

Нестеров С.А.

Санкт-Петербургский политехнический университет Петра Великого, г. Санкт-Петербург, Россия.

**АНАЛИЗ СТАТИСТИКИ ВЫПОЛНЕНИЯ ТЕСТОВЫХ ЗАДАНИЙ В СРЕДЕ
ДИСТАНЦИОННОГО ОБУЧЕНИЯ MOODLE**

АННОТАЦИЯ

В статье рассматриваются средства анализа статистики выполнения тестовых заданий в среде дистанционного обучения Moodle и варианты их применения в ходе преподавания дисциплины «Базы данных».

КЛЮЧЕВЫЕ СЛОВА

Дистанционное обучение; Moodle; анализ тестовых заданий; базы данных.

Nesterov S.A.

Peter the Great Saint-Petersburg Polytechnic University, St-Petersburg, Russia

ANALYSIS OF QUIZ STATISTICS IN LMS MOODLE

ABSTRACT

The paper describes statistical measures of the test tasks quality, which learning management system (LMS) Moodle calculates. Examples of practical use of these indicators in teaching the discipline "Database Systems" are given.

KEYWORDS

E-learning; Moodle; test statistics; database systems.

В 2016/2017 учебном году на кафедре «Системный анализ и управление» СПбПУ, в соответствии с принятыми в университете требованиями к объёму материала, изучаемого на иностранных языках, начато поэтапное внедрение преподавания ряда специальных дисциплин на английском. Среди них оказалась и дисциплина «Базы данных», читаемая бакалаврам 3-го курса. Предварительная оценка знаний студентов показала, что уровень языковой подготовки студентов сильно различается, поэтому было принято решение в этом году читать на английском языке только отдельные модули лекционного курса, предоставляя студентам доступ к аналогичным материалам на русском через портал дистанционного обучения СПбПУ, который использует среду дистанционного обучения Moodle. Кроме того, через СДО Moodle проводится промежуточное тестирование уровня знаний.

В связи с тем, что в таком формате курс проводится впервые, промежуточным тестам уделяется особое внимание, так как по их результатам можно вовремя определить, насколько хорошо студенты усвоили темы, и есть ли разница в понимании материалов, читавшихся на русском и на английском. Планируется регулярно сравнивать результаты промежуточной оценки знаний студентов проводимой на русском (тесты, письменные проверочные работы, результаты сдачи лабораторных работ) с результатами промежуточного тестирования на английском, и при необходимости, уделить дополнительное время отдельным темам. Использовать автоматизированные тесты в среде Moodle представляется удобным по ряду причин:

- студенты уже работают с материалами курса в этой среде;
- на проведение компьютерного теста требуется относительно немного аудиторного времени – тест в 15 небольших заданий с выбором вариантов ответа можно провести за 10 минут;
- результаты теста автоматически обрабатываются, преподавателю доступны не только итоговые результаты, но и статистика по выполнению отдельных заданий.

В качестве базового учебника для модулей курса «Базы данных», читаемых на английском, был выбран учебник Kroenke D., Auer D. Database Processing: Fundamentals, Design, and Implementation издательства Pearson Education [1]. Учебник относится к разряду классических, выдержавших много переизданий: в 2016-м году вышло 14-е издание, тогда как первое было выпущено в 1977. При этом, каждое издание существенно обновляется и в части излагаемого теоретического материала, и в

части лабораторного практикума (например, в последнем издании лабораторные работы по SQL Server даны уже для версии SQL Server 2014). Кроме того, издательством поддерживается комплект материалов для преподавателя, включающий презентации, учебные базы данных и наборы тестовых заданий, которые можно импортировать в Moodle.

