

DOI: 10.21821/2309-5180-2017-9-3-663-673

MODEL MANAGEMENT OF THE PROCESS CRITERIA-ORIENTED TESTING AT TRAINING OF SPECIALISTS IN THE NAVY

D. A. Pechnikov

Naval Academy named after Admiral of the Fleet of the Soviet Union N.G. Kuznetsov,
St. Petersburg, Russian Federation

The article discusses the model of management of the implementation of the system of criterion-oriented test tasks with a single correct answer, that provides a stop test when the required reliability of the evaluation. The development and application proposed in this article models of assessment results the criterion-oriented test provides a transition from the point estimates of the empirical relative frequency to estimate probability of correct actions that have quantifiable confidence interval (CI). The latter provides the application of these models to solve the problem of automation of management of process criteria-oriented testing in order to ensure its operational efficiency by stopping the procedure for making the current of reference (TOR) after receiving the evaluation of a specified accuracy. The article also discusses the use of models of criterion-oriented evaluation in computer systems testing, it is proved that these models improve the results of current control. In order to carry out military training as a civilian Navy and the Navy requires appropriate training facilities. The basis of such a base used was the existing training samples of civil and military equipment and weapons, which were previously reported in the naval and civil Maritime educational institutions in sufficient quantity and provide the necessary level of theoretical and practical training of future specialists in the fleet. The result is a sharp reduction in the 1990s of the branches of the military-industrial complex, developing equipment and weapons for the fleet, the latest models of this equipment and weapons are now being manufactured not series, but individual instances. The production of educational and training samples for Maritime training institutions is not necessary at all. In these circumstances, as the only means to ensure that the professional military and the naval-training preparation, can only be considered virtual analogs of models of military and civilian equipment and weapons developed by means of information technology, i.e. naval training simulation systems (TSS). The development and application proposed in this article models of interval estimation results of Equator-based testing provides a transition from the point estimates of the relative frequency to estimate probability of correct actions that have the amount directly estimate DEE. The latter provides the application of these models to solve the problem of automation of process criteria-oriented testing in order to ensure its efficiency by stopping the procedure for making TK after radiation evaluating a given accuracy.

Keywords: criterion-based test; a test task; the test results; point estimate; interval estimate; confidence interval; the given error, the marines.

For citation:

Pechnikov, Denis A. "Model management of the process criteria-oriented testing at training of specialists in the navy." *Vestnik Gosudarstvennogo universiteta morskogo i rechnogo flota imeni admiral S. O. Makarova* 9.3 (2017): 663–673. DOI: 10.21821/2309-5180-2017-9-3-663-673.

УДК 371.3:681.51

МОДЕЛИ УПРАВЛЕНИЯ ПРОЦЕССОМ КРИТЕРИАЛЬНО-ОРИЕНТИРОВАННОГО ТЕСТИРОВАНИЯ ПРИ ПОДГОТОВКЕ СПЕЦИАЛИСТОВ НА ФЛОТЕ

Д. А. Печников

Военно-морская академия имени Адмирала Флота Советского Союза Н. Г. Кузнецова,
Санкт-Петербург, Российская Федерация

В статье рассматриваются модели управления процессом выполнения системы одиночных критериально-ориентированных тестовых заданий с единственным верным ответом, обеспечивающие остановку тестирования при достижении необходимой достоверности оценки его результатов в процессе подготовки специалистов для флота. Разработка и применение предлагаемых в данной статье моделей оценивания

результатов критериально-ориентированного тестирования обеспечивает переход от точечных оценок эмпирической частоты к оценкам вероятности правильного действия, которые имеют количественно оцениваемый доверительный интервал. Последнее обеспечивает применение этих моделей для решения задачи автоматизации управления процессом критериально-ориентированного тестирования в целях обеспечения его оперативности путем остановки процедуры предъявления текущего задания после получения оценки заданной достоверности. Также в статье рассматривается применение моделей критериально-ориентированного оценивания в компьютерных системах тестирования; доказано, что данные модели повышают результаты текущего контроля. Для того, чтобы осуществлять профессиональную подготовку специалистов как гражданского морского флота, так и ВМФ, необходима соответствующая учебно-материальная база. Основу такой базы раньше всегда составляли действующие учебные и тренировочные образцы гражданской и военной техники и вооружения, которые ранее поступали в военно-морские и гражданские морские учебные заведения в достаточном количестве и обеспечивали необходимый уровень как теоретической, так и практической подготовки будущих специалистов флота. В результате резкого сокращения в 1990 гг. отраслей военно-промышленного комплекса, разрабатывающих технику и вооружение для флота, новейшие образцы этой техники и вооружения теперь изготавливаются не сериями, а отдельными экземплярами. Вопрос о производстве учебных и тренировочных образцов для морских учебных заведений не стоит вообще. В этих условиях в качестве единственного средства, способного обеспечить военно-профессиональную и морскую подготовку, могут рассматриваться только виртуальные аналоги образцов военной и гражданской техники и вооружения, разрабатываемые средствами информационных технологий, т. е. флотские тренажерно-обучающие системы, что указано в международных морских конвенциях.

Ключевые слова: критериально-ориентированный тест; тестовое задание; результаты тестирования; точечная оценка; интервальная оценка; доверительный интервал; заданная погрешность, морской флот.

Для цитирования:

Печников Д. А. Модели управления процессом критериально-ориентированного тестирования при подготовке специалистов на флоте / Д. А. Печников // Вестник Государственного университета морского и речного флота имени адмирала С. О. Макарова. — 2017. — Т. 9. — № 3. — С. 663–673. DOI: 10.21821/2309-5180-2017-9-3-663-673.

