

## ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ПРОЦЕССА ОБУЧЕНИЯ И ТЕСТИРОВАНИЯ

УДК 681.3

**СОЛОБУТО Лариса Вадимовна**

доцент кафедры интеллектуальных информационных систем  
Черноморского государственного университета им. Петра Могилы.

**Научные интересы:** информационные технологии в образовании.

**e-mail:** lar.solobuto@gmail.com

### ВВЕДЕНИЕ

Проблема оптимизации процесса проверки знаний непосредственно связана с повышением качества подготовки специалистов в высших учебных заведениях. Результат зависит, в значительной мере, от того, насколько продуманным будут вопросы теста, а также, будет ли оптимальным количество вопросов. Многочисленные публикации, посвященные вопросам тестирования, уделяют много внимания вопросам содержания теста, трудности заданий, методам оценки, гораздо реже рассматривается вопрос о достаточной длине теста [1-3]. В тех случаях, когда длине теста уделяется достаточное внимание, используется громоздкий математический аппарат, что сужает область применения методики и мобильность в вопросах использования тестов с различными характеристиками [4].

**Целью работы** является создание мобильной методики определения длины тестов, которая бы позволяла простыми средствами и за короткое время рассчитать параметры теста, при необходимой точности и надежности.

### ПОСТАНОВКА ПРОБЛЕМЫ

Если рассмотреть учебный процесс с точки зрения теории информации, то потери информации в связке «преподаватель–студент–тест» происходит на этапе «студент» из-за невнимательности, плохой подготовки, забывчивости последнего. Преподаватель передает некоторое количество информации студенту, а затем, студент, отвечая на вопросы, передает усвоенную часть информации экзаменатору (в качестве экзаменатора

можно рассматривать тест). Студент, отвечая или не отвечая на вопросы, показывает, какая часть информации была им усвоена. Разница между количеством информации, которое было передано преподавателем студенту и количеством информации, которое было передано студентом экзаменатору – потери в процессе передачи информации. Таким образом, студент выступает в роли канала информации, в котором происходят (или не происходят) потери. Исходя из этого, можно применить положения теории информации к процессу тестирования.

Истинный уровень знаний при тестировании определяется с какой-то вероятностью, что в свою очередь, зависит от качественных характеристик теста. Поскольку тест имеет определенную длину, то результирующая оценка является случайной величиной, определяемой вероятностными характеристиками.

При формировании теста, следует учесть, что, исходя из положений теории вероятности и теории информации, если вопросы теста не связаны между собой, то вероятность правильной оценки знаний для всех вопросов равна произведению вероятностей, что в итоге уменьшает общую вероятность  $P(XY)$ . Если же вопросы теста (события, в терминах теории вероятности) связаны между собой, то вероятность получения истинной оценки будет определяться формулой  $PXY = P(X) + P(Y) - P(XY)$ . Следовательно, построив тест так, чтобы вопросы теста основывались на знании предыдущего материала, повышаем вероятность правильной оценки знаний.

### РЕШЕНИЕ ПРОБЛЕМЫ

Рассмотрим разработанную методику определения длины теста с использованием статистической программы SPSS, которая предоставляет разработчикам широкие возможности. Возьмем за основу тест из 30 вопросов, из которых на 20 даны правильные ответы. По 100-балльной системе тест оценивается на 66 баллов (минимально допустимая оценка для успешной аттестации). Шкалу переменных, которая в данном случае является интервальной, разобьем на 7 частей, (каждая часть содержит некоторый интервал вопросов теста, а частотой является количество правильных ответов в этом интервале) и распределим правильные ответы так, чтобы кривая распределения являлась нормальной. Используя программу, получим график распределения и определим основные статистики, чтобы убедиться в нормальности полученного распределения. Более наглядная проверка распределения на нормальность проводится при помощи графика «Вероятность-Вероятность», который показывает связь функций наблюдаемого и теоретического кумулятивных распределений или графика «Квантиль-Квантиль», где каждое наблюдаемое значение сравнивается со значением, ожидаемым при нормальном распределении. В случае, когда все точки исследуемого распределения лежат на диагональной линии – распределение

