

# МОДЕЛИ ПЕДАГОГИЧЕСКОГО ТЕСТИРОВАНИЯ

**Сергей Дуплик,**

Казанский государственный техниче-  
ский университет им. А.Н. Туполева  
swd1999@mail.ru

**Вводится понятие модели педагогического тестирования как метода предъявления тестовых заданий обучаемому и оценивания результатов тестирования. Дается описание известных моделей, их достоинств и недостатков. Предлагаются возможные алгоритмы некоторых моделей.**

## **Введение**

Одним из способов быстрой проверки знаний является тестирование. В настоящее время, в связи с развитием информационных технологий, дистанционного обучения и, в частности, адаптивных обучающих систем, тестирование может применяться как средство идентификации личности обучаемого для построения индивидуальной последовательности обучения, когда каждый обучаемый проходит учебные курсы в том порядке и в том объеме, которые наиболее соответствуют его уровню подготовленности.

В технологии дистанционного обучения, при отсутствии непосредственного контакта обучаемого с преподавателем, тестирование становится одним из основных средств контроля знаний, поэтому особенно остро встает проблема создания качественных тестов, которые могли бы быстро, объективно и адекватно измерять уровень знаний обучаемых.

Тесты представляют образовательные процессы не в идеальном, а в реальном свете. В настоящее время всё большее внимание уделяется тестам как быстрому и удобному способу оценки знаний. Однако кажущаяся простота создания тестов и возросший спрос на них породили множество некачественных материалов, называемых «тестами».

При разработке тестов нужно учитывать два момента:

1. Разработка тестов, способных адекватно оценить знания обучаемых — это не просто составление заданий и объединение их в тест. Тест — это система заданий, в которой каждое задание должно удовлетворять определенным критериям.

2. Нельзя сводить проверку знаний к одному лишь тестированию. С помощью тестов нельзя проверить, например, доказательство теорем, и, конечно же, никакое тестирование не заменит живой беседы обучаемого с преподавателем. Поэтому достаточно часто тестирование

применяется как первый этап сдачи экзамена, по результатам которого осуществляется допуск ко второму этапу — собеседованию с преподавателем.

### Виды тестирования

Разработка тестовых заданий и обработка результатов тестирования подробно изложены в работах Аванесова<sup>1</sup> и Неймана<sup>2</sup>. Однако не менее важное значение имеет порядок предъявления заданий обучаемому и метод определения его уровня знаний по результатам тестирования, т.е. *модель тестирования*. Необходимо отметить, что рассматриваемые модели применяются для педагогического тестирования, т.е. оценки знаний отличаются от моделей, применяемых в психологических тестах для определения характеристик личности, поэтому назовём их *моделями педагогического тестирования*.

В адаптивных обучающих системах тестирование может применяться как средство идентификации личности для построения индивидуальной последовательности обучения.

При этом можно различать три вида тестирования: предварительное, текущее и итоговое.

*Предварительное* тестирование применяется перед началом обучения и направлено на выявление предварительных знаний обучаемого по ряду дисциплин, которые ему предстоит изучать. Сюда же могут включаться психологические тесты для определения индивидуальных характеристик личности обучаемого, ко-

торые учитываются в ходе обучения для настройки на работу с конкретным обучаемым. По результатам предварительного тестирования строится предварительная последовательность изучения учебных курсов.

*Текущее* тестирование — это контроль или самоконтроль знаний по отдельному элементу учебного курса, например, разделу или теме. По его результатам строится последовательность изучения тем и разделов внутри курса, а также может осуществляться возврат к темам, которые были изучены недостаточно хорошо.

И наконец, *итоговое* тестирование — это контроль знаний по курсу в целом или по совокупности курсов. По его результатам корректируется последовательность изучения учебных курсов.

### Девять моделей тестирования

В статье рассматривается девять моделей педагогического тестирования. Ниже приводится краткое описание каждой из них. Во всех моделях предполагается оценивание уровня знаний обучаемого по 100-балльной шкале (0 — нет знаний, 100 — полное знание). Метод перевода оценки из 100-балльной в обычную 5-балльную шкалу описан в работе<sup>2</sup>.

