

## ВЫБОР ОПТИМАЛЬНОГО ВРЕМЕНИ ТЕСТИРОВАНИЯ

**Владимир Ким**

Уссурийский государственный педагогический институт  
vskim@mail.ru

Обсуждается проблема определения оптимального времени (длительности) тестирования. Время тестирования зависит от числа и меры трудности заданий. Главным критерием является максимум дифференцирующей способности теста. При проведении тестирования необходимо учитывать время суток и день недели. Приведены рекомендации по определению числа заданий теста и времени его проведения.

**Ключевые слова:** тест, время тестирования, длительность тестирования, число заданий теста.

### Актуальность проблемы

Педагогические тесты преследуют вполне определённую цель, например, ранжировать испытуемых по уровню достижений. На достижение этой цели влияет как качество тестовых заданий, так и их количество. Общее время тестирования определяется двумя главными факторами — количеством и трудностью заданий. Должно ли это время быть ограниченным или не ограниченным — определяется конкретной ситуацией, в которой применяется тест.

Многие авторы согласны с тем, что правильное определение времени тестирования является важным фактором, влияющим на качество результатов теста. Например, А.Н. Майоров<sup>1</sup> отмечает, что каждый тест имеет оптимальное время тестирования, уменьшение или превышение которого снижает качественные показатели теста. В.С. Аванесов считает тест кратковременным испытанием, а потому полагает, что время педагогического тестирования

Теория

ИЗДАНИЕ

1

*Майоров А.Н.*  
Теория и практика  
создания тестов для  
системы образования. —  
М.: Интеллект-центр,  
2001. 296 с.

2  
 Аванесов В.С.  
 Проблема становления  
 системы педагогических  
 измерений.  
 Педагогические  
 измерения, 2008. № 2  
<http://testolog.narod.ru>

3  
 Аванесов В.С.  
 Основы научной органи-  
 зации педагогического  
 контроля в высшей шко-  
 ле. — М., МИСиС, 1989.  
 167 с.

4  
 Nunnally J.C.  
 Psychometric theory. New  
 York: McGraw-Hill.  
 1978.

5  
 Вспомним, как резко  
 меняется качество игры  
 шахматиста в зависимо-  
 сти от того, играет он  
 двухчасовую партию  
 или пятиминутный  
 блиц.

не должно превышать примерно сорока минут<sup>2</sup>. Ранее он же отмечал, что время выполнения теста является важным системообразующим фактором при разработке теста<sup>3</sup>. Из зарубежных авторов этому вопросу уделял внимание классик зарубежной психометрической науки J. Nunnally<sup>4</sup>. Он ввёл понятия *power test* и *speed test*. Они переводятся, соответственно, как «тест без ограничения времени» и «тест с ограничением времени». Если за отведённое время попробовать ответить на все задания успевают порядка 90% испытуемых, то такой тест этот автор относит к числу *power test*. Если процент испытуемых, успевающих ответить на все задания теста, меньше, то это будет *speed test*.

Повышенное внимание этому вопросу уделяется по той причине, что неверно установленное время тестирования не позволяет тесту достичь своей цели — проверить, насколько испытуемый подготовлен. Отсюда — актуальность рассматриваемой темы.

## Влияние времени на результаты тестирования

К чему приведёт, например, слишком малое время тестирования? Слабые учащиеся не справятся с тестом потому, что не успеют даже разобраться, о

чём их спрашивают, в то время как сильные — потому, что не имели достаточно времени на выполнение заданий. У всех испытуемых будут примерно одинаково низкие индивидуальные баллы, то есть произойдет уменьшение дисперсии исходных тестовых баллов испытуемых, а это означает и снижение дифференцирующей способности теста. Результаты такого теста не будут объективно отражать уровень подготовленности учащихся<sup>5</sup>.

Так же неблагоприятно влияет на тестирование и слишком большое время выполнения теста. В этом случае мы также получим негативное воздействие на измерительные качества теста. В частности, сильные учащиеся, досрочно завершив тестирование, в оставшееся время начнут вести себя неадекватно требованиям процесса тестирования. Другие испытуемые будут долго сидеть над заданиями, не решаясь выбрать ответ. Это вызовет у них утомление, снижение концентрации внимания, расслабление, что также снижает точность оценок тестирования.

