

УДК 331.103.3

## СТРАТЕГИИ ПРОХОЖДЕНИЯ ТЕСТОВ ЗНАНИЙ, ВЫЯВЛЕННЫЕ МЕТОДОМ ХРОНОМЕТРИИ ПРОСМОТРА ВАРИАНТОВ ОТВЕТА

Д.Н. ЩЕРБИНА

e-mail: dnsherbina@sfedu.ru

Южный федеральный университет

*По результатам апробации методики тестирования знаний с хронометрией просмотра вариантов ответа были выявлены четыре стратегии выполнения тестов в условиях образовательного процесса. По мере снижения уровня знаний это стратегии: (1) извлечение ответа из памяти; (2) сопоставление вопроса и вариантов ответа; (3) угадывание; (4) поиск ответа во внешних источниках. С каждой из этих стратегий может быть сопоставлена оценка знания в диапазоне от 0 до 100 %. Показана перспективность разработанной автоматизированной модели идентификации стратегии прохождения тестов на основе анализа обнаруженных и описанных устойчивых паттернов в траектории выполнения заданий.*

**Ключевые слова:** компьютерное тестирование знаний, педагогические измерения, дистракторный анализ, электронные обучающие системы, moodle.

## STRATEGIES FOR PASSING THE KNOWLEDGE TESTS, IDENTIFIED BY THE DISTRACTOR VIEW CHRONOMETRY

D.N. SHERBINA

Southern Federal University

*Approbation of a novel knowledge testing methodology with the distractor view chronometry revealed four strategies for passing tests in a higher school. By decreasing the level of knowledge these strategies are: (1) recalling the answer from memory; (2) comparison of questions and distractors; (3) guessing; (4) searching for answers in external sources. Each of these strategies can be scored at the single knowledge scale ranging from 0 to 100 %. Author show the prospects of development of the automated strategy identification model based on the analysis of detected and described sustainable patterns of task execution.*

**Key words:** computer testing, knowledge tests, pedagogical measurements, distractor analysis, LMS, moodle.

doi:10.18522/2218-2268-2015-4-112-121

### Введение

Традиционная процедура педагогического измерения знаний, которая включает чтение вопроса и выбор одного или нескольких вариантов ответа [Беспалько, 2002; Osterlind, 2004; Аванесов, 2006], может значительно варьировать по составу элементарных когнитивных операций [Morrison, Free, 2001; Thompson, O'Loughlin, 2015].

Современные теории тестирования нацелены на сведение результатов тестирования к единой шкале способностей [Drasgow et al., 1995; DeMars, 2010]. При этом стратегия студентов при прохождении теста зависит от индивидуальных особенностей и уровня знаний [McCoubrie,

2004; Conole, Warburton, 2005; Переверзев, 2008]. Разные стратегии отражаются в особенностях траектории перемещения фокуса внимания по компонентам тестового задания.

Правильная интерпретация траектории прохождения теста с учетом стратегии студента увеличивает информативность теста по сравнению со стандартным подходом, когда информация об ответе на каждый вопрос ограничивается оценкой верно/неверно.

Целью данной работы было эмпирически выявить стратегии выполнения тестов в условиях образовательного процесса. Для этого были использованы результаты апробации методики тестирования знаний с хронометрией просмотра вариантов ответа.

## Методика

### Контингент

Апробация проводилась на группе студентов 4-го курса педагогического факультета ЮФУ, изучающих физиологию человека. Студенты выполняли тесты для пересдачи по отдельным темам:

- Возбудимые ткани
- Нервная ткань
- Мышцы
- Основы регуляции
- Общая физиология ЦНС
- Частная физиология ЦНС
- Сердце, сосуды
- Кровь
- Дыхание
- Пищеварение
- ВНД
- Сенсорные системы

Изначально студенты проходили тесты, скомпонованные из этого же банка вопросов. Но для тех, кто был неуспешен в первом тестировании, была предложена попытка пересдачи по тем темам, по которым у них не хватало баллов. Таким образом студенты уже имели опыт ответа на часть вопросов. Кроме того, у студентов на руках были учебные материалы с перечнем всех вопросов, поэтому при желании они могли ознакомиться со всеми вопросами заранее.

Все участники были хорошо знакомы с процедурой тестирования в рамках электронной обучающей системы и дали информированное согласие на участие в нем.

### Процедура

Тестирование проводилось с помощью программного обеспечения Moodle для организации электронного обучения [<https://moodle.org/>]. Формирование вопросов и проведение процедуры тестирования осуществлялось в соответствии с настройками по-умолчанию. Для перехода к следующему вопросу студенты нажимали клавишу «Далее», при этом они могли ориентироваться по панели навигации. По окончании теста они могли просмотреть количество набранных баллов и содержание вопросов с отмеченными правильными и неправильными ответами.