Объем предлагаемого с учебником банка тестовых заданий достаточно большой – порядка 100 заданий на главу учебника. Несмотря на то, что использование чужих заданий в тестах преподавателю не всегда удобно, в данном случае для тестов на английском это представляется оправданным. По крайней мере, текст тестовых заданий писался носителями языка и можно избежать проблем, вызванных языковыми неточностями в формулировках. Но правильность формулировки с лингвистической точки зрения не решает всех проблем. И для тестовых заданий как на русском, так и на английском важно провести анализ качества самих заданий, что можно сделать, изучая предоставляемые средой Moodle статистические показатели.

Статистические показатели качества тестовых заданий

В теории педагогических измерений для оценки качества тестовых заданий широко используется анализ статистических показателей, полученных в результате пилотного прохождения теста группой студентов. Некоторые из них используются СДО Moodle [2-5]. В частности, это *индекс лёгкости* (англ. facility index), который показывает, какая часть студентов ответила правильно на анализируемый вопрос. Для *i*-го тестового задания он определяется по формуле

$$F_i = \frac{\bar{x}_i}{x_i(\max)}, \quad (1)$$

где \bar{x}_i – среднее значение набранных студентами баллов за *i*-е задание, $x_i(\max)$ – максимально возможная оценка за *i*-е задание. Этот коэффициент в отчетах Moodle приводится в процентном представлении.

Хороший тест должен включать задания разного уровня сложности. При этом, заданий с индексом легкости близким или равным как 1, так и 0, лучше избегать. В первом случае задание является излишне простым – на него все ответили правильно, во втором случае – излишне сложным. В то же время, при анализе этого показателя надо учитывать уровень подготовки группы, проходившей тестирование.

На рисунке 1 представлены результаты прохождения одинаковых тестов группами с разным уровнем подготовки [4].

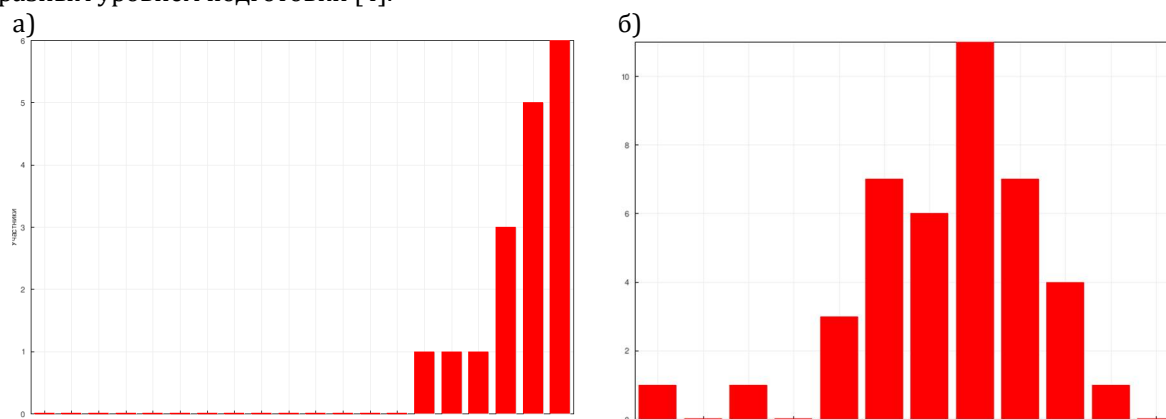


Рис.1 Распределение оценок при прохождении теста:
а) «сильная» группа студентов; б) «слабая» группа

Рис. 1-а) соответствует более сильной группе студентов, рис. 1-б) – более слабой группе. На представленных гистограммах по горизонтальной оси откладываются диапазоны оценки в баллах, по вертикальной – количество студентов с оценкой в указанном диапазоне. Данная диаграмма формировалась системой, исходя из числа студентов и разброса получаемых оценок, настроить шаг изменения оценки не предлагалось, поэтому на гистограммах разное количество столбцов. Но общую картину они показывают: после приведения к пятибалльной шкале, в первой группе основная часть оценок была 5, во второй – 3 и 4. Из 87 тестовых заданий в использованном банке, индекс легкости равный 100% в первом случае получился у 47 заданий, во втором случае – только у двух. Исключить из банка заданий больше половины, исходя только из результатов оценки первого теста, было бы неверно.

