



КАЗАХСКИЙ НАЦИОНАЛЬНЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ АБАЯ

Факультет Педагогики и психологии

Кафедра начального образования

Классическая теория и методики конструирования тестов

Дисциплина: «Технология оценивания в начальном образовании»

Алматы, 2025

Классическая теория и методики конструирования тестов.

Основные этапы конструирования теста

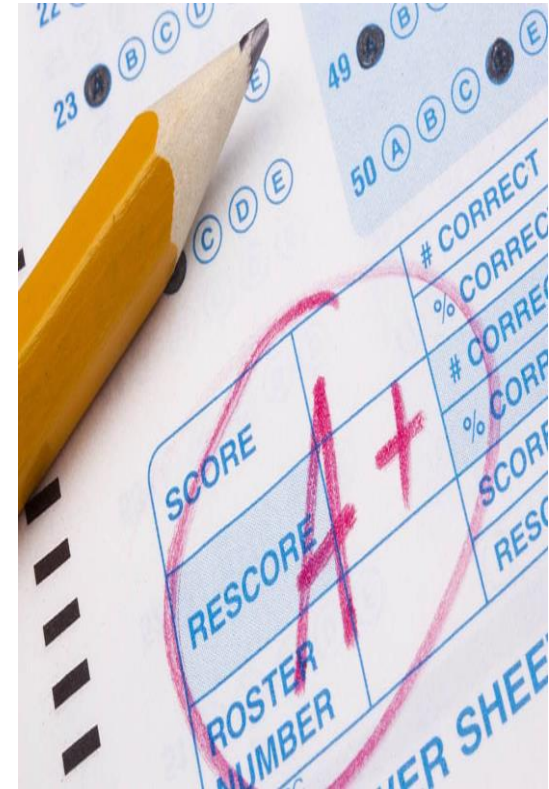
1. Определение цели тестирования, выбор теста и подхода к его созданию.
2. Концептуальный выбор конструкта (переменной измерения).
3. Анализ содержания учебной дисциплины и планирование содержания теста, априорный выбор длины теста и времени его выполнения, разработка спецификации теста.
4. Определение структуры теста, форм заданий и стратегии их расположения в тесте.
5. Создание предтестовых заданий.
6. Отбор заданий в тест и их ранжирование согласно выбранной стратегии предъявления на основании априорных авторских оценок трудности заданий.
7. Экспертиза формы предтестовых заданий и содержания теста.
8. Коррекция заданий и теста по результатам экспертизы.
9. Разработка методики апробационного тестирования, инструкций для учеников и преподавателей, проводящих апробацию теста.



Классическая теория и методики конструирования тестов.

Основные этапы конструирования теста

10. Формирование репрезентативной выборки апробации.
11. Проведение апробационного тестирования.
12. Проверка результатов выполнения теста (автоматизированная или ручная), подготовка эмпирических данных тестирования к виду, удобному для обработки и проведения анализа.
13. Статистическая обработка результатов выполнения теста (автоматизированная с помощью специального программного обеспечения).
14. Анализ и интерпретация результатов обработки в целях улучшения качества теста. Проверка соответствия характеристик теста научно обоснованным критериям качества.
15. Коррекция содержания и формы заданий на основании данных предыдущего этапа. Чистка теста и добавление новых заданий для оптимизации диапазона значений параметра трудности и улучшения системообразующих свойств заданий теста. Оптимизация длины теста и времени его выполнения на основании статистических оценок характеристик теста. Оптимизация порядка расположения заданий в тесте.
16. Повторение этапа апробации для выполнения очередных шагов по повышению качества теста.
17. Интерпретация данных обработки, установление норм теста и создание шкалы для оценки результатов испытуемых.



Классическая теория и методики конструирования тестов.

Основное предположение классической теории тестов.

Предположение о существовании истинного балла (True Score) является основополагающим в классической теории тестов, наряду с терминами «сырой балл» и «Наблюдаемый балл» оно широко используется в процессе разработки тестов [28, 47, 68].

Среди различного рода ошибок при тестировании можно выделить два типа:

←
систематические

←
случайные



Классическая теория и методики конструирования тестов.

Матрица наблюдаемых результатов выполнения теста

Испытуемые	Здания $1...j...n$	Индивидуальный балл (x_{ij})
$\begin{matrix} 1 \\ \dots \\ i \\ \dots \\ N \end{matrix}$	$X_{ij} = \{$	$x_{ij} = \sum_{j=1}^n L_{ij}$
Количество правильных ответов на задания (R_i)	$R_i = \sum_{j=1}^n x_{ij}$	$\sum_{i=1}^N L_{ij} = \sum_{j=1}^n R_i$

Классическая теория и методики конструирования тестов.