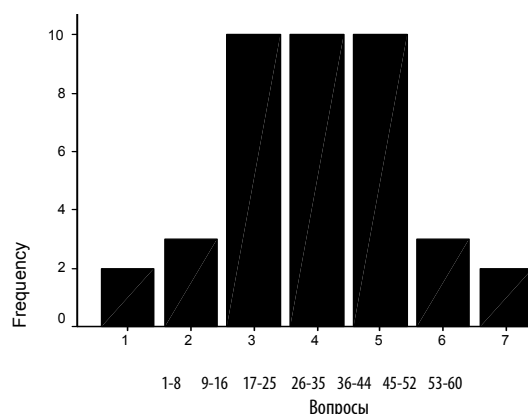
является нормальным. Если увеличить количество вопросов теста, а остальные характеристики оставить без изменения (соотношение между правильными и неправильными ответами, общее количество баллов, а также количество интервалов) то, при условии одинакового вида распределения (в данном случае – нормального), средняя распределения не изменится (Рис. 1а), она примет другое значение только при изменении типа распределения. Эта закономерность дает возможность сравнивать тесты любой длины, т.к. средняя является надежной характеристикой распределения, отражает свойства рассматриваемой совокупности и часто применяется при сравнении различных выборок [1].

**Замечание:** Поскольку мы имеем дело с реальными тестами, у которых каждый интервал соответствует определенному количеству вопросов теста, количество правильных ответов в интервале (или частота данного интервала) не может быть больше количества вопросов, поэтому не всегда удастся получить какой-либо тип распределения в чистом виде. В распределении, что на рис. 1б, невозможно добиться меньшего эксцесса (при данных параметрах), но, проведя исследование распределения на нормальность при помощи теста Лиллифора, который является модификацией теста Колмогорова-Смирнова (Рис. 2), получим  $\text{Sig.} = 0,075$ .

NORMAL60		
N	Valid	40
	Missing	0
Mean		4,00
Std. Error of Mean		,226
Median		4,00
Mode		3 <sup>a</sup>
Std. Deviation		1,432
Variance		2,051
Skewness		,000
Std. Error of Skewness		,374
Kurtosis		-,118
Std. Error of Kurtosis		,733

a. Multiple modes exist. The smallest value is shown

а)



б)

Рисунок 1 – Характеристики нормального распределения правильных ответов теста из 60 вопросов: а) статистические характеристики; б) график распределения

Хотя выбор определенного уровня значимости является достаточно произвольным, на практике окон-

чательное решение обычно зависит от традиции, имеющейся в данной области исследований.

**Tests of Normality**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
NORMAL60	,132	40	,075	,953	40	,094

a. Lilliefors Significance Correction

*Рисунок 2 – Результат проверки теста из 60 вопросов на соответствие нормальному распределению правильных ответов*

Обычно во многих областях результат 0,05 является приемлемой границей статистической значимости, следовательно, при Sig. < 0,05 распределение значимо отличается от нормального [5]. Поскольку в нашем случае Sig. = 0,075, распределение можно считать нормальным, но вероятность ошибки достаточно велика и ее можно уменьшить, увеличивая длину теста (согласно центральной предельной теореме, которая

говорит, что при большом объеме выборки любое распределение приближается к нормальному независимо от начального распределения), а согласно теории статистики, любое распределение приближается к нормальному, если выборка насчитывает свыше 100 наблюдений [5].

**One-Sample Test**

	Test Value = 4				
	t	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
				Lower	Upper
BALL60	,000	1,000	,00	-,57	,57

*Рисунок 3 – Результат исследования теста из 60 вопросов с нормально распределенными правильными ответами при помощи Т-теста одной выборки*

Такое внимание к тестам с нормальным распределением правильных ответов объясняется тем, что они дают максимально значимый результат Sig. (2-tailed) = 1,000 и показывают отсутствие ошибки (t = 0,000) в предположении, что результатам такого теста можно доверять (рис. 3).