При работе с тестами всегда нужно учитывать надёжность результатов тестирования. Под *надёжностью* тестовых результатов понимается характеристика, показывающая точность измерения знаний заданиями теста. Нужно отметить, что речь идёт не

---

1

Аванесов В.С.  
Композиция тестовых заданий. М.: Центр тестирования. 2002.

---

2

Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. М., 2000.

о надёжности *теста*, а о надёжности *результатов тестирования*, т.к. на неё сильно влияет степень однородности различных групп обучаемых, уровень их подготовленности и ряд других факторов, связанных не с самим тестом, а с условиями проведения процесса тестирования. Методы расчёта коэффициента надёжности приведены в упоминавшихся выше работах Аванесова и Неймана.

### 1. Классическая модель

Данная модель является самой первой и самой простой. Имеется  $n$  заданий по определённой области знаний, по нескольким областям знаний или части области знаний (разделу, теме и т.п.). Из этого множества заданий случайным образом выбирается  $k$  заданий ( $k < n$ ), которые предлагаются обучаемому. Обучаемый выбирает или вводит правильный, по его мнению, ответ. Результат ответа на каждое задание оценивается как «правильно» или «неправильно». Результатом тестирования является процент правильных ответов обучаемого.

Достоинство: простота реализации. Недостатки:

1. Из-за случайности выборки нельзя заранее определить, какие задания по сложности достанутся обучаемому. В итоге одному обучаемому могут достаться  $k$  лёгких заданий, а другому —  $k$  сложных.

2. Оценка зависит только от количества правильных ответов и не учитывает сложность заданий.

Классическая модель из-за своих недостатков имеет самую низкую надёжность, т.к. отсутст-

вие учёта параметров заданий часто не позволяет объективно оценить знания обучаемого.

В настоящее время происходит уход от использования данной модели к более совершенным и эффективным моделям, например к адаптивному тестированию.

### 2. Классическая модель с учётом сложности заданий

Имеется  $n$  заданий по определённой области знаний, по нескольким областям знаний или части области знаний (разделу, теме и т.п.). Каждое задание имеет определённый уровень сложности  $T_i$ ;  $i=1, n$ .

Из этого множества заданий случайным образом выбирается  $k$  заданий ( $k < n$ ), которые предлагаются обучаемому. Результат ответа на каждое задание оценивается как «правильно» или «неправильно».

При подсчёте результата тестирования учитывается сложность вопросов, на которые обучаемый дал правильный ответ. Чем выше сложность вопроса, тем выше будет результат тестирования. Для вопросов, на которые был дан неправильный ответ, сложность не учитывается.

Недостаток: из-за случайности выборки нельзя заранее определить, какие задания по сложности достанутся обучаемому. В итоге одному обучаемому могут достаться  $k$  лёгких заданий, а другому —  $k$  сложных.

Модели с учётом сложности заданий позволяют более адекватно подойти к оценке знаний. Но случайность выбора заданий не

позволяет добиться параллельности тестов по сложности, т.е. одинаковости суммарных характеристик сложности заданий, что снижает надёжность тестирования.

### **3. Модель с возрастающей трудностью**

Данная модель описывается в работе 1. Имеется  $n$  заданий по определённой области знаний, по нескольким областям знаний или части области знаний (разделу, теме и т.п.). Каждое задание имеет определённый уровень трудности  $T_i$ ;  $i = 1, n$ .

Имеется  $m$  уровней трудности. В тесте должны присутствовать задания всех уровней трудности. Из этого множества заданий случайным образом выбирается  $k$  заданий ( $k < n$ ). Выбранные задания сортируются по возрастанию трудности, после чего предлагаются учащемуся. Количество заданий по каждому уровню должно быть одинаковым, либо распределение заданий по уровням сложности должно подчиняться нормальному закону.

Результат тестирования определяется аналогично модели с учётом сложности. Данная модель обеспечивает параллельность тестов по сложности, т.е. надёжность результатов тестирования ещё выше, чем в предыдущих моделях.

### **4. Модель с разделением заданий по уровням усвоения**

Данная модель описывается в работах Аванесова<sup>3</sup> и Неймана<sup>4</sup>.

