

ПЕД
измерения

ПАРАЛЛЕЛЬНЫЕ ВАРИАНТЫ ТЕСТА: ДИДАКТИЧЕСКИЙ АСПЕКТ

**Александр Севрук,
Светлана Максимова**

Пермский региональный институт
педагогических информационных технологий (ПРИПИТ)
svet_syr@mail.ru

В статье рассматриваются способы разработки параллельных заданий на основе модульно-фасетной технологии. Предложен программный продукт, позволяющий автоматизировать процесс подготовки и генерации дидактически параллельных вариантов теста.

Ключевые слова: *модульно-фасетная технология, параллельные задания, дидактическая сложность*

В настоящее время в литературе, посвящённой разработке тестов, мало внимания уделяется вопросам автоматизации подготовки множеств параллельных заданий в тестовой форме.

Задание в тестовой форме — это лаконично сформулированное и рассчитанное на выполнение в течение небольшого промежутка времени контролирующее задание, предполагающее однозначный ответ. Правильность выполнения такого задания устанавливается сопоставлением ответа испытуемого с эталоном¹.

Слово «тест» может использоваться в двух смыслах. Во-первых, тест можно рассматривать как систему заданий в тестовой форме, предназначенную для выявления уровня учебных достижений испытуемых. Во-вторых, тест можно рассматривать как испытание.

Параллельные задания — термин, обозначающий равноценность заданий для оценок учебных достижений. Параллельность можно рассматривать в дидактическом или в статистическом смысле. Статистически параллельные тесты дают статистически одинаковые результаты на группе испытуемых. Параллельность заданий или тестов в дидактическом смысле обеспечивается родственным содержанием, одинаковыми постановками заданий и одинаковой мерой трудности.

1

*Севрук А.И.,
Максимова С.Е.*
Тестовые технологии
для учителя. Пермь.
Изд-во: ПРИПИТ, 2008.

Дидактическая сложность есть свойство задания, свидетельствующее о возможности его выполнения. В разных измерительных шкалах сложность можно характеризовать различными признаками. В порядковой шкале дидактическую сложность² можно, например, задавать такими величинами как «сложное», «средней сложности», «лёгкое». Задание можно считать «абсолютно сложным», если для его выполнения необходимо знать и уметь больше, чем это было обеспечено учебными материалами и самим процессом обучения. Но если задание содержит учебные объекты пройденного материала, то можно говорить, по крайней мере, о потенциальной возможности его выполнения учащимися. И только в зависимости от содержания задания и его конкретной постановки становится возможным говорить о его дидактической сложности.

Конечно, дидактическая сложность задания — не единственный фактор успешности его выполнения конкретным испытуемым. Успешность выполнения задания связана также с уровнем подготовленности испытуемого. Чем выше уровень подготовленности испытуемого, тем выше его шансы на выполнение задания.

В теории обычно рассматриваются эмпирические показатели сложности. Чем меньшее

количество испытуемых справилось с заданием, тем большей эмпирической сложностью обладает это задание. Вместе с тем уже на этапе разработки теста важно иметь возможность управлять сложностью составляющих его заданий. Это позволяет активизировать умственную деятельность испытуемых.

Сложность задания можно определять на основе предполагаемого числа и характера умственных операций (К.К. Tatsuoка, В.П. Беспалько и др.), а также изменяя информационную насыщенность задания и др.

Ни один из обсуждаемых факторов, очевидно, нельзя признать ведущим. Роль этих факторов возрастает или, наоборот, убывает в зависимости от подготовленности испытуемого. То, что для одного из них ново, для другого — известный факт. Рассмотрим эти факторы подробнее.

Уровень мыслительных операций, необходимых для выполнения задания, связан не только с предметной областью знаний, но и с общими учебными умениями, которые носят надпредметный характер. В психолого-педагогических исследованиях применяют различные классификации уровней развитости мышления (Б. Блум, Г.С. Ковалева, Э. Молль и др.).

В таксономии Б. Блума используется 6 уровней подготовленности:

Теория

2

Северук А.И.,
Юпина Е.А.
Мониторинг качества преподавания. М.:
Педагогическое общество России. 2003.

ПЕД	
	измерения

1. Знание (Knowledge).
2. Понимание (Comprehension).
3. Применение (Application).
4. Анализ (Analysis).
5. Синтез (Synthesis).
6. Оценка (Evaluation).

