

КВАЛИМЕТРИЧЕСКИЙ ПОДХОД В ОБРАЗОВАНИИ

УДК 37.022.371.263

М. В. Кочукова,
В. В. Свиридов,
Н. М. Ткачева

СИСТЕМНЫЙ ПОДХОД К РАЗРАБОТКЕ СОДЕРЖАНИЯ БАЗ ТЕСТОВЫХ ЗАДАНИЙ ДЛЯ ПРОВЕРКИ ОСТАТОЧНЫХ ЗНАНИЙ

Аннотация. В статье проведен анализ существующих подходов к разработке баз тестовых заданий, изложены основные требования к данным заданиям. Подчеркивается важность системного подхода к формированию содержания баз тестовых заданий и предлагается методика разработки их содержания для проверки остаточных знаний, которая позволяет строить системные тесты. Результаты тестирования позволяют утверждать, что разработанная авторами методика позволит создавать содержательно валидные тесты, способные служить сравнительно более точным инструментом контроля остаточных знаний.

Ключевые слова: остаточные знания, тестирование, системный подход.

Abstract. The paper analyses different approaches to the development of test item banks (TIBs) and shows the main requirements to the latter, the importance of the systematic approach to this work being emphasized. It also presents the procedure of content design of achievement tests, which allows to develop systematic tests. The methods created by the authors can be used to develop valid final achievement tests.

Index terms: permanent knowledge, testing, the systematic approach.

1. Постановка задачи

Управление качеством образования – ключевой элемент любой современной образовательной системы – с необходимостью требует эффективных средств объективного контроля учебных достижений обучаемых. Именно поэтому за рубежом широко распространено, а в России внедряется в практику учебного процесса *тестирование* школьников, студентов и слушателей разнообразных образовательных учреждений.

Исследователи проблем тестирования и разработчики тестовых заданий (ТЗ) основное внимание уделяют требованиям к *форме* тестовых заданий, методам *автоматического* формирования тестов из баз тесто-

вых заданий (БТЗ), вопросам *математической* обработки результатов тестирования и интерпретации последних [2, 10, 11, 14, 15, 24]. Важнейшие вопросы *содержания* тестов, их валидности, не только по форме, но и по существу, чаще всего оказываются в тени.

В настоящей работе предлагается метод разработки содержания БТЗ для проверки остаточных знаний, который позволяет строить действительно системные тесты, проверяющие не только и не столько память обучаемого, но понимание им фундаментальных закономерностей и взаимосвязей данной области знания – как теоретического, так и прикладного.

2. Введение в проблему

Под тестированием понимается стандартизированная процедура объективной оценки образовательных достижений испытуемого или отдельных качеств его личности [26]. Основным инструментом тестирования служит программно-дидактический тест – *целостная система* стандартизированных по форме ТЗ определенного смысла и заданной меры трудности, ориентированная на конкретный результат, позволяющая с необходимой уверенностью и объективностью оценить степень обученности объекта тестирования путем обработки качества его заключений в течение ограниченного интервала времени [Там же]. Отметим, что эта формулировка сама по себе предполагает выполнение следующих требований к БТЗ: целостности и системности; стандартности ТЗ по форме; смысловой (содержательной) валидности; примерно одинаковой меры трудности всех заданий теста; надежности.

С наибольшей детальностью к настоящему времени проработаны требования к форме ТЗ [4, 23]. Сложнее обстоит дело с определением надежности полученной оценки степени обученности, поскольку априори невозможно определить меру трудности того или иного ТЗ. Но, пожалуй, самая неопределенная ситуация сложилась с пониманием и реализацией требования *содержательной валидности* ТЗ, которую мы в данной работе будем понимать как степень адекватности результатов тестирования испытуемых их реальному владению данной областью знания на заданном уровне. Под «реальным уровнем» здесь понимается оценка, данная квалифицированным экспертом или полученная с помощью технологии, позволяющей объективно ранжировать ТЗ и установить их согласованность как между собой, так и с характеристикой испытуемых, которая, собственно, подлежит определению [16].

