.
- Предсказание - метод Posterior Expectation (Likelihood Weighting).
- Результат для тестируемого - вероятность принадлежать каждому из уровней и уровень с максимальной вероятностью.

Качество предсказания оценивали по двум метрикам:

Точность - доля наблюдений, в которых уровень по байесовской сети совпал с уровнем по оценке из КФА.

Уверенность - средняя по наблюдениям максимальная вероятность находиться на каком-либо уровне.

Способность	Точность	Уверенность
CM1	94%	0,889
CM2	94%	0,895
CM	94%	0,893
CP1	95%	0,902
CP23	95%	0,899
CP4	94%	0,898
CP	95%	0,903
CT11	92%	0,873
CT12	94%	0,886
CT21	93%	0,882
CT22	91%	0,858
CT	93%	0,895

Автоматизация ОС

Инициативный проект "**Нейрокреативность**":

Автоматическая оценка креативности с помощью машинного обучения без привлечения экспертов (работа с изображением как с целостным объектом).

Почитать: статья "Оценивание креативности на основе анализа изображений с помощью нейронных сетей" в журнале "Искусственный интеллект и принятие решений" (1 номер 2021 года)

Посмотреть: Запись семинара «Измерения и анализ данных в психологии и образовании» на YouTube-канале МГППУ

Работа над упаковкой продукта

Подали заявку в Фонд развития прикладных исследований:

работы по презентации инструмента, подготовка промо-материалов, упаковка инструмента 4К для проведения тестирований, разработка автоматизированной системы сбора заявок, создания личных кабинетов, анализа результатов и формирования отчетов.

и вообще, валидизация – бесконечный процесс

разработка тестов – тем более

Вопросы на обсуждение

Вопросы на обсуждение

1. Вопрос для обсуждения: для мониторинга допустима более низкая надежность. А для достижения достаточной надежности в более детальной ОС нужно увеличивать количество индикаторов. Но пользователям теста не нравятся длинные сценарии. Как выстроить методологически четкую процедуру контроля длины заданий?
2. Мы считаем, что чем подробнее обратная связь, тем полезнее тест. Вы согласны?
3. Что нужно включать в ОС? Примеры заданий? Психометрические свойства инструмента?
4. Что делать с плохо структурированными теориями?
5. Как вы боретесь с локальной зависимостью?
6. Как определять уровни? Как устанавливать пороги?

А может быть, вы хотите чем-то поделиться из своего опыта? Нам интересно!

Спасибо за внимание!

iuglanova@hse.ru tg: ugl_anova

dgracheva@hse.ru tg: grachdasha

svtarasov@hse.ru tg: tarassv