Чем выше уровень мыслительных операций или действий, требуемых для выполнения задания, тем оно сложнее. Рассмотрим несколько примеров, отражающих возможность использования классификации Б. Блума при составлении заданий в тестовой форме. На первом уровне «Знание» от испытуемого требуется воспроизвести факты, названия изучаемых явлений и предметов. Фактически речь идет о воспроизведении запомнившейся информации. В следующем примере задания в тестовой форме предполагается узнать формулу серной кислоты из трех предложенных формул:

1. СЕРНАЯ КИСЛОТА:

- 1) H_2S ;
- 2) H_2SO_4 ;
- 3) HCl ;
- 4) H_2SO_3 .

На втором уровне «Понимание» предполагается более глубокая проверка тех знаний, которые воспроизводятся. При устном ответе хорошим способом проверки понимания является задание привести собственные примеры. Пониманию отвечает также умение отличать объекты по каким-либо признакам. Например:

2. ПАРЫ ЭЛЕМЕНТОВ, ОТНОСЯЩИЕСЯ К ОДНОЙ ГРУППЕ В ПЕРИОДИЧЕСКОЙ ТАБЛИЦЕ (любое число правильных ответов):

- 1) Р и S;
- 2) Н и В.

На уровне «Применение знаний» предполагается практическое применение знаний и умений. Например, решение любого квадратного уравнения освоенным способом можно отнести к данному уровню учебных достижений. Такого рода достижения можно отнести к так называемому уровню «выполнение задачи по образцу».

На уровне «Анализ» предполагается выделение элементов задания, которые далее могут использоваться для преобразования информации к новому виду. По сути, речь идет о новизне постановки задачи с, возможно, уже известными учебными элементами. С чем большим количеством оригинальных постановок заданий справился испытуемый, тем лучше его академическая подготовка. Например, задание по химии (любое число правильных ответов):

3. ЭЛЕКТРОННЫЕ ФОРМУЛЫ ЭЛЕМЕНТОВ, ПЕРИОД КОТОРЫХ СООТВЕТСТВУЕТ ЭЛЕКТРОННОЙ СХЕМЕ АТОМА (РИС. 1):

- 1) $1S^22S^22P^63S^23P^1$;
- 2) $1S^22S^22P^63S^23P^63D^4$.

В этом задании оба ответа верные. Выполнение этого задания может осуществляться разными способами. Однако в любом случае испытуемому придется проанализировать схему строения атома, электронные формулы элементов, приведённые в ответах, сопоставить схему с каждой из формул. Заметим, что многократное выполнение испытуемым заданий подобного типа переводит это задание в разряд заданий, требующих выполнения «по образцу».

В школьной математике к этому уровню подготовки относятся прикладные задачи, в которых требуется составление алгебраических уравнений (например, задачи на движение). Анализ текста также можно отнести к 4 уровню сложности (по Б. Блуму).

Умение анализировать и синтезировать признаки, имеющиеся в задании, — следующий, более высокий уровень подготовленности испытуемого. Рассмотрим эти умения на примере.

4. ЕСЛИ ПЛОЩАДЬ КВАДРАТА ВЫРАЖЕНА ЧЁТНЫМ ЧИСЛОМ, ТО ЕГО СТОРОНА МОЖЕТ БЫТЬ ВЫРАЖЕНА ЧИСЛОМ (любое число правильных ответов):

- 1) чётным;
- 2) дробным;
- 3) иррациональным.

Выполнение этого задания может осуществляться разны-

ми способами. Однако в любом случае испытуемому придется проанализировать несколько вариантов чётных чисел, которыми может быть выражена площадь квадрата, каждый раз проверяя верность утверждений, получаемых поочередным замыканием вопроса ответами. По сути, эта проверка и есть синтез, т.е. смысловое соединение элементов задачи. Отметим, что замена последнего ответа на вариант «нечётным» заметно упрощает задание, поскольку в первоначальном варианте необходимо «поработать» с иррациональными числами разного типа (с радикалом и без).

Умение дать оценку, сделать заключение пока еще редко применяется в тестах учебных достижений. Задания в тестовой форме на оценку или общий вывод вряд ли возможны без какой-нибудь дополнительной информации, причём обладающей несомненной новизной для испытуемого. Такая информация может содержать художественный или технический текст, схему устройства, репродукцию картины и т.д. На основе одной и той же дополнительной информации может быть построено несколько последовательных заданий. Такого рода задания уместны, пожалуй, в тестах повышенной сложности, используемых для селекции испытуемых.

