

Классификация тестов по целям использования в системе образования:

классификация по книге
Норманна Е. Гронлунда

- **определяющий** (знания или поведения обучаемого в начале обучения)
-

- **формирующий** (прогресса достигнутого в обучении)

- **диагностический** (трудности обучения и их источники)

- **суммирующий** (достижения в конце обучения)

Основные этапы конструирования педагогического теста

I. Этап целеполагания.

В зависимости от цели (текущий, итоговый и пр.) использования тестового инструментария определяются его вид, ресурсные возможности, источники финансирования, бюджет, круг привлекаемых специалистов и пр.

II. Этап подготовительный.

Анализ и отбор содержания проверяемого материала. Подготовка структуры работы, определение типов и количества используемых заданий, времени проверки, системы оценивания, условий проведения и проверки, рекомендаций по подготовке к тестированию.

III. Этап составления теста.

Разработка тестовых заданий. Экспертиза и доработка составленных заданий.

Построение выборки для апробации заданий и тестов.

Компоновка заданий для апробации. Апробация тестовых заданий.

Определение и расчет показателей качества тестовых заданий. Анализ результатов апробации и составление вариантов теста. Экспертиза и доработка вариантов теста.

IV. Этап проведения теста и анализ полученных результатов.

Стандартизация теста. Нормирование теста. Подготовительные работы по оснащению. Отслеживание проведения. Обработка результатов, получение статистических данных для анализа. Анализ результатов.

Подготовка итогового отчета.

Валидность – способность теста служить поставленной цели измерения

Виды:

- содержательная;
- критериальная;
- конструктивная.

Способы оценки валидности:

подсчет корреляции с внешним критерием;
экспертиза.

Репрезентативность – полнота охвата проверяемого материала.

Способы оценки: экспертиза.

Классическая теория тестов. Достоинства и недостатки

- Статистическое обоснование качества теста
- Представление о репрезентативности выборки
- Расчет основных статистических характеристик
- Нормальное распределение

Показатели качества тестовых заданий

В соответствии с классической теорией тестов показателями качества являются статистические характеристики:

- **трудность задания –**

определяется как доля учащихся, справившихся с заданием;

- **дискриминативность, или дифференцирующая способность –**

определяется как способность отделять испытуемых с высоким общим баллом от тех, кто получил низкий балл.

Трудность задания

Характеристика тестового задания, выраженная в процентом отношении количества испытуемых верно выполнивших задание от количества испытуемых репрезентативной выборки.

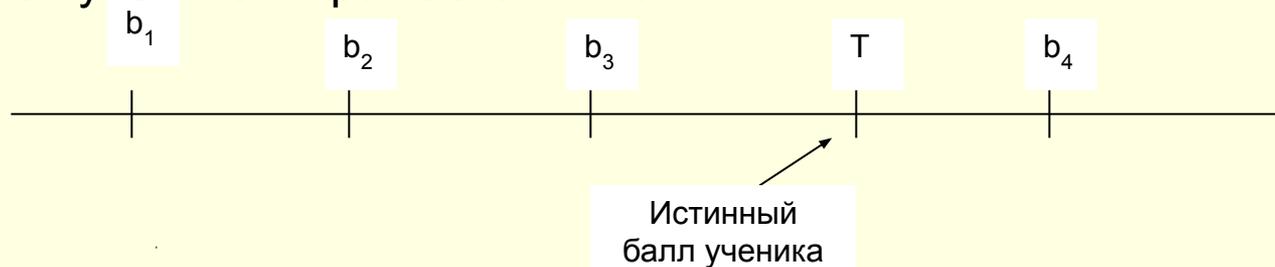
Репрезентативной называют выборку, в которой пропорционально представлены все элементы генеральной совокупности (т.е. представлены испытуемые из различных типов образовательных учреждений, разных социальных слоев).

