

КВАЛИМЕТРИЧЕСКИЙ ПОДХОД В ОБРАЗОВАНИИ

УДК 371.263

Е. К. Хеннер,
Т. С. Ознобихина

ОЦЕНКА ПРОЧНОСТИ ЗНАНИЙ НА ОСНОВЕ СОПОСТАВЛЕНИЯ РЕЗУЛЬТАТОВ РАЗЛИЧНЫХ ВИДОВ ТЕСТИРОВАНИЯ

Аннотация. В настоящее время в системе образования широкое распространение получили тестовые технологии контроля знаний. Однако именно этот метод контроля подвергается острой критике, в том числе и в связи с его использованием в ЕГЭ. Целью данной работы стало изучение ограничений, присущих тестированию. В статье предложен способ оценки прочности знаний на основе сопоставления результатов различных видов тестирования. На примере тестирования большой группы студентов показано, что результаты, полученные с использованием закрытых и открытых тестов, существенно различаются. Произведенный авторами статьи сравнительный анализ позволяет сделать вывод о том, что открытые тесты диагностируют прочность знаний более адекватно, чем закрытые.

Публикация содержит описание лишь единичного эксперимента, повторение которого по той же методике, другими исследователями в других ситуациях даст возможность понять, насколько является правилом различие между результатами двух способов тестирования, и более отчетливо выявить их преимущества и недостатки.

Ключевые слова: тесты, тестовые технологии контроля знаний, прочность знаний.

Abstract. At present, the testing techniques of knowledge estimation are widely spread in educational system. However, this method is seriously criticized including its application to the Unified State Examinations. The research is aimed at studying the limitations of testing techniques. The authors recommend a new way of knowledge solidity estimation bases on the comparative results analysis of various kinds of tests. While testing the large group of students, the authors found out that the results of the closed and open tests substantially differ. The comparative analysis demonstrates that the open tests assessment of the knowledge solidity is more adequate than that of the closed ones. As the research is only based on a single experiment, the authors recommend using this

method further, substantiating the findings concerning the differences in tests results, and analyzing the advantages and disadvantages of the tests in question.

Index terms: testing, testing techniques of knowledge assessment, solidity of knowledge.

Введение

Использование тестовых технологий контроля знаний стало в настоящее время неотъемлемой составляющей образовательного процесса. Из многих существующих видов тестов (см., например, [1]) на практике чаще всего используются так называемые тесты закрытого типа – совокупность альтернативных тестовых заданий с выбором из числа предложенных вариантов одного правильного ответа. Важным достоинством таких тестов считается объективность, понимаемая как независимость процесса тестирования от субъективизма преподавателей, наличия одинаковых условий и правил оценки для всех учащихся, подвергаемых тестированию. Однако именно этот метод контроля подвергается острой критике, в том числе и в связи с его использованием в ЕГЭ. Мы полностью разделяем точку зрения В. С. Аванесова, утверждающего: «Важно понять простую истину: сами по себе – тесты полезный метод, но их применение в практике может оказаться как полезным, так и вредным» [1]. В связи с этим имеет смысл изучение ограничений, присущих тестированию.

Пользуясь разработанной нами простой методикой, описанной ниже, мы пытались уточнить, как соотносятся результаты тестирования с такой важной характеристикой знаний, как прочность. Понятия, связанные со знаниями, являются предметом скорее психологии и физиологии, нежели педагогики. На операциональном уровне нам вполне достаточно определения, данного К. К. Платоновым и Г. Г. Голубевым: «Прочные знания – это закрепленные в памяти глубокие понятия основной сущности предметов и явлений». Существенное уточнение, которое есть далее в том же источнике: «Когда из памяти легко извлекается главное содержание знаний, то уже нетрудно воспроизвести в сознании отдельные компоненты, от крупных их совокупностей до мелких подробностей» [4, с. 79]. Отметим, что в педагогической науке формирование прочных знаний всегда считалось важнейшей задачей обучения на любом уровне системы образования.