Теория

12/08/08

ПЕД	
	измерения

Информационная насыщенность задания

Чем больше информационная насыщенность задания, чем больше количество дидактических единиц в нем использовано, тем оно сложнее. Сложность информационно насыщенных заданий связана с необходимостью оперативно удерживать и использовать одновременно множество учебных терминов, понятий, правил, законов и т.д. Эффект повышенной информационной насыщенности, например, несут задания закрытой формы с произвольным количеством правильных ответов. В этом случае необходимо анализировать все имеющиеся ответы, отсутствует эффект узнавания, что характерно для заданий с одним правильным ответом. Пример задания по математике в начальной школе (любое число правильных ответов):

5. ЕСЛИ СТОРОНА КВАДРАТА СОВПАДАЕТ СО СТОРОНОЙ ТРЕУГОЛЬНИКА, ТО:

1) периметр квадрата меньше суммарного периметра двух треугольников;

2) суммарный периметр двух квадратов теперь больше суммарного периметра трёх треугольников.

Для успешного выполнения задания необходимо знать элементарные фигуры,

понятие периметра, уметь вычислять периметр, уметь оперировать элементарными геометрическими образами. Это задание, скорее всего, можно отнести и к заданиям с необычной (нестандартной) формулировкой, о которых пойдёт речь далее.

Необычность (нестандартность) формулировки

Нетривиальная постановка всегда требует размышления, поиска путей решения задачи. Все задания творческого типа, позволяющие уйти от заданий, проверяющих «механическую» память или действий по образцу, как правило, сложны для значительной части испытуемых. Использование большого числа таких заданий многократно увеличивает сложность теста, ограничивая сферы его применения. Вместе с тем использование заданий, обладающих новизной постановки, заданий, способствующих активизации творческих способностей испытуемого, необходимо рассматривать как одну из ценностей тестов. Именно такие тесты способны давать оценки учебным достижениям в контексте развивающего образования.

Важно подчеркнуть, что необычность, нестандартность формулировки — состояние

временное. То, что необычно в первый раз, постепенно перестаёт быть таковым при повторных использованиях.

Источники дидактической сложности при использовании нестандартных формулировок заданий в тестовой форме:

- содержательная гетерогенность (интеграция разнородных учебных элементов);
- операционная гетерогенность (мыслительные операции разного типа);
- латентность способа выполнения задания.

Приведём пример задания, в котором использована содержательная гетерогенность и латентность способа выполнения задания (любое количество правильных ответов).

6. ЕСЛИ СТОРОНА КВАДРАТА ВЫРАЖЕНА ИРРАЦИОНАЛЬНЫМ ЧИСЛОМ, ТО:

1) диагональ всегда выражается иррациональным числом;

2) площадь может быть выражена рациональным числом.

В задание включены базовые понятия алгебры и геометрии. Но, во-первых, числа необходимо связать с геометрическими свойствами фигур, а во-вторых, неясен способ выполнения задания. Как показывают последние исследования PISA, задания, в которых неясен способ выполнения, вызывают большие затруднения у российских участников.

Сложность тестового задания с выбором правильного ответа можно регулировать также количеством дистракторов. Например, задание с четырьмя ответами, среди которых только один — правильный, нам представляется проще, чем задание с двумя ответами, среди которых может быть любое количество правильных. В случае, когда количество правильных ответов произвольно, правильность выбора ответа соотносится с эталоном. Засчитывается ответ, совпадающий с эталоном. В примере двух ответов, среди которых любое число правильных ответов, возможны четыре варианта распределения правильных ответов: только первый правильный, только второй, оба правильные или ни одного. А эталон только один. Отсюда возникает вероятность угадывания $\frac{1}{4}$. Однако в случае одного правильного ответа (из четырёх) может проявиться эффект «узнавания», что снижает эффективность использования дистракторов.

Вероятность угадывания можно понизить, увеличивая количество ответов. Но в случае с одним правильным ответом увеличение количества дистракторов делает задание громоздким. А использование уже трёх ответов с произвольным количеством правильных

Теория

12/08/08

ПЕД
измерения

ответов снижает вероятность угадывания абсолютно правильного ответа вдвое, до 0,125. Отметим, что одинаковая расчётная вероятность угадывания всех правильных ответов в заданиях с одним и произвольным количеством правильных ответов не означает их равноценности. Задания с произвольным количеством правильных ответов являются более сложными уже потому, что вместо выделения (часто узнавания) одного ответа необходимо проанализировать все предложенные ответы.

Со сложностью заданий связан вопрос о введении для них весовых коэффициентов (например, чем сложнее задание, тем больше баллов начисляется за его выполнение). Однако переменность весов тестовых заданий, т.е. оценивание разных заданий разным количеством баллов имеет смысл только в относительно коротких тестах, задания которых носят творческий характер и требуют увеличенного времени для их выполнения. В тестах с большим числом заданий средней трудности нет необходимости использования заданий с разными весами, поскольку практически всегда достигается высокая дифференциация испытуемых по набранным баллам. А этого вполне достаточно для обеспечения надёжных измерений учебных достижений.

