УДК 378.662.168:51

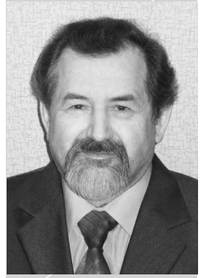
Экспертиза качества результата тестирования

С.Д. Старыгина[1], **Н.К. Нуриев**[1]

[1]Казанский национальный исследовательский технологический университет, Казань, Россия

Поступила в редакцию 16.11.2018

С.Д. Старыгина



Н.К. Нуриев

Аннотация

При оценке качества усвоенных знаний студента через тестирование, каждый раз возникает вопрос: а насколько объективен полученный результат этого тестирования? Очевидно, что объективность результата студента зависит как от точности теста, как инструментального средства, так и от продолжительности времени (в разумных пределах), отпущенного на этот тест. В работе предложена обоснованная методика оценки объективности (качества) результата тестирования, которое можно легко применить на практике.

Ключевые слова: точность теста, объективность результата, продолжительность тестирования, экспертиза качества, качество теста.
Key words: test accuracy, objectivity of the result, test duration, quality examination, quality test.

Введение. В рамках учебного курса, тест является измерительной системой, от которой зависит точность этого теста и объективность результата тестирования студента в целом.

Точность теста (E) является латентной (скрытой) характеристикой, которую можно оценить только опосредованно, то есть как значение функционала, зависящего от множества наблюдаемых экспертом показателей. К таким показателям относятся: VAL – валидность (адекватность и пригодность) комплекса предложенных вопросов (заданий); REL – релевантность комплекса заданий, то есть спрашивается ли в них то, что изложено в рамках курса; REP – репрезентативность, то есть равномерно ли представлены задания из всех разделов изучаемого курса; KSM – качество полноты и целостности комплекса вопросов теста, то есть представлены ли в тесте одинаковое количество вопросов на знание «фактов» и на знание «связей», в изучаемой предметной области. Объективность результата теста (Z), также является латентной характеристикой и функционально зависимой от двух показателей: точности теста (E) как измерительного средства и вероятности (P) выбора продолжительности (T) без ущерба объективности результата при тестировании студента. Необходимость выбора T без ошибки, исходит из того, что по своей природе люди имеют разный темперамент и психологический склад, то есть многие «медлительные», обладая необходимыми знаниями для ответов на вопросы теста, не могут быстро сосредоточиться, медленно думают, не уверены в себе, то есть им необходимо дать больше времени для ответа на вопросы теста, так как речь идет не о тестировании на скорость. С другой стороны, тестирование не может продолжаться бесконечно долго. Очевидно, что с увеличением продолжительности T, с некоторого момента времени результаты тестирования студента значимо не улучшатся (будет исчерпан запас усвоенных знаний на данный момент развития). Исследования показывают [1-3], что если эксперту для ответа на вопросы теста открытого типа требуется время S, то студенту для полного раскрытия своих знаний требуется $T = 3*S$. В этом случае на основе большего количества статистических данных можно доказать, что вероятность ошибки при выборе значения $T = 3*S$ будет не более чем $P = 0{,}05$.

В целом, все изложенное на формальном уровне, можно записать как следующую параметрическую каскадную модель, представленную через два функционала:

$$E = F1\ (VAL, REL, PER, KSM);$$
$$Z = F2\ (E, P),$$

где Z – показатель объективности, то есть единый показатель качества результата тестирования студента. Например, студент в результате тестирования получил $B = bal$ (количество баллов оценивается от 0 до 100). При этом точность этого результата равна E, а надежность выбора времени $T = 3*S$ без ошибки 95%.

В этой ситуации, на практике возникает задача: требуется в рамках курса, оценит (в метриках) качество, то есть объективность результата тестирования студента.

Методика оценки качества (точности) теста

Для оценки качества теста были приглашены шесть экспертов, которые независимо друг от друга должны провести эту экспертизу. Результаты экспертизы представлены в табл. 1.

По данным (табл. 1) в едином круге построена диаграмма Кивиата для демонстрации качества теста по разным критериям (рис. 1).