Различают пять **уровней усвоения учебного материала**:

1. Нулевой уровень (Понимание) — это такой уровень, при котором учащийся способен понимать, т.е. осмысленно воспринимать новую для него информацию. Фактически речь идёт о предшествующей подготовке учащегося.

2. Первый уровень (Опознавание) — это узнавание изучаемых объектов при повторном восприятии ранее усвоенной информации о них или действиях с ними, например, выделение изучаемого объекта из ряда предъявленных объектов.

3. Второй уровень (Воспроизведение) — это воспроизведение усвоенных ранее знаний от буквальной копии до применения в типовых ситуациях. Примеры: воспроизведение информации по памяти, решение типовых задач по образцу.

4. Третий уровень (Применение) — это такой уровень усвоения информации, при котором учащийся способен самостоятельно воспроизводить и преобразовывать усвоенную информацию для обсуждения известных объектов и применения её в нетиповых ситуациях. При этом учащийся способен генерировать новую для него информацию об изучаемых объектах. Примеры: решение нетиповых задач, выбор подходящего алгоритма из набора ранее изученных алгоритмов для решения конкретной задачи.

5. Четвёртый уровень (Творческая деятельность) — это такой уровень владения учебным материалом темы, при котором учащийся способен создавать новую информацию, ранее не известную

3

Соловов А.В. Проектирование компьютерных систем учебного назначения: учебное пособие. Самара: СГАУ, 1995.

4

Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. М., 2000.

никуму. Пример: разработка нового алгоритма решения задачи.

Уровень представления обозначается символом  $a$  и может меняться от 0 до 4.

Задания составляются для каждого из пяти уровней. Сначала проводится тестирование с использованием заданий по уровню 0, затем по уровню 1, 2 и т.д. Перед переходом с уровня на уровень вычисляется степень владения учебным материалом на данном уровне и определяется возможность перехода на следующий уровень.

Для измерения степени владения учебным материалом на каждом уровне используют коэффициент:

$$K_{\alpha} = \frac{P_1}{P_2},$$

где:  $P_1$  — количество правильно выполненных **существенных операций** в процессе тестирования;

$P_2$  — общее количество **существенных операций** в тесте.

Под **существенными** понимают те операции, которые выполняются на проверяемом уровне  $a$ . Операции, принадлежащие к более низким уровням, в число существенных не входят.

Исходя из этого:  $0 \leq K_{\alpha} \leq 1$ .

Таким образом, уровень усвоения учебного материала может быть использован для оценивания качества знаний у обучаемого и выставления оценки. Рекомендуются следующие критерии для выставления оценки:

$K_{\alpha} < 0.7$  — неудовлетворительно;

$0.7 \leq K_{\alpha} < 0.8$  — удовлетворительно;

$0.8 \leq K_{\alpha} < 0.9$  — хорошо;

$K_{\alpha} \geq 0.9$  — отлично.

При  $K_{\alpha} < 0.7$  следует продолжать процесс обучения на том же уровне.

### 5. Модель с учётом времени ответа на задание

В данной модели при определении результата тестирования учитывается время ответа на **каждое** задание. Это делается для того, чтобы учесть возможность самостоятельного ответа на задания: обучаемый может долго искать ответ в учебнике или других источниках, но в итоге его оценка всё равно будет низкой, даже если на все вопросы он ответил правильно. С другой стороны, если он не пользовался подсказками, а долго думал над ответами, это означает, что он недостаточно хорошо изучил теорию, а в результате даже при правильных ответах оценка будет снижена. Аналогичный подход предлагается в [5].

Учёт времени ответа может производиться, например, по формулам (результат ответа на  $i$ -е задание теста):

$$R_i = \begin{cases} 1 - \frac{t_{ome} - t_{max}}{100} & \text{— если ответ} \\ & \text{правильный;} \\ 0 & \text{— если ответ} \\ & \text{неправильный.} \end{cases}$$

Если  $R_i > 1$ , то  $R_i = 1$ .

Если  $R_i < 0$ , то  $R_i = 0$ ,

где:  $t_{ome}$  — время ответа на задание;

$t_{max}$  — время, в течение которого уменьшение оценки не происходит.

