

Модели и процедуры оценки результатов компьютерного тестирования знаний

Печников Андрей Николаевич
Заслуженный деятель науки РФ, д.п.н., д.т.н.,
профессор, профессор кафедры педагогики
Военная академия связи имени Маршала Советского Союза С.М. Буденного,
194064, Санкт-Петербург, Тихорецкий проспект, д. 3, (812) 247-93-72
pan287@users.mns.ru

Туровская Алиса Олеговна
к.п.н, научный сотрудник ЗАО “Фирма “Пассат”,
198514, Санкт-Петербург, ул. Разводная, д. 17, (812) 427-20-21
a_iline@mail.ru

Туктаров Ринат Равильевич
старший преподаватель факультета военного обучения,
Государственный морской университет имени адмирала Ф.Ф. Ушакова
353918, Новороссийск, пр. Ленина, д. 93, (8617) 71-74-63
tukrinat@mail.ru

Аннотация

В статье рассмотрены недостатки существующих методов оценки результатов критериально-ориентированного тестирования знаний и предложены модели и процедуры, обеспечивающие объективность и дифференцированность оценки результатов тестирования знаний.

In article shortcomings of existing methods of an assessment of results of the criteria focused testing of knowledge are considered and models and the procedures providing objectivity and differentiation of results of testing of knowledge are offered.

Ключевые слова

критериально-ориентированное тестирование; измерение и оценивание результатов тестирования.

criterion-oriented testing; measuring and evaluation of testing results.

Введение

Компьютерное тестирование сегодня является одним из наиболее распространенных методов и средств педагогического контроля знаний обучающихся. Характеристики компьютерных средств тестирования (КСТ) и методики их применения широко обсуждаются в педагогическом сообществе. При этом обсуждаются инструментальные средства создания КСТ [1], сравниваются различные виды КСТ [2], оцениваются влияния видов и сложности тестовых заданий на результаты тестирования [3,4], описываются различные методические приемы применения КСТ [5,6,7,8]. Гораздо меньше публикаций направлено на совершенствование самих процедур критериально-ориентированного тестирования. К числу этих процедур и относится процедура оценки результатов тестирования, которая рассматривается ниже.

Тестирование определяется как “один из наиболее стандартизованных и объективных методов контроля и оценивания знаний, умений и навыков испытуемого, который лишен таких традиционных недостатков других методов контроля знаний, как неоднородность требований, субъективность экзаменаторов, неопределенность

системы оценок и т.п.” [9]. Однако указанное положение об объективности педагогического тестирования не может быть полностью распространено на критериально-ориентированные тесты, позволяющие оценить уровень подготовленности испытуемых в конкретной предметной области по несвязанному с процедурой тестирования внешнему критерию. Так, из основных форм закрытых тестовых заданий в современных КСТ объективно оцениваются только задания в форме “одиночный выбор”, когда испытуемый должен выбрать один вариант ответа из нескольких предложенных. В остальных формах заданий процедура объективного измерения результатов тестирования заменяется процедурой субъективной идентификации этих результатов по правилам, установленным преподавателем или разработчиком теста. К таким формам закрытых критериально-ориентированных тестовых заданий относятся: 1) форма “множественный выбор”, когда испытуемый должен выбрать несколько вариантов ответа из нескольких предложенных; 2) форма “соответствие”, когда тестируемому нужно упорядочить два списка таким образом, что бы они соответствовали друг другу; 3) форма “упорядоченный список”, когда тестируемому нужно упорядочить список в определенном порядке.

Общепринятые методы обработки результатов выполнения тестовых заданий таких форм отсутствуют. Так для оценки результатов тестирования в форме “множественный выбор” рассматривается ряд неформальных рекомендаций и формальных моделей.

Неформальные рекомендации сводятся к формулировке некоторого правила назначения балла оценки. В.С. Аванесов в [10] рекомендует за полностью правильное решение дать три балла, за каждую ошибку снимать один балл. Если ошибок больше трех, то давать 0 баллов. М.Б. Челышкова [11] рекомендует за полностью выполненное задание с выбором нескольких верных ответов давать 1 балл и 0 баллов за, хотя бы один, неверный ответ. В.Ю. Переверзев в [12] описывает метод “частичного балла” (*partial credit*), в котором за каждый правильно выбранный ответ дается 1 балл, за неправильно выбранный ответ – 0 баллов. Штрафные баллы в этом методе не предусмотрены. По мнению В.С. Кима [13] использование заданий с выбором одного верного ответа предпочтительней и лучше заменять одно задание с выбором нескольких верных ответов на несколько заданий с выбором одного верного ответа.