Методика эксперимента

Используемая нами методика основана на сопоставлении результатов двух способов тестирования, проведенного на базе одного и того же

учебного материала для одних и тех же групп студентов. Эти способы – классический закрытый тест с пятью вариантами ответа в каждом из тестовых заданий и соответствующий ему открытый с той же последовательностью заданий, но требующий записи краткого ответа – любого, какой испытуемые считают верным (одного ответа, если испытуемые усматривают такую возможность). В открытом тесте сохраняется смысловое наполнение вопросов закрытого теста, но вопросы переформулированы и отредактированы таким образом, чтобы не нарушать логики и стилистики русского языка.

В нашем эксперименте приняли участие 149 студентов 1-го курса трех (непрофильных информатике) факультетов Пермского государственного университета, завершающие изучение дисциплины «Информатика», которая входит в обязательную часть образовательных программ. Сводный тест охватывал темы «Текстовый процессор», «Табличный процессор», «Базы данных», «Системы счисления» и «Измерение информации». Тестирование проводилось по группам (примерно по 25 чел.), во избежание списывания – по 8 вариантам, в которых содержалось по 30 тестовых заданий. Каждое задание оценивалось по двухбалльной шкале (0 – неверно, 1 – верно). Через 2–3 дня проходило повторное тестирование, на котором студенты получали те же самые варианты (без предупреждения об этом), однако если при первом тестировании использовался открытый тест, то при втором – закрытый. Студентам было объявлено, что для получения зачета необходимо успешно выполнить обе контрольные работы.

Приведем примеры заданий двух видов тестирования. В открытом тесте одно из заданий было сформулировано следующим образом: «Приведите пример графической информации» (разумеется, не все вопросы столь примитивны, это лишь иллюстрация). В пустом поле для записи студент должен был дать произвольный ответ. В закрытом тесте вариант этого задания выглядел так:

«Что может служить примером графической информации?

- 1) текст песни;
- 2) таблица значений тригонометрических функций;
- 3) изображение на экране монитора;
- 4) стихотворение в сборнике;
- 5) разговор по телефону».

Конечно, при открытом способе тестирования не требовалось, чтобы студент в качестве ответа вписал именно 3-й ответ из варианта закрытого

теста – по существу, любой верный ответ засчитывается. Решение о правильности ответов принимал эксперт (преподаватель), объективность гарантировалась как опытом принявших участие в эксперименте преподавателей, так и характером опорного материала, на основе которого при корректно поставленном вопросе несложно отличить правильный ответ от неправильного.

В целях улучшения качества теста эксперименту предшествовала апробация базы тестовых заданий. Апробационное тестирование проводилось предварительно за семестр до эксперимента для студентов, у которых к этому времени изучение указанных ранее тем завершилось. Трудность каждого тестового задания вычислялась по формуле:

$$q_j = \frac{R_j}{N},$$

где j – номер тестового задания;

R_j – количество испытуемых, выполнивших j -е задание верно;

N – количество испытуемых [7, с. 299].

По результатам этого тестирования были заменены задания с $q_j < 0,2$ как слишком легкие. Например, вопрос, относящийся к электронным таблицам: «Сколько ячеек в диапазоне С2: Е4?» – был заменен на вопрос: «Что происходит при перемещении или копировании с абсолютными ссылками в электронной таблице?». После корректировки было проведено повторное апробационное тестирование.

Результаты тестирования и их статистическая обработка

Сводные «сырые» результаты двух видов тестирования изображены на рисунках 1 и 2 в виде диаграмм, отражающих связь между числом испытуемых и количеством полученных баллов. Без всякой статистической обработки видно, что между ними существует значительное различие.

Более наглядными результаты тестирования становятся при их группировке по «карманам», в которые попадают результаты от 0 до 5, от 5 до 10 и т. д. На рис. 3 эти результаты сведены воедино.

В целом студенты показали значительно лучший результат по итогам теста закрытого типа, чем открытого: средний балл по закрытому тесту 19,6, по открытому – 13,8.

В обоих случаях распределения баллов с достаточной надежностью соответствуют гипотезе о нормальном распределении (проверка велась в соответствии с критерием Колмогорова – Смирнова [6, с. 145], с достоверностью 0,95).