Использование характеристики «трудность задания»

- Статистическая трудность задания необходима для определения места задания в тесте. Если из 100 учащихся первое задание выполнили 30 учащихся, а второе – 60, то второе задание менее трудное и его надо поставить в начале теста.
- Значение трудности задания для сильных и слабых групп меняется, т.е. зависит от выборки. Значения трудности меньше 20 и больше 80 считают критическими и в нормативно-ориентированные тесты такие задания включать не рекомендуют.
- Для критериально-ориентированных тестов значение трудности задания не существенна.

Правильность расположения заданий в тесте

Профиль ответов испытуемого – результаты выполнения теста в виде дихотомических оценок 0 и 1. Если учащийся выполнил три из четырех заданий и каждое выполненное задание дает 1 балл, а не выполненное дает 0 баллов, то его профиль ответов ученика выражается 1110.



Правильная структура знаний соответствует профилю, где все нули следуют за всеми единицами.

Если картина обратная, то это говорит о случайном характере ответов и отсутствии систематизированных знаний, или не квалифицированном отборе содержания заданий для теста.

Если в итоговом тесте встречаются задания, которое выполняют все учащиеся или не выполняет никто, то их исключают.

По профилю можно регулировать в обучающих (тематических) тестах:

- правильность расположения заданий в тесте (по нарастанию трудности);
- правильность составленной инструкции по проведению тестирования (минимизирование случайных факторов: волнение, заторможенность, обстановка во время тестирования);
- подбор наилучшей формы представления заданий.

Для тематических (текущей проверки) тестов задания необходимо располагать по нарастанию трудности.

Для итоговых тестов возрастание трудности – параметр необязательный. По мнению специалистов для всего курса это требование невыполнимо.

Если в итоговом тесте встречаются задания, которое выполняют все учащиеся или не выполняет никто, их исключают, поскольку они не несут информацию о степени подготовленности учащихся, в тематических тестах такие задания могут присутствовать.

Дискриминативность

Характеристика тестового задания, определяющая насколько точно можно провести различие по определенному измеряемому признаку (подготовленности) между экзаменуемыми с высокими и низкими способностями.

Задание, на которое одинаково хорошо могут ответить экзаменуемые как с высокими, так и с низкими способностями (подготовкой), не обладает хорошей дифференцирующей способностью, поскольку не дает никакой информации об относительных уровнях результатов.

Метод крайних групп, или «метод 27»

Дискриминативность вычисляют как разность долей испытуемых из сильной (27%) и слабой (27%) групп, правильно выполнивших задание.

$$D = \frac{N_{1\text{сильн}}}{N_{\text{сильн}}} - \frac{N_{1\text{слаб}}}{N_{\text{слаб}}}$$

$N_{\text{сильные}}$ – общее количество испытуемых в сильной группе,

$N_{1\text{сильные}}$ – количество учащихся в сильной группе верно выполнивших задание,

$N_{\text{слабые}}$ – общее количество испытуемых в слабой группе,

$N_{1\text{слабые}}$ – количество учащихся в слабой группе верно выполнивших данное задание.

Например: дискриминативность задания = 0,74 была получена так:

- В сильной группе 320 учащихся, из которых выполнили задание 300.
- В слабой группе 280 учащихся, выполнивших данное задание – 160.
- $300/320 - 160/280 = 0,74$

Значение дискриминативности заданий

Изменяется от -1 до +1. В тест должны попасть задания со значением дискриминативности выше 0,2.

Задание со значением близким к 1 правильно разделяет учащихся, говорит о том, что большинство сильных учащихся справились с заданием, а слабым это не удалось.

Нулевое значение говорит о том, что доли справившихся с заданием в сильной и слабой группах будут равны, задание нуждается в корректировке.

Задание с отрицательным значением чаще всего требуется удалить, либо существенно переработать.

ТИПИЧНЫЕ НЕДОСТАТКИ НИЗКОДИСКРИМИНАТИВНЫХ ЗАДАНИЙ:

- излишняя сложность или запутанность формулировки;
- неоднозначно понимаемое условие;
- подсказка в условии;
- опора на память, а не на мыслительные процессы при выполнении задания;
- наличие двух или более правильных ответов;
- наличие «терминологической или логической ловушки» в условии или ответах.