Следующий используемый показатель – *стандартное отклонение* (англ. standard deviation).

Оно характеризует разброс значений оценок, полученных за данное задание теста. Если для какого-то задания этот показатель равен 0, это означает, что все тестируемые получили за этот вопрос одинаковую оценку. Такой вопрос следует признать неудачным. В литературе отмечается [3], что в соответствии с требованиями педагогической теории измерений, задания со значением стандартного отклонения менее 0,3 лучше исключать из теста, так как они не обладают достаточной дифференцирующей способностью, то есть не способны разделить сильных и слабых учащихся.

Балл случайного угадывания (англ. random guess score) – оценка, которую мог бы получить студент при случайном угадывании ответов. Например, если требуется выбрать один вариант ответа из четырех возможных, это значение будет равно 25%.

Намеченный вес (англ. intended question weight, IQW) – вес, который назначается тестовому заданию при формировании сценария теста. Если за одно задание назначено 3 балла из 10 возможных за тест в целом, то намеченный вес составляет 30%. Если обозначить через $x_p(\max)$ и $x_p(\min)$ максимально и минимально возможные оценки за задание, а T_{\max} и T_{\min} – максимальную и минимальную возможные оценки за тест (в текущей версии $x_p(\min)$ и, соответственно, T_{\min} всегда равны 0), получим формулу расчета этого показателя:

$$IQW_p = 100\% \cdot \frac{x_p(\max) - x_p(\min)}{T_{\max} - T_{\min}}. \quad (2)$$

Здесь надо пояснить, что этот параметр рассчитывается не для конкретного вопроса, а для «позиции» задания в тесте. Например, тест может формироваться так, что на каждую позицию случайно выбирается один из нескольких альтернативных вопросов на заданную тему. И если индекс легкости может рассчитываться и для отдельного вопроса, и для «позиции» в тесте, то намеченный и эффективный вес – это характеристики только «позиции» в тесте.

В данной статье, так же как в документации Moodle v.2.0 и выше [5] для обозначения номера вопроса будет использоваться индекс i (все множество вопросов – I), для обозначения позиций в тесте – индекс p и множество P , для обозначения попытки сдачи теста – s и S , соответственно. Количество набранных студентом баллов за тест будет обозначаться T_s , оценка студента за конкретное задание теста – $x_p(s)$, оценка студента за все остальные задания $X_p(s)$:

$$T_s = \sum_{p \in P} x_p(s) \quad (3)$$

$$X_p(s) = T_s - x_p(s)$$

Дисперсия показателя будет обозначаться $V()$, например:

$$V(x_p) = \frac{1}{S-1} \sum_{s \in S} (x_p(s) - \bar{x}_p)^2. \quad (4)$$

Ковариация будет обозначаться через $C()$, например:

$$C(x_p, X_p) = \frac{1}{S-1} \sum_{s \in S} (x_p(s) - \bar{x}_p)(X_p(s) - \bar{X}_p). \quad (5)$$

Вернемся к рассмотрению рассчитываемых Moodle показателей.

Эффективный вес (англ. effective question weight, EQW) характеризует фактическую долю конкретного задания («позиции» в тесте) в итоговой оценке студентов за тест.

$$EQW_p = 100\% \cdot \frac{\sqrt{C(x_p, T)}}{\sum_{p \in P} \sqrt{C(x_p, T)}}. \quad (6)$$

В идеале эффективный вес должен быть равен назначенному. Имея эти значения, преподаватель может скорректировать назначенные веса заданий теста.

