

Тема 10.2. “Принципы конструирования тестів”

И.А. Морев

ТРИДЦАТЬ ТРИ ПРИНЦИПА КОНСТРУИРОВАНИЯ ТЕСТА, СОЗДАНИЯ ТЕСТОВЫХ ЗАДАНИЙ И ВЫБОРА ТЕХНОЛОГИИ ТЕСТИРОВАНИЯ

// <http://www.eunnet.net/proceedings/>

(Открытый университет ДВГУ, г. Владивосток)

Педагогическое тестирование имеет довольно хорошо разработанную теоретическую базу, которая описана в научной и научно-методической литературе. К сожалению, нам не удалось встретить ни одного источника, где была бы изложена вся совокупность принципов, которым должны следовать разработчик тестовых заданий и создатель интерфейса программных дидактических средств для тестирований. Вероятно, здесь она публикуется впервые.

1. **Валидность теста.** Результаты тестирования группы претендентов должны соответствовать объективным характеристикам, данным претендентам их руководителями, коллегами, преподавателями. В соответствии с измеряемыми характеристиками учащихся принято различать виды валидности.

2. **Надежность теста** и технологии тестирования. Результаты тестирования подобных групп претендентов с помощью одного теста должны быть одинаковыми (подобными) и не зависеть от времени. Это второй по важности принцип, которому должны быть подчинены усилия составителя теста. Чаще всего надежность можно охарактеризовать лишь умозрительно, поскольку повторное тестирование одной группы претендентов по одному тесту приводит к росту результатов за счет эффектов памяти, самообучаемости и пр.; очень трудно найти подобные группы претендентов и доказать, что это так; на результаты оказывают влияние предыстория групп претендентов, технические сложности исключения предварительного общения между ними и пр.

Надежность - принцип, предъявляемый к инструменту технического измерения.

3. **Релевантность теста** и технологии тестирования. Выберем два одинаковых по валидности теста (или, точнее, две технологии тестирования). С точки зрения дидактики, если повторное тестирование с использованием первого теста (первой технологии) приводит к более выраженному улучшению результатов, чем с использованием второго теста (второй технологии), то первый тест (первая технология) - лучше. Пусть результаты в обоих случаях улучшаются по-разному за счет разной выраженности эффекта самообучаемости претендентов, привлекательности процедуры и др. С точки зрения принципа надежности оба теста (обе технологии) не надежны. Чем лучше тест (технология тестирования) с точки зрения дидактики, тем он (она) хуже, исходя из принципа надежности. Поэтому в педагогике, в отличие от техники, лучше говорить не о надежности теста (технологии), а о релевантности теста (технологии), т. е. степени соответствия принципам дидактики. Релевантность теста (технологии тестирования) - мера его способности к

инициации процессов самообразования. Релевантность можно выразить численно, взяв отношения прироста усредненных результатов при повторном тестировании с использованием разных тестов (технологий).

4. **Научная достоверность.** В тест включаются только те элементы знания и связи между ними, которые являются объективно истинными.

5. **Обобщенность.** Тема тестирования не должна быть излишне обобщенной.

6. **Значимость.** Значимое ТЗ отражает структурный информационный элемент дисциплины либо связь между структурными элементами, без которых знания становятся неполными, с пробелами.

7. **Соответствие современному знанию.**

8. **Соответствие источникам знания.** Все ТЗ должны включать адрес источника, который может быть включен и в демонстрируемую часть текста ТЗ в случаях, когда имеются разночтения в рекомендованной литературе.

9. **Репрезентативность.** Репрезентативный тест содержит совокупность ТЗ, отражающую все структурные элементы содержания дисциплины и их связи. Репрезентативная (представительная) БТЗ содержит совокупность ТЗ, достаточную для ведения вариативного контроля по любой из имеющихся в структуре дисциплины учебных тем. Репрезентативная БТЗ может включать ТЗ по дополнительным учебным материалам для проведения олимпиад и пр.

10. **Вариативность.** Для достижения достаточной вариативности теста в рамках учебной дисциплины необходимо, чтобы БТЗ содержала тестовые задания, соответствующие разным учебным пособиям и разным способам изложения учебного материала.

11. **Сложность.** Для каждого ТЗ в процессе контроля определяют усредненную выраженную числом сложность. Сложность задания № j определяют двумя способами, например так:

$q_j = 1 - p_j$, где p_j - относительное количество претендентов, выполнивших

данное задание. Либо как “логит”, равный $\lg\left(\frac{q_j}{p_j}\right)$. Претендент, правильно выполняющий сложные задания, с большой вероятностью правильно выполнит и легкие.

12. **Трудоемкость.** Она характеризует количество элементарных операций (и мыслительных, и физических, и расчетных), которые необходимо выполнить в процессе работы над заданием. Отношение трудоемкостей заданий можно определить через усредненное отношение времен их выполнения.

13. **Взвешенность.** Для придания результату тестирования большей объективности все ТЗ после сеанса “взвешивают”. Затем общий результат претендента вычисляют как сумму весов выполненных заданий. Иногда за вес ТЗ принимают его относительную сложность, т. е. долю претендентов, не сумевших его выполнить. Иногда вес приписывают заданиям волонтеристски, опираясь на преподавательский опыт. Оба подхода к вычислению весов не лишены недостатков.