Пример инструкции для текущего тестового контроля

*Тест по теме: «...» состоит из ? частей и
включает ? заданий....*

Внимательно прочитай условие задания и выбери
один правильный ответ.

Номер правильного ответа впиши в свободное место
под номером вопроса в бланке для ответа.

Исправления исключаются. Время выполнения всех
заданий ? минут.

Пример инструкции для ИТОГОВОГО ТЕСТОВОГО КОНТРОЛЯ

На выполнение экзаменационной работы по информатике отводится 4 часа (240 минут), включая работу за компьютером.

Работа состоит из 3 частей, включающих 37 заданий.

Часть 1 включает двадцать четыре задания с выбором ответа. К каждому заданию дается четыре ответа, из которых только один правильный. Задания выполняются на специальном бланке или на компьютере.

Часть 2 состоит из восьми заданий с кратким ответом (к этим заданиям вы должны самостоятельно сформулировать и ввести ответ). Задания выполняются на специальном бланке или на компьютере.

Часть 3 состоит из пяти заданий. Для выполнения заданий этой части вам необходимо написать развернутый ответ в произвольной форме. При выполнении заданий Части 3 вы можете (по желанию) использовать компьютер, предоставленный Центром тестирования и оснащенный соответствующим программным обеспечением.

Выполняйте задания в том порядке, в котором они даны. Если какое-то задание вызывает у вас затруднение, пропустите его и постарайтесь выполнить те, в ответах на которые вы уверены. К пропущенным заданиям можно будет вернуться.

За каждый правильный ответ в зависимости от сложности задания дается один или более баллов. Баллы, полученные вами за все выполненные задания, суммируются. Постарайтесь выполнить как можно больше заданий и набрать как можно больше баллов.

Желаем успеха!

Основные статистические характеристики тестов

- частотное распределение
- мода
- дисперсия
- стандартное отклонение
- коэффициенты корреляции

Матрица результатов тестирования

Номер тестируемого i	Номер задания j										Индивидуальный балл 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
11	0	0	0	0	0	0	0	0	0	0	0(удалить)
12	1	1	1	1	1	1	1	1	1	1	10(удалить)
Число правильных ответов R_j для 10 заданий	9	8	7	6	5	5	3	4	2	1	

Упорядоченная матрица результатов тестирования

Номер тестируемого (i)	Номер задания (j)										Индивидуальный бал (X _i)
	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	0	1	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
Число правильных ответов (R _j) для 10 заданий	9	8	7	6	5	5	3	4	2	1	

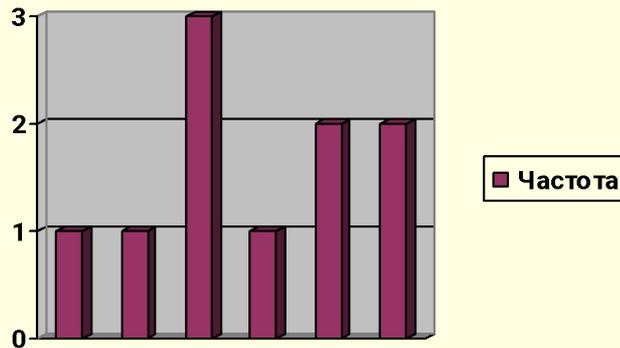
Характеристика «частотное распределение»

Ранжированный ряд

Балл	1	2	4	4	4	5	6	6	9	9
Номер	3	2	5	6	8	7	1	10	4	9
Ранг	1	2	3	3	3	4	5	5	6	6

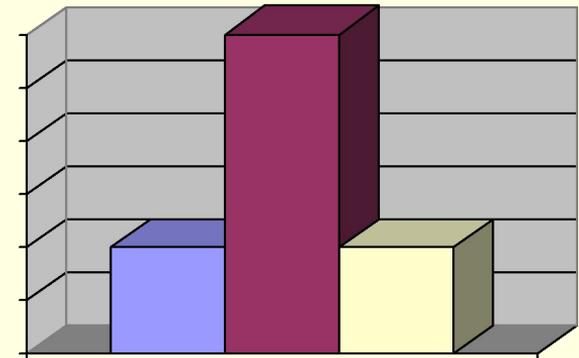
Частотное распределение

Балл	1	2	4	5	6	9
Частота	1	1	3	1	2	2



Сгруппированное частотное распределение

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



Основные понятия: мода, унимодальное и бимодальное распределение, дисперсия, стандартное отклонение

Наиболее часто встречающееся значение среди результатов выполнения теста - мода.

