

**О требованиях текущего педагогического контроля к методам и моделям критериально-ориентированного тестирования**

**Аннотация.** В статье на основании анализа данных педагогики, психологии и квалиметрии обосновываются состав требований текущего педагогического контроля к методам и моделям критериально-ориентированного тестирования. Уточняется содержание требований достоверности, диагностичности, оперативности, многокритериальности и формализуемости данных тестирования при их реализации методами тестирования, а не инструментального наблюдения.

**Ключевые слова:** текущий педагогический контроль, критериально-ориентированное тестирование, достоверность, диагностичность, оперативность, многокритериальность, формализуемость.

**A. Pechnikov, D. Pechnikov**

**About requirements of the current pedagogical control to methods and models of criterion-oriented testing**

**Annotation.** The article substantiates the composition of the requirements of the current corrective pedagogical control to the methods and models of criteria-based testing based on the analysis of data of pedagogics, psychology, qualimetry. Specifies the content requirements of accurate news, diagnosticate, efficiency, multicriteria, formalize in line with test data when they implement the methods of testing, and not instrumental observations.

**Key words:** current pedagogical control, criterion-oriented testing, reliability, diagnostics, operativeness, multicriteriality, formalizability.

**Введение**

Сегодня компьютерные системы тестирования (КСТ) рассматриваются в качестве одного из перспективных средств реализации текущего педагогического контроля. В направлении развития возможностей КСТ в сфере текущего контроля деятельности обучаемых выполнен ряд работ [1; 2; 3; 4; 5; 6 и др.], которые устраняют отдельные недостатки имеющихся КСТ. Однако эти работы не объединены единой концепцией, что исключает возможность разработки КСТ полностью соответствующей целям текущего педагогического контроля. Создание таких КСТ необходимо предполагает наличие той системы требований, которые предъявляет педагогика к функциональным возможностям этих технических устройств и методам реализации, возложенных на них функций. Такая система требований на сегодняшний день отсутствует.

Требования текущего педагогического контроля к методам контроля и техническим средствам их реализации наиболее подробно представлены в [7; 8; 9; 10]. Они включают в себя требования достоверности, диагностичности, оперативности, многокритериальности и формализуемости данных и процедур контроля, а также требование формулировки оценки в традиционной 4-балльной шкале. Однако эти требования ориентированы на реализацию инструментального контроля и не учитывают специфику тестового контроля, поскольку авторами [7; 8; 9; 10] методы тестирования не рассматриваются и вообще не упоминаются. Таким образом существует необходимость адаптации требований текущего педагогического контроля к возможностям и техническим средствам тестирования.

В тестологии рассматриваются критериально-ориентированный и нормативно-ориентированный подходы к оцениванию результатов тестирования. В [1; 2; 4; 5; 11; 12] показано, что из этих двух альтернатив целям текущего педагогического контроля соответствуют только методы критериально-ориентированного тестирования. Поэтому ниже будут рассматриваться только те требования, которые тестология предъявляет к методам критериально-ориентированного тестирования.

#### **Анализ требований тестологии к методам и средствам критериально-ориентированного тестирования**

В тестологии все выборочные задания разделяют на тестовые и псевдотестовые.

Тестовое задание или совокупность таковых классифицируют как тестовое, если оно удовлетворяет следующим требованиям [13; 14; 15]: 1) требованию определенности содержания, исключающему использование в тесте контрольного материала, который не входит в содержание соответствующей учебной дисциплины; 2) требованию взаимосвязи, предполагающему принадлежность частных заданий к одной предметной области и наличие между ними логических связей, адекватно отражающих логику предметной области; 3) требованию валидности, которое для критериально-ориентированных тестов включает требования содержательной, конструктивной и критериальной валидности; 4) требованию надежности теста. Выполнение данных требований предполагает проведение классификационной проверки самих выборочных учебных заданий, т.е. проведение ряда экспериментальных тестирований, результаты которых составляют статистический массив данных для оценки теста. Выборочные учебные задания, прошедшие классификационную проверку, классифицируются как тестовые, а не прошедшие или не удовлетворяющие вышеприведенным требованиям – как задания в тестовой форме (псевдотестовые задания).