Мы рассмотрели вопрос о регулировании дидактической сложности заданий в тестовой форме. Теперь перейдём к вопросу об автоматизации подготовки дидактически параллельных заданий с использованием компьютера. Весьма эффективным способом получения параллельных формулировок заданий является фасетная технология.^{3,4} Фасетную технологию, на наш взгляд, правильнее было бы назвать фреймовой, вслед за фреймовым типом систем знаний, применяемых в теории искусственного интеллекта. Обычно фасетные конструкции применяются для построения содержания тестовых заданий. Однако использование фасетной конструкции одновременно в вопросе и в ответах многократно усиливает возможность создания параллельных заданий, что может быть использовано при построении компьютерных баз тестовых материалов.

База теста — система знаний, предназначенная для компьютерной генерации дидактически параллельных вариантов теста учебных достижений.

Хороший эффект для автоматического создания параллельных тестовых заданий и тестов даёт разработанный нами пакет программ GenTests, который позволяет создавать базы тестовых материалов с использованием модульно-фа-

3

Аванесов В.С.
Форма тестовых заданий. М.: Центр тестирования 2006.

4

Пестряева С.
Вопросы теории и методики разработки параллельных заданий в тестовой форме по английскому языку.// ПИ 2007. №3.

сетной технологии и генерировать тесты для бланкового и компьютерного тестирования. В его состав входят три основные программы:

- BaseInput — для создания баз тестовых материалов на основе модульно-фасетной технологии;
- Testing — для генерации параллельных вариантов тестовых заданий и тестов для проведения компьютерного тестирования.
- B-Test — для генерации параллельных вариантов тестовых заданий и тестов для проведения бланкового тестирования.

Суть модульно-фасетной технологии состоит в следующем. База тестовых материалов представляет собой набор модулей. Модуль — это некоторая обособленная, внутренне целостная составляющая дидактической системы. В базах тестовых материалов модулем может выступать та её часть, которая используется для генерации параллельных (по содержанию) и дидактической сложности) тестовых заданий.

Фасет — часть тестового задания, которую можно изменять (варьировать) при сохранении основной идеи задания. В следующих примерах фасет выделен фигурными скобками.

7. РУССКИЙ ПИСАТЕЛЬ {...} ВЕКА:

- 1) Ф.М. Достоевский;
- 2) А.Н. Радищев;
- 3) А.Н. Толстой.

Использование фасета в данном модуле позволяет сформировать несколько параллельных тестовых заданий, отличающихся друг от друга наименованием века. Если предполагается создавать задания с одним правильным ответом, то вместо многоточия можно вставлять численные наименования веков: 18, 19 или 20. Очевидно, в каждом таком задании будет свой правильный ответ. Здесь фасет использован в тексте задания. Как уже упоминалось выше, фасеты можно использовать и при подготовке списков ответов. В следующем примере использованы фасеты ответов:

8. ЧЁТНОЕ ЧИСЛО:

- 1) {12};
- 2) {15}.

Понятно, что вместо чисел 12 и 15 можно использовать другие чётные и нечётные числа.

В следующем примере фасеты заложены в задании и ответах, так как поставленный там вопрос можно легко трансформировать в другие подобные вопросы, используя идею фасета: «Русский исследователь Центральной Азии», «Русский исследователь Севера» и т.д., соответственно будут подбираться варианты ответов к вопросу.

9. РУССКИЙ ИССЛЕДОВАТЕЛЬ ДАЛЬНЕГО ВОСТОКА:

- 1)*Хабаров;
- 2)*Арсеньев;

ПЕД	
	измерения

- 3)* Беринг;
- 4) Пржевальский;
- 5) Дежнёв;
- 6) Семёнов;
- 7) Командоров;
- 8) Сахалинов;
- 9) Камчатов.

Схематично этот модуль можно изобразить в виде:

Задание

- {фасет 1}
- {фасет 2}
- {фасет 3}

В нашем примере каждый фасет состоит из трёх дидактических единиц. Таким образом, имеем в общей сложности 9 ответов, объединённых в 3 фасетные группы. Первая фасетная группа содержит три правильных ответа, следующие группы содержат только дистракторы. Испытания показали, что последние 3 ответа имеют высокую привлекательность для тех учащихся, которые плохо знают имена исследователей Дальнего Востока, но знают названия некоторых географических объектов, созвучных данным фамилиям.