Индекс дискриминации (англ. discrimination index) обозначается в документации D_p и рассчитывается как:

$$D_p = 100\% \cdot \frac{C(x_p, X_p)}{\sqrt{V(x_p)V(X_p)}}. \quad (7)$$

Этот коэффициент показывает, насколько взаимосвязаны правильность ответа на данный вопрос и остальные вопросы теста. Предполагается, что для хорошего тестового вопроса, студенты с высокими оценками за него, также будут иметь более высокие оценки и за тест в целом. На это укажет положительное значение коэффициента. При этом отмечается [5], что если индекс лёгкости вопроса отличен от 50%, D_p не может быть 100%. Если индекс лёгкости близок к 0% или 100%, то индекс дискриминации всегда будет очень маленьким, а при равенстве индекса лёгкости 0% или 100% D_p будет не определён.

Эффективность дискриминации (англ. discriminative efficiency) – еще один показатель, который по назначению во многом аналогичен индексу дискриминации, но лишен указанного выше недостатка. Он рассчитывается по формуле

$$DE_p = 100\% \cdot \frac{C(x_p, X_p)}{C_{\max}(x_p, X_p)}, \quad (8)$$

где $C(x_p, X_p)$ рассчитывается по формуле (5), а для расчёта $C_{\max}(x_p, X_p)$, множества значений $x_p(s)$ и $X_p(s)$ предварительно сортируются по возрастанию, как будто первый студент получил самые низкие баллы x_p и X_p , а последний студент получил самые высокие x_p и X_p .

Индекс дискриминации и эффективность дискриминации рассчитываются как для позиции в тесте, так и для отдельного вопроса.

Предлагаемый вариант использования средств оценки тестовых заданий

Как уже отмечалось выше, оценивать планируется выполнение заданий с формулировками на английском и на русском, причем тестирование будет проводиться с интервалом 1-2 недели между тестами. Банк заданий на русском использовался уже в течение 3 лет и по его заданиям в системе дистанционного обучения есть статистика. Задания на английском будут использоваться в первый раз.

Если результаты сдачи студентами тестов на русском и английском будут близки, это будет указывать на то, что уровень владения иностранным языком в целом достаточен для изучения данного курса.

Существенно более плохие результаты тестов на английском потребуют анализа: причина в сложности формулировки вопросов или в недостаточном понимании материала соответствующих тем. Здесь, по всей видимости, надо будет использовать сравнение статистики выполнения заданий по одной теме, а также дать сравнительную оценку сложности заданий с точки зрения предметной области.

Наоборот, более успешное прохождение тестов на английском будет указывать на достаточный уровень языковой подготовки и более низкий уровень сложности заданий теста на английском. Анализ таких показателей, как коэффициент легкости и балл случайного угадывания поможет выявить неудачные вопросы, чтобы убрать их из банка заданий.

На момент подготовки материала был проведен первый тест с заданиями на английском. Тест включал 15 вопросов, продолжительность теста 15 минут. Вопросы выбирались случайным образом из двух банков заданий по 20 вопросов в каждом: семь вопросов брались из первого банка (по первой главе учебника Kroenke D., Auer D. Database Processing: Fundamentals, Design, and Implementation. Chapter 1: Introduction) и восемь вопросов – из второго (по второй главе учебника Chapter 2: Introduction to Structured Query Language). Использовались банки заданий, предоставляемые издательством Pearson Education в комплекте материалов для преподавателя. Студентам данные задания ранее не демонстрировались. В банк заданий брались только вопросы с выбором одного ответа из нескольких и вопросы с выбором «верно/неверно».

Тест проводился с ограничением по времени – 10 минут на 15 вопросов. Тест сдавался двумя учебными группами, в общей сложности 45 человек. Средний балл составил 2,77 из 5 возможных. Тест проходил в компьютерном классе в присутствии преподавателя. В ходе проведения теста пользоваться какими-либо материалами, а также словарями и компьютерными программами перевода не разрешалось. Один студент был остановлен при попытке скрыто посмотреть учебные материалы со смартфона. Т.к. эта попытка была через примерно 5 минут после начала теста (около 50% времени на попытку) студенту было предложено досрочно завершить тест, попытка была учтена.