Для определения зависимости между двумя выборками, визуально представленной на рис. 4, использовался критерий Пирсона. По результатам статистической обработки гипотеза о наличии зависимости между двумя выборками подтвердилась с достоверностью 0,95. Коэффициент линейной корреляции (коэффициент Пирсона) равен 0,56: в педагогических исследованиях принято считать, что это означает наличие (среднюю степень) корреляции.

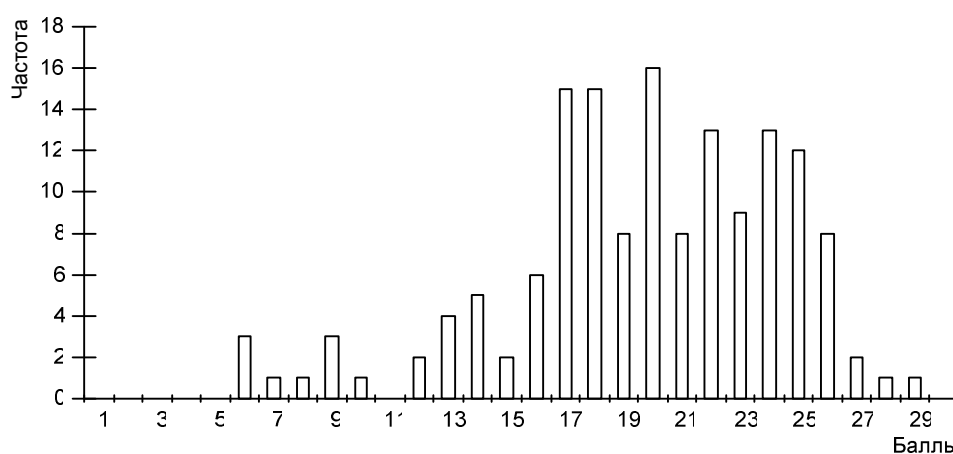


Рис. 1. Распределение баллов закрытого тестирования

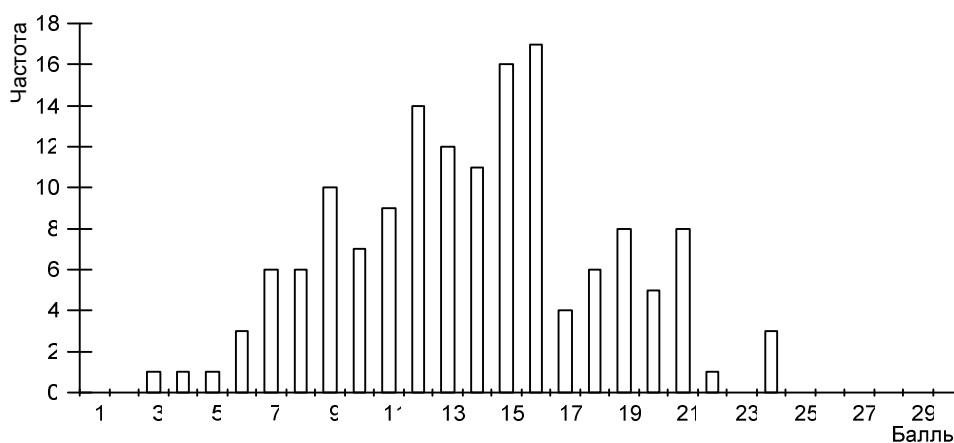


Рис. 2. Распределение баллов открытого тестирования

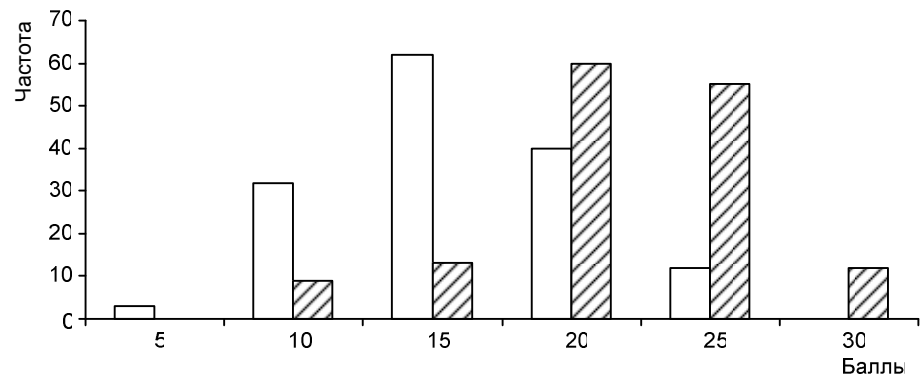


Рис. 3. Сводное распределение частот баллов за оба тестирования:
 □ – открытое тестирование; ▨ – закрытое тестирование

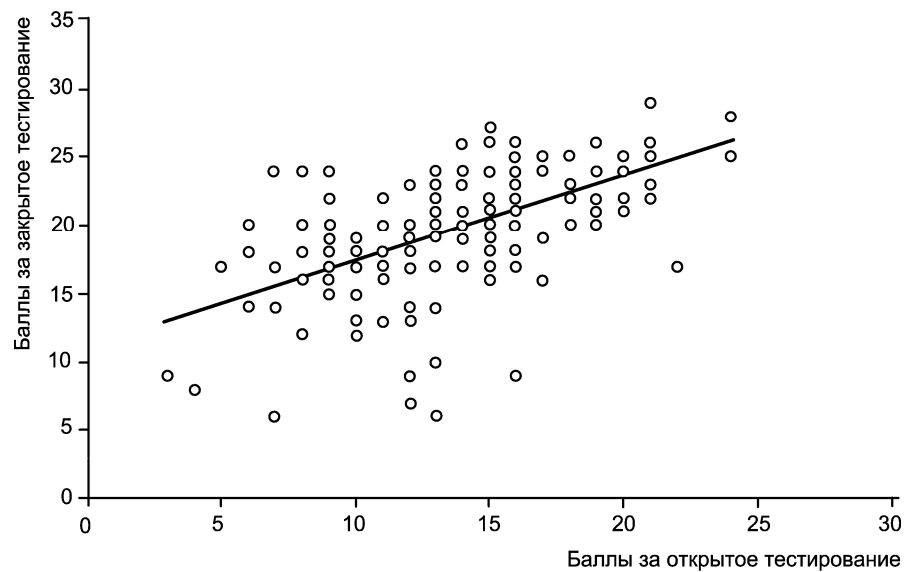


Рис. 4. Отражение связи между баллами за открытое и закрытое тестирование

Представляет интерес динамика изменения корреляции между результатами открытого и закрытого тестирования при последовательном отсечении лучших результатов. Существует гипотеза, что чем лучше подготовлены студенты, тем меньше результаты зависят от способов контроля.

Соответствующие значения коэффициента корреляции приведены в таблице (лучшие результаты отбрасывались по данным открытых тестов).

Коэффициент линейной корреляции для усеченных выборок

Объем выборки, чел.	Коэффициент корреляции
149 (генеральная)	0,57
140	0,52
130	0,47
120	0,42
110	0,37
100	0,29

Интерпретация результатов

На наш взгляд, описанные выше результаты иллюстрируют различия между двумя типами знаний: прочными знаниями (в соответствии с данным в начале статьи определением), и тем типом знаний, которые проявляются лишь при наличии подсказки, содержащейся, в частности, в тестах закрытого типа. Знания второго типа можно в противоположность прочным назвать поверхностными, или «мягкими знаниями» (термин *soft knowledge* широко используется в англоязычной литературе по искусственному интеллекту). «Мягкие знания допускают множественные расплывчатые решения»; «поверхностные знания – совокупность эмпирических ассоциаций и отношений между понятиями предметной области» [2, с. 277]. Повторим, что во все времена было принято считать, что прочность усвоения знаний – одно из важнейших требований к процессу и результатам обучения. «Результатом прочного усвоения знаний является образование у учащихся устойчивых структур знаний, отражающих объективную реальность, когда учащиеся умеют актуализировать и использовать полученные знания» [5, с. 219].