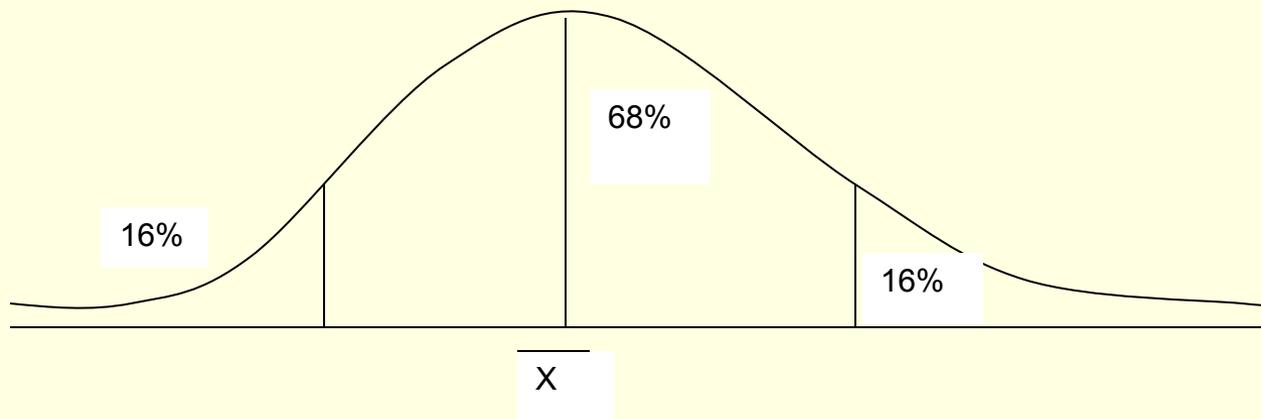
Для нашего случая модой является 4 балла, поскольку встретилось чаще, чем другие значения (3 раза).

Если встречается два значения моды, то распределение называют бимодальным.

Нормальное распределение результатов должно быть унимодальным (с одним значением моды) и симметричным. Бимодальное распределение говорит о неудачно построенном тесте, требующим внимательного анализа других характеристик для выявления причин.

Нормальное распределение результатов

На рисунке представлено распределение по нормальному закону.



Среднее значение баллов

$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N} = \frac{6 + 2 + 1 + 9 + 4 + 4 + 5 + 4 + 9 + 6}{10} = 5$$

Характеристики «дисперсия» и «стандартное отклонение»

Дисперсия отражает меру неоднородности результатов по тесту и вычисляется по формуле:

$$S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N - 1}$$

Низкая дисперсия говорит о слабом разделении тестируемых по уровню подготовки, а излишне высокое значение дисперсии говорит об искаженной картине распределения, а значит проблемах в тесте.

Стандартное отклонение вычисляется по формуле:

$$S_x = \sqrt{S_x^2}$$

Характеристика «корреляция»

Корреляция устанавливает связь между исследуемыми объектами.

В случае необходимости определения связи между заданиями в одном тесте используют коэффициент корреляции Пирсона:

$$\varphi_{jl} = \frac{p_{jl} - p_j p_l}{\sqrt{p_j q_j * p_l q_l}}$$

где p_{jl} – доля тестируемых, верно выполнивших оба задания, т.е. получивших по 1 баллу за оба задания;

p_j – доля тестируемых, правильно выполнивших j -е задание; $q_j = 1 - p_j$;

p_l – доля тестируемых, правильно выполнивших l -е задание; $q_l = 1 - p_l$

Коэффициенты корреляции для итоговых тестов должны быть в пределах (0; 0,3).

Высокие значения говорят о зависимости заданий друг от друга.