Введение

Под контролем качества тестирования обычно понимается совокупность мероприятий, проводимых над объектом тестирования в процессе разработки для постоянного получения исчерпывающей информации о соответствии объекта тестирования поставленным требованиям [1], [2]. В качестве перспективного методологического средства реализации контроля при подготовке специалистов морского транспорта и ВМФ сегодня рассматриваются методы тестологии, а в качестве технического средства — компьютерные системы тестирования (КСТ) при подготовке специалистов для флота. Выделяют входной, текущий и итоговый виды контроля.

Входной и итоговый контроль предваряют и завершают учебный процесс подготовки морских специалистов. Его результаты могут и должны учитываться при проектировании обучения как судоводителей, так и курсантов других морских специальностей, но они не ориентированы на корректуру уже начавшегося процесса. Эти виды контроля могут быть реализованы средствами как нормативно-ориентированного, так и критериально-ориентированного тестирования в подготовке специалистов для морского транспорта. В отличие от них, *текущий* контроль реализуется непосредственно в процессе обучения морским специальностям, а его важнейшей функцией является функция обратной связи, которая позволяет преподавателю получать сведения о ходе процесса усвоения у каждого курсанта и адаптировать процесс обучения к их потребностям и будущей морской специализации [3]. Адаптировать процесс морского образования и конкретно обучения будущих специалистов флота к потребностям обучаемых на основе данных нормативно-ориентированного тестирования невозможно. Для этого нужны оценки, идентифицирующие соотношения вида «усвоил — не усвоил» между теми конкретными обучаемыми и теми конкретными познавательными объектами, в отношении которых реализуется рассматриваемый процесс обучения в морском вузе. Поэтому для решения задач текущего контроля при подготовке специалистов мор-

ского и речного транспорта могут использоваться только методы критериально-ориентированного тестирования.

Между уровнями развития методов и технических средств критериально-ориентированного тестирования имеет место противоречие. В числе морских КСТ, представленных в Интернете, преобладают программные средства, ориентированные на решение задач критериально-ориентированного тестирования (Adit Testdesk, М-Тест, EasyQuizzy, MultiTester System, RSP, The Examiner testing system, FastTEST professional, C-Quest, CONTEST, SunRav TestOfficePro, HyperTest, UniTest System, TestMaster, AVELife TestGold Studio, ACT-Тест Plus, e-University, RichTest, Tester, Грамотей-КЛАСС, «Контроль знаний», «Система проверки знаний», «Универсальный тестовый комплекс», «Экзаменатор», «Аттестация», УСАТИК и т. д.). Преобладание критериально-ориентированных КСТ обусловлено тем, что для контроля уровня обученности более привычной и естественной является задача, решаемая в рамках критериально-ориентированного тестирования [3]. Однако именно этот вид тестирования не имеет развитых методов оценивания полученных результатов.

Методы классической теории тестов и современной тестовой теории IRT (Item Response Theory) [4] преимущественно ориентированы на ранжировку обучаемых по уровню обученности вне зависимости от достаточности или недостаточности достигнутого ими уровня усвоения рассматриваемой предметной области. Вопросы проектирования и оценки результатов выполнения критериально-ориентированных тестов, устанавливающих соотношение «усвоил – не усвоил» между обучаемым и конкретной совокупностью элементов содержания обучения, решаются по остаточному принципу. В результате «... отсутствие стандарта или хотя бы общепринятых норм и правил создания компьютерных систем тестирования для подготовки морских специалистов 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» [5, с. 59].

Ориентированность на решение задач текущего контроля предъявляет к методам и средствам критериально-ориентированного тестирования в морском образовании ряд специфических требований. В числе этих требований одним из ведущих является требование оперативности. Задачи текущего контроля при подготовке специалистов на флоте имеют целью выработку адресных обучающих воздействий и в идеале должны решаться в реальном масштабе времени непосредственно в процессе обучения и подготовки. Поэтому в соответствии с требованиями международных морских конвенций процесс критериально-ориентированного тестирования должен занимать минимум времени. Оперативность процедур обработки и отображения результатов действий обучаемых проблем не вызывает, поскольку полностью обеспечивается за счет быстрого действия аппаратно-программных средств КСТ, а вот оперативность самой процедуры тестирования в современных КСТ не обеспечивается.

Эти обстоятельства определяют необходимость разработки и внедрения в программное обеспечение КСТ, используемых для подготовки специалистов гражданского флота, методов и процедур, которые непосредственно направлены на выполнение требования обеспечения оперативности тестирования. Другими словами, для полноценного достижения целей текущего контроля при подготовке специалистов морских специальностей необходимо минимизировать продолжительность процедуры критериально-ориентированного тестирования, обеспечивая при этом заданную достоверность получаемых результатов.

Методы и материалы

Время выполнения теста определяется как произведение $T = tn$ двух параметров: 1) t — скорость теста (test speededness) [2], оцениваемая временем выполнения одного тестового задания (ТЗ), которое для заданий закрытой формы обычно составляет 30 – 40 с на ТЗ; 2) n — длина теста (test length), измеряемая числом ТЗ в тесте. Скорость тестирования обусловлена психофизиологическими возможностями обучаемых, а потому не может быть существенно сокращена. Сокра-

цена может быть только длина теста. Поэтому выполнение требования оперативности процедур критериально-ориентированного тестирования при подготовке специалистов в морских учебных заведениях связано, прежде всего, с минимизацией числа тестовых заданий, которые 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 – незачет»). На минимизацию числа выполняемых ТЗ ориентирован ряд методов и моделей, которые, по своей сути, базируются на двух подходах.