Если рассмотреть тесты, правильные вопросы в которых распределены асимметрично (левосторонняя и правосторонняя асимметрия), а также тесты с максимально возможными значениями эксцесса, то можно сделать вывод, что наибольшее отклонение от среднего, полученного при нормальном распределении, наблюдается в тесте с правосторонней асимметрией в расположении правильных ответов. Это наиболее неблагоприятное распределение правильных ответов в тесте, поэтому дальнейшие исследования проводим на базе именно таких распределений. Исследования проводились при помощи Т-теста для одной выборки.

Обоснование использования выбранного метода.

Существует два основных метода определения различий между группами на основании средних значений: Т-тесты и дисперсионный анализ. Т-тесты предназначены для установления различий между двумя группами, а дисперсионный анализ позволяет анализировать одновременно любое количество групп. Поскольку на данном этапе исследования стоит задача определения оптимальной длины теста путем сравнения теста, имеющего наиболее неблагоприятное распределение правильных ответов с тестом, имеющим классическое нормальное распределение, рационально остановиться на Т-тесте. Используя Т-тест для одной выборки можно выяснить, отличается ли значительно реальное среднее значение какого-либо теста от среднего значения теста с нормально распределенными правильными ответами (который дает максимальное значение точности и надежности).

При исследовании асимметричных тестов, состоящих из 30, 60 и 90 вопросов, были получены результаты с низкой статистической значимостью. Это означает, что результаты исследования нельзя рассматривать как достоверные и такому тесту доверять нельзя. Дальнейшее увеличение длины теста (120 вопросов) позволило получить достаточно значимый результат, но при этом точность оказалась низкой, поэтому количество вопросов было увеличено и каждая группа вопросов была исследована.

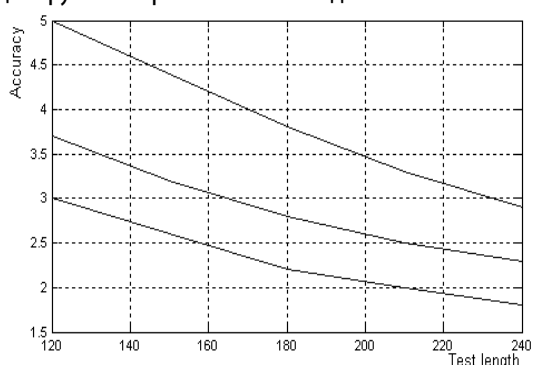


Рисунок 4 – Зависимость точности определения итогового балла от длины теста

Исследовались тесты различной длины (120, 150, 180, 210, 240) и были выявлены зависимости точности оценивания знаний студента (с различной степенью надежности: 90%, 95%, 99%) от длины теста. Тесты из 150, 180, 210 и 240 вопросов дают результаты, которые удовлетворяют нулевой гипотезе (результатам теста можно доверять), но отличаются точностью и размером доверительного интервала. Эти параметры определяются при помощи Т-тестов. Для каждого значения вероятности (90%, 95%, 99%) программа рассчитывает границы интервала, в пределах которого лежит случайная величина (оценка знаний студента). Поскольку график распределения каждого теста исследуется в однородных условиях (шкала всех графиков разбита на 7 интервалов, т.е. интервал, в котором лежит действительный балл, показан в относительных единицах), для каждого теста необходимо пересчитать относительные единицы в количество баллов. Таким образом, для каждого теста можно определить погрешность в баллах. Строим график зависимости (рис. 4). Анализируя полученный график, можно подобрать необходимую

длину теста для каждого конкретного случая (необходимой точности и надежности).

**Замечание.** За основу были приняты результаты тестирования с минимально допустимой оценкой (66 баллов), т.к. при исследовании тестов, оцененных более высоким количеством баллов, были получены более высокие качественные характеристики, а именно: больший уровень статистической значимости и меньшая погрешность при определении действительного балла тестирования. Это дает возможность, при тестировании сильных групп студентов, использовать тесты меньшей длины. Какую именно длину теста необходимо использовать в каждом конкретном случае, можно, используя данную методику, достаточно быстро посчитать.