$t_{max}$  установлено для того, чтобы обучаемый имел возможность прочитать вопрос и варианты от-

ветов, осмыслить их и выбрать правильный, по его мнению, ответ. Параметр  $t_{max}$  может задаваться как константа для всех заданий теста или вычисляться для каждого отдельно взятого задания в зависимости от его сложности, т.е.  $t_{max}^i = f(T_i)$ , т.к. логично предположить, что для ответа на сложное задание требуется больше времени, чем на простое задание. Другая возможная зависимость параметра  $t_{max}$  — от индивидуальных способностей обучаемого, которые должны быть определены ранее.

Результат тестирования:

$$R = \frac{\sum_{i=1}^k R_i}{k}; R \in [0,1]$$

Модели с учётом времени ответа на задание также позволяют повысить надёжность результатов тестирования, особенно в сочетании с моделью с учётом сложности заданий.

### 6. Модель с ограничением времени на тест

Имеется  $n$  заданий по определённой области знаний, по нескольким областям знаний или части области знаний (разделу, теме и т.п.).

Из этого множества заданий случайным образом выбирается  $k$  заданий ( $k < n$ ) и указывается максимальное время для прохождения **теста** (ответа на **все** оценные задания). Для оценки результатов тестирования берутся только те задания, на которые успел ответить обучаемый за дан-

ное время. Сам тест может быть построен по моделям 1–4.

В настоящее время данная модель используется достаточно широко.

В некоторых работах рекомендуется обязательно сортировать задания по возрастанию сложности и устанавливать такое время тестирования, за которое на все задания теста не сможет ответить ни один, даже самый сильный обучаемый. Такой подход предлагается применять при тестировании на бланках, когда обучаемый видит перед собой сразу все задания. Суть его в том, что когда обучаемый ответит на все задания, а время у него ещё останется, он может начать проверять свои ответы, сомневаться, а в итоге может исправить правильные ответы на неправильные. Поэтому рекомендуется или ограничивать время на тест, или забирать бланк сразу после ответа на все задания теста.

### 7. Адаптивная модель

Данная модель является продолжением классической модели с учётом сложности заданий.

Адаптивным называется тест, в котором сложность заданий меняется в зависимости от правильности ответов испытуемого. Если обучаемый правильно отвечает на тестовые задания — сложность последующих заданий повышается, если неправильно — понижается. Также есть возможность задавания дополнительных вопросов по темам, которые обучаемый знает не очень хорошо, для более тонкого выяснения уровня знаний в данных областях. Таким образом,

можно сказать, что адаптивная модель напоминает преподавателя на экзамене: если обучаемый отвечает на задаваемые вопросы уверенно и правильно, преподаватель достаточно быстро ставит ему положительную оценку. Если обучаемый начинает «плавать», то преподаватель задаёт ему дополнительные или наводящие вопросы того же уровня сложности или по той же теме. И наконец, если обучаемый с самого начала отвечает плохо, оценку преподаватель тоже ставит достаточно быстро, но отрицательную.

Данная модель применяется для тестирования обучаемых с помощью компьютера, т.к. на бумажном бланке невозможно заранее разместить столько вопросов и в том порядке, сколько и в котором они должны быть предъявлены обучаемому.

Тестирование обычно начинается с заданий средней сложности, но можно начинать и с лёгких заданий, т.е. идти по принципу повышения сложности (см. модель 3).

Тестирование заканчивается, когда обучаемый выходит на некоторый постоянный уровень сложности, например отвечает подряд на некоторое критическое количество вопросов одного уровня сложности.

Достоинства:

1. Позволяет более гибко и точно измерять знания обучаемых.

2. Позволяет измерять знания меньшим количеством заданий, чем в классической модели.

3. Выявляет темы, которые обучаемый знает плохо, и позво-

ляет задать по ним ряд дополнительных вопросов.

Недостатки:

1. Заранее неизвестно, сколько вопросов необходимо задать обучаемому, чтобы определить его уровень знаний. Если вопросов, заложенных в систему тестирования, оказывается недостаточно, можно прервать тестирование и оценивать результат по тому количеству вопросов, на которое ответил обучаемый.

2. Возможно применение только на ЭВМ.

