

## **Метод управления длиной системы одиночных тестовых заданий с единственным верным ответом**

Палкин Константин Сергеевич  
адъюнкт кафедры кораблевождения, Военный институт (военно-морской)  
ВУНЦ ВМФ “Военно-морская академия”,  
Ушаковская наб., д. 17/1, г. Санкт-Петербург, Россия, 197045; тел. +79312027052;  
[palkinks@mail.ru](mailto:palkinks@mail.ru)

Печников Андрей Николаевич  
доктор педагогических наук, доктор технических наук, профессор, профессор  
кафедры кораблевождения, Военный институт (военно-морской)  
ВУНЦ ВМФ “Военно-морская академия”,  
Ушаковская наб., д. 17/1, г. Санкт-Петербург, Россия, 197045; тел. +79119221982;  
[pan287@users.mns.ru](mailto:pan287@users.mns.ru)

Печников Денис Андреевич  
кандидат технических наук, доцент, доцент кафедры кораблевождения, Военный  
институт (военно-морской)  
ВУНЦ ВМФ “Военно-морская академия”,  
Ушаковская наб., д. 17/1, г. Санкт-Петербург, Россия, 197045, тел. +7921780580724;  
[19pda72@bk.ru](mailto:19pda72@bk.ru)

### **Аннотация**

В статье рассматривается способ управления процессом выполнения системы одиночных критериально-ориентированных тестовых заданий с единственным верным ответом, обеспечивающий остановку тестирования при достижении заданной достоверности оценки его результатов.

The article considers the control method of the system of single criterion-oriented test tasks gated with a single correct answer, providing the process to stop testing when the specified accuracy assessment of the results

### **Ключевые слова**

критериально-ориентированный тест; тестовое задание; результаты тестирования; точечная оценка; интервальная оценка; доверительный интервал; заданная погрешность; достаточное число тестовых заданий  
criterion-oriented test; test task; results of testing; the point estimate; interval estimate; confidence interval; given error; a sufficient number of test items

### **Введение**

Ниже термин “тестовое задание” трактуется как “учебное задание специфической формы, элемент теста, минимальная законченная составляющая единица теста, по которой испытуемый в ходе выполнения теста совершает отдельное действие, а его результат регистрируются в первичном протоколе в форме отдельного ответа” [1]. Под термином “тест” понимается “совокупность тестовых заданий, позволяющая по результатам их выполнения объективно измерить уровень подготовленности испытуемых по конкретным разделам определенной области знания” [1]. Тесты и методы оценки результатов их выполнения разделяются на нормативно-ориентированные и критериально-ориентированные.

Тест нормативно-ориентированный – “частный случай относительного теста, позволяющий сравнивать учебные достижения (уровень подготовленности) отдельных испытуемых друг с другом или относительно предварительно полученных

тестовых норм” [1]. Тест критериально-ориентированный – “частный случай теста, предназначенного для абсолютного тестирования, позволяющий оценить, преодолели ли испытуемый определенный порог усвоения учебного материала рассматриваемой учебной программы или ее части. При этом результаты тестирования сравниваются с некоторым критерием уровня подготовленности” [1].

В неперiodических [2,3,4,5,6,7,8,9,10] и в многочисленных периодических [11,12,13,14,15,16,17,18,19,20 и т.д.] специализированных изданиях обсуждаются преимущественно методы, модели и процедуры нормативно-ориентированного тестирования. При этом, как показывает проведенный в [21] анализ представленных в Интернет компьютерных систем тестирования (КСТ), в их числе, наоборот, абсолютно преобладают программные средства, ориентированные на решение задач критериально-ориентированного тестирования (Adit Testdesk, M-Тест, EasyQuizzy, MultiTester System, RSP, The Examiner testing system, FastTEST professional, C-Quest, CONTEST, SunRav TestOfficePro, HyperTest, UniTest System, TestMaster, AVELife TestGold Studio, АСТ-Тест Plus, e-University, RichTest, Tester, Грамотей-КЛАСС, ПОЛСТАР, “Контроль знаний”, “Система проверки знаний”, “Универсальный тестовый комплекс”, “Экзаменатор”, “Аттестация”, УСАТИК и т.д.). Наличие такого противоречия обусловлено тем, что “для педагогического контроля уровня обученности более привычной и естественной является задача, решаемая в рамках критериально-ориентированного подхода” [8], но именно этот подход не имеет развитой теории построения процедур и оценивания результатов тестирования.

Методы классической теории тестов и современной тестовой теории IRT (*Item Response Theory*) преимущественно ориентированы не на выяснение отношений “усвоил - не усвоил” между обучаемым и конкретной совокупностью элементов содержания обучения, а на выяснение отношений вида “знает лучше - знает хуже” между обучаемыми, т.е. на ранжировку обучаемых по уровню обученности в рассматриваемой предметной области вне зависимости от достаточности или недостаточности достигнутого ими уровня усвоения. Вопросы проектирования и оценки результатов выполнения критериально-ориентированных тестов в теории тестирования решаются по остаточному принципу. В результате “отсутствие стандарта или хотя бы общепринятых норм и правил создания компьютерных систем тестирования привело к тому, что практически каждая из них является “вещью в себе” и не имеет описания алгоритмов функционирования, в частности, описания реализуемых процедур оценивания результатов тестирования” [22].

Эти обстоятельства определяют необходимость разработки и внедрения в программное обеспечение КСТ методов и процедур, которые непосредственно направлены на достижение целей критериально-ориентированного тестирования как средства текущего педагогического контроля.

В общем случае под педагогическим контролем понимается “система научно-обоснованной проверки результатов образования, обучения и воспитания” [23] или “функция управления образовательным процессом, осуществляемая с целью получения достоверной информации о ходе и результатах проводимой воспитательной и обучающей деятельности” [24]. В педагогическом контроле выделяются входной, текущий и итоговый виды контроля.

Входной и итоговый виды контроля предваряют и завершают обучение. Результаты этих видов контроля могут и должны учитываться при проектировании обучения, но не могут изменять уже начавшийся образовательный процесс. Поэтому эти виды контроля могут быть реализованы как средствами нормативно-ориентированного, так и критериально-ориентированного тестирования. В отличие от них, текущий контроль реализуется непосредственно в процессе обучения, а его “важнейшей функцией является функция обратной связи, которая позволяет преподавателю получать сведения о ходе процесса усвоения у каждого учащегося и адаптировать процесс обучения к их потребностям” [25].

В педагогике обратная связь определяется как “положительное или отрицательное воздействие результатов обучения на процесс освоения новых знаний и умений” [26]. В этой дефиниции под результатами обучения могут рассматриваться только оценки соотношения “усвоил - не усвоил” между обучаемым и изученным ранее содержанием обучения. Сравнительные оценки успешности обучаемых здесь не информативны, поскольку не способны оказать никакого влияния на последующий образовательный процесс. Другими словами, адаптировать процесс обучения к потребностям обучаемых на основе данных нормативно-ориентированного тестирования (Петя знает лучше Вани, а Ваня — лучше Кати) невозможно. Поэтому для решения задач текущего педагогического контроля могут использоваться только методы и средства критериально-ориентированного тестирования, целенаправленно ориентированные на идентификацию соотношения “усвоил - не усвоил” между теми конкретными обучаемыми и теми конкретными познавательными объектами, в отношении которых реализуется рассматриваемый процесс обучения.