Как обеспечить содержательную валидность теста? В литературе указываются следующие критерии этого качества [18]:

1) полнота отображения материала учебной программы [7] или, для аттестационного тестирования, требования образовательного стандарта. Проблема состоит в том, что и учебная программа, и образовательный

стандарт дают сильно сокращенное описание области знаний. Поэтому, как правило, в БТЗ включаются задания на проверку таких элементов предметной области, которые формально в программе (стандарте) не названы, но, по мнению разработчиков, необходимы для полноценного владения материалом;

2) значимость содержания ТЗ: этот критерий указывает на необходимость включения в тест только тех элементов содержания, которые являются наиболее важными, ключевыми, обязательными. Необходима уверенность в том, что задания теста охватывают все важные аспекты предметной области в правильной пропорции [3]. Практический вопрос, встающий перед разработчиками БТЗ, заключается в том, как разделить важные и неважные аспекты предметной области, тем более, как вычислить «правильную пропорцию»;

3) предметная чистота ТЗ: ее понимают как отсутствие пересечений содержания заданий одной предметной области с содержанием другой [7]. Этот критерий представляется несколько сомнительным в связи с провозглашенным и проводимым в настоящее время в России переходом к образовательным стандартам третьего поколения, которые подразумевают умение использовать *все* полученные знания во *всей* их совокупности для решения задач, возникающих в ходе профессиональной деятельности.

Какими же способами можно обеспечить содержание тестовой базы, адекватное контролируемой системе знаний? Общий ответ известен: необходим системный подход к формированию содержания БТЗ [2, 3, 10].

3. Проблема выбора подхода к разработке содержания БТЗ

Проблеме конструирования содержания теста посвящены труды многих авторов. В большинстве своем они предлагают изначально опираться на определение целей тестирования, их конкретизацию и классификацию [11, 12, 14, 24], которые позволяют «видеть релевантность заданий по отношению к этим целям» [11]. От выяснения целей авторы переходят к формированию содержания, в котором степень представленности тем, разделов дисциплины пропорциональна важности раздела, определяемой экспертами, и числу часов, отведенных на них в программе. Согласно А. Н. Майорову, эти два критерия позволяют определить вес каждой темы [14]. П. Клайн, М. Б. Чельшкова, М. В. Кларин предлагают вес ТЗ и длину теста преподавателям выбирать самостоятельно, но экспертиза качества содержания теста, в результате которой должны быть выделены «укрупненные единицы знаний», повышающие «полноту отображения», и обнаружены «какие-либо грубые упущения», должна проводиться сторонними экспертами [11, 12, 24]. Более определенные указания об отборе содержания теста даются в работах специалистов АСТ-Центра. Они предлагают разбивать весь учебный материал на элементарные дидактические

понятия, и на их основе создавать ТЗ [4]. С нашей же точки зрения, адекватный тест должен охватывать не формальную полноту дидактических единиц, а более глубокие уровни знания. Для этого при проектировании ТЗ необходимо опираться на системный подход.

Данная проблема рассматривалась В. С. Аванесовым, который указал на необходимость системности содержания ТЗ, пояснив, что такие ТЗ должны быть связаны между собой общей структурой знаний; каждое ТЗ должно проверять свою часть в общей системе знаний [3]. К сожалению, никаких методических указаний для проектирования таких ТЗ при этом приведено не было.

Существует интересное предложение Н. Абовского, исходящее из того, что соблюдение определенных требований к формально-логической структуре ТЗ позволит автоматически добиться системности тестового материала [1]. Однако эта идея плохо согласуется с духом системного подхода, согласно которому системность определяется не свойствами отдельных элементов, а эмерджентно возникает при их взаимодействии [1, 3]. Кроме того, в обсуждаемой ситуации системность теста не может формироваться самопроизвольно, но должна отражать системность контролируемой области знания.