Например, используя методику, можно заключить, что, создав тест длиной 180 вопросов мы получаем ошибку при определении полученного балла  $\pm 2$  вопроса. Учитывая, что тест длиной 180 вопросов должен содержать 120 правильных ответов (для оценки 66 балла), то погрешность  $\pm 2$  вопроса составляет 1,6% от общего количества правильных ответов (для надежности 90%). Дальнейшее увеличение длины теста дает незначительное увеличение точности оценки. Эти результаты согласуются с исследованиями, проведенными иными методами, других авторов [7].

Важной особенностью метода является учет обратных связей между вопросами теста. Все учебные темы одной дисциплины связаны между собой, то есть в более поздних по времени изучения темах используется информация из ранее изученных без ее конкретизации, т.е. предполагается, что обучаемый знает, какой смысл вкладывается в то или иное определение или понятие, но, если в процессе изучения нового материала, преподаватель (руководствуясь наличием обратной связи) делает акцент на ранее изученном, то происходит повторение материала, но уже на более высоком уровне, тем самым, закрепляя уже изученный материал. Поскольку особенностью человеческой памяти является забывание информации, если она не повторяется в течение длительного промежутка времени, то чем меньше будет промежуток времени между информационно связанными темами, тем лучше будет усвоен материал. Этот принцип повторения должен учитываться и при составлении теста, оценивая ответ

на вопрос связанный с предыдущей темой более высоким баллом.

Исходя из вышеизложенного, можно определить количество информации, которое было получено студентом. Использована формула Хартли

$$I = \log N_0 = n \log m, \quad (1)$$

где,  $m$  – количество дискретных сообщений информации,  $n$  – количество символов в каждом сообщении (сообщением является информация, усвоенная студентом).

Более подробно этот вопрос был рассмотрен ранее [8].

## ВЫВОДЫ

Созданная методика обеспечивает получение быстрых и точных результатов при определении длины тестов. Это позволяет оперативно создавать различные тестовые композиции и применять их для контроля знаний студентов как по общеобразовательным, так и по специальным дисциплинам, независимо от характера постановочных задач.

Еще одно достоинство методики заключается в том, что позволяет формировать тесты из множества простейших вопросов. С одной стороны это позволяет исключить элемент случайности в ответах, а другой – позволяет подсчитать количество усвоенной информации (если предварительно разбить тему для изучения на простейшие блоки и подсчитать количество информации, что была передана студенту).

## ЛИТЕРАТУРА:

1. Avanesov V.S. Nauchnye osnovy testovogo kontrolja znanij. – М.: Issledovatel'skij centr, 1994. – 135 s.
2. Avanesov V.S. Osnovy nauchnoj organizacii pedagogического kontrolja v vysshej shkole. – М.: Issledovatel'skij centr, 1987. – 120 s.
3. Avanesov V.S. Testy v sociologическом issledovanii. – М.: Nauka, 1982. – 135 s.
4. Safoncev S.A. Opyt primeneniya tehnologii sozdanija didaktического testa //Pedagogическая diagnostika. – 2003. – №2. – S.83-104.
5. Gmurman V.E. Teorija verojatnostej i matematическая statistika. – М.: Vysshaja shkola, 1999. – 479 s.
6. Bjujul' A., Cjofel' P. SPSS: iskusstvo obrabotki informacii. – М.: DiaSoft, 2005. – 602 s.
7. Hlebnikov V.A. Kak nadezhno izmerjat' uchebnye dostizhenija //Pedagogическая diagnostika. – 2003. – №1. – S.41-46.
8. Solobuto L.V. Sozdanie informacionnoj modeli uchebnogo processa //Vestnik Hersonskogo nacional'nogo tehničeskogo universiteta. – 2006. – №1 (24). – S.438-443.

**Рецензент:** д.т.н., проф. Рябенкий В.М.,  
Национальный университет кораблестроения им. адм. Макарова, Николаев.