УДК 37.012; 519.876.5

Монахов Вадим Валериевич,
Воропаев Роман Алексеевич,
Бушманова Вера Алексеевна,
Бушманова Елена Алексеевна,
Фриш Владимир Сергеевич,
Васильева Анита Витальевна

ОСОБЕННОСТИ РАЗРАБОТКИ КОМПЬЮТЕРНЫХ ТЕСТОВ, ЗАДАЧ И ВИРТУАЛЬНЫХ ЛАБОРАТОРИЙ

Аннотация

На основе опыта использования программного комплекса distolymp в интернет-олимпиадах школьников по физике и при проведении занятий со студентами проведен анализ общих особенностей разработки компьютерных заданий. Обсуждена специфика и общие черты трех принципиально отличающихся типов заданий: тестов, теоретических задач и заданий на основе виртуальных лабораторий по физике. Обсуждается погрешность измерений, надежность и валидность заданий, влияние сложности заданий на корректность измерений. Особое внимание уделено обсуждению особенностей заданий на основе моделей виртуальных лабораторий.

Ключевые слова: тест, тестирование, дистанционное обучение, образование, физика, моделирование, интернет-олимпиада, BARSIC, distolymp, погрешность измерений баллов, надежность, валидность, IRT, программное обеспечение.

В настоящее время основной методикой дистанционной проверки результатов обучения является проверка с помощью тестов – например, в системах дистанционного обучения (СДО) Moodle, Sakai, BlackBoard и др. Такие типы заданий позволяют проверить знания и очень ограниченный набор умений и навыков. Однако в физике важную роль играет эксперимент, требующий специфических умений и навыков. Кроме того, тестовая форма проверки заметно отличается от традиционной письменной формы решения задач, и встает вопрос о создании компьютерного аналога задач, но с автоматической проверкой правильности их решения.

Нами разработан программный комплекс distolymp, обеспечивающий возможность проверки трех типов заданий: тестов, теоретических задач и заданий на основе виртуальных лабораторий по физике. Этот комплекс используется для проведения интернет-олимпиад школьников по физике [1–4] и постоянно совершенствуется.

Многолетняя практика проведения интернет-олимпиад в широком возрастном диапазоне (с 7 по 11 классы) на территории всей России и ближнего зарубежья позволила выявить особенности составления и применения упомянутых выше трех типов заданий.

1. АНАЛИЗ ОБЩИХ ОСОБЕННОСТЕЙ РАЗРАБОТКИ КОМПЬЮТЕРНЫХ ЗАДАНИЙ (на примере особенностей разработки тестов)

Приведем важнейшие общие особенности разработки компьютерных заданий на примере тестов, поскольку теория и практика тестирования наиболее разработана.

Тест – набор кратких, четко сформулированных заданий, выполнение которых требует небольшого количества времени. На первый взгляд, кажется, что достаточно по-

© Монахов В.В., Воропаев Р.А., Бушманова Е.А., Бушманова В.А., Фриш В.С., Васильева А.В., 2013

добрать набор вопросов на заданные темы, и, если в подготовленных заданиях нет ошибок, тест можно использовать. Для того чтобы не было списывания или заучивания ответов на конкретные вопросы, создается банк с большим числом заданий, из которых тестируемый получает заданное число заданий.

В связи с появлением возможности самостоятельно создавать тесты в системах дистанционного обучения и различного рода программах создания компьютерных тестов, появилось и появляется большое количество таких «тестов». У них имеются два очень серьезных недостатка, полностью лишающих их ценности:

– результаты, полученные с помощью таких тестов, плохо воспроизводимы (дают большой разброс по количеству полученных баллов), то есть у них низкая *надежность*;

– эти результаты нельзя связать с уровнем усвоения изученного материала, приобретенного знаниями, умениями, навыками или с уровнем способностей.

Описываемая далее методика [5] обычно применяется только к тестам. Однако она может быть применена и *к любым видам заданий*, в том числе к задачам и моделям. Ниже излагается очень небольшая часть материала, важная для проведения первоначального статистического анализа качества заданий.

Очень важной характеристикой теста является *погрешность* измерения результатов. Например, ее можно измерить методом ретестирования, то есть учащиеся дважды проходят один и тот же тест (ретест) через некоторое время, когда вопросы, полученные ими при первом прохождении, подзабылись. В другом варианте, более сложном в плане подготовки заданий, но часто единственно возможном организационно (например, для итогового тестирования), используется разбиение теста на два эквивалентных набора вопросов и сравнивается различие результатов, полученных для этих двух подтестов (параллельных вариантов).

Оценку стандартной погрешности S_E заданий теста обычно (см. [6]) делают по формуле

$$S_E = S_X(1 - R_{xx'})^{1/2}, \qquad (1)$$

где S_X – стандартное отклонение (среднеквадратичная вариация) баллов группы участников, $R_{xx'}$ – коэффициент корреляции между параллельными вариантами X и X'. Однако данная формула применима только для случая нормального распределения участников по набранным баллам [6]. Поэтому при распределении, отличном от нормального, она дает только грубую оценку погрешности.