Матрица результатов тестирования

Номера испытуемых i	Номера заданий j										Индивидуальные баллы (множество X_i)
	1	2	3	4	5	6	7	8	9	10	
1	1	1	1	1	1	1	0	0	0	0	6
2	1	1	0	0	0	0	0	0	0	0	2
3	0	0	0	0	0	0	0	1	0	0	1
4	1	1	0	1	1	1	1	1	1	1	9
5	1	0	1	0	1	1	0	0	0	0	4
6	1	1	1	0	0	0	0	1	0	0	4
7	1	1	1	1	0	1	0	0	0	0	5
8	1	1	1	1	0	0	0	0	0	0	4
9	1	1	1	1	1	1	1	1	1	0	9
10	1	1	1	1	1	0	1	0	0	0	6
Число правильных ответов (множество)	9	8	7	6	5	5	3	4	2	1	50

Классическая теория и методики конструирования тестов.

Матрица результатов тестирования после удаления

Номера испытуемых i	Номера заданий j										$\sum$
	1	2	3	4	5	6	8	7	9	10	
3	0	0	0	0	0	0	1	0	0	0	1
2	1	1	0	0	0	0	0	0	0	0	2
5	1	0	1	0	1	1	0	0	0	0	4
6	1	1	1	0	0	0	1	0	0	0	4
8	1	1	1	1	0	0	0	0	0	0	4
7	1	1	1	1	0	1	0	0	0	0	5
1	1	1	1	1	1	1	0	0	0	0	6
10	1	1	1	1	1	0	0	1	0	0	6
9	1	1	1	1	1	1	1	1	1	0	9
4	1	1	0	1	1	1	1	1	1	1	9
R_j	9	8	7	6	5	5	4	3	2	1	50



Классическая теория и методики конструирования тестов.

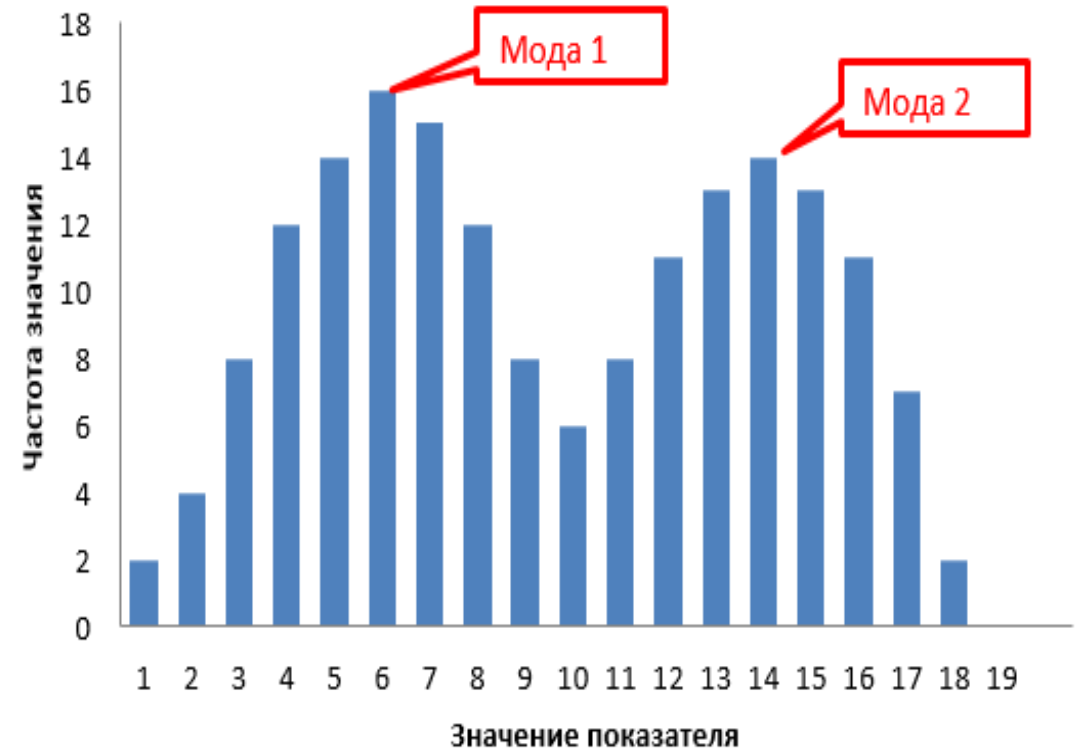
Распределение сгруппированных частот результатов учащихся

Интервал баллов	Частота
1-3	2
4-6	6
7-9	2

Классическая теория и методики конструирования тестов.