Утомление обусловлено чувством усталости, которое проявляется процессами торможения в клетках коры головного мозга. В состоянии утомления испытуемый способен показать лишь малую долю своих истинных способностей.

Тестировать его в этом случае бесполезно, так что мы не добьемся цели тестирования.

А.Н. Майоров<sup>6</sup> приводит следующие эффекты проявления утомления:

1) на поведенческом уровне — приводит к уменьшению скорости и точности работы;

2) на физиологическом уровне — приводит к повышению инерции в динамике нервных процессов;

3) на психологическом уровне — ведёт к нарушению качеств внимания, процессов памяти, степени адекватности функционирования интеллектуальных процессов;

4) происходят сдвиги в эмоционально-мотивационной сфере.

### Оптимальное время тестирования

Оптимальное время тестирования — это время от начала процесса тестирования до момента наступления утомления<sup>7</sup>.

Как же определить оптимальное время тестирования? Вопрос не простой. Начнём с практических рекомендаций. Разработчики педагогических тестов на выполнение одного задания обычно отводят 30-60 секунд. Если задания соответствуют простому «узнаванию» (первый уровень таксономии Блума), то, как показывает наш опыт, вполне достаточно и

меньшего времени, 5–10 секунд. Но по мере продвижения на верхние уровни таксономии Блума время тестирования должно увеличиваться.

Количество заданий теста и время тестирования — тесно связанные и в определённом смысле эквивалентные характеристики, но определяющим всё же является время тестирования, поскольку оно задаёт порог утомления, за которым тест начинает терять свои измерительные свойства.

Теоретически рассчитать это время невозможно, поэтому рекомендуется использовать эмпирические данные по результатам первичной апробации теста. При наличии опыта, ещё на этапе разработки тестового задания, можно примерно оценить время его выполнения. Суммарное время по всем заданиям даст общее предполагаемое время тестирования. В.С. Аванесов закладывает этот показатель в свою спецификацию теста.

### Две точки зрения

Ж. Nunnally и В.С. Аванесов, предлагают определять время эмпирически, используя разные варианты времени тестирования, на параллельных выборках испытуемых и фиксации того временного периода, где дисперсия тестовых результатов достигает максимального

Теория

2008

6

Майоров А.Н.  
Теория и практика создания тестов для системы образования. — М.: Интеллект-центр, 2001. 296 с.

7

Аванесов В.С.  
Материалы лекций.

ПЕД  
измерения

значения. (рис.1). А.Н. Майоров считает, что оптимальное время тестирования соответствует не максимуму дисперсии, а моменту начала её увеличения<sup>8</sup>.

Рассмотрим детальнее эти точки зрения. На рисунке показана гипотетическая зависимость дисперсии тестовых результатов от времени тестирования.

Предположим, что все испытуемые одновременно начинают и одновременно заканчивают процесс тестирования при условии, что досрочное завершение не допускается. Очевидно, что при очень малом времени тестирования все испытуемые одинаково не выполнят ни одного задания, то есть дисперсия должна равняться нулю. При очень больших значениях времени тестирования

при том же числе заданий больше вероятность появления высоких баллов у многих испытуемых, особенно если у них появятся возможности списывать или пробовать подставлять поочерёдно готовые ответы в задания. В этих случаях повысится средний арифметический тестовый балл, но понизится дисперсия. При оптимальном времени тестирования (согласно В. Аванесову, это  $t_3$ ) дисперсия будет максимальной. А. Майоров, считает, что оптимальное время тестирования соответствует точке  $t_2$ .