Надёжность результатов тестирования в данном случае самая высокая, т.к. осуществляется приспособление под уровень знаний конкретного обучаемого, что обеспечивает более высокую точность измерений.

### **Возможный алгоритм адаптивной модели тестирования**

Данный алгоритм является достаточно простым и позволяет варьировать только уровень сложности, не учитывая статистику ответов на предыдущие вопросы. На каждом шаге тестирования по каждому уровню сложности обучаемому даётся два задания, и по результатам ответов на них определяется уровень сложности для следующих заданий. Данное количество заданий (два) позволяет более адекватно оценивать уровень знаний, чем одно задание, на которое обучаемый может отгадать или случайно забыть ответ, и в то же время не даёт большого количества сочетаний вариантов ответов, как в случае трёх

и тем более большего количества заданий.

Пусть имеется  $m$  уровней сложности. Вводится коэффициент  $k_T = \frac{100}{m}$ .

Обозначим:  $t$  – текущий уровень знаний обучаемого,  $t_n$  – нижний уровень знаний,  $t_b$  – верхний уровень знаний. Все уровни знаний будем измерять от 0 до 100 (0 – нет знаний, 100 – абсолютное знание).

1. Установить  $t = 50$ ;  $t_n = 0$ ;  $t_b = 100$ .

2. Вычислить текущий уровень сложности  $T = \frac{t}{k_T}$ .

3. Выдать два задания сложности  $T$ . Пусть  $k_{пр}$  – количество правильных ответов;  $K_{пр} \in [0, 2]$ .

Пересчёт уровня знаний: если  $k_{пр} = 2$ , то  $t_n = t$ ;  $t_b = t_b + 0.5t$ . Если  $t_b > 100$ , то  $t_b = 100$ . Если  $k_{пр} = 1$ , то  $t_n = t_n / 4$ ;  $t_b = t_b + 0.1t$ . Если  $t_b > 100$ , то  $t_b = 100$ . Если  $k_{пр} = 0$ , то  $t_n = t_n / 2$ ;  $t_b = t$ .

$$4. t_1 = \frac{t_b + t_n}{2}.$$

5. Если  $|t - t_1| < \varepsilon$ , то уровень знаний равен  $t_1$ , выход.

6. Перейти к шагу (2).

$\varepsilon$  устанавливается исходя из необходимой точности оценки знаний. Однако с уменьшением  $\varepsilon$  возрастает число вопросов, необходимых для включения в тест.

## 8. Модель тестирования по сценарию

Данная модель также является продолжением классической модели. Данная модель реализует-

ся в Интегрированной системе управления ВУЗом (ИСУ ВУЗ), разрабатываемой в Университете управления ТИСБИ (г. Казань)<sup>5</sup>.

Существенным недостатком является непараллельность тестов для различных обучаемых, т.к. нельзя заранее определить, какие задания по сложности и по каким темам достанутся обучаемому. Поэтому при сценарном тестировании преподаватель перед тестированием формирует сценарий тестирования, где может указывать:

- количество заданий по каждой теме, которые должны быть включены в тест;
- количество заданий каждого уровня трудности, которые должны быть включены в тест;
- количество заданий каждой формы, которые должны быть включены в тест»;
- время прохождения теста;
- другие параметры.

Сценарий может создаваться по любому объёму учебного материала: разделу, предмету, специальности и т.д.

Формы тестовых заданий и требования к ним подробно описываются в упоминавшейся выше работе В.С. Аванесова. Существует четыре формы тестовых заданий:

1. Задания с выбором, которые делятся на 3 подгруппы: задания с выбором одного правильного ответа (или одновариантные задания), задания с выбором нескольких правильных ответов (или многовариантные задания), задания с выбором наиболее правильного ответа.

2. Задания открытой формы.

3. Задания на установление соответствия.

4. Задания на установление правильной последовательности.

Непосредственно при тестировании выборка заданий каждого уровня сложности, по каждой теме, каждой формы и т.д. производится случайным образом из общей базы заданий, поэтому каждый обучаемый получает свои задания. Получаемые тесты для всех обучаемых являются параллельными, т.е. имеют одинаковое число заданий и одинаковую суммарную сложность. Но в отличие от модели с возрастанием сложности, которая также обеспечивает параллельность, здесь разработчик теста решает сам, сколько и каких заданий должно быть предъявлено по каждой теме, следовательно, обеспечиваются абсолютно одинаковые условия тестирования для всех обучаемых.

