

Психометрические инструменты для взаимного оценивания в онлайн-курсах

Ведущий аналитик Центра психометрических исследований
в онлайн-образовании НИУ ВШЭ

Дарья Кравченко



Оценивание в онлайн-курсах

Итоговое оценивание (of learning)

- вывод об освоении содержания
- большой объем
- каждый модуль, уровень курса

Формирующее оценивание (for learning)

- используется для обучения
- меньший объем
- уровень одного урока, концепции

соотнесено с учебными целями и результатами

Инструменты оценивания

Тесты с заданиями разного типа

Взаимное оценивание

Задания на программирование

Самооценивание

Взаимное оценивание

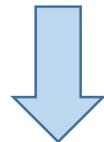
Взаимное оценивание предполагает самостоятельное конструирование ответа студентом и его последующую проверку другими студентами.

Преимущества и важность взаимного оценивания

- оценка работ, например, эссе, дизайн – проекты
- простота организации
- быстрота получения данных
- относительно небольшие затраты на реализацию
- возможность охвата больших групп

Проблема

субъективность в оценках (намеренное завышение или занижение баллов, низкий уровень подготовленности учащихся, различное понимание вопросов и критериев оценки)



отсутствие уверенности в качестве оценивания

Психометрика — это наука об измерениях в образовании и психологии. В основе психометрики лежит математическая статистика: **статистическое моделирование латентных конструктов и переменных.**

Латентный — значит, невидимый, ненаблюдаемый.

Наблюдения:
тесты, опросники, задачки, чтобы на основе ответов — верных или неверных — и с помощью специально разработанного математического аппарата установить связь между нашими наблюдениями и интересующими нас латентными переменными. — это наука об измерениях в образовании и психологии.

Задания с автоматизированной оценкой (тесты)

- эффект студента
- эффект задания

Задания на взаимное оценивание

- эффект студента
- эффект задания
- **эффект эксперта**

Как эффект сказывается на оценке?

Эффект студента	Эффект задания	Эксперт эксперта	Итого
4	0	-1	3
4	0	0	4
4	0	-1	5

Классическая теория

Согласованность взаимных оценок студентов

Конструктивная валидность

между средней оценкой по экспертам (студентам) с оценками по тестам

Критериальная валидность

между средней оценкой по экспертам (студентам) с итоговой оценкой за курс

Современная теория

Задача: измерить эффект эксперта на оценку

Психометрическое решение – применение трехфасетной модели Раша:

- трудность задания
- подготовленность студента
- строгость/снисходительность эксперта

Исследование в рамках классической теории

1 онлайн-курс «Философия культуры»

868 работ по курсу

Всего было исследовано 2604 взаимных оценки

3 студента-эксперта

студенты-эксперты отбирались платформой

Результаты (классическая теория)

- Взаимное оценивание вносит большой вклад в итоговый балл и имеет прогностическую силу
- Оценки экспертов в первом задании на 15% объясняют итоговый балл по курсу
- Согласованность критериев 0,53

Исследование в рамках современной теории

1483 работ студентов:

868 работ по курсу «Философия культуры»

615 работ по курсу на английском языке «Understanding
Russians: Contexts of Intercultural Communication»

Всего было исследовано 4464 взаимных оценки

3 студента-эксперта

студенты-эксперты отбирались платформой

Результаты (современная теория)

- Строгость экспертов не соответствует подготовленности студентов.
- Эксперты склонны завышать баллы.
- Есть существенно более строгие студенты.