Компьютерная программа (B-Test или Testing) в данном случае может сформировать в общей сложности $3 \cdot 3 \cdot 3 = 27$ троек ответов, отличающихся друг от друга хотя бы одним из ответов. Кроме того, можно получить $27 \cdot 6 = 162$ варианта ответов на одно и то же задание,

отличающихся либо содержанием ответов, либо их расположением относительно друг друга. В данном случае компьютерный генератор на основе случайного выбора будет формировать тестовые задания с одним правильным ответом, поскольку все правильные ответы (помечены *) расположены в одной фасетной группе.

Все такие модули можно использовать либо по отдельности (для генерации нескольких заданий), либо объединять в блоки, тем самым можно повысить вариативность базы теста. Однако блок модулей должен удовлетворять определённым требованиям, прежде всего требованию однородности всех модулей блока между собой как по тематике, так и по дидактической сложности.

При компьютерной генерации задания теста из блока базовых модулей реализуются три случайных процесса. Во-первых, выбирается один из модулей блока. Во-вторых, выбираются по одному ответу из каждой фасетной группы уже выбранного модуля. Наконец, производится расстановка ответов к заданию — в случайном порядке.

Приведём пример блока на основе предыдущего задания, составленного из двух модулей с тремя фасетными группами, по 3 ответа в каждой группе.

10. РУССКИЙ ИССЛЕДОВАТЕЛЬ ДАЛЬНЕГО ВОСТОКА:

- 1) Хабаров
- 2) Арсеньев
- 3) Беринг
- 4) Пржевальский
- 5) Дежнёв
- 6) Семёнов
- 7) Командоров
- 8) Сахалинов
- 9) Камчатов.

11. РУССКИЙ ИССЛЕДОВАТЕЛЬ ЦЕНТРАЛЬНОЙ АЗИИ

- 1) Семёнов
- 2) Крузенштерн
- 3) Лаптев
- 4) Пржевальский
- 5) Дежнёв
- 6) Баренцев
- 7) Командоров
- 8) Архангельский
- 9) Ляховский.

При создании формулировок ответов в модуле важно исключить создание конечных заданий с противоречивыми ответами.

Отметим, что из данного блока модулей можно создать $27 \cdot 2 = 54$ задания, отличающихся хотя бы одним ответом. С учётом случайного порядка следования ответов количество оригинальных заданий увеличивается еще в шесть раз. Использование таких блоков модулей кратно увеличивает мощность базы по охвату учебного материала, подлежащего контролю.

В контексте развивающего обучения важно, чтобы задания

в тестовой форме проверяли не только факты, но и побуждали к размышлению. Приведём возможные постановки заданий с тремя ответами, в которых фасеты использованы как в вопросе, так и в ответах.

В ОТЛИЧИЕ ОТ ПРОИЗВЕДЕНИЙ {Л.Н. ТОЛСТОГО}, В ПРОИЗВЕДЕНИЯХ А.П. ЧЕХОВА:

- 1) выражается недовольство обывательским существованием;
- 2) изображены богатейшие слои населения;
- 3) герои ищут смысл жизни.

Очевидно, что ответы (правильные и неправильные) можно подобрать только для конкретного содержимого фасета в вопросе. Если вместо Л.Н. Толстого использовать имя другого писателя, то, возможно, понадобятся другие списки ответов. Точно так же можно организовать фасет на том месте задания, где упоминается другой писатель — А.П. Чехов. Различные сочетания пар писателей, поэтов, других творческих личностей заметно увеличивает количество конкретных формулировок вопроса к заданию. Таким образом, создается аналог типового задания на сопоставление по критерию «сходство-различие» Задания данного типа требуют определенных размышлений, если ответы не являются очевидными фактами, которые

Теория

12/0000

ПЕД	
	измерения

легко запомнить. Главное, что ещё необходимо отметить — это высокое разнообразие конечных формулировок тестовых заданий, которое возникает за счёт использования фасетной технологии при формулировке вопроса, и при формулировках ответов.

Вариативность тестовых заданий является сильной стороной фасетной технологии. Чем больше вариантов фасетных формулировок, тем выше вариативность базы тестовых материалов. Следование этому правилу позволяет расширить

дидактическое содержание тестов учебных достижений.

Рассмотренные подходы к созданию и использованию тестов широко используются в Пермском крае как педагогами нашего педагогического вуза при обучении студентов, слушателей курсов профессиональной переподготовки и повышения квалификации, школьников по программам заочного дополнительного образования, так и школьными учителями, прикрепленными к экспериментальным площадкам вуза.