Пересдача проходила с помощью усовершенствованной методики, в которой для просмотра вопроса и вариантов ответа требовалось на-

водить курсор мыши на соответствующую область экрана. Методика позволяет измерять время просмотра каждого варианта ответа и паттерн перемещения между ними.

Все тесты состояли из 20 вопросов, выбранных из банка вопросов объемом около 100 вопросов по каждой теме. 4–5 вопросов в тесте для пересдачи уже входили в первоначальный тест, поэтому некоторые из них были уже знакомы тестирующимся. Кроме того, некоторые студенты проходили пересдачу по два раза. Время прохождения и тип теста определялись академической необходимостью (выделялись часы в определенные дни в соответствии с расписанием занятий). Все вопросы оценивались одинаково в 1 балл и при оценке по пятибалльной системе имели вес 0,25 балла каждый. Содержание вопросов, а также их индексы легкости и индексы дискриминации не анализировались.

Попытки пересдать тест могли повторяться и длились от 6 до 20 мин. С правильными ответами студенты могли ознакомиться по завершению теста. Желавшие могли пересдать эти же тесты повторно. При этом они отвечали на уже знакомые им вопросы, но с измененным порядком вариантов ответа. Такие реализации являлись моделью прохождения теста с извлечением знаний из памяти.

В учебном классе, где проводилась апробация, протоколировалось время сеансов работы с приложениями с помощью системы учета распределения бюджета времени самостоятельной работы студентов за компьютером [Щербина, 2012]. Эта информация использовалась для анализа активности при переключении в окна других приложений (см. раздел «Результаты»).

### Анализ

Для подбора критериев выделения стратегий сравнивались параметры хронометрических показателей для успешных и неуспешных ответов.

### Результаты

Тестирование проходило в присутствии преподавателя, однако многие студенты прибегали к помощи посторонних источников – учебных пособий, загруженных в смартфон, или поисковых сервисов в Интернете. Наличие ошибок свидетельствует о том, что в ходе тестирования у них не было гарантированного источника правильных ответов для всех вопросов. Итого-

вые оценки прохождения тестов для пересдачи были преимущественно положительные. Например, медиана оценок за тест «Возбудимые ткани (для пересдачи)» по шкале 0–100 % составила 85 %, средняя величина – 81,59 %, стандартное отклонение – 10,06 %. По другим тестам наблюдалась похожая статистика.

### Стратегии прохождения тестовых заданий

Стратегии оценивались экспертом на основании хронограмм выполнения отдельных заданий. Хронограмма представляет собой графическое представление длительности выполнения отдельных операций по решению задания на единой шкале времени (рис. 1). На хронограмме наглядно видно, как разворачивались события по ходу времени. Для единообразия виды событий по оси ординат помечены кодами: qN – вопрос с числовым индексом, t – текст вопроса, aN – вариант ответа с числовым индексом. Поскольку при тестировании вопросы отображались по одному на странице, то числовой индекс вопроса всегда равен 0. Если тестируемый переключался в другое окно и прерывал выполнение теста, то на хронограмме отображался красный прямоугольник высотой во весь график. Момент клика мышкой визуализируется вертикальной темно-красной чертой.

В процессе тестирования при перемещении курсора мыши с одного варианта ответа на другой, первый из них плавно закрывался непрозрачным экраном, а экран, закрывающий второй, плавно исчезал (становился прозрачным). Таким образом, на хронограмме можно оценить, сколько и когда студент мог видеть текст вопроса или одного из вариантов ответа.

Для удобного сравнения длительности просмотра разных вариантов ответа между собой справа от хронограммы представлена гистограмма суммарного времени просмотра каждого из вариантов ответа.

Также на рисунке представлено содержание вопроса в том виде, как оно выглядит при просмотре результатов тестирования в электронной обучающей системе. Аналогичным образом вопрос выглядел и в ходе тестирования, за исключением того, что отсутствовали отметки о правильных/неправильных ответах, а также что одновременно можно было просматривать или текст вопроса, или один из вариантов ответа при наведении курсора мыши.

В целом, выполнение одного тестового задания делилось на стадии:

- 1) восприятие вопроса;
- 2) просмотр вариантов ответа;
- 3) выбор правильного варианта путем клика мышкой;
- 4) переход к следующему заданию.

Все многообразие траекторий прохождения тестовых заданий возникало из-за вариабельности 2-й и 3-й стадий. Эти стадии могли повторяться в разном порядке, формируя характерные паттерны прохождения. Ниже мы представляем выделенные нами стратегии прохождения тестовых заданий с примерами характерных паттернов, представленных в виде хронограмм.

### Стратегия извлечения ответа из памяти

При этой наиболее ожидаемой стратегии паттерн выполнения задания (рис. 1А) характеризуется быстрым восприятием вопроса, быстрым просмотром неправильных вариантов ответа, выбором правильного варианта ответа и переходом к следующему заданию. То есть выполнением минимального количества стадий в достаточно быстром темпе.