14. **Системность.** Задания в БТЗ связываются между собой общей структурой знаний. Такая связь может быть определена умозрительно либо методами факторного анализа.

15. **Комплексность и сбалансированность.** Следует гармонично сочетать в тесте задания на проверку знаний теоретического материала (понятия, законы, закономерности, гипотезы, факты, структурные компоненты теории), методов научной и практической деятельности, умений решать типовые задания.

16. **Соответствие содержания и формы.**

17. **Разрешающая способность.** От сбалансированности ТЗ по сложности и трудоемкости зависит способность теста дифференцировать претендентов в соответствии с их свойствами. Эта способность теста, выраженная числом, называется разрешающей способностью.

18. **Дидактическая направленность.** Тест (технология тестирования), в соответствии с принципами дидактики, должен (должна) не только дифференцировать и измерять свойства претендентов, но и обладать свойством инициирования их самообучения и проявлять их стремление к повышению качества знаний, умений, навыков.

19. **Оптимальность.** В тест, как правило, включаются только те ТЗ, содержание которых не дублируется и равномерно покрывает, в своей совокупности, учебный материал. Часть ТЗ исключают из теста, когда известно, что претенденты не знают соответствующей части материала или, наоборот, знают ее гарантированно хорошо; претенденты не имеют навыков работы и поэтому не справятся с трудоемкими заданиями. Тест считают более оптимальным, если с его помощью можно измерить такие же характеристики и провести дифференцирование претендентов за меньшее время.

20. **Темперированность сложности.** Последовательность заданий теста такова, что каждое последующее Тз, по мнению составителя, труднее предыдущего.

21. **Темперированность трудоемкости.** Последовательность заданий теста такова, что каждое последующее ТЗ, по мнению составителя, более трудоемко, чем предыдущее. Применяется, когда оптимизируют тесты для измерения работоспособности, внимательности и подобных характеристик претендентов.

22. **Свобода выбора.** Принцип составления теста, пришедший на смену принципу “возрастающей сложности”. Принцип свободы выбора в технологиях компьютерного тестирования проявляется в том, что претендентов не ограничивают в выборе последовательности выполнения заданий.

23. **Вариативность.** Каждое обращение к БТЗ приводит к генерации нового варианта теста в соответствии с алгоритмом, ключ которого может формироваться в соответствии с регистрационными данными, датой и временем сеанса. Алгоритм, ориентирующийся при составлении варианта теста на личность претендента, может исключать возможность повторения ТЗ при повторном тестировании.

24. **Адаптивность.** Содержание теста адаптируют для придания тесту большей разрешающей способности в известной группе претендентов. Для адаптации теста к свойствам группы претендентов и каждого претендента в отдельно-

сти необходимо наличие в БТЗ такого набора классифицированных заданий, что возможно построение вариантов теста, нацеленных на измерение конкретных свойств.

25. **Полнота.** В банке ТЗ должны быть отражены абсолютно все структурные элементы содержания дисциплины и связи между ними.

26. **Целесообразность.** Содержание теста зависит от цели тестирования.

27. **Верифицируемость результата.** Тест может содержать равномерно распределенные среди всех заданий такие ТЗ, для которых характерны разная форма и равное содержание. Установив корреляцию результатов их выполнения, можно установить, как претендент выполнял задания - вдумчиво либо методом «случайного тыка».

28. **Сбалансированность.** Количество разделов БТЗ и включенных в них заданий должно коррелировать с содержанием, относительными объемами информации разделов и формой дисциплины.

29. **Стилистическая непротиворечивость.** ТЗ формулируются в привычной для претендентов форме с использованием словарных оборотов из рекомендованных им учебных пособий. Исключения - сюрпризные задания, включаемые в тест для противодействия возникновению в поведении претендентов эффекта «ровной дороги».

30. **Логическая и семантическая непротиворечивость.** Среди логических требований - определенность предмета измерения. Чем меньше пересечений дисциплин и их разделов, тем чище, определеннее выражается в тесте их содержание.

31. **Дистрактивность.** Варианты выполнения заданий формулируются так, чтобы правильные варианты были похожи на неправильные и наоборот. Формулировки правильных вариантов должны точно соответствовать материалу учебного курса.

32. **Избыточность.** Случается так, что в группу претендентов попадают учащиеся, учившиеся по разным программам и учебникам. Тест составляется один для всех и поэтому, чтобы сравнять шансы, предпринимаются следующие шаги: в тест включаются ТВ, составленные в формулировках разных учебников, примерно в равных пропорциях; эквивалентные по информации задания группируются, и участники информируются, что необходимо и достаточно правильно выполнить только одно задание группы; количество ТВ в тесте превышает норму, которую мог бы выполнить усредненный участник сеанса.

33. **Деловая игра.** Игровая технология позволяет сделать процесс тестирования привлекательным и достичь высоких успехов за счет естественной индивидуализации; возникновения соревнования: коллективности; публичности.

Используемые аббревиатуры:

ТЗ - тестовое задание;

БТЗ - банк (батарея) тестовых заданий.