Кроме того, (1) дает только усредненную погрешность теста, а она в разных областях итоговых баллов разная и может отличаться в 2–3 раза от погрешности в области средних набранных участниками баллов. Измерение погрешности в зависимости от набранных баллов (ее называют условной) чаще всего проводят по методу Трондайка [6]:

$$S_E = S_{X-X'}, \qquad (2)$$

где $S_{X-X'}$ – среднеквадратичная вариация разницы баллов X и X', набранных в параллельных вариантах, при этом считается, что средние значения $<X>$ и $<X'>$ в точности равны. Обычно значения усредняют на некотором интервале. Основной недостаток метода – низкая точность при небольшом числе испытуемых. Однако метод очень устойчив, и его точность известна из классической теории измерений.

Профессионально составленный тест изучается на достаточно большой пробной группе тестируемых, и только после этого применяется для измерения результатов учащихся. Как уже говорилось, погрешность измерений является его важнейшей характеристикой. Она непосредственно связана с *надежностью* R_{XX} теста, которая задается теоретической формулой [5]:

$$R_{XX} = 1 - \frac{S_e}{S_X}, \qquad (3)$$

где S_e – стандартное отклонение погрешности измерения (среднеквадратичная вариация погрешности итогового балла по всем участникам тестирования). Однако значение S_e неизвестно, поэтому часто для оценки надежности теста используются его парал-

лельные формы. Для тест-ретестового варианта надежность теста равна коэффициенту корреляции между двумя параллельными тестами [5].

Коэффициент корреляции R_{XY} между величинами X и Y при измерении для n участников равен

$$R_{XY} = \frac{c_{XY}}{S_X S_Y} = \frac{\sum_{j=1}^{n} (X_j - < X >)(Y_j - < Y >)}{S_X S_Y}, \quad (4)$$

где $<X>$ и $<Y>$ – средние значения величин по всем измерениям (в нашем случае – усреднение по всем участникам), j – номер участника.

В случае нахождения корреляции между параллельными формами теста X_j – балл, набранный за первую форму теста участником j, а Y_j – набранный им за вторую форму теста. При расщеплении теста на два полутеста h и h' формула несколько меняется [5]:

$$R_{XX} = \frac{2R_{hh'}}{1 + R_{hh'}}, \quad (5)$$

где $R_{hh'}$ – коэффициент корреляции между полутестами.

Однако формула (5), часто упоминаемая в учебниках, на практике почти никогда не используется, так как она обычно дает сильно заниженные значения надежности. Вместо нее применяют коэффициент α-Кронбаха (K_α) [5]:

$$\alpha = \frac{k}{k-1} \frac{\sum c_{ii'}}{S_X{}^2} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^{k} s_i{}^2}{S_X{}^2} \right), \quad (6)$$

где k – число вопросов (заданий) в тесте, $S_X{}^2$ – дисперсия по всему тесту, $c_{ii'}$ – ковариации между вопросами с номерами i и i', $s_i{}^2$ – дисперсия для отдельного вопроса с номером i.

Хотя формула (6) строго выводится только для тестов с выбором одного варианта из нескольких, она может хорошо служить для оценки надежности и других типов заданий. Набор заданий с $K_\alpha < 0{,}2$ считается неудовлетворительным, $K_\alpha > 0{,}6$ – надежным. Необходимо отметить, что значение K_α относится к конкретной группе тестируемых, и если в ней подготовленность участников заметно отличается от реальных групп, то надежность теста, полученная для пробной группы, может сильно отличаться от значения для реальной группы.

На этапе подборки заданий теста измерение погрешности или коэффициента надежности всего теста не дает возможности улучшить характеристики теста, так как не позволяет выяснить, какие конкретные задания необходимо удалить или заменить. Иногда пользуются следующим приемом: по одному удаляют из теста задания и смотрят, имеется ли задание, при удалении которого надежность всего теста возросла наибольшим образом. Удаляют его из теста и повторяют процедуру до тех пор, пока не удастся удалить ни одного задания без понижения надежности.

Также используют проверку на взаимную корреляцию заданий по формуле (4), при этом в качестве X и Y выступают баллы X_{i_1} и X_{i_2}, полученные участниками за задания с номерами i_1 и i_2. Задания с отрицательным коэффициентом корреляции отбрасываются.

Очень важной характеристикой заданий является *валидность* – способность измерения именно той характеристики учащихся, для которой предназначен тест. Существует большое количество подходов к определению и измерению валидности [5].

Валидность теста зависит от валидности входящих в него заданий. Этой характеристике тестов и других видов заданий начинающие разработчики компьютерных программ часто не придают значения, в результате чего ценность многих компьютерных средств обучения оказывается близкой к нулю и даже отрицательной, так как подобные компьютерные инструменты просто дискредитируют идею использования компьютеров в образовании.