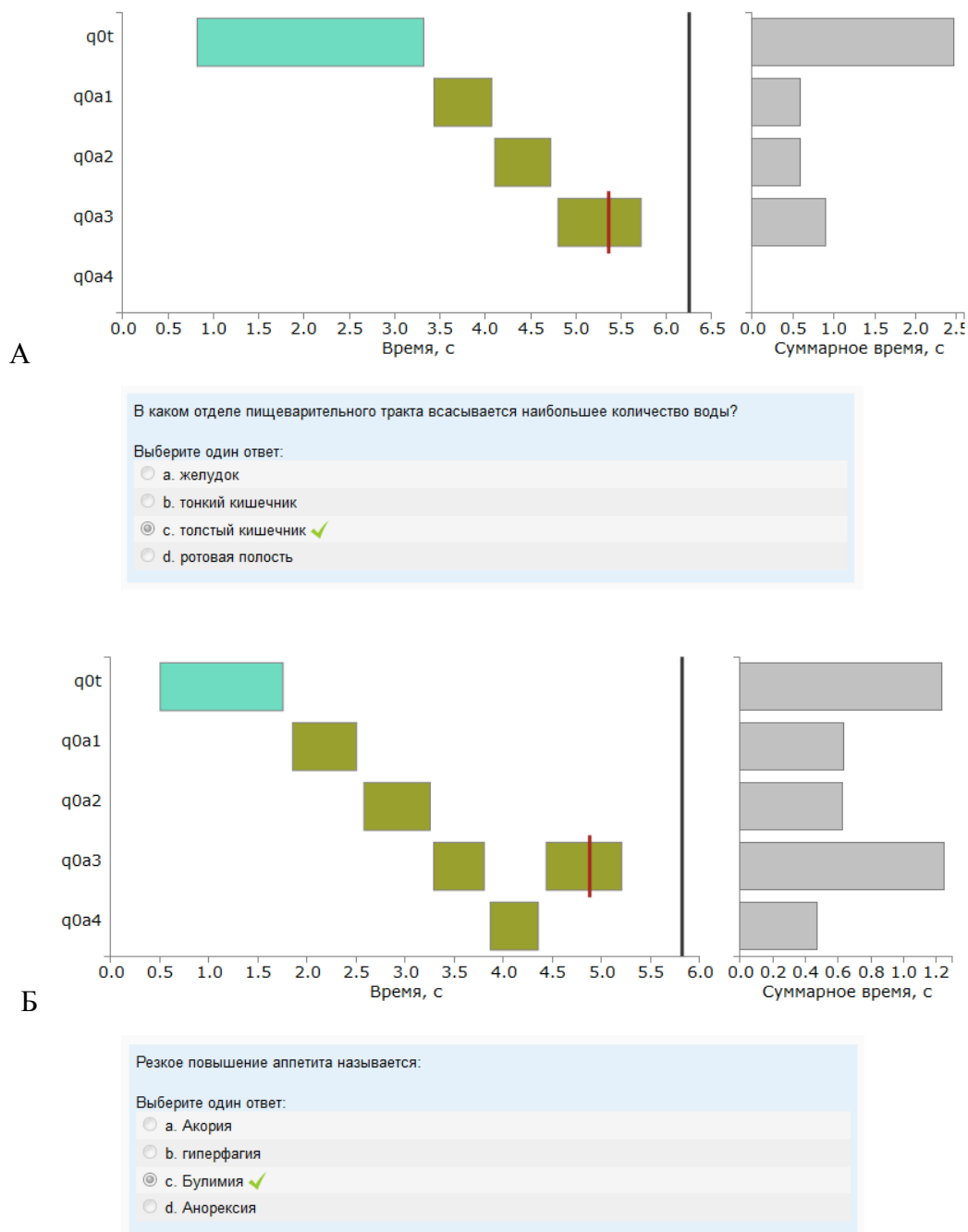
Такой паттерн выполнения задания можно считать оптимальным, и, соответственно, знание ответа на вопрос оценить в 100 %.

В примере, представленном на рисунке, студент проходил тест со второй попытки и, вероятно, уже знал ответ на вопрос.

Если информации, полученной при прочтении вопроса недостаточно, то студенты часто прибегают к предварительному просмотру всех вариантов ответа. При таком беглом просмотре студент не принимает решения о выборе правильного ответа, а лишь получает дополнительную информацию о вопросе. Для краткости мы будем называть паттерн быстрого просмотра всех вариантов – сканирование (рис. 1Б).

Включение сканирования в процедуру ответа на вопрос вызвано стремлением получить максимум информации перед тем, как приступить к принятию решения. Такое стремление объясняется разными причинами:

- 1) облегчение извлечения из памяти, когда нет готового ответа, а только смутные припоминания;
- 2) проявление осторожности из-за незнания или свойств личности;
- 3) привычка, приобретенная в ходе предыдущих тестов знаний.