Интегральную оценку качества теста можно вычислить как среднее геометрическое, то есть

$$E = F1\ (VAL, REL, PER, KSM);$$
$$E = \sqrt[4]{0{,}75 \cdot 0{,}78 \cdot 0{,}8 \cdot 0{,}8} = 0{,}782.$$

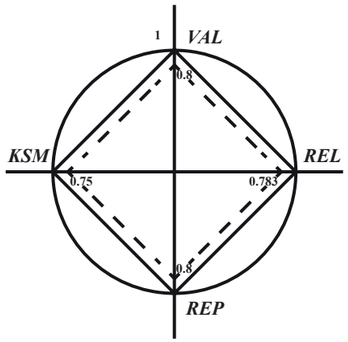
Таким образом, качество теста в рассматриваемом курсе 78% из 100% возможных, например, в принятой шкале в вузе, качество теста курса оценивается как «отлично». Результаты работы экспертной группы, полученные в ходе анкетирования должны пройти обязательную проверку на согласованность. Если мнения экспертов окажутся несогласованными, то есть мнения существенно отличаются внутри группы, результаты признаются непригодными для вынесения содержательных суждений о предмете экспертизы, а сама экспертиза не состоявшейся. Подобная ситуация может возникнуть либо по причине значительного отличия в уровне квалификации приглашенных экспертов, либо вследствие отсутствия общепризнанных критериев оценки обсуждаемой проблемы в сообществе специалистов. В первом случае затруднение легко преодолевается путем формирования новой группы экспертов, а во втором признается, что проблема созрела только для дискуссии, но не для экспертизы.

Таблица 1. Сводная таблица экспертных оценок качества теста учебного курса

Критерии \ Эксперты	1	2	3	4	5	6	Среднее значение
VAL	0,9	0,8	0,6	0,9	0,7	0,9	0,8
REL	0,8	0,7	0,8	0,7	0,7	1	0,783
REP	0,8	0,7	0,7	0,8	0,8	1	0,8
KSM	0,7	0,8	0,7	0,6	0,7	1	0,75

ИНЖЕНЕРНОЕ ОБРАЗОВАНИЕ
24'2018

Рис. 1. Диаграмма Кивиата для визуализации качества теста учебного курса



Общепринятым методом проверки согласованности мнений является метод, основанный на вычислении коэффициента множественной ранговой корреляции Кендалла-Смита коэффициент конкордации и проверки его статической значимости.

Для проведения процедуры проверки оценки, представленные экспертами, ранжируются: самой высокой оценки присваивается ранг 1, следующей – 2 и т.д. Одинаковым оценкам присваиваются одинаковые ранги, равные среднему арифметическому их порядковых номеров. Такие ранги называются связанными. Сводные таблицы ранжирования представлены в табл. 2.

Для вычисления коэффициента Кендалла-Смита K воспользуемся известным соотношением:

$$K = \frac{\sum_{i=1}^{n}(\sum_{j=1}^{m} r_{ij} - \bar{r})^2}{\frac{1}{12}(m^2(n^3 - n) - m \cdot \sum_{j=1}^{m} T_j)},$$

где r_{ij} – ранг i-ого показателя у j-ого эксперта:

$$\bar{r} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij}}{n},$$

где n – число оцениваемых показателей; m – число экспертов в составе группы.

$$T_j = V_j^3 - V_j,$$

где V_j – количество одинаковых связанных рангов, выставленных j-ым экспертом.

Используя расчетные соотношения, получим:

Таблица 2. Сводная карта ранговых оценок

Эксперты / Критерии	1	2	3	4	5	6	∑
VAL	1	1,5	4	1	3	4	14,5
REL	2,5	3,5	1	3	3	2	15
REP	2,5	3,5	2,5	2	1	2	13,5
KSM	4	1,5	2,5	4	3	2	17

$$\bar{r} = (14,5 + 15 + 13,5 + 17)/4 = 15$$

$$T_1 = V_1^3 - V_1 = (2)^3 - 2 = 8 - 2 = 6$$

$$T_2 = [(2)^3 - 2] + [(2)^3 - 2] = 12$$

$$T_3 = [(2)^3 - 2] = 6$$

$$T_4 = 0$$

$$T_5 = (3)^3 - 3 = 27 - 3 = 24$$

$$T_6 = (3)^3 - 3 = 27 - 3 = 24$$

$$\sum_{j-1}^{6} T_j = 72$$

$$K = \frac{6,5}{\frac{1}{12}(36 \cdot 60 - 6 \cdot 72)} = 0,045$$

.