Поскольку задачи текущего контроля решаются непосредственно в процессе обучения и имеют целью выработку адресных обучающих воздействий, то процесс критериально-ориентированного тестирования должен занимать минимум времени. Иначе говоря, к самому процессу критериально-ориентированного тестирования, а также процедурам обработки и презентации его результатов должно быть предъявлено требование оперативности.

Оперативность процедур обработки и отображения результатов действий обучаемых проблем не вызывает, поскольку полностью обеспечивается за счет быстродействия аппаратно-программных средств КСТ. А вот оперативность самой процедуры тестирования соответствующего обеспечения не имеет. Поэтому условия и методы ее реализации требуют более подробного рассмотрения.

### **Краткая история и современное состояние проблемы**

Время выполнения теста определяется как произведение  $T = tn$  двух параметров: 1)  $t$  — “скорость теста (*test speededness*)” [23], оцениваемая временем выполнения одного тестового задания (для заданий закрытой формы обычно составляет 30-40 секунд на задание); 2)  $n$  — “длина теста (*test length*)”, измеряемая числом тестовых заданий в тесте. Скорость тестирования  $t$  обусловлена психофизиологическими возможностями обучаемых, а потому не может быть существенно сокращена. Другое дело, длина  $n$  теста. На минимизацию этого параметра ориентирован ряд методов, которые, по своей сути, базируются на одном и том же подходе.

Суть общепринятого подхода к решению задачи минимизации числа заданий в тесте состоит в определении (выборе) такого наименьшего числа заданий, при котором соотношения “больше” или “меньше” между оценкой результатов тестирования и принятым критерием определяются с достаточной достоверностью.

У истоков рассматриваемого подхода стояли такие корифеи тестологии как Ф.М. Лорд (*F.M. Lord*) и М.Р. Новик (*M.R. Novick*). Они в 1968 году сформулировали положение о том, что “если результаты тестовой проверки квалификации представлены набором дихотомически оцениваемых (0-1) заданий, взятых из некоторого множества таких заданий, ... то решение задачи определения числа тестовых заданий, которые необходимо выполнить для получения достоверной оценки, дает биномиальное распределение (при большом числе заданий), либо гипергеометрическое распределение (при небольшом числе заданий)” [27].

Положение о перспективности использования для решения задачи минимизации длины теста биномиального распределения далее развивал целый ряд исследователей, но конкретный толчок к полноценной формулировке задачи оптимизации длины теста дал Т. Криволл (*T. Kriewall*). В 1969 году в [28] он

высказал положение о том, что длина теста функционально связана с результатами тестирования, а значит может быть оптимизирована (минимизирована) с учетом индивидуального уровня подготовки обучаемого. Основываясь на этом положении, в 1972 году Д. Милман (*J. Millman*) в [29] предложил биномиальную модель оценки длины теста (биномиальную модель теста, метод Милмана). Аналогичные подходы в том же году опубликовали К. Клауер (*K. Klauer*) [30] и Т. Криволл (*T. Kriewall*) [31].

Метод Д. Милмана базируется на четырех предположениях: 1) тест состоит из случайной выборки дихотомически оцениваемых заданий; 2) тестовые задания независимы друг от друга; 3) вероятность правильного ответа оцениваемого испытуемого - величина постоянная для всех тестовых заданий; 4) ошибки подчиняются биномиальной модели.

Основу метода составляет биномиальная модель вида

$$P(L) = C_n^x p^x q^{n-x} \quad \text{при} \quad C_n^x = \frac{n!}{x!(n-x)!}, \quad (1)$$

где:  $P(L)$  – оценка вероятности события  $L$  появления  $x$  успехов в серии из  $n$  испытаний при рассматриваемых истинных значениях вероятностей  $p$  и  $q$ ;  $p$  – вероятность безошибочного выполнения задания;  $q = 1 - p$  – вероятность выполнения задания с ошибкой;  $x$  – число заданий, выполненных без ошибок;  $(n - x)$  – число заданий, выполненных с ошибкой;  $n$  – общее число заданий.

По модели (1) в [29] рассчитана приведенная ниже таблица (см. табл. 1).

Таблица 1

**Возможный процент неправильно оцененных студентов [29]**  
(*Percent of students expected to be misclassified*)

Passing Score	No. of Test Items	STUDENT'S TRUE LEVEL-OF-FUNCTIONING*								
		40	50	60	70	75	85	90	95	
1 out of 1	1	40	50	60	70	75	85	90	95	
2 out of 2	2	16	25	36	40	56	28	19	10	
3 out of 3	3	6	13	22	34	42	39	27	14	
4 out of 4	4	3	6	13	24	32	48	34	19	
4 out of 5	5	9	19	34	53	63	16	8	2	
5 out of 6	6	4	11	23	42	53	22	11	3	
6 out of 7	7	2	6	16	33	44	28	15	4	
7 out of 8	8	1	4	11	26	37	34	19	6	
8 out of 9	9	-	2	7	20	30	40	23	7	
8 out of 10	10	1	5	17	36	53	18	7	1	
10 out of 12	12	-	2	8	25	39	26	11	2	
12 out of 15	15	-	2	9	30	46	18	6	1	
16 out of 20	20	-	1	5	24	41	17	4	-	
20 out of 25	25	-	-	3	19	38	16	3	-	
24 out of 30	30	-	-	2	16	35	15	3	-	
32 out of 40	40	-	-	1	11	30	14	2	-	
40 out of 50	50	-	-	-	8	26	12	1	-	
48 out of 60	60	-	-	-	6	23	11	1	-	
60 out of 75	75	-	-	-	4	19	9	-	-	
80 out of 100	100	-	-	-	2	15	7	-	-	

В этой таблице приведены: 1) варианты оценок проходного балла (*passing score*), которые при рассматриваемой длине теста (*No. of test items*) соответствуют принятому критерию 0,8 (80%) успешности; 2) значения вероятности (в %) неправильной оценки студентов, которые набрали проходной балл, приведенный в

левом столбце таблицы, но имеют иной истинный балл (*student true level-of-functioning, true score*), под которым понимается “гипотетический балл, который получил бы испытуемый, если бы выполнил все задания из банка заданий для данного теста” [8,23,29]. Вертикальная пунктирная линия отображает принятый критерий успешности (80%). Оценки, которые приведены левее этой линии, обозначают вероятность неправильной положительной оценки испытуемого (зачет вместо незачета), а правее этой линии — неправильно выставленной отрицательной оценки (незачет вместо зачета).