**Рис. 1.** Хронограмма выполнения задания с выбором правильного ответа. А – быстрое решение; Б – с предварительным просмотром всех вариантов ответа (сканированием). По оси абсцисс время от начала выполнения задания, с. По оси ординат – коды просматриваемых элементов вопроса:  $q_0$  – порядковый номер вопроса,  $t$  – текст вопроса,  $a_1 \dots a_4$  – варианты ответа. Справа – суммарное время просмотра каждого из элементов вопроса, с. Ось ординат одна и та же. Снизу – представление содержания вопроса в электронной обучающей системе с отметками о правильности выбранных вариантов ответа.

Первое объяснение самое вероятное при условии, что вопрос четко сформулирован и что один из предложенных вариантов явно совпадает с правильным ответом. В таком случае наличие сканирования в ответе дает основание оценить знание как неуверенное с количественной оценкой 90 %.

Окончательный вывод о причине сканирования можно сделать за счет перекрестного анализа с траекториями прохождения других вопросов. Так, например, если ответы данного студента даже на простые вопросы (те, на которые отвечает большинство студентов в группе) содержат в себе сканирование, то это является его

устойчивой характеристикой, а не отражает неуверенное знание именно этого вопроса.

### Стратегия сопоставления вопроса и вариантов ответа

Если при сканировании вопросов возникает необходимость уточнить вопрос, то сканирование может прерваться на повторный просмотр вопроса (рис. 2). В первом из представленных примеров (рис. 2А) студент, читая варианты ответа, понял, что ему надо выбрать между видами нервных систем, и в этот момент он решил уточнить информацию, извлеченную им при чтении

текста вопроса. Судя по правильному ответу, он мог сразу ответить на вопрос правильно, однако повторное обращение к тексту вопроса свидетельствует, что после первого прочтения он не был готов дать окончательный ответ. После этого он продолжил просмотр вариантов ответа и, сделав правильный выбор, закончил решение тестовой задачи. Возможна более простая интерпретация, что в первом прочтении студент был недостаточно внимателен и поспешил перейти к просмотру ответов, не уяснив окончательно смысл вопроса.