Нас интересуют требования в КСТ как техническим средствам текущего корректирующего контроля, обеспечивающим создание тестовых заданий и проведение тестирования испытуемых. С этих позиций все приведенные выше требования тестологии относятся не к КСТ, а к самим тестам, которые

разрабатываются и предъявляются испытуемым с их применением. Исключение составляет требование критериальной валидности теста.

Термин “критериальная валидность (внешняя валидность, валидность, критериально определяемая)” определяется как “валидность инструмента тестирования, оцениваемая посредством определения связи между значениями, полученными по этому тесту, и некоторым независимым, не относящимся к тесту критерием” [1; 2; 5]. Из требования критериальной валидности необходимо следует, что процедура вывода результатов критериально-ориентированного тестирования в КСТ должна обеспечивать: 1) объективность и дифференцированность первичных результатов тестирования; 2) формулировку результатов сравнения данных тестирования с внешним показателем в системе предпочтений преподавателя как эксперта в предметной области соответствующей учебной дисциплины и лица, принимающего решения (ЛПР) по управлению процессом обучения.

Последние требования к методам инструментального наблюдения в [7; 8; 9; 10] не предъявлялись, что определяет необходимость анализа перечисленных выше требований текущего педагогического контроля к средствам контроля на предмет их корректности и достаточности в случае применения тестового подхода.

### **Требования текущего педагогического контроля к методам и средствам критериально-ориентированного тестирования**

**1. Требование достоверности** определяет, что “результаты контроля должны объективно и точно отражать уровень обученности обучаемых и обеспечивать корректность сравнения показателей, достигнутых по различным объектам и видам их учебной деятельности” [7].

Используемые в этой формулировке понятия достоверности, объективности и точности имеют много трактовок. В целях выполнения семантического принципа однозначности терминологии примем в отношении этих понятий трактовки, представленные в ГОСТ Р 51170-98 [17]. Выбор именно этого руководящего документа определяется тем, что он устанавливает трактовки понятий, описывающих требования к исходным данным любого процесса управления.

В [17] все свойства данных делятся на две группы: формально-технические и социально-психологические свойства. Достоверность данных определяется как “свойство данных не иметь скрытых ошибок. ... В свойстве достоверности выделяют безошибочность данных и их истинность” [17]. Под безошибочностью данных понимается “свойство данных не иметь скрытых случайных ошибок” [17], а под истинностью – “свойство данных не иметь искажений, внесенных намеренно человеком” [17]. Безошибочность и истинность оцениваются вероятностью ошибки. Под термином “вероятность ошибки” понимается “вероятность наличия хотя бы одной ошибки в данных определенного объема, для которого эта ошибка

искажает содержание данных” [17]. Безошибочность и истинность определяются путем оценивания.

“Оценивание (параметра) – операция определения на основе выборочных данных числовых значений параметров распределения, принятого в качестве статистической модели генеральной совокупности, из которой извлечена выборка. ... Результат этой операции может быть выражен как одним числовым значением, так и доверительным интервалом” [18]. Первый вариант результата называют точечным, а второго – интервальным оцениванием. “Точечное оценивание параметра – получение оценки параметра в виде одного численного значения. Интервальное (доверительное) оценивание параметра – получение оценки параметра в виде доверительного интервала” [18]. Под “точностью (результата проверки)” понимают “близость результата проверки к принятому значению величины” [18]. Из этого определения следует, что точность может быть обеспечена только путем интервального (доверительного) оценивания результатов любой проверки. “Понятие точности, когда его относят к результатам проверки, включает в себя комбинацию случайных компонентов и общего компонента систематической ошибки или смещения” [18]. “Случайные ошибки в данных обусловлены, как правило, ненамеренными искажениями содержания сведений человеком, сбоями технических средств и ошибками программ ЭВМ. При анализе истинности данных рассматривают намеренные искажения данных человеком – источником сведений” [17]. Из сравнения последних двух дефиниций следует, что комбинация случайных компонентов отражает свойство безошибочности, а смещение (общий компонент систематической ошибки) является проявлением свойства истинности.