Мода - это такое значение, которое встречается наиболее часто среди результатов выполнения теста.

Например, для данных матрицы, представленной в табл. 10, модой является балл «4», потому что он встречается чаще (три раза) любого другого значения балла.



Классическая теория и методики конструирования тестов.

Среднее выборочное (среднее арифметическое) определяется суммированием всех значений совокупности баллов и последующим делением на их число. Для индивидуальных баллов $X_1, X_2, \dots, X_N$ группы N испытуемых среднее значение $\bar{X}$ будет $\longrightarrow$

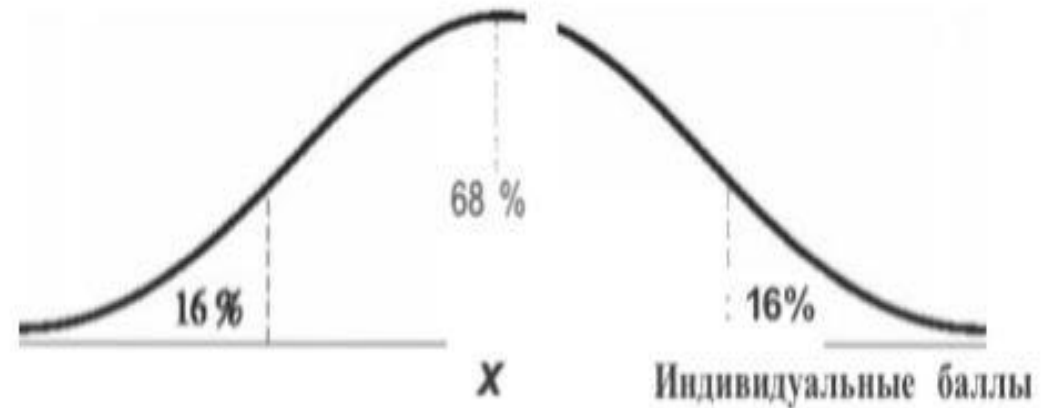
$$\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N}$$

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$$

Классическая теория и методики конструирования тестов.

Нормальная кривая - изобретение математиков в сглаженном, идеальном виде описывает реальный полигон частот. На практике никогда не была и не будет получена совокупность данных, распределенных точно по нормальному закону, просто иногда полезно, допуская определенную ошибку, утверждать, что распределение эмпирических данных близко к нормальной кривой.

Нормальное распределение унимодально и симметрично, т.е. половина результатов, расположенная ниже моды, в точности совпадает с другой половиной, расположенной выше, а мода и среднее значение равны.



Нормальная кривая распределения индивидуальных баллов

Классическая теория и методики конструирования тестов.

Размах измеряет на шкале расстояние, в пределах которого изменяются все значения показателя в распределении. Например, для распределения индивидуальных баллов, представленных в табл. 10, размах равен $9 - 1 = 8$. Вариационный размах легко вычисляется, но при характеристике распределения баллов по тесту используется крайне редко. Во-первых, размах является весьма приближенным показателем, так как не зависит от степени изменчивости промежуточных значений, расположенных между крайними значениями в распределении баллов по тесту. Во-вторых, крайние значения индивидуальных баллов, как правило, ненадежны, поскольку содержат в себе значительную ошибку измерения. В этой связи более удачной мерой изменчивости считается дисперсия.



Классическая теория и методики конструирования тестов.

Таблица 15. Пример значений коэффициента корреляции между заданиями

	1	2	3	4	5	6	8	7	9	10
1	1,0000	0,6667	0,5092	0,4082	0,3333	0,3333	0,4082	0,2182	0,1667	0,1111
2	0,6667	1,0000	0,2182	0,6124	0,0000	0,0000	0,1021	0,3273	0,2500	0,1667
3	0,5092	0,2182	1,0000	0,3563	0,2182	0,2182	0,3563	00,476	0,2182	0,5092
4	0,4082	0,6124	0,3563	1,0000	0,4082	0,4082	0,1667	0,5345	0,4082	0,2722
5	0,3333	0,0000	0,2182	0,4082	1,0000	0,6000	0,0000	0,6547	0,5000	0,3333
6	0,3333	0,0000	0,2182	0,4082	0,6000	1,0000	0,0000	0,2182	0,5000	0,3333
8	0,4082	0,1021	0,3563	0,1667	0,0000	0,0000	1,0000	0,3563	0,6124	0,4082
7	0,2182	0,3273	00,476	0,5345	0,6547	0,2182	0,3563	1,0000	0,7638	0,5092
9	0,1667	0,2500	0,2182	0,4082	0,5000	0,5000	0,6124	0,7638	1,0000	0,6667
10	0,1111	0,1667	0,5092	0,2722	0,3333	0,3333	0,4082	0,5092	0,6667	1,0000
Суммы	3,3385	3,1392	1,3888	4,2417	4,0478	3,6114	1,3436	4,5346	4,6495	3,2915

Классическая теория и методики конструирования тестов.