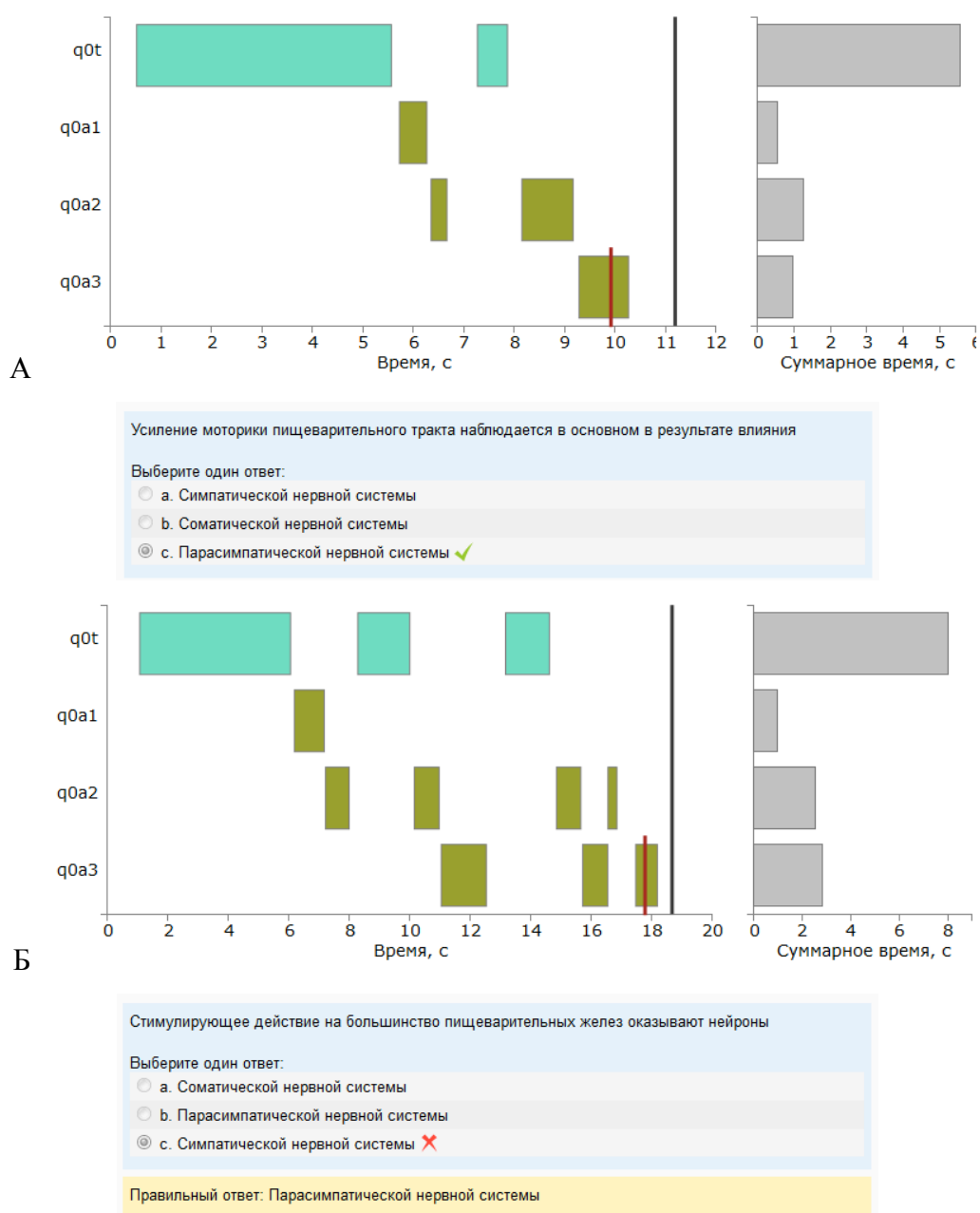


Рис. 2. Хронограмма выполнения задания с дополнительным просмотром текста вопроса и вариантов ответа. А – правильный ответ, Б – неправильный ответ. Обозначения, как на рис. 1

Во втором из представленных примеров (рис. 2Б) сопоставление вопроса и вариантов ответа более очевидно. Студент сосредоточивается на сравнении двух наиболее правдоподобных дистракторов, на просмотр каждого из которых затрачивает более 2,5 с, что примерно втрое превышает время, затраченное на просмотр первого нерелевантного дистрактора. Он дважды повторно обращается к тексту вопроса, а также поочередно просматривает два варианта ответа, будто пытаясь угадать, какой из них является правильным. Однако, в конце концов, делает неверный вывод и ошибается. В данном случае мы можем предположить, что студент знаком с физиологической терминологией – он знает, что бывают симпатическая и парасимпатическая нервная системы. При этом ошибка в ответе указывает на отсутствие конкретных знаний по данному вопросу. Наличие остаточных знаний позволило студенту повысить вероятность случайного угадывания с 0,33 до 0,5, поскольку к концу решения задания он сосредоточился лишь на двух дистракторах из трех.

Данные примеры иллюстрируют то, что появление в паттерне решения тестовой задачи дополнительных просмотров текста вопроса и вариантов ответа служит индикатором активности по сопоставлению информации из текста вопроса с информацией из вариантов ответов. Причинами появления такой активности могут быть:

- 1) недостаточное понимание вопроса при первом прочтении;
- 2) мысль, на которую натолкнул вариант ответа, заставляющая посмотреть под другим углом на поставленный вопрос, для чего требуется уточнить какую-нибудь упущенную при первом прочтении деталь;
- 3) попытка угадать правильный ответ на основе семантического анализа содержимого вопроса и дистракторов, полагаясь на остаточные или общие знания.

Последнюю из перечисленных причин следует считать наиболее вероятной, поскольку именно она будет систематически выделяться при анализе. Остальные причины могут иметь место лишь эпизодически и могут быть легко скорректированы самим испытуемым в процессе прохождения теста. Например, после одного-двух

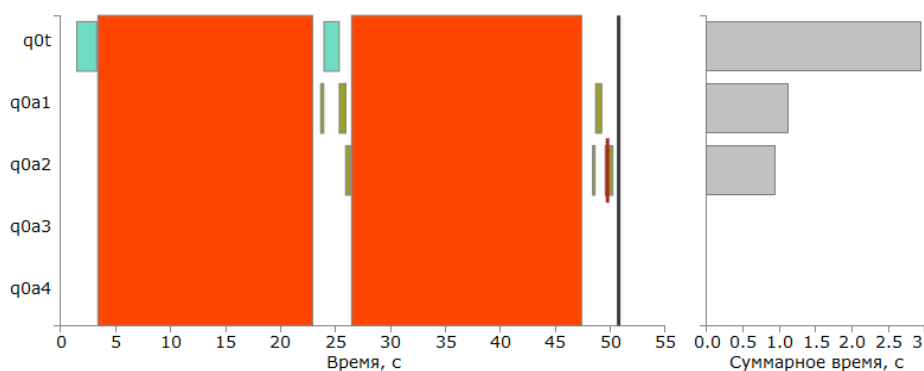
поспешных прочтений текста вопроса студент снизит скорость чтения вопроса до оптимального уровня так, чтобы этого времени хватило для извлечения необходимой информации из текста и не приходилось прибегать к повторному чтению.