Таким образом, при хорошей общетеоретической разработанности положений системного подхода применительно к задаче проектирования содержания БТЗ [1, 3, 9] не существует разработанных конкретных методик. Разработчикам предлагаются ценные, но слишком общие указания.

4. Методика системного подхода к проектированию содержания БТЗ для проверки остаточных знаний

Широкое распространение в последнее время получили такие формы оценки знаний учащихся, как ЕГЭ для выпускников школ и Федеральный интернет-экзамен в сфере профессионального образования (ФЭПО) для студентов вузов и техникумов. Поскольку число учащихся, вовлеченных в ФЭПО, измеряется сотнями тысяч, а в ЕГЭ – миллионами [16], то проблема валидности используемых БТЗ оказывается чрезвычайно актуальной. Поэтому одна из основных задач, решавшихся нами в данной работе, заключалась в разработке методики, позволяющей проектировать содержательно валидные БТЗ для использования в целях, декларируемых организаторами ЕГЭ и ФЭПО.

К сожалению, при анализе результатов ЕГЭ выясняется, что в настоящем своем виде он неэффективен с точки зрения точности измерения [12, гл. 11]: задания части А оказываются чрезмерно легкими, части В – чрезмерно трудными, а части С – запредельно трудными для большей части выпускников школ. Эти проблемы порождены прежде всего двойственностью целей ЕГЭ. С одной стороны, ЕГЭ – это проверка школьных

знаний на уровне минимальной компетентности. Для такой проверки следует предъявлять испытуемым предельно простые задания [2, 3]. С другой стороны, ЕГЭ должен выполнять роль дифференцирующего входного тестирования в вуз. Но тогда в нем должны присутствовать задания разных уровней сложности, причем без таких вопиющих разрывов этого уровня, как между частями А, В и С. Очевидно, указанные цели ЕГЭ принципиально противоречивы. Пока органы управления образованием не внесут кардинальных коррективов в целеполагание процедуры ЕГЭ, не имеет смысла обсуждать содержание используемых в нем заданий (каждое из которых, взятое в отдельности, может быть вполне разумным и адекватным). Исходя из этих соображений, мы будем рассматривать только задания ФЭПО.

Что должно контролироваться Федеральным интернет-экзаменом? Существуют два основных мнения: либо знания «остаточные», либо «элементарные».

С нашей точки зрения, идея проверять элементарные знания применительно к ФЭПО не очень хороша. По регламенту ФЭПО, тестированию подлежат студенты, которые закончили изучать данную дисциплину не более чем за год до тестирования. На деле временной разрыв между изучением материала и тестированием может быть гораздо больше года – для дисциплин, осваиваемых в течение нескольких семестров.

В педагогической психологии известны «кривые забывания», которые показывают, что большая часть полученных в ходе обучения знаний быстро и неизбежно забудется [20]. Что весьма существенно, динамика забывания практически не зависит от «элементарности» или «сложности» знаний: и бессмысленный набор односложных слов, и высокохудожественный текст забываются одинаково быстро [20]. Таким образом, то, что можно проверить у обучаемых через несколько месяцев после изучения курса, – это не элементарные знания, отбираемые по признаку простоты, а остаточные знания, отбираемые по признаку необходимой включенности в *систему* знаний данной предметной области.

В предлагаемой нами методике используется системный подход к проектированию БТЗ, то есть подбор такого содержания ТЗ, который отвечал бы требованиям системности знаний. «Важно иметь задания, связанные между собой общей структурой знаний. Это возможно в тех случаях, когда каждое задание проверяет преимущественно свою часть в общей системе знаний» [3]. Кроме того, на наш взгляд, задания в рамках системного подхода должны проверять знание взаимосвязей и взаимозависимостей «частей в общей системе знаний».