Известные формальные методы и модели [14,15] либо не имеют ясного обоснования, либо, вообще, приводят к некорректным результатам.

Например, в [15] эталонный ответ и результаты тестирования представляются соответственно в виде векторов $X = (x_1, x_2, \dots, x_i, \dots, x_n)$ и $Y = (y_1, y_2, \dots, y_i, \dots, y_n)$, в которых элементы x_i, y_i равны 1, если их выбор верен, и равны 0, если их выбор не верен. Различия между этими векторами оцениваются “манхэттенским расстоянием”, рассчитываемым по формуле

$$r = \frac{1}{k} \sum_{i=1}^{i=n} |x_i - y_i|, \quad (1)$$

где: k – количество верных пунктов ответа в векторе X .

В [15] указано, что рассчитанные по (1) оценки изменяются в интервале $[0,1]$. При этом имеют место следующие соответствия: 1) $r = 0$ - все пункты верных ответов выбраны; 2) $r = 1$ - тест полностью не пройден; 3) $0 < r < 1$ - тест пройден частично. Работоспособность модели (1) в [15] демонстрируется на примере, где исходным данным $X = (0,1,0,1,0)$ $Y = (0,1,0,0,0)$ по (1) соответствует оценка $r = 0,5(0+0+0+1+0) = 0,5$. Но если рассмотреть пару $X = (0,1,0,1,0)$, $Y = (1,0,1,0,1)$, то формула (1) дает результат $r = 0,5(1+1+1+1+1) = 2,5$, который лежит за пределами интервала $[0,1]$ изменения r .

Аналогичные несоответствия имеют место и в отношении формальных методов и моделей оценок тестовых заданий в формах “соответствие” и “упорядоченный список”.

Для устранения подобных несоответствий процедур измерения результатов критериально-ориентированного тестирования предлагается использовать данные теории распознавания образов [16].

В [17] обосновано, что ответы как на перечисленные выше формы закрытых тестовых заданий (“множественный выбор”, “соответствие” и “упорядоченный список”), так и производные от них формы, могут быть сведены к двум базовым формам ответов: МНОЖЕСТВО (неупорядоченное множество) элементов или СПИСОК (упорядоченное множество). Поэтому дальнейшее описание предлагаемых процедур измерения и оценивания результатов тестирования приводится в отношении последних двух базовых форм тестовых заданий.

Обработка ответов обучающегося типа МНОЖЕСТВО

Постановка задачи оценки ответа типа МНОЖЕСТВО: Пусть дано (см. рис. 1): 1) множество V исходных альтернатив тестового задания; 2) множество M_0 , характеризующее ответ обучающегося; 3) множество M_3 , характеризующее эталонный ответ. Пусть для M_0, M_3 выполнено условие

$$(M_0 \subseteq V) \wedge (M_3 \subseteq V), \quad (2)$$

где: $V = \{a_i\} (i = \overline{1, v}; v < \infty)$, $M_0 = \{a_i\} (i = \overline{1, n}, n \leq v)$, $M_3 = \{a_i\} (i = \overline{1, m}, m \leq v)$.

Требуется оценить ответ испытуемого в традиционной 4-балльной шкале

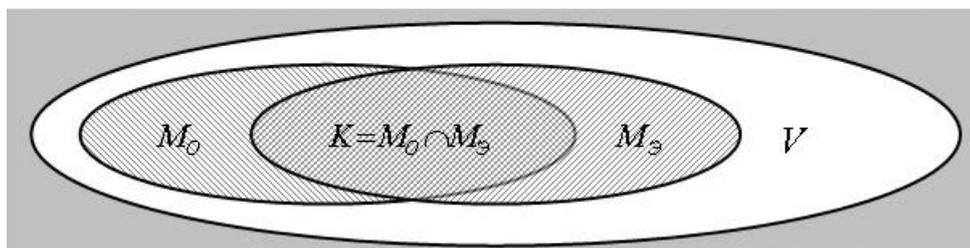


Рис. 1. Диаграмма Эйлера-Венна для множеств V, M_0, M_3

Решение задачи: Введем функцию $r(M_i, M_j) \in [0, 1]$, которая бы являлась мерой различия множеств M_i и M_j и удовлетворяла следующим требованиям: 1) $r = 0$, если M_0 и M_3 состоят из одних и тех же элементов, т.е. $M_0 = M_3 = M_0 \cap M_3 = M_0 \cup M_3$ (если $M_0 = \emptyset$ и $M_3 = \emptyset$, то $M_i = M_j$); 2) $r = 1$, если $M_0 \cap M_3 = \emptyset$; 3) $0 < r < 1$, если $M_0 \cap M_3 \neq \emptyset$.