Визуальная оценка значения коэффициента конкордации свидетельствует о несогласованности экспертных оценок. Тем не менее, чтобы в этом убедиться, проверим гипотезу согласованности статистики по критерию χ^2. Для этого воспользуемся формулой:

$$\chi^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(r_{ij} - \bar{r})^2}{\frac{1}{12}(m \cdot n(n-1) - \frac{1}{n-1}\sum_{j=1}^{m} T_j)}.$$

В рассматриваемом случае

$$\chi^2 = \frac{6.5}{\frac{1}{12}(24 \cdot 3 - \frac{72}{3})} = 1,62$$

.

Сравним эту величину с табличным значением критерия для уровня значимости α = 0,05 степень свободы f = 3; $\chi^2_{таб}$ = 7,8 > 1,62. Вывод – мнение экспертов не согласованное, относительно «отличного» качества теста. Разумеется, это снижает надежность оценки теста, то есть мнение экспертов разошлись, и верить этой оценке нельзя.

Методика оценки продолжительности тестирования студента

Рассмотрим задачу: на основе статистических данных требуется оценить значение величины T (продолжительность времени тестирования студента). Значение величины T зависит от значения детерминированной величины S (сложность – трудоемкость теста, которую оценит эксперт) и случайной величины X, где значение случайной величины X равно продолжительности времени реакции студента для ответов на комплекс вопросов теста.

На основе статистического материала установим закон распределения случайной величины X. Для этого на специально сформированной шкале (рис. 2) отложим продолжительности реакций всех студентов в группе.

Для идентификации закона распределения рассмотрим экспериментальные данные, которые сформировались в системе MOODLE в течении 15 лет. В эксперименте участвовало 50 групп. Средняя численность студентов в одной группе 25 человек. Усредненные данные представлены на рис. 3.

Согласно данным из графика, частота (эмпирические вероятности) добровольного выхода студента из процесса (процедуры тестирования) будут следующие (X – случайная величина – время выхода по завершению теста), $P(X < 1,25*S) = 0$, то есть вероятность, того, что студент завершит тест и выйдет из процесса тестирования раньше, чем $1,25*S$ равна нулю, где S – сложность теста. $P(X < 1,5*S) = 0,2222$, то есть вероятность того, что студент завершит тест и выйдет из процес-

Рис. 2. Шкала с отметками значений X (продолжительности реакции студента для ответа на вопросы теста)

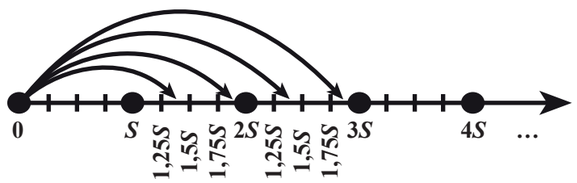
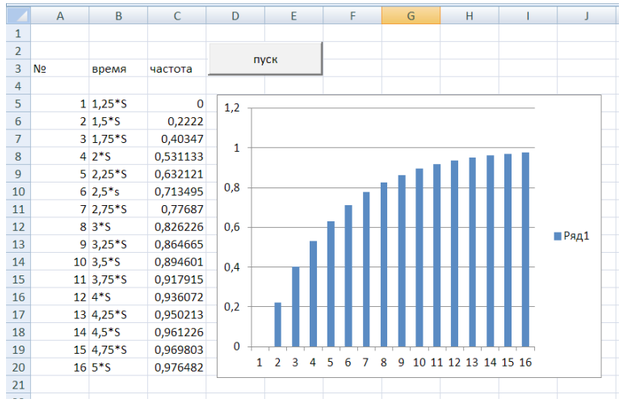


Рис. 3. Результат обработки экспериментальных данных (эмпирический закон распределения величины X)



№	время	частота
1	1,25*S	0
2	1,5*S	0,2222
3	1,75*S	0,40347
4	2*S	0,531133
5	2,25*S	0,632121
6	2,5*s	0,713495
7	2,75*S	0,77687
8	3*S	0,826226
9	3,25*S	0,864665
10	3,5*S	0,894601
11	3,75*S	0,917915
12	4*S	0,936072
13	4,25*S	0,950213
14	4,5*S	0,961226
15	4,75*S	0,969803
16	5*S	0,976482

са тестирования раньше, чем 1,5*S равна 0,22 (22%).