Реализация метода предполагает анализ представленной выше таблицы (см. табл. 1) и выбор на основании приведенных в ней данных той длины теста, вероятности ошибок которой тестирующий считает допустимыми. Этот анализ рекомендуется проводить в следующей последовательности [29]: 1) задается величина критериального балла; 2) задается максимально допустимая вероятность для неправильной положительной аттестации при определенном истинном балле ниже критериального балла; 3) задается максимально допустимая вероятность для неправильной неаттестации при определенном истинном балле выше критериального балла; 4) путем экспертной оценки выбирается та длина теста, которая наиболее полно соответствует предъявленным требованиям. Последняя процедура в методе Милмана являлась эвристической, что вызвало необходимость ее формализации.

Для этого задача определения минимальной длины теста была сформулирована как задача выбора такой длины теста, которая бы при сравнении достигнутой вероятности  $p$  успешного выполнения теста с критериальным значением  $p_0$  этой вероятности обеспечивает исключение заданных величин ошибок 1 рода ( $\alpha$ ), когда знающий материал студент получает отрицательный результат (незачет), и 2 рода ( $\beta$ ), когда не знающий материала студент получает зачет.

Поиск решения такой постановки задачи развивался в двух направлениях. В рамках первого направления искомые параметры биномиальной модели Милмана определялись на основе использования закономерностей самого биномиального распределения. Второе направление предполагало поиск этих же параметров модели Милмана путем аппроксимации биномиального распределения нормальным.

Первый законченный вариант реализации первого подхода [32] был предложен Р.Р. Вилкоксом (*R.R. Willcox*) в 1980 году. Р.Р. Вилкокс показал, что решение рассматриваемой задачи может быть сведено к поиску параметров, обеспечивающих выполнение представленных ниже соответствий [32]:

$$p_0 > \hat{p} = \frac{x - (n - x)}{n(m - 1)}, \quad (2)$$

где:  $\hat{p}$  - оценка вероятности угадывания правильных ответов обучаемым;  $m$  – число дистракторов (альтернатив) в тестовых заданиях;  $x$  – число заданий, выполненных без ошибки;  $n$  - общее число заданий;

$$\left. \begin{aligned} p_1 &= (1 - \alpha)(p_0 - \delta^*) + \beta(1 - p_0 + \delta^*); \\ p_2 &= (1 - \alpha)(p_0 + \delta^*) + \beta(1 - p_0 - \delta^*) \end{aligned} \right\}, \quad (3)$$

где  $p_1, p_2$  - оценки вероятности  $p$  правильного выполнения заданий соответственно для наименьшего ( $p_0 - \delta^*$ ) и наибольшего ( $p_0 + \delta^*$ ) из допустимых значений  $p_0$ ;  $\delta^*$  - величина максимально допустимого отклонения  $p_0$ ;

$$\left. \begin{aligned} P_1 &= \sum_{x=0}^{x=n_0-1} C_n^x p_1^x (1-p_1)^{n-x} \leq \alpha; \\ P_2 &= \sum_{x=n_0}^n C_n^x p_2^x (1-p_2)^{n-x} \leq \beta \end{aligned} \right\}, \quad (4)$$

где:  $P_1, P_2$  - так называемые оценки “правильного решения” (*correct decision*) соответственно для  $p_1, p_2$  из (3);  $n_0$  - искомая минимальная длина теста;  $\alpha, \beta$  - соответственно значения значимости ошибок 1 и 2 рода принимаемых решений.

Далее Р.Р. Вилкоккс в качестве решения задачи рассматривает число  $n_0$  тестовых заданий, при котором обе оценки  $P_1, P_2$  в (4) при  $\delta^* \leq 0,1$  близки к 1 ( $P_1 \geq 0,9 \text{ и } P_2 \geq 0,9$ ). Оставаясь верным биномиальному распределению, он предлагает компьютерную программу итерационного подбора  $P_1, P_2$  при заданных значениях параметров  $\delta, \alpha, \beta, \delta^*$ . На основе таких расчетов он публикует в [32] ряд таблиц, одна из которых приведена ниже (см. табл. 2).

**Таблица 2**

**Варианты минимальной длины теста [32]**  
(*Minimum Test lengths Using Correction for Guessing*  
*Formula Scores,  $p_0 = 0,8$  and  $P_{1,2} = 0,9$* )

$\alpha$	$\beta$	n (m=4)	n (m=5)
0	.15	35	35
0	.30	65	93
0	.40	159	281
.02	.15	54	44
.02	.30	51	56
.02	.40	113	205
.05	.15	419	188
.07	.15	>1500	>1400

Реализуя второй подход, С. Фанер в [33] приводит соотношения, которые аналогичны (2-4), но дальше отходит от биномиальной модели и использует ее аппроксимацию нормальным распределением

$$\left. \begin{aligned} \Phi \left\{ \frac{n_0 - np_1}{\sqrt{np_1(1-p_1)}} \right\} &= 1 - \alpha; \\ \Phi \left\{ \frac{n_0 - np_2}{\sqrt{np_2(1-p_2)}} \right\} &= \beta \end{aligned} \right\}, \quad (5)$$

на основе которой получает аналитическое решение вида

$$n_0 = \left[ \frac{z_{1-\alpha} \sqrt{np_1(1-p_1)} + z_{1-\beta} \sqrt{np_2(1-p_2)}}{p_2 - p_1} \right]^2, \quad (6)$$

где  $z_{1-\alpha}, z_{1-\beta}$  - квантили стандартного нормального распределения, соответствующие  $(1-\alpha)$  и  $(1-\beta)$  соответственно.

Результаты аппроксимации биномиальной модели (4) нормальным распределением С. Фанер представляет в виде таблицы (см. табл. 3), в которой приводит результаты применения своего метода (расчет по формуле (6), в табл. 3 – столбец “ $n(6)$ ”) и метода Р.Р. Вилкокса (расчет по формуле (4), в табл. 3 – столбец “ $n(4)$ ”), а также оценки  $\Delta p = p_2 - p_1$  соответствующих зон неопределенности (интервалов безразличия).

Таблица 3

**Варианты минимальной длины теста [33]**  
(*Minimum Number of Items for  $\alpha = \beta = 0,05$* )

$p_1$	$p_2$	$p_2 - p_1$	$n(4)$	$n(6)$
0.30	0.70	0.40	13	9
0.35	0.65	0.30	19	17
0.40	0.60	0.20	43	40
0.45	0.55	0.10	–↑	163
0.55	0.95	0.40	13	6
0.60	0.90	0.30	19	12
0.65	0.85	0.20	43	29
0.70	0.80	0.10	–↑	121

–↑ The binomial table used restricted to  $n \leq 50$

Комментируя полученные результаты, С. Фанер отмечает, что для обоих методов “сужение интервала безразличия ведет к необходимости включения в тест неприемлемо большого числа заданий. Двигаясь в обратном направлении, мы для тестов приемлемой длины в 30–50 заданий получаем достаточно большую (около 0,2) ширину зоны безразличия” [33]. Сравнивая результаты (см. табл. 3) применения метода Р.Р. Вилкокса и своего метода, он приходит к выводу, что “аппроксимация биномиальной модели нормальным распределением дает несколько заниженные оценки длины теста, но имеемые различия практического значения не имеют” [33].