### *Стратегия поиска ответа вовне*

При апробации время на прохождение одного теста было ограничено 20 минутами, поэтому длительные паузы между чтением вопроса и выбором ответа при анализе интерпретировались как поиск ответа с привлечением внешних источников. Например, это мог быть документ с фондом оценочных средств или учебное пособие, просматриваемые на экране смартфона. Паттерн части вопросов у некоторых студентов содержал длительную паузу от 15 до 40 с. Во время паузы не просматривался ни один из элементов вопроса или же курсор мыши оставался наведенным на текст вопроса, что должно было облегчать поиск. Фактически в данном случае студенты проходили тестирование не на имеющиеся знания, а на умение найти ответ на вопрос за короткое время. Они читали вопрос, затем приобретали знания для ответа на него, используя доступные средства, и после этого приступали к выбору в соответствии со стратегией извлечения ответа из памяти.

Если поиск происходил на этом же компьютере, например в поисковом сервисе в Интернете, то это явно отображалось на диаграмме (рис. 3). В данном случае после первого чтения вопроса студент в течение 10 с набирал поисковый запрос в поисковой системе, и затем в течение 9 с знакомился с результатами поиска. Потом он переключился в окно тестирования и за пару секунд повторно прочитал вопрос, а также глянул первые два предлагаемых варианта ответа. После этого он снова переключился в поисковую систему, перешел по ссылке в Википедию, где 17 секунд искал ответ. В итоге он выбрал правильный ответ и перешел к выполнению следующего задания.

Примечательно, что при получении информации из внешнего источника у студента не было причин просматривать другие дистракторы, поскольку найденная информация четко совпала с предлагаемым вариантом ответа.



В какой сосуд выбрасывается кровь из правого желудочка?

Выберите один ответ:

a. в полые вены

b. в легочный ствол ✓

c. в легочные вены

d. в аорту

15:26:05 Сердце, сосуды (для передачи) - Chromium  
 15:26:15 Тонус сосудодвигательного центра при повышении артериального давления в рефлексогенных зонах - Поиск в Google - Chromium  
 15:26:25 правый желудочек - Поиск в Google - Chromium  
 15:26:34 Сердце, сосуды (для передачи) - Chromium  
 15:26:38 правый желудочек - Поиск в Google - Chromium  
 15:26:41 Без имени - Chromium  
 15:26:42 Правый желудочек - Википедия - Chromium  
 15:26:59 Сердце, сосуды (для передачи) - Chromium

Рис. 3. Хронограмма выполнения задания с поиском ответа в Интернете. Обозначения, как на рис. 1. Внизу – заголовки активных окон

### Стратегия угадывания

Известной особенностью тестов с готовыми вариантами ответа является возможность случайного угадывания правильного варианта, обратно пропорциональная количеству дистракторов. Поэтому показателем отсутствия знаний по теме является не нулевое количество баллов, а соответствующее уровню случайного угадывания (при 4 вариантах ответа – 25 % от максимального).

Крайний случай угадывания – выбор первого попавшегося варианта – имеет характерный паттерн хронограммы (рис. 4А). Косвенные признаки того, что это угадывание: (1) малое время просмотра варианта ответа до клика мышью – недостаточное для восприятия смысла текста; (2) отсутствие просмотров других вариантов. Случайное угадывание можно ожидать при низкой мотивации выполнения задания или при нехватке времени, если время, отведенное на тест, ограничено.

При отсутствии дефицита времени угадывание может сопровождаться просмотром разных ответов (рис. 4Б). В этом случае угадывание сходно со стратегией сопоставления вопроса и вариантов ответа, но только с негативным исходом. В условиях отсутствия знаний по теме угадывание отличается формальным просмотром нескольких вариантов и принятием решения о выборе до просмотра всех предложенных вариантов.

Наличие в тесте паттернов угадывания имеет важные последствия для оценки правильных ответов с паттерном сопоставления вопроса и вариантов ответа. Даже если при выполнении задания с признаками сомнений был выбран правильный ответ, но ответы на соседние задания по этой же теме имеют явные признаки случайного выбора, то данный правильный ответ можно диагностировать как угадывание.