Оценка за тест приводилась к оценке по пятибалльной шкале. Все вопросы принимались равнозначными, таким образом, каждый правильный ответ приносил 0,33 балла (оценки приводятся с округлением до второго знака после запятой). Распределение оценок за тест представлено на рис. 2.

Минимальная оценка 1 балл – 3 вопроса правильно из 15 – была у одного человека (студент, досрочно закончивший тест из-за попытки списывания). Следующий результат с конца – 1,33 бала (4 правильных ответа из 15) – также у одного студента, далее – 1,67 баллов (5 правильных ответов из 15) – у пяти студентов. Максимальная оценка 4,33 – 13 вопросов правильно из 15 – была у 4 человек. Наиболее распространенная оценка 3 балла (9 правильных заданий) – у 9 учащихся.

Если анализировать данные о времени, затраченном на тест, то 22 студента из 45 использовали все 10 минут. Но здесь, возможно, кто-то просто дожидался окончания счетчика времени, не отправляя тест досрочно. У остальных студентов время выполнения теста от 4 мин. 54 сек. и далее. Причем наименьшее время показал один из четырех студентов, набравших

наибольший балл. Все прочие студенты, набравшие наибольший балл, также закончили тест раньше 10 минут.

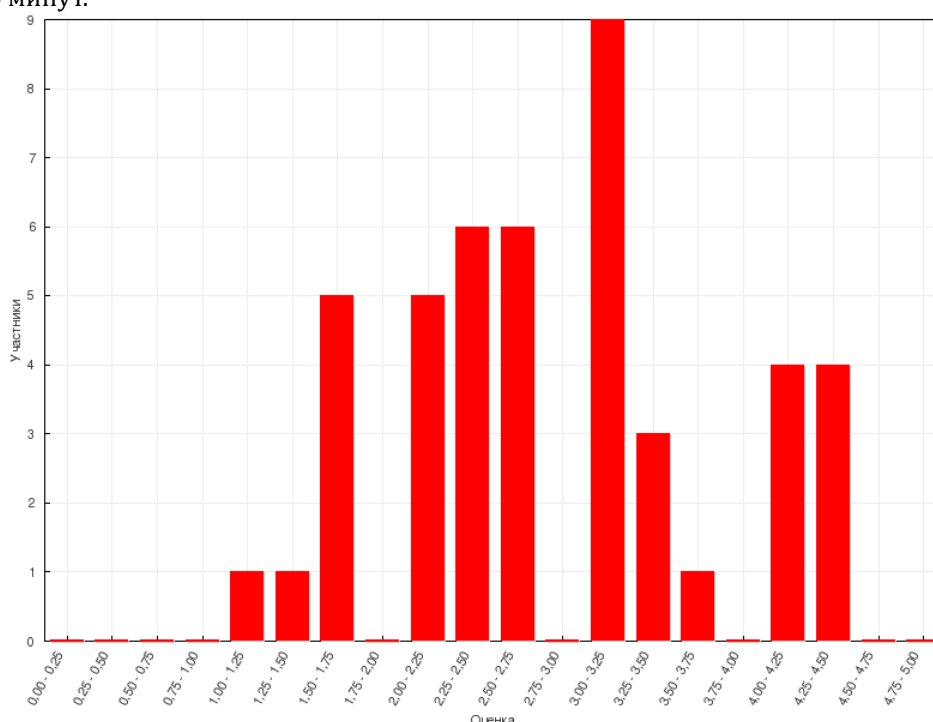


Рис.2. Распределение оценок за тест

Не приводя подробный анализ статистических характеристик тестовых заданий можно отметить, что сами задания, по всей видимости, соответствуют уровню текущей подготовки студентов. В частности, из 40 заданий коэффициент легкости 100% был только у двух. Но оба эти задания были типа "верно / неверно", а этот тип задания всегда легче вопросов с множественным выбором. Средний коэффициент лёгкости – 56.25% для вопросов по первой главе и 55.56% для вопросов по второй главе. Значения коэффициента 0% нет ни у одного задания, т.е. на каждое задание у каких-то студентов были правильные ответы.