Бисериальный коэффициент корреляции используется в том случае, когда один набор значений распределения задается в дихотомической шкале, а другой - в интервальной. Под эту ситуацию подпадает подсчет корреляции между результатами выполнения каждого задания (дихотомическая шкала) и суммой баллов испытуемых по заданиям теста (интервальная или квазиинтервальная шкала). С помощью подсчета значений бисериального коэффициента корреляции оценивается валидность, иногда называемая показателем дифференцирующей способности (дискриминативности) заданий теста.

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{\sigma_y} \cdot \sqrt{\frac{n_1 \cdot n_0}{n(n-1)}},$$

Классическая теория и методики конструирования тестов.

Оценка трудности тестовых заданий в классической теории тестов осуществляется по формуле:

$$R_j = \frac{P_j}{N_j}$$

Где p_j - доля правильных ответов на j -е задание;
 R_j - количество учеников, выполнивших j -е задание верно; N - число учеников в тестируемой группе; j - номер задания теста ($j = 1, 2, \dots, n$). Трудность задания нередко выражают в процентах. Для этого оценку, полученную по формуле, умножают на 100 %.



Классическая теория и методики конструирования тестов.

Дискриминативностью (discriminatory power) называется способность задания дифференцировать учеников на лучших и худших.

Высокая дискриминативность - важная характеристика удачного тестового задания, она определяет меру валидности задания, его адекватность целям создания теста.

Статистика

Правильных ответов

68%

Дискриминативность

0,5 

Выполнено

25

Классическая теория и методики конструирования тестов.

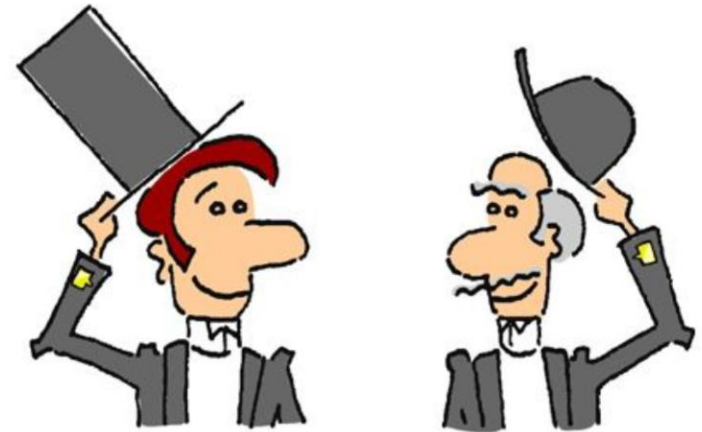
Этические нормы в тестировании:

1. Справедливость и равенство:

Этические нормы требуют, чтобы тесты были справедливыми и равноправными для всех, независимо от их культурных, социальных и этнических особенностей. Тесты не должны содержать предвзятость и дискриминацию.

2. Прозрачность и открытость:

Этические нормы требуют, чтобы тесты были разработаны и проведены в открытом и прозрачном процессе. Участники тестирования должны быть информированы о целях, условиях и оценочных процедурах.



Классическая теория и методики конструирования тестов.

Этические нормы в тестировании:

3. Конфиденциальность и приватность:

Защита данных и личной информации участников тестирования - важнейший этический аспект.

Результаты тестов должны быть доступны только тем лицам, для которых они предназначены.

4. Информированное согласие: Перед проведением тестов участники должны давать информированное согласие на участие и быть осведомленными о возможных последствиях результатов тестирования.



Классическая теория и методики конструирования тестов.

Этические нормы в тестировании:

5. Честность и честные практики:

Этические нормы требуют от исследователей и тестирующих соблюдать честность во всех аспектах тестирования, включая сбор данных, анализ и интерпретацию результатов.

6. Обеспечение безопасности:

При проведении физических тестов, например, в медицинских областях, важно обеспечить безопасность участников и минимизировать риски для их здоровья.

7. Учет потребностей специальных групп: Этические нормы также требуют учета потребностей людей с ограниченными возможностями и адаптацию тестов для них.



Классическая теория и методики конструирования тестов.