Такую меру различия между M_0 и M_3 предлагается представить в виде

$$r = 1 - \frac{n(M_i \cap M_j)}{n(M_i \cup M_j)} = 1 - \frac{k}{l} = 1 - \frac{k}{m + n - k}, \quad (3)$$

где: $n(M_i \cap M_j) = k$ – число элементов в множестве $K = M_0 \cap M_3$ (число одинаковых элементов в M_0 и M_3), $n(M_i \cup M_j) = l = n + m - k$ – число элементов в множестве $L = M_0 \cup M_3$ (общее число неповторяющихся элементов множеств M_0 и M_3).

Мера (3) отвечает всем предъявленным к ней требованиям и представляет собой меру различия, коэквивалентную мере сходства Жаккара [18,19].

Практически расчет ее значения сводится:

1) к подсчету (см. табл. 1) числа m решений вида “да” обучающегося, числа n решений вида “да” в эталонном ответе и числа k совпадений решений вида “да” в ответах обучающегося и эталонном ответе;

2) подстановке полученных значений m, n, k в (3).

Таблица 1

Форма представления ответа типа МНОЖЕСТВО

Альтернативы (элементы G_v множества V)	1	2	...	i	...	v
Ответ обучающегося (элементы G_v^O множества M_O)	нет	да	...	да	...	нет
Эталонный ответ (элементы G_v^Y множества M_Δ)	да	да	...	нет	...	да

В теории системно-информационного анализа [20] введено понятие неупорядоченности q как меры различия какого-либо выбранного параметра x_j в отношении эталона порядка x_{jm} , которая стремится к нулю при $x_j \rightarrow x_{jm}$. Оценка (3) соответствует этому понятию, что позволяет организовать процедуру оценивания измеренных результатов по схеме, предложенной в [21]:

1. Абсолютная неупорядоченность ответа испытуемого оценивается как:

$$q = r^2. \tag{4}$$

2. Рассчитывается оценка N в традиционной 4-балльной шкале:

$$N = \log_2 \frac{S}{q}, \tag{5}$$

где: $S=2^5 q_{6z}=32q_{6z}$ — коэффициент, адаптирующий количественное выражение балла оценки к традиционной 4-балльной шкале; q_{6z} – максимальная оценка q , не вызывающая снижения балла оценки (более подробно см. [21]).

Применение формул (2-5) обеспечивает возможность настройки балла оценки на систему предпочтений преподавателя за счет изменения q_{6z} .

Обработка ответов обучающегося типа СПИСОК

Постановка задачи оценки ответа типа СПИСОК: Пусть выборочные ответы X_O вида СПИСОК и соответствующий им эталон X_Δ определены на исходном множестве V следующим образом:

$$\left. \begin{aligned} &V = X_\Delta \cup M; \\ &(X_\Delta = \{a_1, a_2, \dots, a_i, \dots, a_k\}) \wedge \left(X_\Delta = p^k = \begin{pmatrix} 1, 2, \dots, i, \dots, k \\ a_1, a_2, \dots, a_i, \dots, a_k \end{pmatrix} \right) (i = \overline{1, k}); \\ &(M = \{b_1, b_2, \dots, b_u, \dots, b_s\}) \wedge \left(M \neq p^s = \begin{pmatrix} 1, 2, \dots, u, \dots, s \\ b_1, b_2, \dots, b_u, \dots, b_s \end{pmatrix} \right) (u = \overline{1, s}) \\ &X_O \subseteq V \end{aligned} \right\}. \tag{6}$$

Необходимо определить оценку $N(X_O, X_\Delta)$ ответа X_O испытуемого в традиционной 4-балльной шкале.

Решение задачи: Для решения подобных задач применяется два метода. Метод Кендалла [16] обеспечивает получение оценок $r = [0, 1]$, соответствующих принятой мере различия, но работоспособен только при одинаковых числе и номенклатуре элементов в списках. Метод Фора [16] работоспособен при любом наборе эле-

ментов, но неоднозначен и исключает нормирование оценок, а значит и их представление в традиционной 4-балльной шкале. Ниже предлагается метод, который базируется на модели (3) оценки ответа типа МНОЖЕСТВО, методе Кендалла и процедуре создания пустых элементов, применяемой в методе Фора.