Хотя метод С. Фанера появился раньше и с позиций реализации выглядит привлекательнее, более широкое признание и распространение получил метод Р.Р. Вилкокса, обозначенный как “метод нейтральной зоны Вилкокса” [8]. Причина более широкого распространения этого метода состоит в том, что применение метода С. Фанера корректно не всегда, а только в определенных условиях: 1) при  $np > 5$ ,  $nq > 5$  и  $0,1 < p < 0,9$  [34]; 2) при  $npq > 9$  [35]; 3) при  $n > 100$  и  $npq > 20$  [36].

Однако, на наш взгляд, дело даже не в соблюдении конкретных условий, а в конфликте направленности эффективности рассматриваемого метода и целей, которые должны быть достигнуты в результате его применения. В общем случае нельзя признать корректным применение для решения задачи минимизации какого-либо параметра метода, эффективность которого снижается с уменьшением этого параметра. В рамках рассматриваемой задачи нас интересуют именно минимальные значения числа  $n$  тестовых заданий, т.е. те значения  $n$ , при которых примененный С. Фанером метод наименее всего работоспособен. Из этих же соображений бесполезным в отношении задачи минимизации длины теста представляется также метод А.И. Буравлева и В.Ю. Переверзева [37], которые в 1999 году повторили попытку С. Фанера в аппроксимации биномиального распределения нормальным при оценке оптимальной длины теста и надежности его результатов.

Следующий шаг в развитии подхода Д. Милмана - Р.Р. Вилкокса в 1980-82 годах сделали Р.А. Берк (*R.A. Berk*) и Ван дер Линден (*Wim J. van der Linden*), которые представили в [38,39] аналитическое решение (4).

Они обратили внимание на то, что “в рамках принятой биномиальной модели имеет место линейная зависимость между проходным баллом и соответствующей ему длиной теста” [39] и эта зависимость “может быть использована в процедуре оптимизации длины теста в качестве нового критерия” [39].

Перспективность применения такого нового критерия определялась тем, что, если раньше для обозначения проходного балла как критерия оценки результатов тестирования следовало указывать два вариативных показателя (число  $c$  набранных баллов и число  $n$  тестовых заданий), то теперь стало достаточно указать только один постоянный показатель — отношение  $c/n$ , величина которого остается постоянной при изменении длины теста (числа выполненных тестовых заданий). Было обосновано, что для такого критериального показателя справедливо соотношение

$$\frac{c_{1,2}}{n} = \frac{\ln \frac{1-p_1}{1-p_2}}{\ln \frac{p_2(1-p_1)}{p_1(1-p_2)}} + \frac{\ln \lambda_{1,2}}{n \ln \frac{p_2(1-p_1)}{p_1(1-p_2)}} \quad \text{при} \quad \begin{cases} \lambda_1 = \frac{\beta}{1-\alpha} & \text{для } c_1; \\ \lambda_2 = \frac{1-\beta}{\alpha} & \text{для } c_2 \end{cases}, \quad (7)$$

где:  $c_{1,2}$  — соответственно значения проходного балла для верхней ( $p_2$ ) и нижней ( $p_1$ ) границ зоны неопределенности (интервала безразличия),

Как отметил Ван дер Линден, это “соотношение интересно тем, что левая его часть представляет из себя относительную оценку проходного балла, а правая является константой, которая не зависит от длины теста и определяется только граничными значениями зоны безразличия. Всякий раз, когда устанавливаются границы зоны безразличия, мы получаем возможность узнать проходной балл для любой длины теста” [39].

Если выбрать для зоны безразличия параметры  $\alpha = \beta = 0,05$ ;  $p_0 = 0,5$ ;  $\delta = 0,1$  ( $p_1 = 0,4$ ;  $p_2 = 0,6$ ) из третьей строки табл. 3 и выполнить (7) для  $n = \overline{1,50}$ , зоны принятия решения вида “усвоил – не усвоил (зачет – не зачет, аттестован – не аттестован и т.п.)” примут вид, представленный на рис. 1.

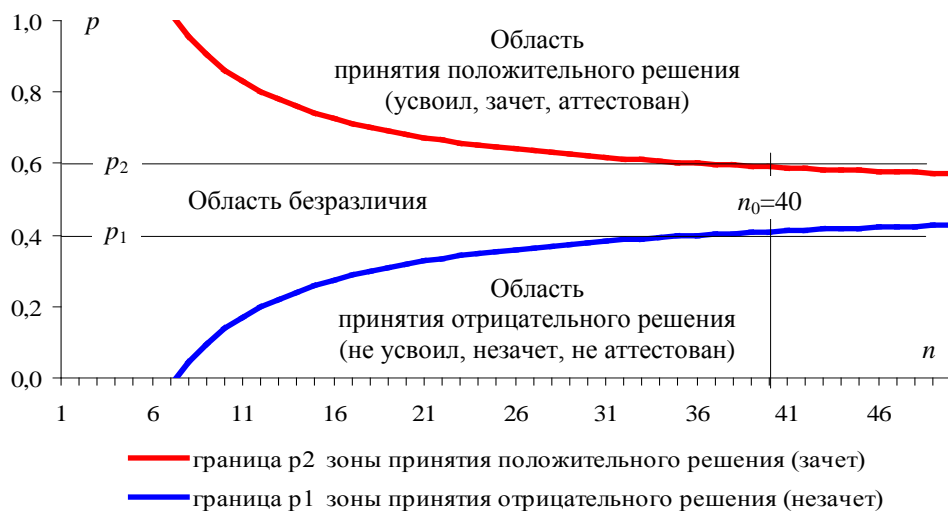


Рис. 1. Области принятия решения “усвоил – не усвоил (зачет- незачет)”



Графики, приведенные на рис. 1, в явном виде демонстрируют правоту Вандер Линдена. Разработанный в [38,39] метод оценки длины теста гораздо информативнее метода нейтральной зоны Вилкокса, поскольку дает возможность представить зоны безразличия и принятия решений в виде двухмерных областей, а не одномерных интервалов.

Двумерное представление рассматриваемых зон, в свою очередь, позволяет не только определять целесообразную длину теста, но и контролировать динамику изменения координат результата выполнения теста обучаемым относительно областей принятия решений, а также фиксировать момент перехода оценки результатов тестирования из области безразличия в одну из областей принятия решений. Последнее определило возможность реализации идеи индивидуального критериально-ориентированного тестирования [8,9,37,38,40].

Идея индивидуального критериально-ориентированного тестирования исходит из того, что “обязывать каждого обучаемого проходить полностью критериально-ориентированный тест не рационально, ... любой испытуемый должен продолжать тестирование до тех пор, пока не станет совершенно ясно, что его “истинный” уровень подготовленности (*true score*) выше или ниже критериального балла (стандарта оценивания)” [38]. Другими словами, цели индивидуального критериально-ориентированного тестирования направлены на выполнение требования оперативности текущего педагогического контроля путем “минимизации числа тестовых заданий, который должен выполнить испытуемый, за счет остановки процесса тестирования после выполнения очередного задания и получения текущих результатов, обеспечивающих принятие достоверного индивидуального решения “аттестован – не аттестован (зачет – не зачет)” [40]. Это позволяет определить индивидуальное критериально-ориентированное тестирование как метод управления длиной критериально-ориентированного теста путем ее минимизации на основе проводимой в режиме реального времени оценки достоверности результатов сравнения текущего и критериального значений принятых показателей вероятности успешного выполнения тестовых заданий.