По сравнению с адаптивной моделью данная модель является менее эффективной, т.к. не настраивается под индивидуальные особенности каждого обучаемого, однако имеет преимущество психологического характера: при тестировании по адаптивной модели обучаемые отвечают на разное количество вопросов и как будто бы находятся в разных условиях. В случае тестирования по сценарию все обучаемые получают одинаковое количество вопросов по каждой теме и по каждому уровню сложности.

Надёжность результатов тестирования сопоставима с надёжностью, получаемой при тестировании с возрастанием сложности.

## 9. Модель на нечёткой математике

Цель введения нечёткой математики — попытка математической формализации нечётких, качественных явлений и объектов с размытыми границами, встречающихся в реальном мире. Нечёткое управление оказывается особенно полезным, когда описываемые процессы являются слишком сложными для анализа с помощью общепринятых количественных методов или когда доступные источники информации интерпретируются качественно, неточно или неопределённо.

Экспериментально показано, что нечёткое управление даёт лучшие результаты по сравнению с получаемым при общепринятых алгоритмах управления. Нечёткая логика, на которой основано нечёткое управление, ближе по духу к человеческому мышлению и естественным языкам, чем традиционные логические системы. Нечёткая логика, в основном, обеспечивает эффективные средства отображения неопределённостей и неточностей реального мира. Наличие математических средств отражения нечёткости исходной информации позволяет построить модель, адекватную реальности<sup>6</sup>.

Данная модель тестирования является развитием любой предыдущей модели, в которой вместо чётких характеристик тестовых заданий и ответов используются их нечёткие аналоги. Примерами могут служить:

- сложность задания («лёгкое», «среднее», «выше среднего», «сложное» и т.п.);

- правильность ответа («правильно», «частично правильно», «скорее неправильно», «неправильно» и т.п.);
- время ответа («маленькое», «среднее», «большое», «очень большое» и т.п.);
- процент правильных ответов («маленький», «средний», «большой», «очень большой» и т.п.);
- итоговая оценка;
- и другие.

Введение нечётких характеристик может помочь преподавателям разрабатывать тесты. Например, преподаватель может достаточно быстро определить, является ли задание сложным или нет. Но сказать точно, насколько оно сложно, например, по 100-балльной шкале, или точно оценить разницу сложностей двух заданий будет для него достаточно трудно. С точки зрения обучаемого, нечёткая оценка его знаний в виде «хорошо», «отлично», «не очень хорошо» и т.п. более понятна ему, чем чёткое количество баллов, которое он набрал в результате тестирования. Пример такой модели описывается в работе<sup>7</sup>.

Модели могут быть комбинированными, например:

- классическая модель, учитывающая сложности заданий, и модель с учётом времени ответа на задание;
- модель с возрастающей сложностью и модель с учётом времени ответа на задание;
- модель с возрастающей сложностью и модель с ограничением времени на тест;

- модель с учётом времени ответа на задание и адаптивная модель;
- модель с учётом времени ответа на задание и модель на нечёткой математике;
- модель с разделением заданий по уровням усвоения и модель с учётом сложности заданий;
- и другие.

## Заключение

Тестирование — быстрый и эффективный способ контроля и оценки знаний. Однако возросший интерес к тестированию приводит к появлению некачественных тестов, которые могут неправильно и необъективно измерять уровень знаний обучаемых. Борьба с этой ситуацией призвана наука, которую называют «педагогические измерения». Кроме правильной разработки тестовых заданий, важна методика предъявления их обучаемому или *модель педагогического тестирования*. В настоящее время всё реже используются старые классические модели, которые не позволяют достаточно быстро и эффективно оценивать знания. В связи с этим разрабатываются и внедряются новые адаптивные модели, которые настраиваются на уровень подготовленности обучаемых, в том числе с использованием нечёткой математики. Как показывает практика, за такими моделями — будущее, и на их развитие нужно ориентироваться при разработке тестовых программ.