Суть *первого подхода* состоит в завершении процедуры тестирования после того, как соотношение «больше» или «меньше» между оценкой результатов тестирования и принятым критерием станет определяться с заданной (достаточной) достоверностью. В рамках этого подхода сформулирована задача выбора такой длины теста, которая бы при сравнении достигнутой вероятности p успешного выполнения теста с критериальным значением p_0 этой вероятности обеспечивала исключение заданных величин ошибок 1-го рода (α), когда знающий материал обучаемый получает отрицательный результат (незачет), и 2-го рода (β), когда незнающий материал обучаемый получает зачет. Наиболее совершенный вариант решения этой задачи был предложен в 1980 – 1982 гг. Р. А. Берком [6] и Ван дер Линденом [7], которые обратили внимание на то, что в рамках принятой биномиальной модели имеет место линейная зависимость между проходным баллом и соответствующей ему длиной теста, и эта зависимость может быть использована в процедуре оптимизации длины теста в качестве нового критерия. Для такого показателя справедливо соотношение вида [11]:

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

где $c_{1,2}$ — соответственно значения проходного балла для верхней (p_2) и нижней (p_1) границ зоны неопределенности (интервала безразличия); α, β — соответственно значения значимости ошибок 1 и 2-го рода принимаемых решений.

Как отметил Ван дер Линден, «... это соотношение интересно тем, что левая его часть представляет собой 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» [7, с. 149–164]. Суть этого положения демонстрируется на рис. 1, где приведены зоны принятия решения вида «усвоил – не усвоил» («зачет – незачет», «аттестован – не аттестован» и т. п.) для области безразличия с параметрами $\alpha = \beta = 0,05$; $p_1 = 0,4$; $p_2 = 0,6$.

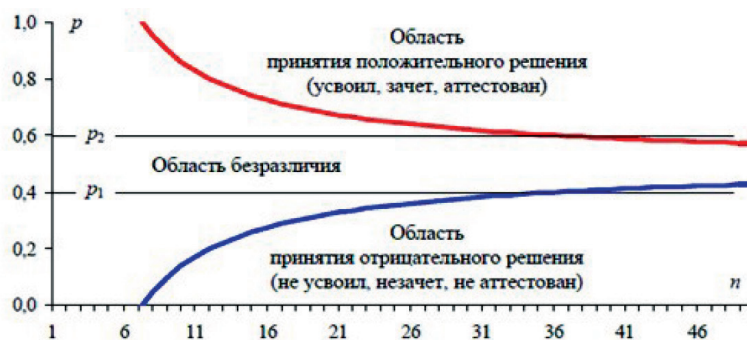


Рис. 1. Области принятия решений вида «аттестован – не аттестован»

Суть *второго подхода*, который частично представлен в [8], состоит в прекращении тестирования после того, как ширина доверительного интервала (ДИ) оценки результатов тестирования

станет меньше некоторого наперед заданного нормативного значения. Здесь задача выбора длины теста формулируется как задача прекращения тестирования после того, как фактическая ширина ДИ последней оценки p успешного выполнения теста становится меньше ее наперед заданного значения. Если ТЗ теста формулируются в закрытой форме с единственным правильным ответом, то в соответствии с [11], [12] для решения задачи минимизации длины теста на основе данных о погрешности полученных оценок результатов тестирования могут быть использованы две модели.

Первая модель, основанная на биномиальном законе распределения, имеет вид [11]:

$$\left. \begin{aligned} \hat{p}_i &= \frac{i-k}{i}; \\ p_i^{\text{верх}} &= \frac{(i-k+1)F_{2(i-k+1), 2k, 1-\varepsilon/2}}{k+(i-k+1)F_{2(i-k+1), 2k, 1-\varepsilon/2}}; \\ p_i^{\text{низ}} &= \frac{(i-k)}{(i-k)+(k+1)F_{2(k+1), 2(i-k), 1-\varepsilon/2}} \end{aligned} \right\}, \quad (2)$$

где $i (i = \overline{1, n})$ — число выполненных ТЗ; n — число ТЗ в тесте; $k (k = \overline{0, i})$ — число ошибок; $\hat{p}_i (p \in (0, 1))$ — оценка вероятности успеха после выполнения i -го ТЗ; $p_i^{\text{верх}}, p_i^{\text{низ}}$ — соответственно верхняя и нижняя границы ДИ вероятности успеха после выполнения i -го ТЗ; $F_{f, g, \alpha}$ — квантиль порядка α распределения F с f, g степенями свободы.

Вторая модель базируется на геометрическом законе распределения и в [11] представлена в следующем виде:

$$\left. \begin{aligned} \hat{p}_i &= \arg \max_p \left[p^{(i-k)+(m-l)} (1-p)^{k+l} \right]; \\ p_i^{\text{верх}} &= \hat{p}_i + u_{1-\varepsilon/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{i}}; \\ p_i^{\text{низ}} &= \hat{p}_i - u_{1-\varepsilon/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{i}} \end{aligned} \right\}, \quad (3)$$

где $(i-k) ((i-k) = \overline{0, i})$ — число правильно выполненных ТЗ; $l, (m-l)$ — соответственно число векторов непрерывного успеха и непрерывной неудачи ($m+l \leq i$); $\hat{p}_i (p \in (0, 1))$ — результат тестирования после выполнения i -го ТЗ (оценка вероятности успеха, получаемая по методу правдоподобия); $L_i = p^{(i-k)+(m-l)} (1-p)^{k+l}$ — функция правдоподобия; $p_i^{\text{верх}}, p_i^{\text{низ}}$ — соответственно верхняя и нижняя границы ДИ вероятности успеха после выполнения i -го ТЗ; $u_{1-\varepsilon/2}$ — квантили стандартного нормального распределения порядка $1 - \varepsilon / 2$.