В области нормативно-ориентированного тестирования сегодня широко обсуждаются процедуры адаптивного тестирования, которое также достигает свои цели путем минимизации числа предъявляемых обучаемому тестовых заданий. Поэтому представляется целесообразным разделить понятия и подходы адаптивного нормативно-ориентированного тестирования и индивидуального критериально-ориентированного тестирования.

Адаптивное тестирование определяется как “совокупность процессов генерации, предъявления и оценки результатов выполнения адаптивных тестов, обеспечивающая прирост эффективности измерений по сравнению с традиционным тестированием” [41]. “Тест адаптивный – это система стандартизованных заданий с известными (заданными) параметрами трудности и дифференцирующей способности. Очередность предъявления каждого задания зависит от успешности выполнения испытуемым предыдущего задания. При верном выполнении предыдущего задания – степень трудности последующего возрастает, и наоборот. ... Самая главная характеристика заданий адаптивного теста – это уровень их трудности” [42]. Л. Крокер (*Linda Croker*) и Дж. Алгина (*James Algina*) трактуют уровень трудности (трудность) тестового задания (*item difficulty*) как “основную статистическую характеристику тестового задания, определяемую долей испытуемых выборки, которые отвечают на это задание правильно” [43]. При этом они указывают, “что это — технический термин, который иногда кажется противоречащим общепринятой трактовке термина “трудность” [43].

Таким образом, в адаптивном тестировании “трудность задания” — это условное наименование некоторого заранее известного параметра, который используется для оптимизации процедуры выбора следующего тестового задания, но

не соответствует смысловому содержанию понятия “трудность” в педагогике и определяется заранее, а, значит, не характеризует деятельность испытуемых, проходящих тестирование.

Как средство текущего педагогического контроля критериально-ориентированное тестирование “направлено на формирование качественных знаний путем своевременного выявления пробелов для столь же своевременного их устранения” [10]. Потому оно должно предоставлять информацию о проблемах (трудностях), возникших при усвоении тестируемых объектов, в полном соответствии с представлениями педагогики.

В педагогике термин “трудность” обычно употребляется в смысле субъективной сложности усвоения изучаемого объекта (для обучаемого) или его предъявления (для преподавателя). Если в соответствии с таким подходом рассматривать результат выполнения каждого из тестовых заданий как адекватное последствие вполне определенного обучающего воздействия (приема или способа обучения) на обучаемого со вполне определенным уровнем обучаемости, то доля ошибок в выполнении тестовых заданий обучаемым, будет характеризовать его трудности в освоении тестируемых познавательных объектов. Соответственно доля обучаемых, которые выполнили определенное тестовое задание неправильно, будет характеризовать трудности и проблемы преподавателя в выборе того обучающего воздействия, которое было реализовано при изучении познавательного объекта, фигурирующего в рассматриваемом тестовом задании.

Таким образом, в сфере текущего педагогического контроля применение методов и моделей адаптивного тестирования некорректно ввиду того, что в адаптивном тестировании трудность задания известна заранее, а в индивидуальном критериально-ориентированном тестировании является целью его проведения. Поэтому все достижения адаптивного тестирования к индивидуализации процедуры критериально-ориентированного тестирования как средству обеспечения оперативности текущего педагогического контроля отношения не имеют.

Основанные на (7) идеи индивидуального критериально-ориентированного тестирования представлены в виде частных моделей [8,9,37,40] и алгоритма последовательного анализа результатов тестирования при дихотомическом оценивании типа “зачет - незачет” [38]. В 2012 году был проведен натурный эксперимент по оценке работоспособности этих моделей и алгоритмов.

Группе из 134 курсантов был предложен тест по знанию основных понятий навигации. Тест включал 50 закрытых заданий с выбором из предложенных 5 альтернатив единственного верного ответа. Каждому испытуемому задания предъявлялись в случайном порядке на КСТ.

После выполнения каждого тестового задания текущие индивидуальные результаты тестирования фиксировались в виде соответствующего модели (7) показателя безошибочности (частоты успеха)

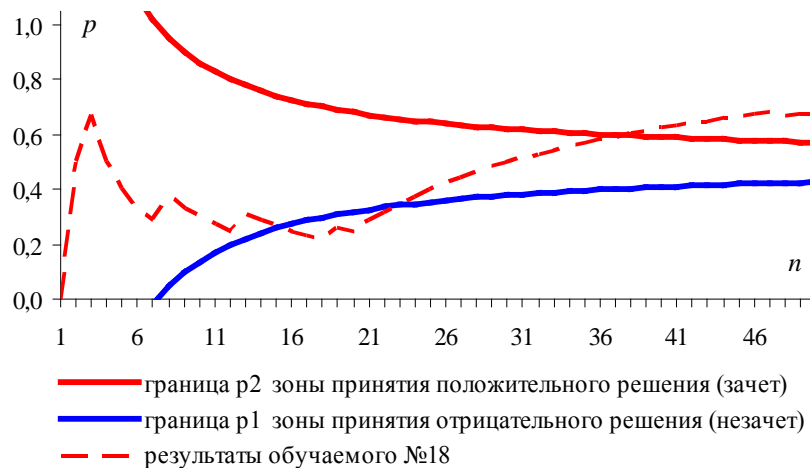
$$p_i^* = \frac{\sum_{i=1}^n x_i}{n}, \quad (8)$$

где:  $p_i^*$  - частота успешного выполнения задания;  $i$  ( $i = \overline{1, n}$ ) – номер тестового задания;  $n$  - число выполненных заданий,  $x_i$  ( $x_i = 0, 1$ ) результат выполнения одиночного задания по условию:  $x_i = 1$  – ошибки нет,  $x_i = 0$  – ошибка есть.

По завершении тестирования траектория полученных результатов наносилась на рассчитанные по модели (7) области безразличия и принятия решений. Расчет этих областей производился с параметрами  $\alpha = \beta = 0,05$ ;  $p_0 = 0,5$ ;  $\delta = 0,1$  ( $p_1 = 0,4$ ;  $p_2 = 0,6$ ), представленными выше на рис. 1.

В результате обработки выяснилось, что основанный на модели (7) алгоритм последовательного анализа результатов тестирования [38] для любой из траекторий абсолютно безошибочно фиксирует момент вхождения и выхода текущего результата тестирования из области безразличия в области принятия решений и наоборот. Другими словами, он вырабатывает сигналы остановки тестирования, который соответствует принятым областям принятия решений. Однако в 10 (7,5 %) исследованных ситуациях тестирования этот сигнал являлся преждевременным, а в 16 (11,9 %) ситуациях сигнал остановки тестирования не вырабатывался вообще.