Разделим весь временной интервал на пять характерных областей I, II, III, IV и V. В области I (очень малые времена тестирования) дисперсия быстро растёт в связи с тем, что время реакции у испытуемых разное и поэтому часть испы-

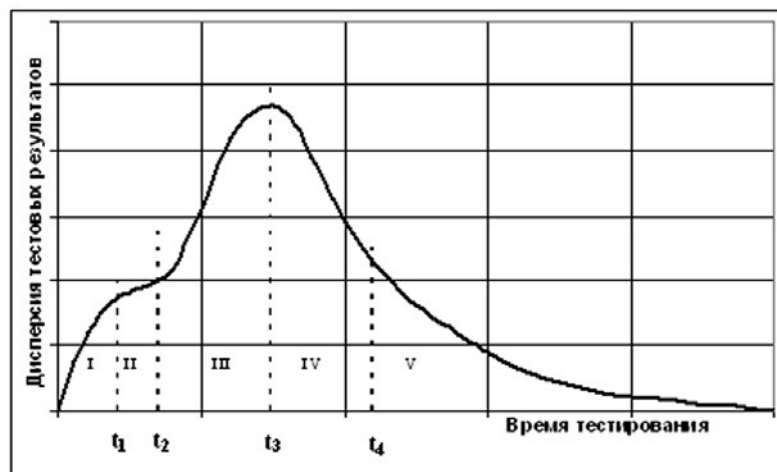


Рис. 1. Зависимость значения дисперсии тестовых

8

Майоров А.Н.  
Теория и практика создания тестов для системы образования. — М.: Интеллект-центр, 2001. 296 с.

туемых начнет успевать справляться с некоторыми заданиями теста. Произойдет дифференциация испытуемых, что и обусловит быстрый рост дисперсии на начальном участке.

Далее, во второй области рост дисперсии замедляется, так как теперь испытуемые с замедленной реакцией тоже начнут успевать выполнять задания. Темп увеличения дифференциация испытуемых уменьшится, то есть замедлится рост дисперсии тестовых баллов.

В третьей области скорость возрастания дисперсии снова увеличится. Это происходит по причине того, что теперь время тестирования достаточно велико и большинство испытуемых успевают полноценно проанализировать задания. Здесь начинает работать другой механизм — дифференциация испытуемых происходит не за счёт различия во времени реакции, а за счёт различия в уровне подготовленности.

В точке  $t_3$  дисперсия достигает максимума и далее, в области IV, начнет снижаться. Уменьшение дисперсии обусловлено усилением утомления испытуемых. В области V утомление становится настолько сильным, что дисперсия тестовых баллов падает практически до нуля.

В области III утомление испытуемых, появившись, начинает воздействовать на дисперсию тестовых баллов, а в точке  $t_3$  становится настолько сильным, что начинает снижать дисперсию.

Из приведённого анализа следует, что точка зрения J. Nunnally и В.С. Аванесова предпочтительнее, во всяком случае, для тестов с нормативно-ориентированной интерпретацией тестовых результатов. Основным доводом в пользу такого вывода является то, что важнейшей задачей теста является дифференциация испытуемых по уровню их подготовленности. В точке  $t_3$  эта дифференциация будет в основном обеспечена различием именно по этому уровню.

Таким образом, для эмпирического определения оптимального времени тестирования необходимо провести серию экспериментов различной длительности. Эти серии надо неоднократно повторять на выборках испытуемых, как можно более близких по своим характеристикам. После обработки собранного статистического материала, необходимо построить функцию, как показано на рис. 1, и определить значение момента времени  $t_3$ . Это и будет оптимальное время тестирования.

Теория

120000

ПЕД  
измерения

9

Черненко В.В., Котенкова Н.А., Лобанова И.В.  
Пряженникова О.А.  
О механизме возникновения систематической погрешности при тестировании уровня интеллектуальных способностей // Мат. Всерос. НТК, посвящённой 300-летию военного, военно-морского и высшего профессионального образования в России. Т. 1. Военно-исторические, военно-педагогические, гуманитарные и социально-экономические вопросы. ТОВМИ им. С.О. Макарова. — Владивосток, 2000. С. 156–158.

10

Колдаева В.Б.,  
Колдаев В.М.  
К вопросу о планировании занятий по теоретическим дисциплинам // Гуманитарные и социально-экономические аспекты обучения и воспитания кадров ВМФ. Сб. научных статей. Вып. 3. — Владивосток, ТОВМИ, 2000.