Оценки (2) и (3) в достаточной мере согласованы. Экспериментально выявленная в [11] динамика этих оценок в общем случае имеет вид, приведенный на рис. 2.

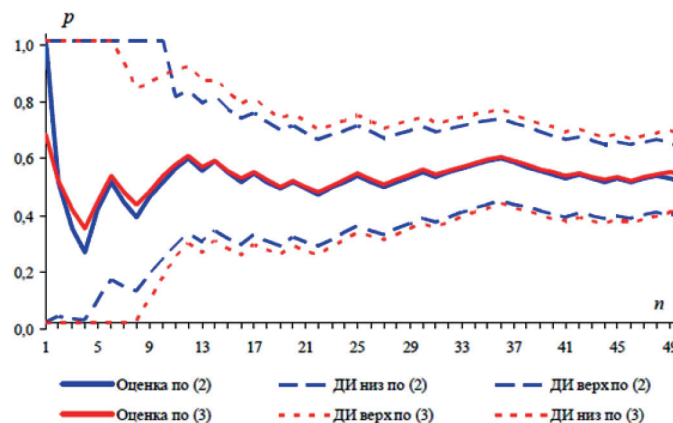


Рис. 2. Динамика изменения результатов тестирования среднеуспешного обучаемого (результат оценивания $0,376 < p = 0,500 < 0,624$ при $\alpha = 0,1$)

Кроме уже рассмотренных различий между первым и вторым подходами, есть еще одно, на которое следует обратить внимание. Это различие состоит в том, что первый подход реализует процедуры измерения и оценивания, а второй подход — только процедуру измерения. Суть этого отличия требует определенного комментария. Понятия «оценивание» и «измерение» формулируются в двух предметных областях квалитологии — в метрологии и в квалиметрии.

В метрологии измерение рассматривается как частный случай оценивания: 1) измерение — сравнение измеряемой величины (конкретного проявления измеряемого свойства) со шкалой этого свойства в целях получения результата измерения (оценки свойства или значения величины) [9]; 2) оценивание — «аналог измерения, применяемый в тех случаях, когда невозможно выполнить измерение (не выделена величина как физическая и не определена единица измерений этой величины) [9, с. 121].

В квалиметрии приняты другие трактовки: 1) измерение — процедура определения единичного показателя качества [10]; 2) оценивание — процедура определения численного значения уровня качества продукции [10]. Единичный показатель качества — это количественная характеристика одного из свойств продукции [13], а уровень качества — относительная характеристика качества, основанная на сравнении значений показателей качества оцениваемого объекта с базовыми значениями соответствующих показателей [13]. Из последних дефиниций следует, что в квалиметрии понятия оценивания и измерения рассматриваются как несоподчиненные.

Эталоны, используемые при измерении и оценивании, также имеют отличия: в оценивании эталон имеет ценностный, системно-социальный характер и меняется вместе с социально-экономическим и научно-техническим прогрессом, т. е. имеет переходящее существование. В измерении эталон постоянен и определяет шкалы измерения единичных показателей качества [10]. Другими словами, в квалиметрии считается, что любое свойство нужно сначала измерить (получить его объективную оценку в некоторой неизменной шкале), а затем оценить полученный результат [14], т. е. сравнить его с тем значением рассматриваемого свойства, которое принято в качестве эталона, и сформулировать свое отношение к результатам этого сравнения.

Необходимость различения первого и второго подходов по признаку «измерение – оценивание» определяется тем фактом, что во всех КСТ принят квалиметрический подход [15]: 1) функции измерения и оценивания разделены; 2) функция измерения реализуется всегда одинаково в одной и той же шкале; 3) параметры оценивания определяются разработчиком теста (преподавателем), который сам выбирает эталоны и шкалы их оценивания. Далее следует отметить, что модели (2) и (3) являются интервальными, а потому обеспечивают сравнение с эталонами как самих результатов тестирования, так и их погрешностей. Другими словами, они способны обеспечить реализацию обоих рассмотренных подходов.

С учетом последнего, всю процедуру управления критериально-ориентированным тестированием при подготовке специалистов для флота можно представить в виде, приведенном на рис. 3. Эта процедура предполагает, что после выполнения испытуемым каждого ТЗ проводится проверка текущего результата тестирования на его соответствие принятому критерию управления. Если такое соответствие имеет место, то тестирование прекращается. В противном случае испытуемый, вплоть до выполнения всего теста, получает следующее ТЗ. В любом случае в качестве окончательного результата тестирования принимается последняя текущая оценка.

Различие в процедуре реализации рассмотренных подходов к минимизации длины теста сводится к различию в принятых критериях управления.