Таким образом, приведенное выше требование достоверности в терминах [17; 18] можно сформулировать как **требование интервального оценивания результатов контроля с заданным уровнем статистической значимости ошибки (1 или 2 рода).**

Однако такая интерпретация достаточна только в отношении методов инструментального наблюдения, построенных на процедурах объективного измерения физически измеряемых параметров деятельности. Результаты таких измерений фиксируются в виде данных, которые и рассматриваются в [17; 18] в качестве единственной формы существования информации. В отличие от методов инструментального наблюдения, результаты тестирования еще нужно представить в виде данных. Последнее положение отражает сформулированное выше требование объективности и дифференцированности первичных результатов тестирования. Рассмотрим это требование более подробно.

*Требование объективности первичных результатов тестирования.* “Объективность педагогических измерений можно определить как такое отражение интересующего свойства личности на числовой шкале, которое адекватно действительному распределению испытуемых по данному свойству”

[19]. С учетом этой трактовки под объективностью результатов тестирования будем понимать независимость процедуры определения (измерения) первичных тестовых баллов за выполнение единичного тестового задания от мнений разработчика теста или тестирующего.

*Требование дифференцированности первичных результатов тестирования.* Первичные результаты выполнения единичного тестового задания и теста в целом должны представляться в определенной шкале. “Шкала диагностическая – это континуум количественных значений, обозначающих определенную степень выраженности измеряемого или диагностируемого свойства” [20]. В метрологии точность измерений – это близость “измеренного значения к истинному значению измеряемой величины” [20]. Понятие точности измерений объединяет понятия правильности и прецизионности измерений. Прецизионность измерений – это “близость между измеренными значениями величины, полученными при повторных измерениях для одного и того же или аналогичных объектов при заданных условиях” [20]. Одним из средств обеспечения высокой прецизионности является разрешающая способность шкалы измерения, определяемая “наименьшим изменением измеряемой величины, которое является причиной заметного изменения соответствующего показания” [20]. Из приведенных дефиниций корректно следующее определение: разрешающая (дифференциальная) способность шкалы измерения первичных результатов тестирования – это минимальное изменение суждений испытуемого, которое может быть зафиксировано в этой шкале. Отсюда требование дифференцированности первичных результатов тестирования – это требование применения для измерений результатов тестирования шкал, способных фиксировать минимальные изменение суждений испытуемого.

Таким образом, **требование достоверности результатов тестирования должно включать:** 1) требование независимости процедуры определения результатов выполнения тестовых заданий от мнения тестирующего (разработчика теста); 2) требование представления результата выполнения любого тестового задания в шкале, разрешающая способность которой соответствует разрешающей способности используемой формы тестового задания; 3) требование детерминации процедуры оценивания результатов выполнения теста (системы тестовых заданий) до начала тестирования и ее неизменности вплоть до завершения обработки результатов тестирования всех испытуемых; 4) требование интервального оценивания результатов тестирования с заданным уровнем значимости теста.

**2. Требование диагностичности** определяет, что “результаты контроля должны обеспечивать количественный и качественный анализ процесса обучения в целях установления причин появления выявленных недостатков и прогнозирования его дальнейшего развития” [7].

Это требование обеспечивает корректирующий характер текущего контроля, результаты которого наряду с характерными для констатирую-

шего контроля обнаружением и оценкой несоответствий в объекте контроля должны обеспечивать выбор адекватных корректирующих действий. В общем случае для устранения ошибки в деятельности необходимо: определить источник ошибки (внешнее действие, в котором ошибка проявилась), а затем причину ее появления (умственное действие, следствием которого стало ошибочное внешнее действие). Приведенное выше требование диагностичности указывает на необходимость выявления только причин появления ошибки и исключает необходимость поиска источника этой ошибки. Этот факт объясняется особенностями инструментального наблюдения, анализируемого в [7; 8; 9].

В общем случае для контроля обучения могут применяться два метода: тестирование и инструментальное наблюдение. Инструментальное наблюдение – это осуществляемый с применением технических средств метод педагогического контроля, основанный на мониторинге параметров деятельности обучаемого и математической обработке полученных данных для построения выводов об истинном уровне его учебных достижений. Метод инструментального наблюдения применяется в тренажерно-обучающих системах (ТОС), где вся среда деятельности обучаемых моделируется аппаратно-программными средствами ТОС. Поэтому в ТОС любое действие обучаемого, которое совершается им за пределами его собственного сознания, реализуется в полностью контролируемой среде. Это позволяет измерять, идентифицировать и документировать каждое действие обучаемого, которое выходит за пределы его сознания. Последнее позволяет исключить в ТОС необходимость диагностики внешних действий и свести функцию диагностики к анализу умственных действий обучаемых.