В заключение можно попробовать дать субъективную оценку первого тестирования. Результат прохождения теста на английском, по всей видимости, ниже, чем был бы для аналогичного теста на русском. Таким образом, недостаточное знание английского языка является сдерживающим фактором. Об этом свидетельствует и отсутствие полностью правильных результатов в данной группе. В то же время, отсутствие студентов, которые вообще не смогли ответить на поставленные вопросы и средний балл 2,77 (выше 50%) говорит о том, что уровень языковой подготовки для данной группы не является непреодолимым барьером.

Помимо возможности оценить уровень подготовки студентов, проведение тестов с заданиями на английском можно рассматривать как этап подготовки к сдаче сертификационных экзаменов, проводимых компаниями-вендорами ПО. Не все производители предлагают экзамены с заданиями на русском, а отсутствие опыта прохождения тестов на английском иногда пугает тех обучающихся, чье знание предметной области в принципе достаточно для успешной сдачи подобных экзаменов.

Литература

1. Kroenke D., Auer D. Database Processing: Fundamentals, Design, and Implementation. Global Edition. – Pearson Education, 2016. ISBN: 9781292107639.
2. Коржик И.А., Протасова И.В., Толстобров А.П. Тестовая система Moodle и качество тестовых заданий. // Современные информационные технологии и ИТ-образование. Сборник избранных трудов VII Международной научно-практической конференции. – М.: ИНТУИТ.РУ, 2012. – С. 187-196.
3. Толстобров А.П., Коржик И.А. Возможности анализа и повышения качества тестовых заданий при использовании сетевой системы управления обучением MOODLE. // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2008. № 2. С. 100-106.
4. Нестеров С.А., Сметанина М.В. Оценка качества тестовых заданий средствами среды дистанционного обучения MOODLE // Научно-технические ведомости СПбГПУ. № 5 (181) 2013. СПб: Изд-во Политехн. ун-та, 2013. С.87-92.
5. Quiz statistics calculations. URL: http://docs.moodle.org/dev/Quiz_statistics_calculations

References

1. Kroenke D., Auer D. Database Processing: Fundamentals, Design, and Implementation. Global Edition. – Pearson Education, 2016. ISBN: 9781292107639.
2. Korzhik I.A., Protasova I.V., Tolstobrov A.P. Testovaya sistema Moodle i kachestvo testovykh zadaniy. // Sovremennyye informatsionnyye tekhnologii i IT-obrazovaniye. Sbornik izbrannykh trudov VII Mezhdunarodnoy nauchno-prakticheskoy konferentsii. – M.: INTUIT.RU. 2012. – p. 187-196.
3. Tolstobrov A.P., Korzhik I.A. Vozmozhnosti analiza i povysheniya kachestva testovykh zadaniy pri ispolzovanii setevoy sistemy upravleniya obucheniyem MOODLE. // Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyy analiz i informatsionnyye tekhnologii. 2008. № 2. p. 100-106.
4. Nesterov S.A., Smetanina M.V. Otsenka kachestva testovykh zadaniy sredstvami sredy distantsionnogo obucheniya MOODLE // Nauchno-tekhnicheskiye vedomosti SPbGPU. № 5 (181) 2013. SPb: Izd-vo Politekhn. un-ta. 2013. p.87-92.
5. Quiz statistics calculations. URL: http://docs.moodle.org/dev/Quiz_statistics_calculations

Поступила 10.10.2016

Об авторе:

Нестеров Сергей Александрович, доцент кафедры «Системный анализ и управление» Санкт-Петербургского политехнического университета Петра Великого, кандидат технических наук, nesterov@saiu.ftk.spbstu.ru.