Аналогично: $P\ (X < 1{,}75{*}S) = 0{,}40$; $P\ (X < 2{*}S) = 0{,}51$, то есть ко времени 2*S, завершив процесс, выйдут чють больше половины студентов и т.д.

Из графика следует, что активное время выхода студентов по завершении теста (продолжительность самообслуживания) начинается с момента времени S и продолжается до момента T – конец тестирования. Исходя из этого, начало координат на графике можно перенести на момент S, так как до момента S никто не завершает тестирование. Как следует из частотной характеристики случайной величины X (интегральная характеристика) средняя продолжительность самообслуживания (тестирования) в активной зоне равна величине T(ср) = S.

Из статистического анализа данных следует, что при уровне значимости α = 0,05 (гипотеза проверялась по критерию χ^2) случайная величина X подчиняется экспоненциальному закону распределения с интенсивностью потока равным $\lambda = 1/T(\text{ср}) = 1/S$, то есть

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

Из этого следует, что поток самообслуживающихся студентов является Пуассоновским потоком.

Из тех же рассмотренных экспериментальных данных известно, что в среднем из группы с 25 студентами, тест на положительную оценку не могут сдать 3,5 студента, и это не зависимо от продолжительности времени T. В целом, это означает, что в среднем примерно 14% студентов сдают тест на «два». Исходя из этой информации и данных графика, находим, что $T = 3{*}S$.

Итак, экспериментально доказано, что продолжительность тестирования T устанавливается по правилам:

1. Эксперты должны оценить S – сложность (трудоемкость в мин/раб) теста.

2. Задать для студентов продолжительность (трудоемкость) тестирования $T = 3{*}S$ (мин/раб) и провести процедуру тестирования.

При этом сложность S (трудоемкость) теста оценивается экспертами через «себя», то есть оценивается продолжительность времени, которое необходимо эксперту на выполнение теста. Допустим, 6 экспертов заполнили таблицу (табл. 3).

Итак, данные точности теста, продолжительности тестирования сведем в одну таблицу (табл. 4).

Из этой таблицы следует, что мнение экспертов о точности теста не согласованно. В этом случае, авторы курса поработали над «ошибками» в тестовой системе и провели повторную экспертизу. Результаты новой экспертизы приводятся в табл. 5.

Приведем пример расчета оценки объективности результата тестирования, допустим, после усвоения учебного курса, проведено тестирование студентов продолжительностью $T = 3{*}S = 3{*}16{,}4 \approx 50$ мин. При этом студент заработал $B = 52$ балла.

В данном случае, согласно методике можно сказать, что результат тестирования $B = 52$, получен при следуюших условиях: точность теста как измерительной системы $E = 0{,}87$. Вероятность ошибки при выборе величины $T = 50$ мин. не более чем $P = 0{,}95$. Показатель объективности результата тестирования вычисляется как среднее геометрическое, то есть

$Z = F2(E, P)$;

$$Z = \sqrt{E \cdot P} = \sqrt{0{,}87 \cdot 0{,}95} = 0{,}92$$

Таким образом, можно утверждать, что результаты тестирования студентов по рассматриваемому курсу обладают 92% качеством, то есть 92% объективностью.

Вывод

Методику оценки качества результатов тестирования можно применить на практике. Разумеется, что процесс расчетов лучше автоматизировать. В нашем вузе система экспертизы качества тестов развернут в сети как сайт. Применение этой методики на практике приводит к некоторой стандартизации тестов и ликвидации низкокачественных тестов в рамках любых курсов.