В качестве типичного примера ситуации выработки преждевременного сигнала на остановку тестирования на рис. 2 представлена траектория результатов обучения обучаемого №18, фактически имевшая место в процессе эксперимента. Этот испытуемый в начале процедуры тестирования при выполнении заданий с 3 по 16 подряд допустил ряд ошибок, в результате которых текущее значение показателя его безошибочности после задания 15 вошло в область принятия отрицательного решения (незачет) и был выработан сигнал остановки тестирования. Сигнал перестал вырабатываться после выполнения задания 23. После выполнения задания 37 сигнал был выработан вновь, но уже как сигнал успешного прохождения теста.



**Рис. 2. Типичный вид неправильно диагностируемой траектории результатов тестирования**

По результатам эксперимента был сделан вывод, что подобные сбои в выработке сигналов на остановку процесса тестирования возникают в оцениваемом алгоритме из-за принятой модели (8) представления текущего значения успешности результатов тестирования. Эта модель адекватно оценивает сложившуюся ситуацию, но не способна обеспечить прогнозирование направления ее развития.

### Постановка и решение задачи

При анализе факторов, которые не учитывались в моделях и алгоритме последовательного анализа результатов тестирования [38] и могли влиять на достоверность решения об идентификации результатов тестирования, выяснилось, что в качестве такого фактора могла выступать только погрешность текущего значения  $p_i^*$  показателя безошибочности, оцениваемого по модели (8).

Было принято решение вместо точечной оценки (8) для расчета текущего значения успешности результатов тестирования использовать предложенную в [44] интервальную оценку  $\hat{p}_i$ , получаемую после выполнения каждого из текущих тестовых заданий путем решения задачи нелинейного программирования вида

$$\left. \begin{aligned} \hat{p} &= \arg \max_{p \in (0,1)} C_n^k p^{n-k} (1-p)^{k+1} \sum_{s=1}^{s=m} p^{h_s} \\ \text{при } p &\in (0,1); i = \overline{1, n}; k = \overline{1, n}; h = \overline{1, (n-k)}, \sum_{s=1}^{s=m} h_s = n-k \end{aligned} \right\}, \quad (9)$$

где:  $k$  – число заданий, выполненных с ошибкой,  $n$  – общее число заданий,  $n-k$  – число заданий, выполненных без ошибок;  $h_s$  – длина вектора непрерывного успеха [44], оцениваемая числом успешно выполненных подряд тестовых заданий.

Соответственно для определения доверительных интервалов оценок вида (9) использовалась формула расчета точного доверительного интервала (ДИ) биномиального распределения [44]

$$\left( p^- = \frac{(n-k)}{(n-k) + (k+1)F_{2(k+1), 2(n-k), 1-\epsilon/2}}, p^+ = \frac{(n-k+1)F_{2(n-k+1), 2k, 1-\epsilon/2}}{k + (n-k+1)F_{2(n-k+1), 2k, 1-\epsilon/2}} \right), \quad (10)$$

где  $p^-$ ,  $p^+$  – соответственно нижняя и верхняя границы ДИ;  $n$  – число испытаний,  $k$  – число ошибок, а  $F_{f,g,\alpha}$  – квантиль порядка  $\alpha$  распределения  $F$  с  $f, g$  степенями свободы;  $k$  – число заданий, выполненных с ошибкой,  $n$  – общее число заданий,  $n-k$  – число заданий, выполненных без ошибок.

Задача определения минимальной длины теста была представлена в традиционном виде как задача выбора такой длины теста, которая бы при сравнении достигнутой вероятности  $p$  успешного выполнения теста с критериальным значением  $p_0 \pm \delta$  обеспечивала заданную достоверность принятого решения.

В качестве условий выполнения требований этой задачи были приняты соотношения, представленные на рис. 3.

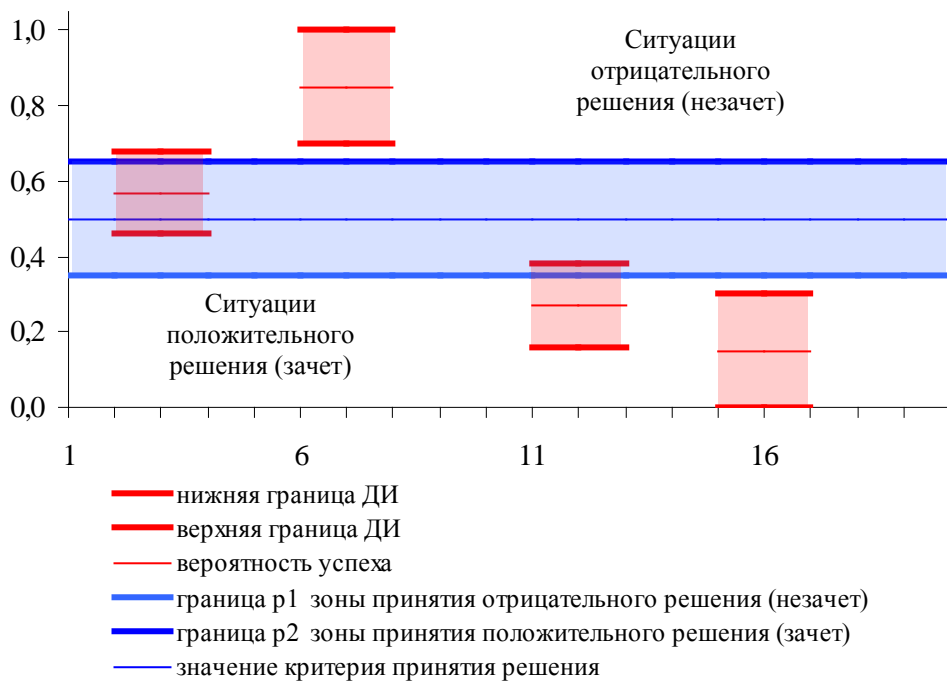


Рис. 3. Области принятия решения “усвоил – не усвоил (зачет- незачет)”

Эти соотношения определяют следующий порядок принятия решений:

1. Сигнал о принятии положительного решения и остановке тестирования вырабатывается в следующих ситуациях:

1) когда нижняя граница  $p_i^-$  ДИ оказывается выше верхней границы ( $p_0 + \delta$ ) области безразличия (см. рис. 3 ситуация 1), т.е. при выполнении условий вида

$$p_i^- > p_0 + \delta ; \quad (11)$$

2) когда ширина  $2\delta$  области безразличия оказывается больше ширины ( $p_i^+ - p_i^-$ ) ДИ, а верхняя граница  $p_i^+$  ДИ оказывается выше верхней границы ( $p_0 + \delta$ ) области безразличия (см. рис. 3 ситуация 1), т.е. при выполнении условий вида

$$\left. \begin{array}{l} p_i^+ - p_i^- \leq 2\delta; \\ p_i^+ > p_0 + \delta \end{array} \right\} ; \quad (12)$$

2. Сигнал о принятии отрицательного решения и остановке тестирования вырабатывается в следующих ситуациях:

1) когда верхняя граница  $p_i^+$  ДИ оказывается ниже нижней границы ( $p_0 - \delta$ ) области безразличия (см. рис. 3 ситуация 1), т.е. при выполнении условий вида