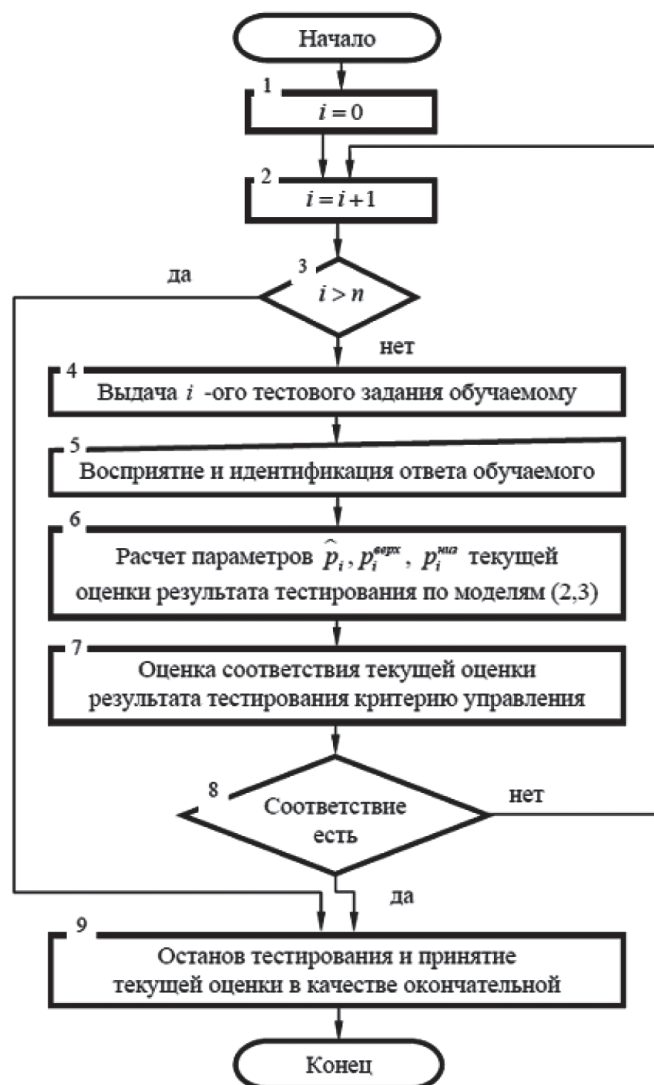


Рис. 3. Процедура управления критериально-ориентированным тестированием при подготовке специалистов для флота

Для реализации *первого* из рассматриваемых подходов вместо расчетов по модели (1) предлагается после выполнения обучаемым каждого ТЗ по моделям (2) и (3) определять верхнюю границу $p_i^{\text{верх}}$ и нижнюю границу $p_i^{\text{низ}}$ ДИ оценки $\hat{p}_i$ и использовать эти данные для выработки управляющих воздействий по критерию

$$I_i = \begin{cases} 1 & \text{при } \begin{cases} p_i^{\text{низ}} > p_0; \\ p_i^{\text{верх}} < p_0, \end{cases} \\ 0 & \text{в ином случае} \end{cases} \quad (4)$$

где I_i — результаты сравнения верхней $p_i^{\text{верх}}$ и нижней $p_i^{\text{низ}}$ границ ДИ текущей оценки $\hat{p}_i$ с нормативным значением p_0 результата тестирования ($I_i = 1$ — останов тестирования; $I_i = 0$ — продолжение тестирования).

Критерий (4) определяет наличие у текущего результата $\hat{p}_i$ требуемой достоверности, но функцию его оценивания не реализует. Формулировка в системе предпочтений преподавателя отношения к значению $\hat{p}_i$ («зачет – незачет») здесь не производится, поскольку функция оценивания во всех КСТ (пример на рис. 4) реализуется отдельно, причем не только в дихотомической, но и в других шкалах.



Рис. 4. Окно задания параметров процедуры оценивания в КСТ «Адаптивная система тестирования» (АСТ)

В соответствии со *вторым* подходом процесс тестирования должен быть остановлен после получения текущего результата $\hat{p}_i$ заданной точности $\delta^{\text{зад}}$, когда границы ДИ начинают располагаться внутри нормативного интервала ($\hat{p}_i \pm \delta^{\text{зад}}$), т. е. имеет место соотношение $(p_i^{\text{верх}}, p_i^{\text{низ}}) \in (\hat{p}_i + \delta^{\text{зад}}, \hat{p}_i - \delta^{\text{зад}})$. В этом случае остановку процесса тестирования следует осуществлять по критерию

$$I_i = \begin{cases} 1 & \text{при выполнении условий} \begin{cases} p_i^{\text{верх}} < \hat{p}_i + \delta^{\text{зад}}; \\ p_i^{\text{низ}} > \hat{p}_i - \delta^{\text{зад}} \end{cases} \\ 0 & \text{в ином случае} \end{cases}, \quad (5)$$

где I_i результат сравнения заданной и достигнутой точности оценок тестирования ($I_i = 1$ — останов тестирования; $I_i = 0$ — продолжение тестирования).

Результаты

Группе, состоящей из 129 курсантов-судоводителей (испытуемых) был предложен тест, который включал 50 закрытых ТЗ с выбором из пяти альтернатив единственного верного ответа. ТЗ предъявлялись в случайном порядке. Воздействия по управлению процессом тестирования по моделям (4) и (6) вырабатывались и фиксировались апостериори. Для расчета управляющих воздействий были приняты следующие модели: 1) модели (2) и (3) с параметром $\alpha = 0,05$; 2) критерий (4) с параметром $p_0 = 0,5$; 3) критерий (5) с параметрами $\delta_1^{\text{зад}} = 0,1$; $\delta_2^{\text{зад}} = 0,15$; $\delta_3^{\text{зад}} = 0,2$. В табл. 1 приведены характеристики выработанных воздействий.