## Экспериментальные результаты

До сих пор мы обсуждали время тестирования как таковое, абстрагируясь от личности испытуемого. Зависимость на рис. 1 — гипотетическая, основанная на умозрительных предположениях. В.В. Черненко<sup>9</sup> приводит экспериментальные данные по временной зависимости степени утомления (рис. 2)<sup>10</sup>. В первые 9 минут эксперимента объём информации, воспринимаемой осознанно как по зрительному, так и по слуховому каналам, достигает своего максимума. Далее, в течение последующих 18 минут плавно, но незначительно снижается, а в последующие 9 минут для зрительного канала достигает первоначального значения, для слухового канала достигает первоначально

значения в течение 18 минут. По истечении 45 минут относительный объём осознанно воспринимаемой информации довольно значительно падает.

Если предположить, что занятия по теоретическим дисциплинам и тестирование требуют равного высокого интеллектуального напряжения, то из приведённой зависимости следует, что утомляемость испытуемых начинает заметно проявляться через время  $t_y$  равное 36 минутам после начала тестирования. Оптимальное время тестирования  $t_3$  соответствует моменту, когда положительный эффект, обусловленный увеличением времени тестирования, будет компенсирован отрицательным воздействием утомления испытуемых. Время  $t_3$  должно быть немного больше времени  $t_y$  — запаздывание обеспечи-

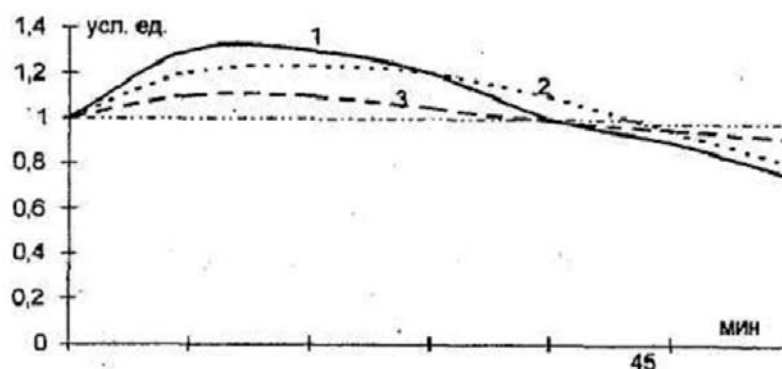


Рис. 2. Изменение относительного объёма воспринимаемой информации по зрительному (1), слуховому (2) каналам в течение занятия

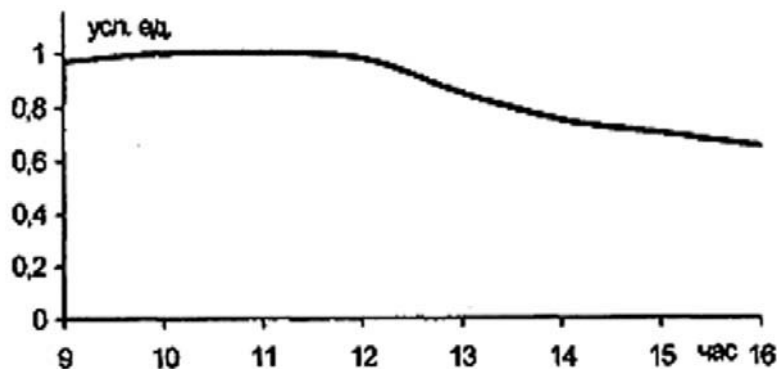


Рис. 3. Изменение относительного объема воспринимаемой информации по часам учебного дня

вает попадание в область максимума дисперсии тестовых результатов. Тогда из этих данных следует, что  $t_3$  примерно равно 40–45 мин. Эти значения находятся в удовлетворительном согласии с рекомендациями ограничивать число заданий теста примерно заданиями<sup>11</sup>. Если на одно задание отводить 30–60 секунд, то общее время тестирования составит примерно 50 минут. А.И. Буравлев и В.Ю. Переверзев<sup>12</sup> показали, что 50 заданий теста с критериально-ориентированной интерпретацией могут обеспечить надёжность результатов, равную 0,90. Иными словами, тест из 50–60 заданий, с одной стороны, обеспечивает достаточно высокую надёжность, а с другой — эффект утомления для такого теста ещё слабо влияет на результаты.

