СТРАТЕГИЯ СОСТАВЛЕНИЯ ТЕСТОВ ДЛЯ БОЛЬШИХ ГРУПП ИСПЫТУЕМЫХ С РАЗЛИЧНЫМ УРОВНЕМ ПОДГОТОВЛЕННОСТИ

Гаврилов А.И., Хайрюзова Е.В., Шапошникова Т.Л.

(Краснодар)

Обсуждаются результаты использования критерия экспоненциального возрастания сложности при отборе заданий для тестирования больших групп испытуемых. Показано, что для выбранных заданий результаты тестирования удовлетворяют похожим распределениям при различном уровне подготовленности групп.

THE STRATEGY OF TESTS COMPOSITION FOR LARGE GROUPS OF EXAMINEES WITH DIFFERENT LEVELS OF KNOWLEDGE

Gavrilov A.I., Gavrilov I.A., Hairjuzova E.V., Shaposhnicova T.L.

(Krasnodar)

The results of usage criterion of exponential increasing difficulty are discussed, especially in selection of tasks for doing tests in large groups of examinees. It is shown that for the chosen tasks the test results describe as similar distributions in the groups of different level.

Применение практики единого государственного экзамена (ЕГЭ) в тестовой форме создает ряд проблем для теоретической тестологии, связанных с нарушением принятых критериев в составлении заданий. В частности, математический аппарат анализа результатов тестирования опирается на гауссово распределение числа правильных ответов среди участников испытания. На практике выполнимость этого требования достигается подбором тестовых заданий по итогам испытаний на выделенной группе участников, принятой за эталонную. Отобранное задание считается "хорошим", если соответствует нормальному распределению. Очевидно, оно будет оставаться таковым только при испытаниях на группах, близких по уровню обученности к эталонной, и существенно отличается при масштабном тестировании в различных районах страны в рамках ЕГЭ. В этой связи остановимся на двух важных вопросах: методике подбора тестовых заданий, пригодных для работы с широким кругом испытуемых, и анализе результатов, не отвечающих гауссовому распределению.

В качестве базового принципа, который требуется подтвердить экспериментально, примем утверждение: если на некоторой совокупности заданий достаточно представительная выборка испытуемых показывает экспоненциальное спадание числа правильных ответов при переходе от одного задания к другому, то и другая выборка аналогичной природы также будет показывать на этих вопросах экспоненциальное спадание, но, быть может, с другой скоростью.

Рассмотрим на конкретном примере метод отбора упомянутых выше заданий и результаты их применения в тестировании. Разумеется, совокупность этих заданий не образует "хорошего" теста в общепринятом смысле. Тестирование проводилось в трех группах студентов технологического университета по одному из разделов общего курса физики. Группы существенно отличались по объему часов, отводимых на изучение раздела, профессиональной ориентации и половому составу. Лекционные занятия в них проводились различными преподавателями. Общая численность тестируемых составляла 74 человека. Им было предложено 27 заданий в тестовой форме открытого типа (без вариантов ответа) широкого спектра сложности: от самых простых до достаточно сложных. Необходимо подчеркнуть, что на первом этапе широкий спектр заданий совершенно необходим, также как и представительность выборки испытуемых. Соответственно, и время тестирования на этом этапе будет значительно большим, чем на последующих этапах, когда число отобранных заданий существенно уменьшится.

В таблице 1 представлен результат тестирования.

Таблица 1.

Номер вопроса	Число правильных ответов	Номер вопроса	Число правильных ответов
1	47	15	29
2	72	16	1
3	72	17	54
4	72	18	53
5	68	19	9
6	20	20	37
7	30	21	9
8	54	22	2
9	23	23	35
10	53	24	36
11	13	25	22
12	29	26	5
13	31	27	21
14	57		

Распределение числа правильных ответов среди участников представлено графически на рис. 1. Выберем из предложенных вопросов те, которые (приблизительно) соответствуют экспоненциальному закону убывания правильных ответов от $N_0 - 1$ до 1

$$N = (N_0 - 1)e^{-\lambda(k-1)}, \tag{1}$$

где $$\lambda = \frac{\ln(N_0 - 1)}{M - 1}. \tag{2}$$

Принятые обозначения:

N_0 — число испытуемых;

$k = 1, 2, \ldots, M$ – номер вопроса в порядке убывания числа правильных ответов;

M – число вопросов;

N – число правильных ответов.

Формуле (1) соответствуют вопросы 2, 14, 1, 20, 13, 15, 9, 6, 11, 19, 21, 26, 22, 16, т.е. всего 14 вопросов. Распределение правильных ответов на выбранные вопросы представлено на рис. 2. Соответственно, по группам испытуемых были получены распределения, приведенные на рис.3 — 5. Все обозначения на графиках аналогичны обозначениям на рис.1 и 2. На рис. 6 собраны все распределения по отобранным вопросам, приведенные к общему числу испытуемых в группах. Из графиков видно, что ответы на выбранные вопросы образуют сходные по характеру распределения, обладающие отличительными чертами, связанными с особенностями подготовленности испытуемых по данному разделу физики.