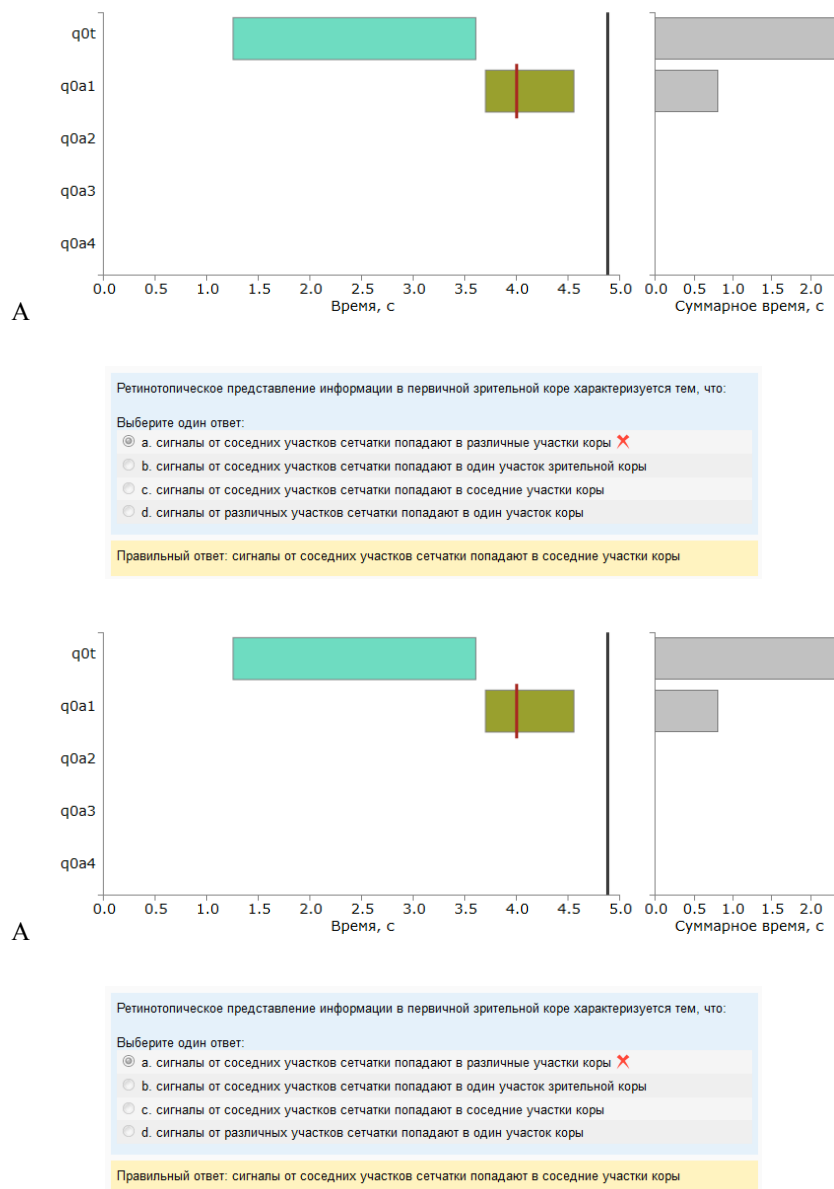


Рис. 4. Хронограмма выполнения задания методом угадывания: А – случайное, Б – с сопоставлением вопроса и вариантов ответа. Обозначения, как на рис. 1

### Обсуждение

По результатам апробации методики тестирования знаний с хронометрией просмотра вариантов ответа были выявлены четыре стратегии выполнения тестов в условиях образовательного процесса. По мере снижения уровня знаний это стратегии:

- 1) извлечение ответа из памяти;
- 2) сопоставление вопроса и вариантов ответа;
- 3) угадывание:
  - а) с опорой на остаточные знания;
  - б) случайное угадывание;
- 4) поиск ответа во внешних источниках.

Можно предложить следующую единую нормализованную шкалу для оценки знаний по результатам выполнения одного тестового задания (рис. 5). Наличие знаний по теме принимается за единицу (100 %) и диагностируется по использованию стратегии извлечения из памяти. Использование менее надежных стратегий сопоставления вопроса и вариантов ответа и угадывания с опорой на остаточные знания оценивается как фрагментарные знания по теме. Конкретное значение в диапазоне от 0 до 1 в этом случае определяется вероятностью наличия знаний, которая рассчитывается как соотношение правильных ответов к общему числу заданий, выполненных с использованием данной страте-





Рис. 5. Шкала оценки знаний с учетом стратегии выполнения теста. По оси ординат – оценка знания в долях от единицы. Уровень случайного угадывания указан для заданий с выбором правильного ответа из 4 вариантов

гии. Случайное угадывание оценивается как отсутствие знаний (0 %). Поиск во внешних источниках при контрольном тестировании знаний расценивается как списывание и может караться отрицательной оценкой. В случае тестирования для самоконтроля, где использование внешних источников допустимо, стратегия поиска вовне может оцениваться положительно как подтверждение знания терминов и умения пользоваться поисковыми сервисами.

Предложенный метод усовершенствованной оценки знаний при большом количестве тестовых заданий должен приближаться к усредненной оценке, полученной традиционным способом. Однако при небольшом количестве заданий более точная оценка за каждое из них должна приводить к значительному повышению статистической мощности. Для экспериментального подтверждения данного предположения требуется проведение дополнительных исследований.