Обсуждаемая методика имеет целью построение ТЗ, связанных между собой общей факторной структурой знаний, – системных ТЗ (СисТЗ). Они должны проверять не только знание элементарных дидактических

единиц (ЭДЕ), но, прежде всего, понимание взаимоотношений между ними, а также свойств, приобретаемых ЭДЕ только в составе данной системы знаний. Поэтому для построения СистЗ необходимо выделить связи, существующие между понятиями. Эти связи выражаются, во-первых, законами, теоремами, правилами и принципами данной предметной области. Во-вторых, они могут отражать более широкие отношения: функциональной зависимости, симметрии, включения или исключения, логического или хронологического следования, генетического наследования и т. д. В-третьих, связи могут иметь характер не только специфический для предметной области, но и более широкий, транспредметный. Действительно, по одной из аксиом теории систем, каждая система входит в качестве подсистемы в систему более высокого ранга, и потому ее понимание требует знания и понимания взаимосвязей данной системы с другими системами.

В соответствии с нашей методикой, на первом этапе выделяются наиболее фундаментальные понятия данной предметной области. Фундаментальность определяется экспертом. Эти понятия рассматриваются как вершины графа [5, 22, 25], ребрами которого служат взаимосвязи между данными понятиями.

По мере построения такого графа обычно выявляется необходимость в дополнительных понятиях, значимость которых на первых этапах была недооценена по тем или иным причинам. У графа могут быть «висячие» ребра, связывающие понятия данной предметной области с понятиями смежных областей. Простейший вариант таких графов – деревья, которыми принято описывать состав систем в терминах отношений «часть – целое» [19].

Основные методические принципы, отражающие как формальные свойства «дерева знаний» предметной области, так и эффективные приемы его конструирования:

- граф легче строить от вершин, а не от ребер;
- граф должен иметь центр, состоящий из одной или двух смежных вершин;
- все вершины графа разбиваются на непересекающиеся ярусы;
- вершины каждого яруса должны быть равноудалены от центра графа;
- наиболее значимые понятия располагаются в вершинах с наибольшим весом.

На втором этапе понятия ранжируются по весу: наибольший вес присваивается тем, из которых можно логически вывести остальные. Расположив таким образом фундаментальные понятия и приняв во внимание известные закономерности предметной области, можно определить «остов» дерева понятий.

На третьем этапе вершины графа соединяются ребрами, которые представляют собой взаимосвязи – отношение, закон, теорема, тождество. В зависимости от того, вершины каких ярусов оно соединяет, ребру приписывается вес. Наибольший вес, естественно, у ребер, соединяющих вершины верхних ярусов.

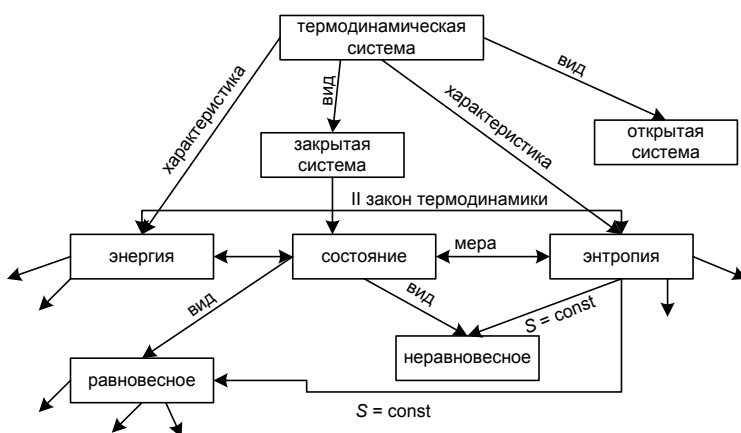
На четвертом этапе для каждого ребра/вершины составляются ТЗ, проверяющие знание соответствующей взаимосвязи/понятия. Максимальное число ТЗ создается для ребер (вершин) с наибольшим весом.