Отрицательные значения говорят об отсутствии предметной чистоты содержания теста, такие задания, как правило, удаляются.

Для тематических тестов корреляция будет достаточно высокой, поскольку такие задания отражают исходное содержание и слабо варьируются.

Коэффициент точечно-бисериальной корреляции

Расчетное значение не выходит за границы интервала $[-1; +1]$

$$(r_{pbis})_j = \frac{(\overline{X}_1)_j - (\overline{X}_0)_j}{S_x} \sqrt{\frac{(N_1)_j * (N_0)_j}{N(N-1)}}$$

$(\overline{X}_1)_j$ - среднее значение индивидуальных баллов тестируемых, выполнивших верно j -е задание теста;

$(\overline{X}_0)_j$ - среднее значение индивидуальных баллов тестируемых неверно j -е задание теста;

S_x – стандартное отклонение по множеству значений индивидуальных баллов;

N – общее число тестируемых;

$(N_1)_j$ – число тестируемых, верно выполнивших j -е задание теста;

$(N_0)_j$ – число тестируемых, неверно выполнивших j -е задание теста;

Выводы

- Анализ значений точечно – бисериальной корреляции позволяет сделать вывод о валидности задания, т.е. насколько задание пригодно для измерения в соответствии с общей целью построения теста.
- Если цель итогового тестирования – дифференцировать по уровню подготовки учащихся, то задания должны четко разделять сильно и слабо подготовленных в тестируемой группе. Значение коэффициента корреляции должно приближаться к 0,5.
- Задание лучше работает на дифференциацию учащихся, если выше значение разности:

$$(\overline{X_1})_j - (\overline{X_0})_j$$

Показатели качества теста

- Надежность
- Валидность

Любые оценки надежности и валидности имеют определенную степень достоверности. На практике надо пользоваться по возможности несколькими методами.

Надежность и валидность - связанные понятия, поэтому при построении теста придется идти на компромиссы. Высокое значение коэффициента надежности достигается при высокой однородности теста. Но для повышения валидности в тест лучше включать задания, содержание которых не связано каким-либо заметным образом. Это приведет к повышению полноты охвата содержания, следовательно, к росту валидности теста.

Надежность

Отражает точность тестовых измерений и устойчивость тестовых результатов к действию случайных факторов. Данная характеристика имеет два значения:

- точность
- устойчивость

Точность измерения означает устойчивость к действию помех и всевозможных случайностей при проведении тестирования. Уменьшить случайные факторы можно, грамотно составленной инструкцией.

Устойчивым тест считают, если получены близкие по значению результаты при повторном тестировании на параллельной форме.

Фундаментальное соотношение классической теории тестов

$$X = T + E; \quad S_X^2 = S_T^2 + S_E^2$$

где X – наблюдаемый балл, T – истинный балл, E – суммарная ошибка, S – дисперсия.

$$\frac{S_T^2}{S_X^2} = 1 - \frac{S_E^2}{S_x^2}$$

Формула для расчета надежности через дисперсию баллов:

$$r_n = 1 - \frac{S_E^2}{S_x^2}$$

На надежность влияют:

1. Внутренняя согласованность теста, которая обеспечивается содержательной однородностью заданий
2. Длина теста, поскольку по мере роста количества заданий увеличивается надежность теста.

Формула Кьюдера-Ричардсона (KR-20)

Формула пригодна для заданий дихотомического типа и ее часто используют на практике

$$(KR - 20) = \frac{k}{k - 1} \left(1 - \sum \frac{PQ}{S_y^2} \right)$$

где P - доля учащихся, верно выполнивших задание, $Q = 1 - P$;

k – количество заданий, S_y^2 - квадрат стандартного отклонения.

Чем больше квадрат стандартного отклонения для всего теста, тем он надежнее.

Количественное значение коэффициента надежности может колебаться в интервале (0,7 – 0,9).

Определение надежности наиболее важно для нормативно-ориентированных тестов предназначенных для итогового контроля качества подготовки учащихся. Для тестов, ориентированных на критерии, определение надежности затруднено.