ИНТЕРЕСНЫЕ ФАКТЫ:

1. Первые тесты появились более ста лет назад:
Первый массовый тест IQ был разработан Альфредом Бине в 1905 году. Это стало отправной точкой для развития тестового тестирования.
2. Создание “Сатурн-5” и тест на уровень IQ: Для набора инженеров, работавших над проектом “Аполлон” и созданием ракеты “Сатурн-5”, был проведен тест на интеллект. Он показал, что многие из них имели выдающиеся умственные способности.



Классическая теория и методики конструирования тестов.

ИНТЕРЕСНЫЕ ФАКТЫ:

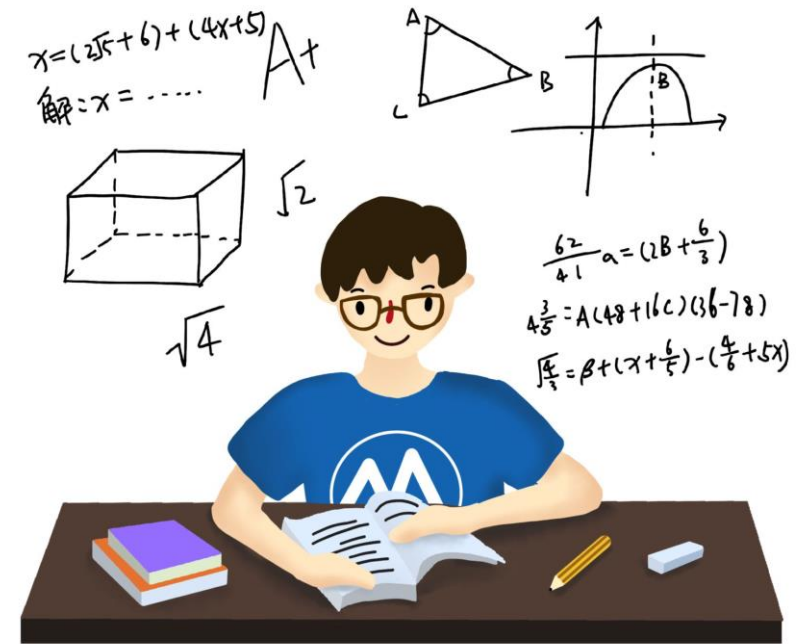
3. Коэффициент корреляции Гуттмана: Янси Гуттман, один из основателей классической теории тестов, внес значительный вклад, разработав коэффициент корреляции Гуттмана, который используется для изучения связи между результатами разных тестов.
4. Метод Гуттмана для создания шкал: Янси Гуттман также разработал метод создания шкал, при котором элементы шкалы упорядочиваются от самых легких к самым сложным, и каждый элемент предполагает знание всех предыдущих.



Классическая теория и методики конструирования тестов.

ИНТЕРЕСНЫЕ ФАКТЫ:

5. Метод дихотомических вопросов: Этот метод, используемый при создании тестов, предполагает, что на каждый вопрос можно ответить либо “да”, либо “нет”. Он часто применяется в тестировании знаний.
6. Критика классической теории: Несмотря на широкое использование, классическая теория тестов получила критику за ограниченность в объяснении сложных многомерных явлений, что привело к развитию более современных методов, таких как теория ответов на элементы.



Классическая теория и методики конструирования тестов.

ИНТЕРЕСНЫЕ ФАКТЫ:

- 7.Эволюция онлайн-тестирования: С развитием интернета и компьютерных технологий, многие тесты перешли в онлайн-формат, что позволяет более гибко проводить и администрировать тестирование.
- 8.Значение в образовании и бизнесе: Тесты играют важную роль в сферах образования и бизнеса, помогая принимать решения о приеме на работу, обучении и карьерном развитии.



Классическая теория и методики конструирования тестов.

ИНТЕРЕСНЫЕ ФАКТЫ:

9. Использование в психологических исследованиях:

Тесты часто используются для изучения психологических явлений, таких как личностные черты, интеллект, стрессовая устойчивость и многие другие аспекты психологии.

10. Сложности при создании валидных тестов:

Создание валидных тестов требует глубокого знания предметной области, статистических методов и психометрики, что делает эту область исследования и разработки весьма сложной и интересной.





Классическая теория и методики конструирования тестов.

ЗАКЛЮЧЕНИЕ:

Классическая теория и методики конструирования тестов основаны на надежности, валидности и справедливости. Они включают в себя разработку вопросов с единственным ответом, статистический анализ и использование теории надежности. Учитывать контекст и цель тестов важно для их эффективности и справедливости.

Классическая теория и методики конструирования тестов.

Спасибо за внимание