На последнем этапе самим ТЗ приписывается вес, соответствующий весу представляемых ими ребер/вершин графа, и определяется успешность прохождения теста по сумме весов успешно выполненных ТЗ.

Подробное изложение методики, к сожалению, слишком объемно для журнальной статьи и будет опубликовано в дальнейшем более подробно.

5. Эксперимент по проектированию содержания БТЗ для проверки остаточных знаний (на примере раздела «Термодинамика» курса «Общей физики»)

Для сравнения результатов тестирования по БТЗ, разработанных по обычной методике и по предлагаемой нами системной методике, был поставлен эксперимент. В качестве первой попытки была создана БТЗ по разделу «Термодинамика», так как это область физики отличается четко выраженной логической структурой. Термодинамику не зря называют «геометрией в физике», к тому же она одинаково представлена в курсе физики различных вузов, и на момент проведения эксперимента студенты, участвующие в нем, ее изучили. Для создания данной БТЗ был разработан граф, фрагмент которого представлен на рисунке.



Граф «Понятия – взаимосвязи» по разделу Термодинамика курса общей физики

БТЗ в полном объеме была предложена 172 студентам Воронежской государственной медицинской академии, закончившим изучение раздела «Термодинамика» курса физики.

На основе данного графа составлена БТЗ, в которой 60% составляли СистЗ (разработанные нами) и 40% задания на знания ЭДЕ (взятые из контрольного теста, используемого на кафедре медицинской физики и разработанные на основе методики В. И. Васильева и Т. Н. Тягуновой).

Уровень учебных достижений каждого студента оценивался независимо преподавателем-экспертом (который вел занятия в группе более одного семестра, принял несколько зачетов и хорошо представляет уровень знаний студентов) по обычной десятибалльной шкале. Нас интересовали, прежде всего, соотношения между

- 1) оценками эксперта и результатами обычного тестирования;
- 2) оценками эксперта и результатами системного тестирования;
- 3) результатами обычного тестирования и результатами по СистЗ.

Указанные соотношения анализировались как для всей выборки, так и для сильных групп (тех, у кого баллы от 10 до 8 включительно) и отдельных слабых студентов (тех, у кого баллы от 7 и ниже).

6. Обсуждение результатов и основные выводы

Для сильных студентов соответствие результатов тестирования оценке эксперта-преподавателя оказалась довольно слабым. Мы объясняем этот результат тем, что, во-первых, количество сильных студентов, принявших участие в тестировании, оказалось слишком малым, чтобы получить устойчивые и статистически значимые результаты. Кроме того, уровень сильных студентов, по всей видимости, заметно превышал уровень трудности большинства заданий. Для таких групп учащихся снижается дискриминирующая способность теста и возрастает значение случайных (индивидуальных) вариаций [16, 27].

Анализ величин корреляционного отношения между оценкой эксперта и результатами обычной методики и между оценкой эксперта и результатами системного тестирования для сильных студентов показал, что при малом их значении связь, хоть и слабая, все-таки существует. Другими словами, тест не позволяет определить хороших студентов, но с большой вероятностью позволяет идентифицировать слабого студента. Если же трудность теста подгонять к уровню решаемости (отношение числа испытуемых, решивших задание по данной теме, к общему числу участников педагогических измерений [17]) слабых студентов, он не даст возможность определить уровень способности сильных.

Сравнивая значения величин коэффициента корреляции и корреляционного соотношения между оценкой эксперта и результатами системного тестирования для слабых студентов, можно заметить, что они раз-

личны (т. е. связь нелинейна), причем величина корреляционного отношения близка к единице. Следовательно, можно с уверенностью сказать, что тесты, построенные на системных вопросах, ближе к оценке эксперта. Для отношения между оценкой эксперта и результатами обычного тестирования величины коэффициента корреляции и корреляционного отношения близки по значению и малы.