В.В. Черненко отмечает, что источником систематической погрешности может стать и пренебрежение суточным и недельным распределением момента начала тестирования.

Зависимость, показанная на рис. 3, свидетельствует о том, что проведение тестирования в послеобеденное время приведёт к снижению результатов по сравнению с утренним тестированием<sup>13</sup>. Пренебрежение этим эффектом может крайне негативно сказаться на результатах нормативно-ориентированного тестирования. Получается, что ранг испытуемого зависит от того, когда его тестировали — утром или вечером.

Аналогичный эффект проявляется при проведении тестирования в различные дни недели. Результаты исследований, представленные на рис. 4, показывают, что день недели

## Теория

### 11

Аванесов В.С.  
Материалы лекций.

### 12

Буравлев А.И.,  
Переверзев В.Ю.  
Выбор оптимальной  
длины педагогического  
теста и оценка надёжно-  
сти его результатов.  
[http://www.e-  
joe.ru/sod/99/2\\_99/  
st160.html](http://www.e-joe.ru/sod/99/2_99/st160.html).

### 13

Черненко В.В.,  
Котенкова Н.А.,  
Лобанова И.В.  
Пряженникова О.А. О  
механизме возникновения  
систематической  
погрешности при тести-  
ровании уровня интел-  
лектуальных способнос-  
тей // Мат. Всерос.  
НТК, посвящённой  
300-летию военного, во-  
енно-морского и высше-  
го профессионального  
образования в России.  
Т. 1. Военно-историче-  
ские, военно-педагогиче-  
ские, гуманитарные и  
социально-экономиче-  
ские вопросы. ТОВМИ  
им С.О. Макарова. —  
Владивосток, 2000.  
С. 156–158.

ПЕД	
	измерения

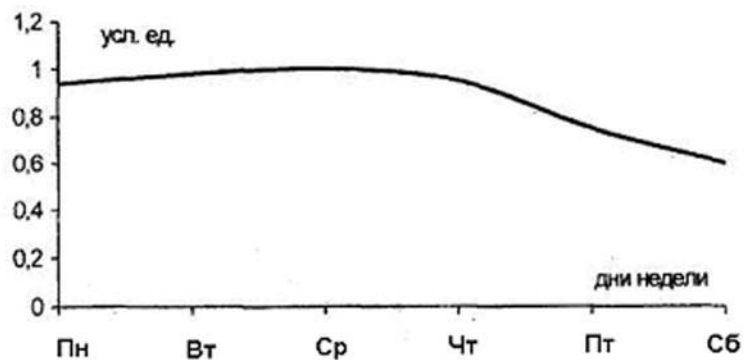


Рис. 4. Относительные объемы воспринимаемой информации по дням недели

может вносить систематическую погрешность в результаты тестирования.

Исследования В.В. Черненко показывают, что указание оптимального времени является необходимым, но недостаточным параметром теста. При определении выборки стандартизации необходимо указывать день недели и часы тестирования. Это особенно важно для тестов с нормативно-ориентированной интерпретацией тестовых результатов.

По данным А. Анастаси, результаты тестирования интеллекта для одного и того же испытуемого в начале недели могут дать показатель, равный 110, а в конце недели — 80. Вполне возможно, что это связано со снижением работоспособности к концу недели.

## Выводы

Таким образом, при определении времени тестирования необходимо учитывать следующие рекомендации:

1) время тестирования определяется по расположению максимума дисперсии тестовых результатов и не должно превышать 60 минут;

2) число заданий теста при таком времени не должно превышать 60 минут, учитывая предположение, что на выполнение одного задания требуется не более одной минуты;

3) тестирование необходимо проводить в первой половине дня;

4) тестирование желательно проводить не позднее середины недели.