Таблица 1

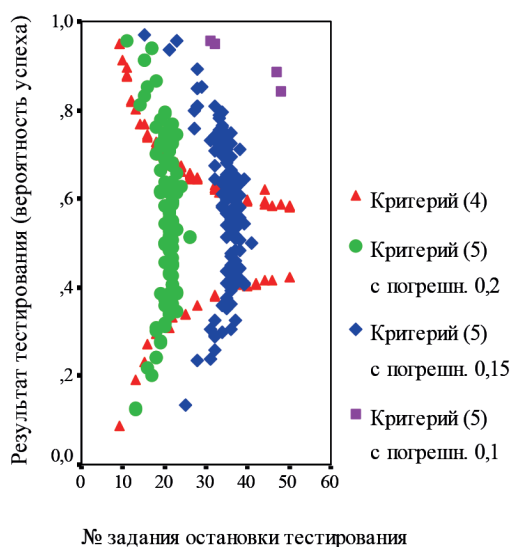
Результаты применения критериев управления тестированием

Результаты применения способа управления	Управление по критерию (5) с допустимой погрешностью			Управление по критерию (4)
	$\delta^{\text{зад}} = 0,1$	$\delta^{\text{зад}} = 0,15$	$\delta^{\text{зад}} = 0,2$	
Число случаев останова тестирования	5	129	129	87
Процент случаев останова тестирования	3,9 %	100,0 %	100,0 %	67,4 %
Среднее отклонение результата от «истинного»	0,005	0,056	0,091	0,037

Наглядно результаты применения критериев (4) и (5) представлены на рис. 5.

Для оценки эффективности альтернативных критериев управления (4) и (5) использовались показатели эффективности, представляющие собой измеряемые на интервале (0;1) показатели вида «чем больше, тем лучше»:

1) частный показатель $k_{ij}^{(1)}$ оперативности управления



№ задания остановки тестирования
Рис. 5. Взаимосвязь результатов тестирования и длины теста при управлении по критериям (4, 5)

$$k_j^{(1)} = 1 - \frac{i_j^{\min}}{n} \quad k_j^{(1)} \in [0, 1], \quad (6)$$

где $j (j = \overline{1, m})$ — номер испытуемого; $i_j^{\min}$ — номер последнего ТЗ, выполненного j -м испытуемым (номер ТЗ, после выполнения которого тестирование было прекращено); n — длина теста (число ТЗ в тесте);

2) частный показатель $k_j^{(2)}$ точности оценки результатов тестирования

$$k_j^{(2)} = 1 - \Delta p_{i_j^{\min}} = 1 - |p_{i_j^{\min}} - p_j^*|, \quad k_j^{(2)} \in [0, 1], \quad (7)$$

где $j (j = \overline{1, m})$ — номер испытуемого; $i_j^{\min}$ — номер последнего выполненного ТЗ, $p_{i_j^{\min}}^*$ — окончательный результат тестирования; p_j^* ($p_j = p_{nj}$) — истинный балл (true score) — результат, полученный после выполнения всех ТЗ теста;

3) интегральный показатель K_j эффективности управления процессом тестирования

$$K_j = k_j^{(1)} k_j^{(2)} = \left(1 - \frac{i_j^{\min}}{n}\right) (1 - \Delta p_{i_j^{\min}}), \quad K_j \in [0, 1]. \quad (8)$$

В табл. 2 приведены обобщенные результаты оценки эффективности критериев (4) и (5) управления тестированием.

Таблица 2

Показатели эффективности критериев управления тестированием

Показатель эффективности	Управление по критерию (5) с допустимой погрешностью			Управление по критерию (4)
	$\delta^{\text{зад}} = 0,1$	$\delta^{\text{зад}} = 0,15$	$\delta^{\text{зад}} = 0,2$	
Оперативности управления (6)	0,007	0,313	0,593	0,313
Точности управления (7)	0,995	0,944	0,909	0,963
Интегральный (8)	0,006	0,294	0,538	0,292

Наглядно эти результаты представлены на рис. 6.

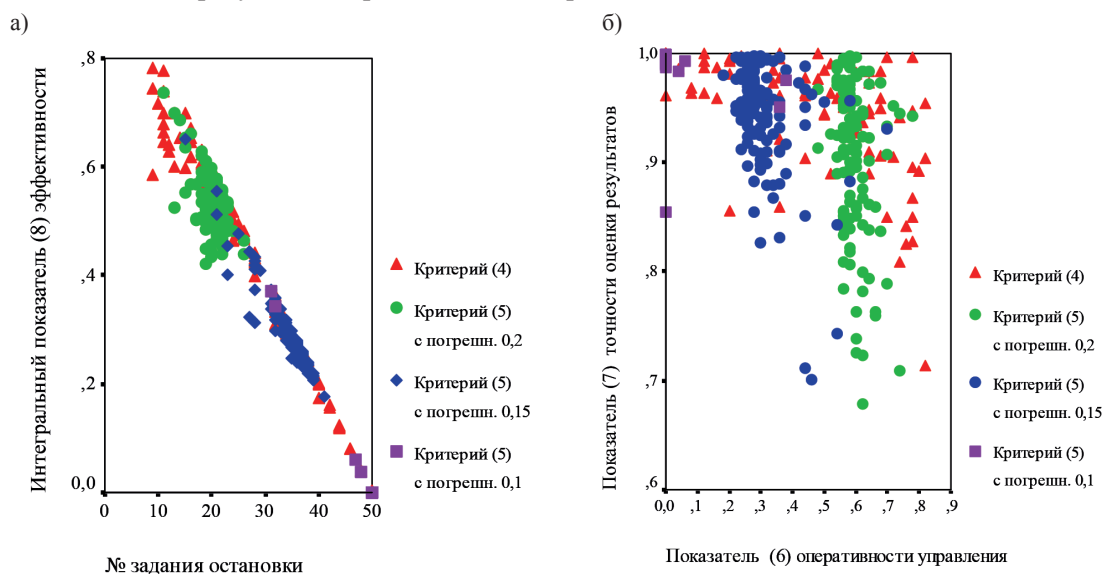


Рис. 6. Точечная диаграмма соответствия:

а — индивидуальных оценок интегральной эффективности управления тестированием длине теста;
б — частных показателей (6) и (7) эффективности управления тестированием

Заключение

Критерии (4) и (5) и реализующая их процедура управления критериально-ориентированным тестированием (см. рис. 1) работоспособны и обеспечивают приемлемую эффективность в подготовке и оценке знаний морских специалистов. Каждый из критериев (4) и (5) имеет свои достоинства и недостатки, которые обусловлены общими закономерностями изменения используемых в них показателей. Последнее определяет целесообразность попытки синтеза этих критериев в один обобщенный критерий. Разработанный в статье новый переход от точечных оценок эмпирической частоты успешного выполнения тестовых заданий к интервальным оценкам вероятности этой успешности, позволяет с достаточной точностью определять погрешность и доверительный интервал полученных оценок при подготовке специалистов для флота.

Для эффективного внедрения средств оперативного критериально-ориентированного контроля знаний в процесс подготовки специалистов для флота и обеспечения возможности совместного анализа и обработки оценок результатов обучения по морским специальностям, полученных компьютерными системами тестирования и преподавателем, целесообразно реализовать следующие практические рекомендации:

– в соответствии с материалами ряда международных морских конвенций, формулировку системы предпочтений преподавателя в отношении качества подготовки специалистов для флота и системы предпочтений преподавателя в отношении значимости единичных тестовых заданий в тесте следует рассматривать в качестве наиболее ответственных этапов подготовки теста к практическому использованию;

– преподавателю, впервые производящему ввод своих систем предпочтений, следует проверить их фактическое соответствие на выборке из пяти – десяти таких результатов выполнения теста, которые покрывают весь спектр возможных результатов (от полностью верных до полностью неверных).

СПИСОК ЛИТЕРАТУРЫ

1. Аванесов В. С. Проблема формы тестовых заданий [Электронный ресурс] / В. С. Аванесов. — Режим доступа: <http://testolog.narod.ru/Theory32.html> (дата обращения: 07.11.2016).
2. Мятлев В. Д. Теория вероятностей и математическая статистика, математические модели / В. Д. Мятлев. — М.: Академия, 2009. — 320 с.
3. Переверзев В. Ю. Моделирование процесса критериально-ориентированного педагогического тестирования курсантов вузов: дис. ... канд. пед. наук / В. Ю. Переверзев. — М.: Военный ун-т, 2000. — 198 с.
4. Печников А. Н. Метод интервальной оценки результатов выполнения системы одиночных тестовых заданий закрытого типа с единственным верным ответом / А. Н. Печников, К. С. Палкин // Образовательные технологии и общество. — 2014. — Т. 17. — № 2. — С. 491–501.
5. Тухтаров Р. Р. Оценка контроля знаний с использованием обучающих систем ВМФ / Р. Р. Тухтаров, А. О. Туровская // Морской сборник. — 2013. — Т. 2000. — № 11. — С. 57–59.
6. Berk R. A. Criterion-referenced measurement: The state of art / R. A. Berk. — Baltimore, MD: The Johns Hopkins University Press, 1980. — Pp. 41–42.
7. Van der Linden Wim J. Passing score and length of a mastery test / Wim J. van der Linden // Evaluation in Education. — 1982. — Vol. 5. — Is. 2. — Pp. 149–164. DOI: 10.1016/0191-765X(82)90015-5.
8. Палкин К. С. Метод управления длиной системы одиночных тестовых заданий с единственным верным ответом / К. С. Палкин, А. Н. Печников, Д. А. Печников // Образовательные технологии и общество. — 2015. — Т. 18. — № 1. — С. 544–560.
9. РМГ 29-2013. Метрология. Основные термины и определения (введен 01.01.2015). — М.: Стандартинформ, 2014. — 121с.
10. ГОСТ 15467-79. Управление качеством продукции. Основные понятия. Термины и определения (с изм. № 1). — М.: Стандартинформ, 2009. — 11 с.
11. Тухтаров Р. Р. Анализ подходов к разрешению противоречий в процедурах измерения и оценивания результатов критериально-ориентированного тестирования в современных компьютерных системах тестирования / Р. Р. Тухтаров // Журнал научных публикаций аспирантов и докторантов. — 2013. — № 10 (88). — С. 239–243.
12. Кошелев М. Н. Компьютерная технология критериально-ориентированного тестирования / М. Н. Кошелев, М. Ю. Мамонтова. — Омск: Изд-во Ур.ГПУ, 2014. — 106 с.
13. Бочкарева Ю. Е. Моделирование тестово-обучающей системы на основе комбинированной модели с использованием критериально-ориентированного тестирования / Ю. Е. Бочкарева, Н. А. Грачева //

Современные исследования в области естественных и технических наук: междисциплинарный поиск и интеграция: материалы науч.-практич. всероссийской конф. (школы-семинара) молодых ученых. —Тольятти: ФГБОУ ВПО «Тольяттинский государственный университет», 2012. — С. 17–20.