Итак, на БТЗ, построенной с применением системного подхода, задача идентификации слабых студентов решается гораздо увереннее. Если учесть, что при проверке остаточных знаний большинство студентов должны, в силу динамики «забывания знаний», выглядеть как слабые, то полученные результаты дают основание надеяться на существенное повышение достоверности оценки остаточных знаний.

Литература

1. Абовский Н. Системный подход к тестированию знаний // *Alma mater* (Вестн. высш. шк.). 2007. № 6. С. 26–27.
2. Аванесов В. С. Композиция тестовых заданий. М.: Адепт, 1998. 217 с.
3. Аванесов В. С. Принципы разработки содержания теста // *Управление школой*, 1999. № 38.
4. Васильев В. И., Киринок А. А., Тягунова Т. Н. Требования к программно-дидактическим тестовым материалам и технологиям компьютерного тестирования. М.: МГУП, 2005. 29 с.
5. Гофмарк А. С. Элементы теории графов: текст лекций. Ташкент: ТИХН, 1980. 32 с.
6. Даль В. И. Толковый словарь русского языка. Современное написание. М.: Астрель, 2003. 984 с.
7. Евтюхин Н. В., Бондарева Т. В., Дубинина Т. В., Сурыгина И. Ю. Современное состояние методов тестирования знаний и умений за рубежом и в России // *Инновации в образовании*. 2004. № 1. С. 27–47.
8. Ефремова Т. Ф. Толковый словарь служебных частей речи русского языка. М.: АСТ, 2004. 814 с.
9. Зорина Л. Я. Дидактические основы формирования системности знаний старшеклассников. М.: Педагогика, 1978. 128 с.
10. Караушев В. Ф., Терентьева Л. В., Тягунова Т. Н. Введение в разработку банков программно-дидактических тестовых заданий. М.: МГУП, 2005. 52 с.
11. Клайн П. Справочное руководство по конструированию тестов: введение в психометрическое проектирование. Киев: Пан Лтд, 1994. 283 с.
12. Кларин М. В. Инновационные модели обучения в зарубежных педагогических поисках. М.: Арена, 1994. 152 с.
13. Корсини Р., Ауэрбах А. Психологическая энциклопедия. СПб.: Питер, 2006. 1095 с.

14. Майоров А. Н. Теория и практика создания тестов для системы образования. М.: Интеллект-центр, 2002. 296 с.

15. Мамонтова М. Ю. Стохастические модели управления качеством результатов обучения: ограничения и перспективы использования // Образование и наука. Изв. УрО РАО, 2009. № 6(63). С. 21–36.

16. Маслак А. А. Измерение латентных переменных в социально-экономических системах: теория и практика. Славянск-на-Кубани: СГПИ, 2007. 424 с.

17. [Электрон. ресурс]. Материалы форума FEPO <http://www.fepo.ru>, 2008–2009гг.

18. [Электрон. ресурс]. 2000-2008 MirSlovari.com: Мир словарей – коллекция словарей и энциклопедий.

19. Новосельцев В. И. Системный анализ: современные концепции. Воронеж: Кварта, 2003. 360 с.

20. Психология памяти / под ред. Ю. Б. Гиппенрейтер и В. Я. Романова. М.: ЧеРо, 2002. 816 с., ил. (Хрестоматия по психологии).

21. Равен Джон. Педагогическое тестирование: проблемы, заблуждения, перспективы. М.: Когито-Центр, 1999. 144 с.

22. Системный подход в современной науке. М.: Прогресс-Традиция, 2004. 560 с.

23. Чельшкова М. Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. 432 с.

24. Чельшкова М. Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. 432 с.

25. Харари Ф. Теория графов. М.: Едиториал УРСС, 2003. 296 с.

26. Шадриков В. Д. Концепция централизованного тестирования в системе непрерывного общего и профессионального образования: принято на науч.-метод. совете Минобразования РФ по проблемам тестирования. Протокол № 4 от 8 декабря 1999 г.

27. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press, 1980. 199 p.