При усреднении полученных оценок для заданий, предъявленных за одно тестирование, следует учитывать возможную смену стратегии в условиях дефицита времени. Например, две трети заданий студент с фрагментарными знаниями может выполнять с использованием стратегии сопоставления вопроса и ответов, а затем из-за нехватки времени перейти к стратегии угадывания. Одним из подходов к обнаружению такого пере-

ключения с одной стратегии на другую может быть использование метода скользящего среднего для отображения сглаженной результативности студента в динамике прохождения теста.

Более точные оценки дают перспективу более информативной оценки знаний с относительно меньшими затратами времени на тестирование. Этот важный результат в условиях интенсификации образования может быть достигнут при внедрении надежной автоматизированной системы идентификации перечисленных стратегий.

Для создания автоматизированной системы требуется разработать математическую модель, способную идентифицировать стратегию прохождения теста с высокой долей вероятности. Исходными данными для работы модели являются показатели хронометрии просмотра текста вопроса и вариантов ответа каждого из заданий в тесте. В перечень параметров должны быть включены количество тех или иных операций, а также их временные характеристики. Для учета вариации временных характеристик, зависящих от содержания заданий, например длины текста, они должны быть заменены на скоростные характеристики.

В предлагаемом нами методе время просмотра определяется по времени нахождения курсора мыши над областью экрана с текстом компонента вопроса. Преимуществом данного подхо-

да является дешевизна и легкая адаптация к необходимости дополнительных манипуляций. Задача быстрого и точного перемещения курсора по экрану не вызывает проблем у здоровых студентов. Добавление маскирующих экранов может удлинять выполнение задания лишь в случае необходимости сравнения вариантов ответа между собой. В случае, если в тексте отдельного варианта ответа достаточно информации для решения выбирать этот вариант или нет, то потери времени практически не происходит.

Альтернативным методом для хронометрии просмотра текста вопроса и вариантов ответа без изменения визуального интерфейса среды тестирования мог бы быть метод отслеживания перемещения и фиксации взгляда (*eye-tracker*), однако его широкое использование ограничено высокой стоимостью необходимого оборудования.

Таким образом, предложенный метод повышения информативности процедуры тестирования в электронной обучающей системе позволит снизить продолжительность тестирования для получения оценки знаний с той же надежностью. Это важно при использовании методик адаптивного тестирования [Челышкова, 2000; Georgiadou et al., 2007; Weiss, 2014], поскольку позволяет быстрее и точнее подстроиться к текущему уровню знаний студента.

### Выражения признательности

Благодарю зав. кафедры физиологии человека и животных Академии биологии и биотехнологии ЮФУ проф. Е.К. Айдаркина и преподавателей С.Н. Кульбу и А.Г. Глумова за возможность апробации новой методики тестирования в ходе учебного процесса.

Работа выполнена в рамках ГЗ по теме №213.01–11/2014–35.

### Литература

Аванесов ВС. Знания как предмет педагогического измерения // Современная высшая школа. 2014; 3: 57–67.

Аванесов ВС. Форма тестовых заданий. М.: Центр тестирования, 2006; 156.

Беспалько ВП. Образование и обучение с участием компьютеров (педагогика третьего тысячелетия). М.: МОДЭК, 2002; 351.

Переверзев ВЮ. Критериально-ориентированное педагогическое тестирование в профессиональном образовании (методология, теория, практика). М.: ФИРО, 2008; 247.

Система управления обучением Moodle <https://moodle.org/> [Веб-ресурс].

Челышкова МБ. Адаптивное тестирование в образовании. М., 2000.

Щербина ДН. Система учета распределения бюджета времени самостоятельной работы студентов за компьютером: Свидетельство о государственной регистрации программы для ЭВМ № 2012617238 от 10.08.2012г.

Conole G, Warburton B. A review of computer-assisted assessment. *ALT-J Research In Learning Technology*. 2005 Mar; 13(1):17–31.

DeMars C. *Item response theory*. New York: Oxford University Press, 2010.

Drasgow F, Levine MV, Tsien S, Williams B, Mead AD. Fitting Polytomous Item Response Theory Models to Multiple-Choice Tests. *Applied Psychological Measurement*. 1995 Jun; 19(2):143–166.

Georgiadou E, Triantafyllou E, Economides A. A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment*, 2007; 5(8).

Haladyna TM, Downing SM, Rodriguez MC. A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*. 2002 Jul; 15(3):309–333.

McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. *Medical Teacher*. 2004 Dec; 26(8):709–712.

Morrison S, Free K. Writing multiple-choice test items that promote and measure critical thinking. *Journal of Nursing Education* 40. 2001; 17–24. PMID:11198905

Osterlind SJ. *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats*. Columbia : University of Missouri-Columbia, 2004.

Thompson AR, O'Loughlin VD. The Blooming Anatomy Tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. *Anatomical sciences education*. 2015 Nov; 8 (6):493–501. PMID:25516150.

Weiss DJ, ed. *New Horizon Testing: Latent Trait Test Theory and Computerized Adaptive Testing*. Elsevier, 2014.