14. Субетто А. И. Сочинения. Ноосферизм: в 13 т. / А. И. Субетто; под ред. Л. А. Зеленова. — СПб.: Кострома: Изд-во КГУ им. Н. А. Некрасова, 2009. — Т. 8. Квалитативизм: философия и теория качества, квалитология, качество жизни, качество человека и качество образования. — Кн. 2. — 334 с.

15. Kaftandjieva F. Methods for setting cut scores in criterion-referenced achievement tests: A comparative analysis of six recent methods with an application to tests of reading in EFL / F. Kaftandjieva. — Cito, Arnhem: EALTA, 2010. — 169 p.

REFERENCES

1. Avanesov, V. S. Problema formy testovykh zadaniy. Web. 7 Nov. 2016 <<http://testolog.narod.ru/Theory32.html>>.
2. Myatlev, V. D. *Teoriya veroyatnostei i matematicheskaya statistika, matematicheskie modeli*. M.: Akademiya, 2009.
3. Pereverzev, V. Yu. Modelirovanie protsessa kriterial'no-orientirovannogo pedago-gicheskogo testirovaniya kursantov vuzov. PhD diss. M.: Voennyi universitet, 2000.
4. Pechnikov, A. N., and K. S. Palkin. "Metod interval'noi otsenki rezul'tatov vypolneniya sistemy odinochnykh testovykh zadaniy zakrytogo tipa s edinstvennym vernym otvetom." *Obrazovatel'nye tekhnologii i obshchestvo* 17.2 (2014): 491–501.
5. Tukhtarov, R. R., and A. O. Turovskaya. "Otsenka kontrolya znaniy s ispol'zovaniem obuchayushchikh sistem VMF." *Morskoi sbornik* 2000.11 (2013): 57–59.
6. Berk, R. A. *Criterion-referenced measurement: The state of art*. Baltimore, MD: The Johns Hopkins University Press, 1980.
7. Van der Linden, Wim J. "Passing score and length of a mastery test." *Evaluation in Education* 5.2 (1982): 149–164. DOI: 10.1016/0191-765X(82)90015-5
8. Palkin, K. S., A. N. Pechnikov, and D. A. Pechnikov. "Metod upravleniya dlinoi sistemy odinochnykh testovykh zadaniy s edin-stvennym vernym otvetom." *Obrazovatel'nye tekhnologii i obshchestvo* 18.1 (2015): 544–560.
9. Russian Federation. Recommendations on interstate standardization RMG 29-2013. Metrology. Basic terms and definitions (introduced from 01.01.2015). M.: StandardInform, 2014.
10. Russian Federation. State Standart GOST 15467-79. Product-quality control. Basic concepts. Terms and definitions. M.: Standardinform, 2009.
11. Tuktarov, R. R. "Analiz podkhodov k razresheniyu protivorechii v protsedurakh izmereniya i otsenivaniya rezul'tatov kriterial'no-orientirovannogo testirovaniya v svoevremennykh komp'yuternykh sistemakh testirovaniya." *Zhurnal nauchnykh publikatsii aspirantov i doktorantov* 10 (88) (2013): 239–243.
12. Koshelev, M. N., and M. Yu. Mamontova. *Komp'yuternaya tekhnologiya kriterial'no-orientirovannogo testirovaniya*. Omsk: Ur.GPU, 2014.
13. Bochkareva, Yu. E., and N. A. Gracheva. "Modelirovanie testovo-obuchayushchei sistemy na osnove kombinirovannoi modeli s ispol'zovaniem kriterial'no-orientirovannogo testirovaniya." *Sovremennye issledovaniya v oblasti estestvennykh i tekhnicheskikh nauk: mezhdistiplinarnyi poisk i integratsiya: materialy nauchno-prakticheskoi vse-rossiiskoi konferentsii (shkoly-seminara) molodykh uchenykh*. Tol'yatti: FGBOU VPO "Tol'yatinskii gosudarstvennyi universitet", 2012: 17–20.
14. Subetto, A. I. *Sochineniya. Noosferizm: V 13 tomakh. Tom vos'moi: Kvalitativizm: filosofiya i teoriya kachestva, kвалitologiya, kachestvo zhizni, kachestvo cheloveka i kachestvo obrazovaniya. Kniga 2*. Edited by L. A. Zelenov. S.-Peterburg-Kostroma: KGU im. N.A. Nekrasova, 2009.
15. Kaftandjieva, F. *Methods for setting cut scores in criterion-referenced achievement tests: A comparative analysis of six recent methods with an application to tests of reading in EFL* / F. Kaftandjieva. Cito, Arnhem: EALTA, 2010.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Печников Денис Андреевич —
кандидат технических наук, доцент
Военно-морская академия имени Адмирала
Флота Советского Союза Н. Г. Кузнецова
197045, Российская Федерация, г. Санкт-Петербург,
Ушаковская набережная, 17/1
e-mail: 19pda72@bk.ru

INFORMATION ABOUT THE AUTHOR

Pechnikov, Denis A. —
PhD., associate professor
Naval Academy named after Admiral of the Fleet
of the Soviet Union N.G. Kuznetsov
17/1, Ushakovskaya naberezhnaya, St. Petersburg,
197045, Russian Federation
e-mail: 19pda72@bk.ru

Статья поступила в редакцию 24 апреля 2017 г.

Received: April 24, 2017.