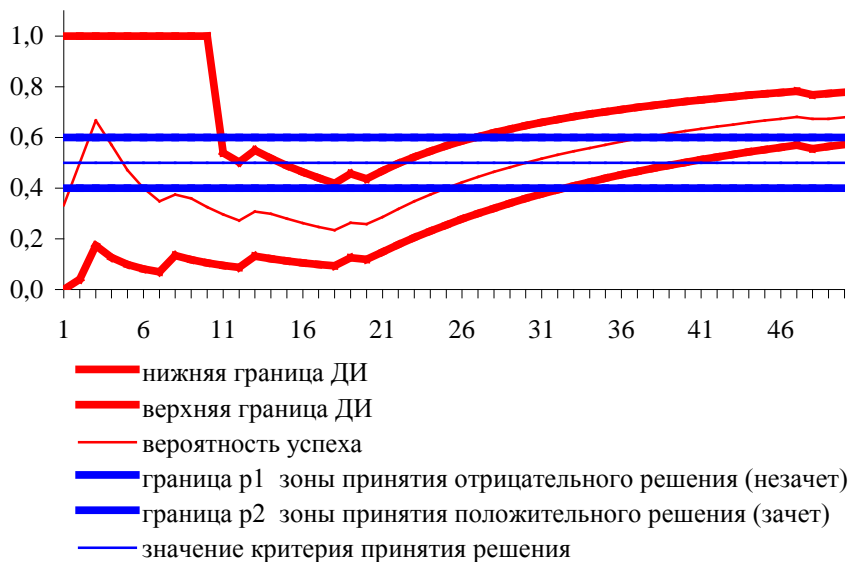
$$p_i^+ < p_0 - \delta ; \quad (13)$$

2) когда ширина  $2\delta$  области безразличия оказывается больше ширины ( $p_i^+ - p_i^-$ ) ДИ, а нижняя граница  $p_i^-$  ДИ оказывается ниже нижней границы ( $p_0 - \delta$ ) области безразличия (см. рис. 3.2), т.е. при выполнении условий вида

$$\left. \begin{array}{l} p_i^+ - p_i^- \leq 2\delta; \\ p_i^- < p_0 - \delta \end{array} \right\} ; \quad (14)$$

3. В остальных случаях вырабатывается сигнал продолжения тестирования.

Такой подход позволил устранить преждевременное принятие решения и выработку сигнала остановки тестирования во всех 10 ситуациях обучающей выборки. В частности, траектория результатов обучения испытуемого №18, представленная на рис. 2, приняла вид, приведенный на рис. 4.



**Рис. 2. Вид траектории результатов тестирования испытуемого №18 при оценке по моделям (9,10)**

В отношении этого испытуемого сигнал положительного решения по результатам тестирования (зачет) и остановка тестирования были осуществлены после выполнения задания 44, когда условия (11) были выполнены в следующем виде:

$$\left. \begin{aligned} p_i^+ - p_i^- &= 0,199 \leq 2\delta = 0,2; \\ p_i^+ &= 0,760 > p_0 + \delta = 0,6 \end{aligned} \right\}.$$

Таким образом, выполнение условий (11-14) при управлении длиной теста по моделям (9,10) обеспечило отсутствие случаев преждевременной остановки тестирования и принятия недостоверных решений, но не исключило необходимость выполнения всего теста и отсутствие решений в отношении 16 (11,9 %) ситуаций обучающей выборки, в которых решение по аттестованию испытуемых не принималось, а сигнал остановки тестирования не вырабатывался.

В отношении 118 (88,9%) случаев, по которым было принято решение и произведена остановка процесса тестирования, была произведена оценка предельных и вероятного значения модуля отклонения значений  $p_j^*$  вероятностей успеха, по которым принималось решение, от соответствующего “истинного” значения  $p_j^{50}$  (“истинного балла”). Результаты  $\Delta p_{\min} = 0 < \overline{\Delta p} = 0,086 < \Delta p_{\max} = 0,263$  позволяют сделать вывод о вполне удовлетворительных прогностических возможностях метода индивидуального критериально-ориентированного тестирования по моделям (9,10) и условиям (11-14).

В 2013 году рассматриваемый метод был проверен на независимой выборке из 87 курсантов на том же тесте и по той же методике. Здесь случаев преждевременной выработки сигнала остановки тестирования не наблюдалось. В 13 случаях решение по аттестованию испытуемых не было принято и остановка тестирования не производилась вплоть до полного выполнения теста. Оценка прогнозных возможностей метода производилась по данным 73 испытуемых и дала результаты  $\Delta p_{\min} = 0 < \overline{\Delta p} = 0,072 < \Delta p_{\max} = 0,211$ , сравнимые с обучающей выборкой.

На основании полученных оценок можно сделать вывод, что предлагаемый метод индивидуального критериально-ориентированного тестирования устраняет некоторые недостатки метода, представленного в [38]. Однако он пока не позволяет полностью отказаться от волюнтаристических решений в отношении траекторий результатов тестирования лежащих в области безразличия.

## Литература

1. Проект отраслевого терминологического стандарта Центра тестирования. Педагогические тесты. Термины и определения. – М.: Центр тестирования и развития “Гуманитарные технологии”, 2009. URL: <http://www.ege.ru/dict/dict1.htm> (дата обращения: 05.01.2015)
2. Войтов А.Г. Тестология гуманитариям. Теория и практика учебного тестирования. – М.: МГУ 2005. - 401с. URL: <http://www.twirpx.com/file/269589/> (дата обращения: 05.01.2015)
3. Ефремова Н.Ф. Тестовый контроль в образовании. - М.: Университетская книга, Логос, 2007. - 263 с. URL: <http://www.twirpx.com/file/1436309/> (дата обращения: 05.01.2015)
4. Ким В.С. Тестирование учебных достижений: Монография. - Уссурийск: Издательство УГПИ, 2007. - 214 с. URL: <http://www.twirpx.com/file/639729/> (дата обращения: 05.01.2015)

5. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов: учебник / пер. с англ. Н.Н. Найденовой, В.Н. Симкина, М.Б. Чельшковой; под общ. ред. В.И. Звонникова, М.Б. Чельшковой. - М.: Логос, 2010. - 668 с. URL: <http://www.twirpx.com/file/1346290/> (дата обращения: 05.01.2015)
6. Михайлычев Е.А. Дидактическая тестология. - М.: Народное образование, 2001. - 432 с. URL: <http://www.twirpx.com/file/276208/> (дата обращения: 05.01.2015)
7. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. - Москва: Прометей, 2000. - 168 с. URL: <http://bookre.org/reader?file=507195> (дата обращения: 05.01.2015)
8. Переверзев В.Ю. Критериально-ориентированные педагогические тесты для итоговой аттестации студентов. - М.: Изд-во НМЦ СПО Минобразования РФ, 1998. - 152с. URL: <http://www.twirpx.com/file/279031/> (дата обращения: 05.01.2015)
9. Переверзев В.Ю. Технология разработки тестовых заданий: справочное руководство - М.: Е-Медиа, 2005. - 272 с.
10. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. - М: Логос, 2002. - 432 с. URL: <http://www.twirpx.com/file/101903/> (дата обращения: 05.01.2015)
11. 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://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
12. Кузнецов А.В. Методика тестирования знаний и устранение случайных ошибок // Образовательные технологии и общество (Educational Technology & Society). 2007. Т. 10. № 1. - С. 271-275. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
13. 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://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
14. Рукшин С.Е., Поздняков С.Н. Технология дистанционной поддержки экспресс-олимпиад, построенных на оценке суждений // Образовательные технологии и общество (Educational Technology & Society). 2010. Т. 13. № 3. С. 374–386. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
15. Кузьмин Д.Н., Гриценко Е.М. Структурно-функциональная модель сетевого динамического тестирования на основе автомата. // Образовательные технологии и общество (Educational Technology & Society). 2011. Т. 14. № 3. С. 337-353. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
16. 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://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
17. 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://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).