В соответствии с (6) список X_O может содержать элементы вида b_u , на которых не может быть определено отношение порядка. Поэтому в общем случае расчет оценки $N(X_O, X_3)$ предлагается производить в 4 этапа.

Этап 1. Оценка неупорядоченности $q_1(X_O, X_3)$ ответа испытуемого (списка X_O) относительно эталона (списка X_3) по номенклатуре элементов.

Списки X_O и X_3 рассматриваются как неупорядоченные множества, а неупорядоченность их номенклатуры оценивается по формулам (3,4).

Этап 2. Оценка неупорядоченности $q_2(X_O, X_3)$ ответа испытуемого (списка X_O) относительно эталона (списка X_3) по упорядоченности элементов.

В соответствии с (6) списки X_O и X_3 имеют вид

$$X_O = \{a_i, b_u\}, \quad (7)$$

$$X_3 = p^k = \begin{pmatrix} 1, 2, \dots, i, \dots, k \\ a_1, a_2, \dots, a_i, \dots, a_k \end{pmatrix} \quad (i = \overline{1, k}), \quad (8)$$

где: $i = \overline{0, t} (t \leq k)$, $u = \overline{0, g} (g \leq s)$.

Произведем в списке X_O замену всех элементов вида b_u на пустой элемент λ (процедура, применяемая в методе Фора), обладающий свойством

$$\forall a_i \in X_O \rightarrow a_i = \lambda, \quad (9)$$

в случае $k < n = t + g$ дополним список X_3 элементами a_{k+1}, \dots, a_n равными λ , а в случае $k > n = t + g$ — список X_O элементами a_{n+1}, \dots, a_k равными λ .

Списки X_O и X_3 примут вид перестановок, которые имеют одинаковый состав элементов, а значит могут быть корректно оценены по методу Кендалла:

1. Определяются коэффициенты сравнения:

$$\Delta_{lk}^i = \begin{cases} +1 & \text{при } x_{il} > x_{ik} \\ -1 & \text{при } x_{il} < x_{ik}, \text{ где } l < k. \\ 0 & \text{при } x_{il} = x_{ik} \end{cases} \quad (10)$$

2. Рассчитывается нормированная оценка Кендалла:

$$r = \frac{1}{2} - \frac{1}{n(n-1)} \sum_{l < k} \Delta_{lk}^j \Delta_{lk}^i. \quad (11)$$

Оценка $r_{cn}(X_O, X_3)$ по формуле (11) формально и по смыслу соответствуют понятию неупорядоченности, откуда в соответствии с (4) $q_2(X_O, X_3) = r_{cn}^2$.

Этап 3. Оценка общей неупорядоченности $q = f(q_1, q_2)$ ответа испытуемого (списка X_i) относительно эталонного ответа (списка X_0).

Общая неупорядоченность ответа вида СПИСОК определяется как

$$q = 1 - (1 - q_1)(1 - q_2). \quad (12)$$

Этап 4. Определение оценки выборочного ответа типа СПИСОК в традиционной 4-балльной шкале.

Оценка рассчитывается путем подстановки в (5) значения q , определенного по (12).

В совокупности с моделью и процедурой оценки тестовых заданий в форме “одиночный выбор” представленные выше модели и процедуры обеспечивают объективность и дифференцированность оценки всех известных форм тестовых заданий.

Использованная литература:

1. Клыков В.Е., Денисевич Т.В., Филатова Н.Н. Оболочка для создания тестирующих программ // Образовательные технологии и общество (Educational Technology & Society). - 2003. - Т. 6.- № 3. - С. 116-120. URL: <http://elibrary.ru/download/31674325.pdf>
2. Галеев И.Х., Иванов В.Г., Аристова Н.В., Урядов В.Г. Сравнительный анализ программных комплексов TestMaker и АСТ-Test // Образовательные технологии и общество (Educational technology & Society) – 2007. - Т. 10 -№ 3.- С.336-360. – URL: <http://elibrary.ru/download/29640481.pdf> .
3. Кинцель Д.А., Кузнецов А.В. Нечисловой подход к моделям тестирования и оцениванию параметров тестов // Образовательные технологии и общество (Educational Technology & Society). - 2007. - Т. 10. - № 1. - С. 276-281. URL: <http://elibrary.ru/download/51665322.pdf>
4. Сафаров Р.Х., Панищев О.Ю. Численное моделирование инвариантности оценки знания относительно трудности тестовых заданий в рамках модели Г. Раша // Образовательные технологии и общество (Educational Technology & Society). - 2012. - Т. 15. - № 1. - С. 424-435. URL: <http://elibrary.ru/download/71724984.pdf>
5. Хохлов Д.Г., Захарова З.Х., Захаров А.Н. Оценка уровня подготовки в системе электронного обучения программированию // Образовательные технологии и общество (Educational Technology & Society). - 2012. - Т. 15. - № 1. - С. 537-554. URL: <http://elibrary.ru/download/68618600.pdf>
6. Сафаров Р.Х., Хрусталева А.В. Типы вопросов тематических тестов по классической механике на основе программного продукта TestMaker // Образовательные технологии и общество (Educational Technology & Society). - 2010. - Т. 13. -№ 1. - С. 281-292. URL: <http://elibrary.ru/download/57753121.pdf>
7. Осадчий В.В. Компьютерная система рейтингового оценивания знаний, как средство повышения уровня знаний студентов. // Образовательные технологии и общество (Educational Technology & Society). - 2013. - Т. 16. - № 2. - С. 361-371. URL: <http://elibrary.ru/download/75320260.pdf>
8. Галеев И.Х. Компьютерный контроль знаний (локально и дистанционно) / И.Х. Галеев, В.Г. Иванов, Д.Л. Храмов, О.В. Колосов; Под ред. И.Х. Галеева. - Казань: Казанский государственный технологический университет, 2005. – 126с.
9. ОСТ Т 1.1. Педагогические тесты, термины и определения. Отраслевой стандарт (первая редакция). – М.: Министерство образования РФ, 2001. – 13с. URL: http://bank.orenipk.ru/Text/t19_135.htm .
10. Аванесов В.С. Форма тестовых заданий. -М.: Центр тестирования, 2005. -156 с. URL: <http://www.twirpx.com/file/626051/> .
11. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. - 432 с. URL: <http://www.twirpx.com/file/101903/> .
12. Переверзев В.Ю. Технология разработки тестовых заданий: справочное руководство. –М.: Е-Медиа, 2005. -265 с.
13. Ким В.С. Тестирование учебных достижений: Монография. - Уссурийск: Издательство УГПИ, 2007. - 214 с. URL: <http://www.twirpx.com/file/639729/>

14. Карпова И.П. Некоторые аспекты качественной оценки ответов тестируемых в системах контроля знаний // Вопросы тестирования в образовании, 2006, № 4 (26), С. 22-31. URL: <http://www.rema44.ru/resurs/papers/karpova/stat9.doc> .
15. Кумаритов А.М., Дубенко Ю.В. Методы и алгоритмы контроля знаний и оценки эффективности автоматизированных обучающих систем на производственном предприятии // Аудит и финансовый анализ, 2009, №2, С. 455-467. URL: http://www.auditfin.com/fin/2009/3/11_01/11_01%20.pdf .
16. Фор А. Восприятие и распознавание образов / Пер. с фр. / Под ред. Г.П. Катгса. - М.: Машиностроение, 1989. - 272 с.
17. Карпова И.П. Исследование и разработка подсистемы контроля знаний в распределенных автоматизированных обучающих системах: Дисс. ... канд. техн. наук. – М.: МГИЭМ, 2002. – 239с. URL: <http://www.twirpx.com/file/344282/> .
18. Сёмкин Б.И., Двойченков В.И. Об эквивалентности мер сходства и различия // Исследование систем. Т. 1. Анализ сложных систем. - Владивосток, 1973. - С 95–104.
19. Андрейчиков А.В., Андрейчикова О.Н. Анализ, синтез, планирование решений в экономике. - М.: Финансы и статистика, 2000. - 368 с. URL: <http://www.twirpx.com/file/657537/>
20. Горский Ю.М. Системно-информационный анализ процессов управления. - Новосибирск: Наука. Сиб. Отд-ние, 1988. – 268с.
21. Печников А.Н. Теоретические основы психолого-педагогического проектирования автоматизированных обучающих систем. - Петродворец: ВВМУРЭ, 1995. - 326с. URL: http://www.pedlib.ru/Books/1/0224/1_0224-1.shtml .