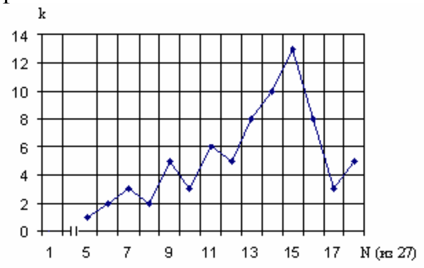

Рис. 1. Распределение правильных ответов из 27 вопросов

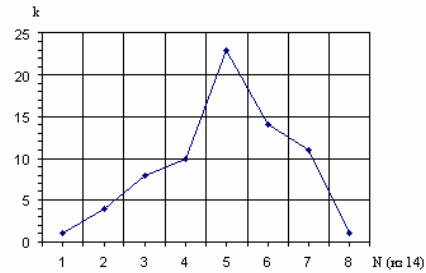

Рис. 2. Распределение правильных ответов из 14 вопросов

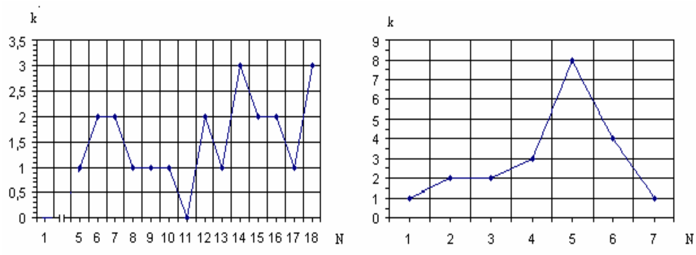

Рис. 3. Группа № 1 (22 человека)

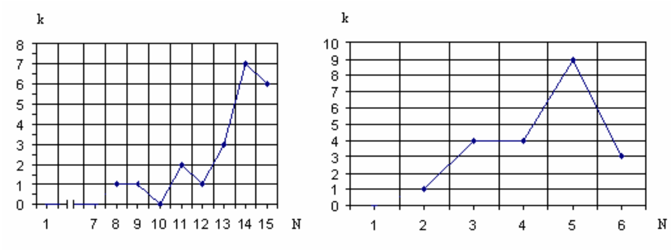

Рис. 4. Группа № 2 (21 человек)

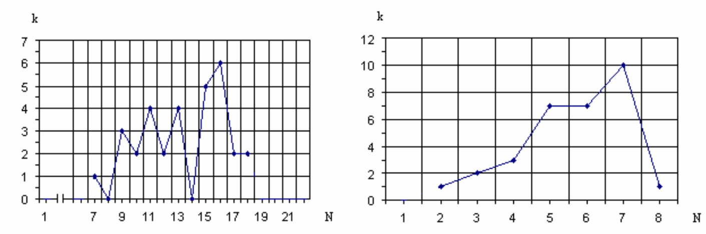

Рис. 5. Группа № 3 (32 человека)

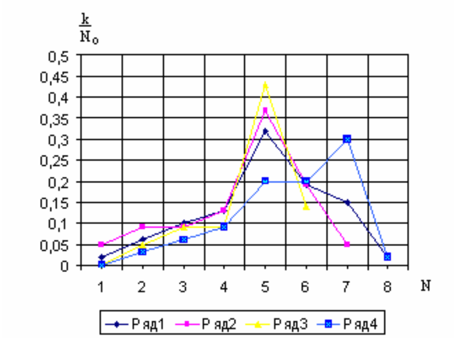

Рис. 6. Приведенные распределения: ряд1 – общий; ряд 2 – группа № 1; ряд3 – группа № 2; ряд 4 – группа №3

Примечательно, что полученные эмпирические распределения достаточно хорошо описываются аналитической зависимостью

$$\frac{k}{N_0} = P_M(k) = AC_M^k p_0^k e^{-\lambda k^2}(1 - p_0 e^{-\lambda k})^{M-k}, \qquad (3)$$

где C_M^k — коэффициенты биноминального разложения; A, p_0, λ

– подгоночные параметры, которые должны быть согласованы с условием нормировки. Параметр p_0 имеет смысл вероятности правильного ответа в данной группе на самые простые вопросы, т.е. характеризует однородность состава группы. Параметр λ характеризует скорость спадания числа правильных ответов, т.е. подготовленность группы.

В рассматриваемых примерах $A \approx 1{,}5$. Для групп №1 и №2 $\lambda \approx 0{,}17$, а для группы №3 $\lambda \approx 0{,}12$. Это свидетельствует о наличии в последней группе студентов с более высокой подготовленностью. В то же время для первых двух групп $P_0 \approx 0{,}9$, а для третей – $P_0 \approx 0{,}8$, что позволяет считать группы №1 и №2 более однородными по составу. Последний тезис находит подтверждение при анализе ответов каждого студента: в третьей группе большее число студентов ошибаются при ответах на легкие вопросы.

Интересно отметить, что не вполне равномерный характер диаграмм распределения отражает объективно наличие в группах дискретного расслоения студентов по уровню подготовленности.

Установленные закономерности позволяют надеться, что дальнейшее развитие предлагаемого подхода позволит упорядочить анализ больших массивов тестовых данных.