18. Оганесян А.Г. Анатомия внешнего независимого оценивания знаний // Образовательные технологии и общество (Educational Technology & Society). 2012. Т. 15. № 4. С. 457-479. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
19. Колгатин А.Г., Колгатина Л.С. Вопросы качества процедур тестирования и интерпретации тестовых результатов в информационно-коммуникационной педагогической среде // Образовательные технологии и общество (Educational Technology & Society). 2013. Т. 16. № 1. С. 575-585. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
20. Осадчий В.В. Компьютерная система рейтингового оценивания знаний как средство повышения уровня знаний студентов // Образовательные технологии и общество (Educational Technology & Society). 2013. Т. 16. № 4. С. 361-371. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).
21. Туктаров Р.Р. Модели и процедуры оценки знаний в компьютерных системах тестирования ВМФ: дис. ... канд. техн. наук. - Новороссийск: ГМУ им. адмирала Ф.Ф. Ушакова, 2013. - 248 с.
22. Карпов В.Э., Карпова И.П. Язык описания системы контроля знаний // Компьютеры в учебном процессе, 2000, № 4, С. 147-155
23. Балыхина Т.М. Словарь терминов и понятий тестологии. – М.: РУДН, 2000. – 86с. URL: <http://www.psyoffice.ru/slovar-s174.htm> (дата обращения: 05.01.2015)
24. Коджаспирова Г.М., Коджаспиров А.Ю. Словарь по педагогике.- М.: ИКЦ «МарТ»; Ростов н/Д: Изд. центр «МарТ», 2005. - 448 с. URL: <http://www.twirpx.com/file/64113/> (дата обращения: 05.01.2015)
25. Талызина Н.Ф. Педагогическая психология. – М.: Изд. центр “Академия”, 1998. - 288с. URL: <http://www.persev.ru/book/pedagogicheskaya-psihologiya> (дата обращения: 05.01.2015)
26. Вишнякова С.М. Профессиональное образование словарь. Ключевые понятия, термины, актуальная лексика. – М.: НМЦ СПО, 1999. – 538с. URL: <http://mexalib.com/view/39189> (дата обращения: 05.01.2015)
27. Lord F. M., Novick M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.
28. Kriewall, Thomas Edward. Application of Information Theory and Acceptance Sampling Principles to the Management of Mathematics Instruction. Unpublished doctoral dissertation, University of Wisconsin, 1969.
29. Millman Jason, 1972, Determining test length: Passing scores and test length for objectives-based tests. Los Angeles: Instructional Objectives Exchange, 33 p.
30. Klauer, K. J. (1972). Zur Theorie und Praxis des binomialen Modells lehrzielorientierter Tests. In K.J. Klauer, R. Fricke, M. Herbig, H. Rupperecht & F. Schott (Hrsg.), Lehrzielorientierte Tests. Beiträge zur Theorie, Konstruktion und Anwendung. Düsseldorf: Schwann.
31. Kriewall, T.E. Aspects of applications of criterion-referenced tests. Illinois School Research, 9, 5-18, 1972.
32. Willcox, R.R. Determining the length of a criterion-referenced test. Applied Psychological Measurement, 1980, 4, pp. 425-446. URL: <https://conservancy.umn.edu/bitstream/handle/11299/100199/v04n4p425.pdf?sequence=1> (дата обращения: 05.01.2015)
33. Fhaner, S. Item sampling and decision-making in educational testing. British Journal of Mathematical and Statistical Psychology, 27, 172-175, 1974.
34. Эддоус М., Стэнсфилд Р. Методы принятия решений / Пер. с англ. под ред. член-корр. РАН И.И. Елисеевой. — М.: Аудит, ЮНИТИ, 1997. — 590 с. URL: <http://www.twirpx.com/file/31403/> (дата обращения: 05.01.2015)



35. Джонсон Н.Л. Одномерные дискретные распределения / Н.Л. Джонсон, С. Коц, А.У. Кемп; пер. 2-го англ. изд. - М.: БИНОМ, Лаборатория знаний, 2012. - 559с.
36. Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов. - 2-е изд., перераб. и доп. - М.: Юнити-Дана, 2004. - 573 с. URL: [http://www.e-joe.ru/sod/99/2\\_99/st160.html](http://www.e-joe.ru/sod/99/2_99/st160.html) (дата обращения: 05.01.2015)
37. Буравлев А.И., Переверзев В.Ю. Выбор оптимальной длины педагогического теста и оценка надежности его результатов // Открытое образование. 1999. №2. С. 23 -26. URL: [http://www.e-joe.ru/sod/99/2\\_99/st160.html](http://www.e-joe.ru/sod/99/2_99/st160.html) (дата обращения: 05.01.2015)
38. Berk R.A. Criterion-referenced measurement: The state of art, Baltimor,MD: Johns Hopkins University Press, 1980.
39. Wim J. van der Linden. (1982). Passing score and length of a mastery test. Evaluation in Education. 1982, Vol. 5, pp. 149-164. URL: <http://doc.utwente.nl/68993/1/Linden82passing.pdf> (дата обращения: 05.01.2015)
40. Переверзев В.Ю. Моделирование процесса критериально-ориентированного педагогического тестирования курсантов вузов: Дис. ... канд. пед. наук. – М.: Военный университет, 2000. - 198с.
41. Чельшкова М.Б. Адаптивное тестирование в образовании (теория, методология, технология). - М: Исследовательский центр проблем качества подготовки специалистов, 2001. - 165 с.
42. Морев И.А. Образовательные информационные технологии. Часть 2. Педагогические измерения: Учебное пособие. – Владивосток: Изд-во Дальневост. ун-та, 2004. – 174 с.
43. Крокер Л. Введение в классическую и современную теорию тестов: учебник / Л. Крокер, Дж. Алгина; пер. с англ. Н.Н. Найденовой, В.Н. Симкина, М.Б. Чельшковой; под общ. ред. В.И. Звонникова, М.Б. Чельшковой. – М.: Логос, 2010. – 668 с. URL: <http://www.twirpx.com/file/1346290/> (дата обращения: 05.01.2015)
44. Печников А.Н., Палкин К.С. Метод интервальной оценки результатов выполнения системы одиночных тестовых заданий закрытого типа с единственным верным ответом // Образовательные технологии и общество (Educational Technology & Society). 2014. Т. 17. № 2. С. 491-501. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 05.01.2015).