Наличие знаний тесно связано с процессом запоминания, которое в психологии принято подразделять на непреднамеренное (непроизвольное) и преднамеренное (произвольное). «Непреднамеренное запоминание – это запоминание без заранее поставленной цели, без использования каких-либо приемов и проявления волевых усилий. ... В отличие от непроизвольного запоминания произвольное (преднамеренное) запоминание характеризуется тем, что человек ставит перед собой определенную цель – запомнить некую информацию – и использует специальные приемы запоминания. Произвольное запоминание представляет собой особую и сложную умственную деятельность, подчиненную задаче запомнить» [3, с. 263].

Сопоставляя результаты двух способов тестирования по степени успешности конкретного студента в том и другом виде тестирования, судить наверняка о доминирующем по данному предмету типе знаний можно лишь в том случае, если испытуемый не справился с закрытым тестом – тут наличие незнание (относительное, по принятому количественному критерию – например, доле правильных ответов). Если студент успешен в открытом тесте (по заданному критерию), то он успешен и в закрытом в рамках принятого критерия успешности (статистически – в конкретных случаях возможны вариации). Если же студент успешен в закрытом тесте, но неуспешен в альтернативном открытом тесте, то есть основания считать его носителем мягких знаний при отсутствии прочных. При этом речь ни в коем случае не идет о его знаниях в целом, а лишь о знаниях по данному конкретному предмету, измеряемых на заданном уровне требований.

Например, если применительно к описанному выше эксперименту установить критерий минимальной успешности в 60% правильных ответов, то при открытом тестировании с работой справились 21% студентов, а при закрытом – 71%. Иначе говоря, половина испытуемых, участвующих в эксперименте, не проявила прочных знаний (по заданному критерию), но обладает их «мягкой» формой.

При последовательном отсеке лучших результатов, показанных при открытом тестировании, корреляция между результатами открытого и закрытого тестирования становится все менее значимой. В нашей интерпретации это означает, что чем ниже уровень знаний испытуемых, тем больше среди них носителей мягких знаний, не обладающих прочными.

Для практики образования принципиально важен вопрос о том, влияет ли доминирующая система контроля на прочность формируемых знаний. По мнению авторов, перекося в употреблении закрытых тестов в ущерб иным формам контроля знаний, особенно в итоговых государственных системах контроля, наиболее важных по своим последствиям для учащихся и преподавателей, способствует формированию мягких знаний в ущерб прочным.

Обращаем внимание на то, что мы описали единичный эксперимент. Представляет интерес его повторение – по той же методике, другими исследователями в других ситуациях. Тогда появится возможность понять, насколько является правилом различие между результатами двух способов тестирования, и более отчетливо выявить их преимущества и недостатки, обнаруженные при проведении нашего эксперимента.

Литература

1. Аванесов В. С. Теория и методика педагогических измерений [Электрон. ресурс] // В. С. Аванесов. Начала теории. URL: <http://testolog.narod.ru/Theory1.html> (дата обращения: 07.10.2011).
2. Барановская Т. П., Лойко В. И., Семенов М. И., Трубилин А. И. Информационные системы и технологии в экономике: учеб. / под ред. В. И. Лойко. М.: Финансы и статистика, 2005. 416 с.
3. Маклаков А. Г. Общая психология: учеб. для вузов. СПб.: Питер, 2008. 583 с.
4. Платонов К. К., Голубев Г. Г. Психология. М.: Высш. шк., 1977. 248 с.
5. Российская педагогическая энциклопедия: в 2 т. / гл. ред. В. В. Давыдов. М.: Большая рос. энцикл., 1998. Т. 2, М-Я. 672 с.
6. Теория статистики: учеб. / под ред. Г. А. Громыко. М.: ИНФРА-М, 2005. 467 с.
7. Чельшкова М. Б. Теория и практика конструирования педагогических тестов: учеб. пособие. М.: Логос, 2002. 432 с.