Как известно, возможности тестирования ограничены рамками констатирующего контроля (обнаружение и оценка несоответствий в объекте контроля). Поэтому **требование диагностичности** для них должно быть расширено до следующего требования: **результаты контроля должны обеспечивать установление как источников ошибок (ошибочных действий), так и причин их появления.**

**3. Требование оперативности** определяет, что “время проведения и обработки результатов контроля должно быть минимальным и обеспечивать непрерывность процесса управления обучением”.

ГОСТ Р 51170-98 определяет оперативность как “свойство данных соответствовать временным характеристикам процесса, который эти данные отображают” [17]. Принцип непрерывного управления, фигурирующий в требовании, основан на концепции НОРД цикла (OODA loop). Эта концепция определяет, что критическим фактором успеха управления любым процессом является его непрерывность и оперативность: “кто быстрее сможет получить ситуационную информацию, осмыслить ее, принять лучшее решение и эффективно его реализовать, тот и получит наилучшие результаты” [21]. В отношении обучения принцип непрерывного управле-

ния определяет, что наилучшая эффективность обучающих воздействий будет достигнута, когда данные о результативности реализованной части воздействий будут поступать непосредственно в процессе реализации этих воздействий, а сами воздействия корректироваться в процессе их реализации. Поэтому **рассматриваемое требование оперативности есть требование минимизации времени тестирования и обработки полученных результатов.**

**4. Требование многокритериальности** определяет, что “результаты контроля должны отражать весь набор требований, которые предъявляются к виду оцениваемой деятельности, с учетом их влияния на достижение поставленной цели” [7].

Применительно к методам инструментального наблюдения это требование означает возможность измерения и оценивания любого выходного контролируемого параметра действия (например, поворот руля вправо на 15 градусов, удар сверху вниз с силой 45 кг и т.п.). Казалось бы, данное требование к тестированию не может быть предъявлено, поскольку результат теста вне зависимости от содержания теста оценивает только два параметра — правильность и время выполнения тестового задания. Однако, если в соответствие каждому из показателей частных целей деятельности поставить тестовое задание, то общая совокупность этих тестовых заданий обеспечит возможность оценки любого множества целей деятельности или требований, предъявляемых к ней. При всем том, требование в основном относится не к КСТ, а к реализуемому с его помощью тесту. Применительно к КСТ **требование многокритериальности следует трактовать как требование возможности оценивания как результатов тестирования (правильности и времени выполнения тестовых заданий), так и тех психофизиологических параметров испытуемых, которые влияют на получение этого результата.**

**5. Требование формализуемости** определяет, что “процедуры контроля должны обеспечивать ввод, компьютерную обработку выходных контролируемых параметров деятельности обучаемых, хранение и отображение комплекса информации, обеспечивающего диагностику, корректуру и оптимизацию процесса обучения” [7].

Для КСТ выполнение этого требования является средством выполнения требования достоверности результатов тестирования. Объективность, дифференцируемость и точность результатов тестирования, могут быть реализованы в КСТ только за счет формализации исходной информации и процедур тестирования. Поэтому **требование формализуемости в отношении методов тестирования следует трактовать как требование описания всех процедур отображения исходных данных и получения результатов тестирования на языке математики и в рамках формальной логики.**

**6. Требование к формулировке результатов оценивания** состоит в том, что “результаты контроля должны формулироваться в традиционной 4-балльной шкале, принятой в практике боевой подготовки ВМФ, и соответствовать сложившимся представлениям профессорско-преподавательского состава”. Это требование является общим как для методов инструментального наблюдения, так и для методов тестирования.

#### **Заключение**

Приведенный вариант требований текущего педагогического контроля учитывает все особенности и возможности критериально-ориентированного тестирования как метода контроля, что обеспечивает возможность разработки и создания эффективных компьютерных систем тестирования.