Таблица 3. Результаты хронометража теста

Эксперты	1	2	3	4	5	6	среднее
Продолжительность теста (мин/раб)	15	17	14	18	16	15	15,83

Таблица 4. Значения метрик теста и тестирования

Название учебного курса		
Точность теста	Продолжительность тестирования	Согласованность экспертов (да/нет)
0,782	15,83	нет

Таблица 5. Результаты хронометража теста

Название учебного курса		
Точность теста	Продолжительность тестирования	Согласованность экспертов (да/нет)
0,873	16,3	да

Материалы статьи докладывались на международной научно-практической конференции «Синергия 2018» по проблемам интегративной подготовки линейных инженеров для предприятий нефтегазового и нефтегазохимического комплексов России

ЛИТЕРАТУРА

1. Нуриев, Н.К. Дидактическая инженерия: разработка регламента педагогического тестирования [Электронный ресурс] / Н.К. Нуриев, С.Д. Старыгина // Образовательные технологии и общество: междунар. электронный журнал. – 2017. – Т. 20, № 4. – С. 478-483. – URL: http://ifets.ieee.org/russian/periodical/journal.html
2. Старыгина, С.Д. Построение математической модели процесса регламентации педагогического тестирования / С.Д. Старыгина, Н.К. Нуриев, Е.А. Печеный // Информационные технологии и математическое моделирование (ИТММ-2017): мат. XVI междунар. конфр. им. А.Ф. Терпухова. – Томск: Изд-во НТЛ, 2017. – С. 223-229.
3. Нуриев, Н.К. Надежность результата теста для оценка качества владения компетенцией / Н.К. Нуриев, С.Д. Старыгина // Современные проблемы безопасности жизнедеятельности: интеллектуальные транспортные системы и ситуационные центры: мат. V междунар. науч.-практ. конф. – Казань: Центр инновационных технологий, 2018. – С. 261-271.

Наши авторы

БЕЛАШ
ОЛЬГА ЮРЬЕВНА

доцент, кандидат технических наук, директор Центра маркетинга Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» (СПбГЭТУ «ЛЭТИ»), почетный работник высшего профессионального образования Российский Федерации
E-mail: marketing @etu.ru

БЛЕСМАН
АЛЕКСАНДР ИОСИФОВИЧ

кандидат технических наук, доцент, заведующий кафедрой «Физика» Омского государственного технического университета, почетный работник сферы образования
E-mail: blesm@mail.ru

БОГОУДИНОВА
РОЗА ЗАКИРОВНА

доктор педагогических наук, профессор кафедры инженерной педагогики и психологии Казанского национального исследовательского технологического университета, «Заслуженный деятель науки Российской федерации», лауреат премии Правительства Российской федерации в области образования
E-mail: rozabog@bk.ru

БОЧКАРЁВ
СЕРГЕЙ КОНСТАНТИНОВИЧ

доцент, кандидат технических наук, ассистент кафедры теории двигателей летательных аппаратов Самарского национального исследовательского университета имени академика С.П. Королева
E-mail: bochkar@ssau.ru

БУГАКОВА
НИНА ЮРЬЕВНА

доктор педагогических наук, профессор, первый проректор Калининградского государственного технического университета, почетный работник науки и техники, заслуженный работник высшей школы РФ
E-mail: bugakovakgtu@mail.ru, bugakova@klgtu.ru

БУДЗИНСКАЯ
ОЛЬГА ВЛАДИМИРОВНА

кандидат экономических наук, доцент Российского государственного университета нефти и газа (НИУ) имени И.М. Губкина
E-mail: budzinskaya@bk.ru

ВИШНЯКОВА
ИРИНА ВЯЧЕСЛАВОВНА

доцент, кандидат технических наук, доцент кафедры Методологии инженерной деятельности Казанского государственного технологического университета
E-mail: kazakova-ulyana@mail.ru

ВОДОПЬЯНОВА
СВЕТЛАНА ВИТАЛЬЕВНА

кандидат технических наук, доцент Казанского национального исследовательского технологического университета
E-mail: vod-sveta@yandex.ru

ВОЛКОВА
ГАЛИНА ЛЕОНИДОВНА

стажер-исследователь отдела исследований человеческого капитала, Институт статистических исследований и экономики знаний, Национальный исследовательский университет «Высшая школа экономики»
E-mail: gvolkova@hse.ru