Логиты	+Эксперты	+Студенты	+Задание	Шкала
10	*****.	*****.	+	(3)
9	.	+	+	
8	.	+	+	
7	.	+	+	
6	.	+	+	
5	*	+	+	
4	**.	+	+	
3	**.	+	+	
2	***.	+	+	---
1	****	+	+	2
0	***	*****.	1	
*	**.	*****.	2 4	---
-1	.	**.	+	1
-2	*	***	+	---
-3	.	***.	+	
-4	.	+	+	
-5	.	+	+	
-6	.	+	+	
-7	.	+	+	
-8	+	+	+	(0)

Логит | * = 4 | * = 4 | +Задание | Шкала

Логиты	+Эксперты	+Студенты	+Задание	Шкала
7	***	+	+	(4)
6	*****.	+	+	---
5	**	+	+	
4	**.	+	+	
3	*	+	+	3
2	**.	+	+	
1	***.	+	+	
0	*	*****.	*	2
-1	*	***.	+	
-2	.	+	+	2
-3	.	+	+	---
-4	*	+	+	4
-5	.	+	+	1
-6	.	+	+	---
-7	.	+	+	
-8	.	+	+	(0)

Логиты | * = 3 | * = 10 | +Задание | Шкала

Ограничения исследования

- Отсутствие возможности проследить намеренность выставления оценок.
- Проведение исследования в рамках двух заданий гуманитарных онлайн-курсов.
- Отсутствие проведения анализа с учетом пола студентов, возраста и уровня мотивации.

Ограничения исследования

- Отсутствие возможности проследить намеренность выставления оценок.
- Проведение исследования в рамках двух заданий гуманитарных онлайн-курсов.
- Отсутствие проведения анализа с учетом пола студентов, возраста и уровня мотивации.

Перспективы

Объяснение эффектов (пол, возраст, подготовленность студентов, язык и другие)



Психометрика как измерительный инструмент качества онлайн-курсов



Возможности

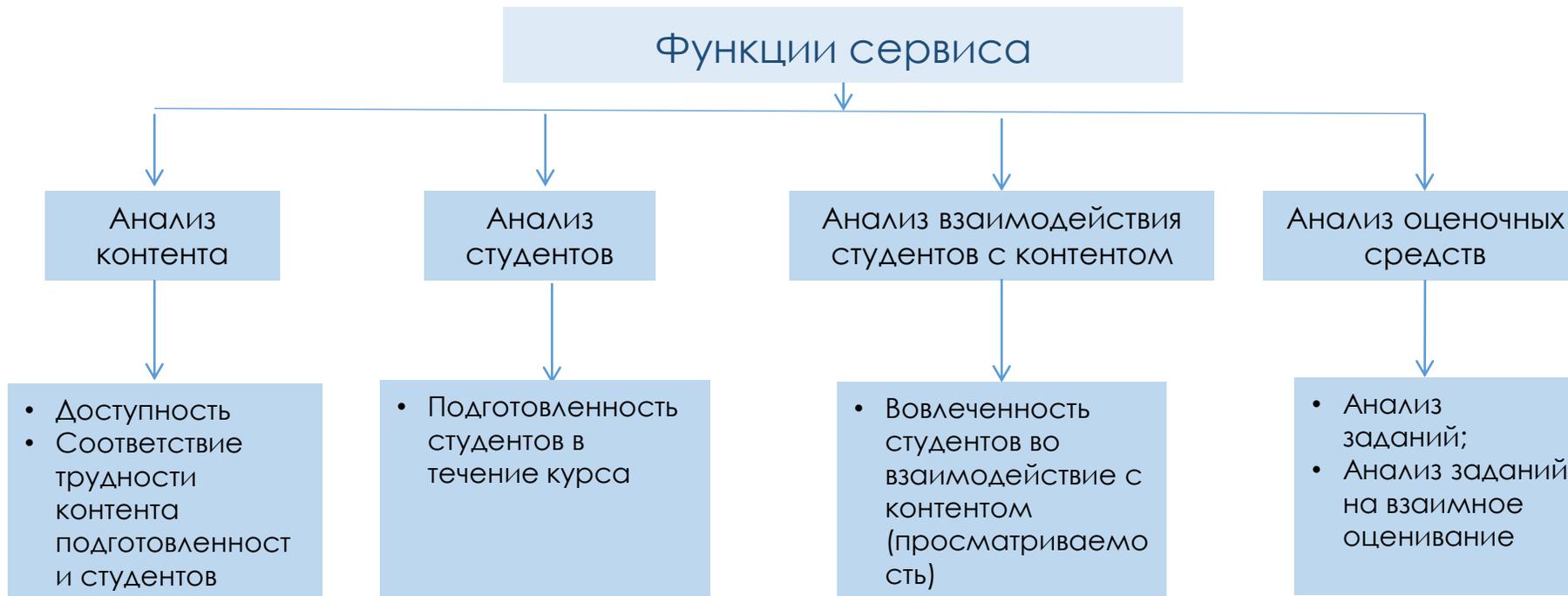
Для ВУЗов:

- базовая оценка качества контента онлайн-курсов
- оценка динамики обучения слушателей онлайн-курсов
- оценка качества тестовых заданий
- обратную связь для регулярной доработки курсов на основе результатов психометрической аналитики

Для студентов:

- обновления и модернизации с помощью диагностики и устранения проблем.

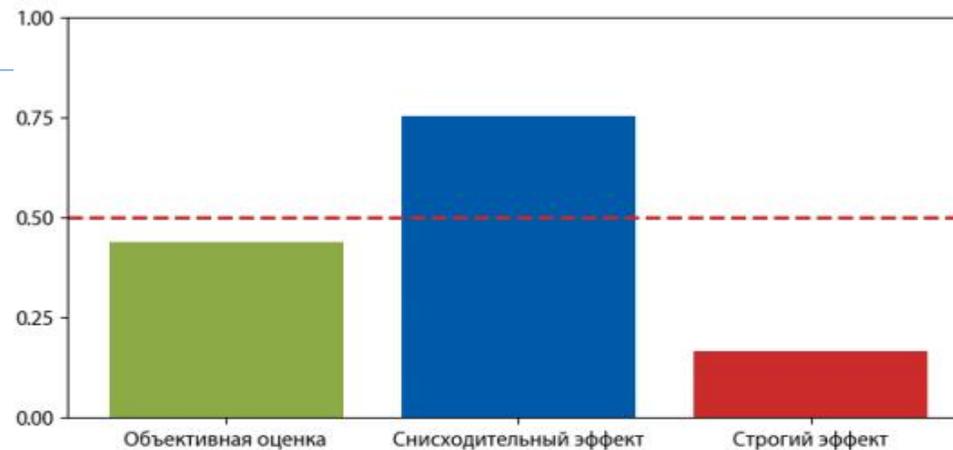
Методика и инструментарий



Анализ заданий на взаимную оценку

4.2. Анализ заданий на взаимную оценку

Параметры вероятности получения оценки студентами



Интерпретация:

Если все три оценки расположены в одной зоне вероятности решения задания (только выше или только ниже среднего значения), то эффект строгости/снисходительности эксперта не приводит к существенным смещениям оценки студента. Если снисходительная или строгая оценка находятся в разных зонах вероятности решения задания по сравнению с объективной оценкой, имеет место необъективность в оценках – существенная часть проверяющих занижает или завышает оценки.

Рекомендации:

Модернизации подлежат задания, в которых присутствует необъективность в оценивании.

Модернизация в части повышения объективности в экспертных оценках осуществляется через:

- использование заданий с ограничением на ответ;
- использование примеров качественного выполнения задания;
- использование примеров оценивания с комментариями для проверяющих;
- использование дихотомической системы оценивания (да/нет) с четким описанием критериев;
- использование дробной системы оценивания (выполнено/частично выполнено/не выполнено) с четким описанием критериев.

Спасибо!

Центр психометрики НИУ ВШЭ:
<https://elearning.hse.ru/psychometrics/>

Ссылка на статью: Кравченко Д. А. [Классический и современный подходы к измерению валидности заданий на взаимное оценивание в MOOK](#) // Вопросы образования. 2018. № 4. С. 99-115.