#### **Библиографический список**

1. Печников А. Н., Туровская А. О., Туктаров Р. Р. Модели и процедуры оценки результатов компьютерного тестирования знаний // Образовательные технологии и общество. 2013. Т. 16. № 4. С. 365 – 371.
2. 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. 2014. Т. 17. № 2. С. 491 – 501.
3. Палкин К. С., Печников А. Н., Печников Д. А. Метод управления длиной системы одиночных тестовых заданий с единственным верным ответом // Образовательные технологии и общество. 2015. Т. 18. № 1. С. 544 – 560.
4. Печников А. Н., Печников Д. А. Решение задач текущего педагогического контроля на основе анализа результатов критериально-ориентированного тестирования // Образовательные технологии и общество. 2015. Т. 18. № 2. С. 489 – 513.
5. Печников А. Н., Печников Д. А. Метод анализа результатов критериально-ориентированного тестирования в целях текущего контроля учебного процесса // Педагогические измерения. 2015. № 4. С. 17 – 39.
6. Печников А. Н., Аванесова Т. П., Шиков А. Н. Альтернативные подходы к проектированию и внедрению компьютерных технологий обучения // Образовательные технологии и общество. 2013. Т. 16. № 2. С. 433 – 446.
7. Печников А. Н. Теоретические основы психолого-педагогического проектирования автоматизированных обучающих систем. Петродворец: ВВМУРЭ им. А.С. Попова, 1995. 326 с.
8. Печников А. Н., Ветров Ю. А. Проектирование и применение компьютерных технологий обучения. Ч.1. Концепция САО и моделирование процессов деятельности. Кн. 1. СПб: БГТУ, 2003. 195 с.



9. Печников А. Н., Шиков А. Н. Проектирование и применение компьютерных технологий обучения: Монография. СПб.: Изд-во ВВМ, 2014. 393 с.
10. Печников А. Н., Аванесова Т. П., Шиков А. Н. Электронное обучение: Учебное пособие. СПб.: ВАС, 2014. 73 с.
11. Печников А. Н. О едином подходе к трактовке компетенций в сфере социального управления и образования // Образование и наука. 2016. № 2 (131). С. 4 – 17.
12. Печников А. Н., Шиков А. Н. Эффективность электронного обучения как проблема педагогической информатики // Педагогическая информатика. 2013. № 3. С. 49 – 59.
13. Аванесов В. С. Композиция тестовых заданий. Учебная книга для преподавателей вузов, учителей школ, аспирантов и студентов педвузов. 4-е изд. М.: Центр тестирования, 2002. 237 с.
14. Аванесов В. С. Научные проблемы тестового контроля знаний / Монография. М.: Исследовательский центр проблем качества подготовки специалистов, 1994. 135 с.
15. Опарина Н. М. Адаптивное тестирование. Хабаровск: Изд-во ДВГУПС, 2007. 95 с.
16. Национальная психологическая энциклопедия. [Электронный ресурс] URL: <https://vocabulary.ru/> (дата обращения: 22.02.2018).
17. ГОСТ Р 51170-98. Качество служебной информации. Термины и определения. М.: Стандартинформ, 2005. 38 с.
18. ГОСТ Р 50779.10-2000. Статистические методы. Вероятность и основы статистики. Термины и определения. М.: Стандартинформ, 2001. 134 с.
19. Аванесов В. С. Проблема объективности педагогических измерений // Педагогические измерения. 2008. №3. С. 3 – 31.
20. Международный словарь по метрологии: основные и общие понятия и соответствующие термины: пер. с англ. и фр. / Всерос. науч.-исслед. ин-т метрологии им. Д. И. Менделеева, Белорус. гос. ин-т метрологии. Изд. 2-е, испр. СПб.: НПО «Профессионал», 2010. 82 с.
21. Boyd, John R. (September 3, 1976). Destruction and Creation. U.S. Army Command and General Staff College. URL: [http://www.goalsys.com/books/documents/DESTRUCTION\\_AND\\_CREATION.pdf](http://www.goalsys.com/books/documents/DESTRUCTION_AND_CREATION.pdf) (дата обращения: 22.02.2018).