Практически важной при создании теста является *диагностическая (конкурентная) валидность* заданий – способность различать уровень подготовки участников по

результатам тестирования. Она находится по корреляции результатов задания с общим результатом теста. То есть в формуле (4) в качестве X и Y выступают баллы X_i, полученные участниками за задание с номером i, и X_Σ – полученные за весь тест.

При этом связь валидности и взаимной корреляции заданий нетривиальна: максимальная валидность теста достигается при таком наборе заданий, в котором они минимально коррелируют между собой и при этом максимально коррелируют с результатами всего теста. Поэтому очевидно, что часто встречающаяся в литературе рекомендация удалять из теста задания с близкой к нулю взаимной корреляцией является ошибочной. Напротив, подозрительны именно сильно коррелирующие задания, если они не являются параллельными вариантами.

Очевидно, что валидность задания сильно зависит от уровня подготовки (и способностей) тестируемых: задание, с которыми справились все тестируемые, будет обладать нулевой валидностью, как и задание, с которыми не справился ни один человек. Например, задания, валидные для заключительного тура крупной олимпиады школьников, окажутся невалидными для группы школьных двоечников, и наоборот. Поэтому подбор заданий должен быть ориентирован на конкретный уровень подготовки тестируемых.

Так, в дистанционных (отборочных) этапах интернет-олимпиады по физике мы, наряду со сложными заданиями в виде теоретических задач и на основе моделей, используем тесты с выбором одного варианта ответов из пяти предложенных, и при этом тестовые задания выбираются очень простыми. Это делается для того, чтобы дифференцировать по результатам наиболее «слабых» участников олимпиады.

Внешняя валидность (валидность по внешнему критерию) – очень важная характеристика теста в психометрике. Она определяется как корреляция между тестом и другим тестом, который гарантированно с высокой точностью измеряет то, что должен измерять данный тест. Но эта характеристика в тестах по физике и математике пока не используется на практике из-за отсутствия близких к идеальным тестов и невозможности предсказания экспертами процента правильного выполнения задания участниками тестирования, хотя в области изучения коэффициента интеллекта (IQ) существуют тесты, близкие к идеальным – тесты Стэнфорд-Бине и Векслера [7]. Так что со временем можно ожидать появления эталонных заданий в области математики, физики и других школьных предметов.

Существует и другой вариант нахождения внешней валидности, когда соответствие заданий критерию (измеряет ли задание то, что надо) определяется по результатам опроса экспертов. Этот вариант является очень трудоемким и не очень объективным (сильно зависит от подбора экспертов и формулировок в критериях оценки).

Ещё один критерий, по которому оценивается задание, – его сложность. Как мы видели, сложность задания сказывается на его валидности. Но она влияет и на многие другие характеристики теста, в частности, – на погрешность измерения результатов. Например, если сформировать банк из заданий с сильно различающейся сложностью и один и тот же тестируемый два раза выполнит тест, то часто будет складываться ситуация, когда в первый раз ему попадутся очень легкие задания, а во второй раз – очень сложные. Из-за этого его результаты будут очень сильно отличаться. Различие набранных баллов будет погрешностью измерений. Очевидно, что различие баллов в пределах данной погрешности не позволяет делать какие-либо выводы о связи различия результатов тестируемых и их уровня подготовки, и полученный рейтинг участников тестирования – вопрос случайности, а не различия в подготовленности или способностях. Поэтому при отсутствии информации о погрешности любые рейтинги и оценки, основанные на результатах тестирования, оказываются недостоверными.

При компьютерном тестировании стали очень популярны адаптивные тесты, в которых, в зависимости от правильности или неправильности ответов испытуемых, выбирается сложность последующих вопросов.

Считается, что подобный подход позволяет сократить время тестирования при той же точности измерений, либо повысить точность при том же времени тестирования. Однако при адаптивном тестировании трудно осуществить корректное сравнение результатов, так как данный алгоритм обладает сильной зависимостью погрешности измерений от фактора случайности (траектории, по которой пошли ответы участника тестирования). Кроме того, при адаптивном тестировании гораздо труднее обеспечить охват всех необходимых разделов, чем в случае теста по нескольким разделам, поскольку число вопросов, получаемых разными участниками, различно. Тем не менее, будущее, вероятно, именно за адаптивным тестированием.

Мы пока не применяем адаптивное тестирование в интернет-олимпиаде, однако планируем провести в будущем эксперименты по использованию адаптивных тестов. Поскольку интернет-олимпиада является добровольным, но очень массовым мероприятием (более 30 тысяч участников ежегодно), а тесты используются только как вспомогательный инструмент в отборочных турах со специально заниженным весовым вкладом в итоговый балл, это открывает большой простор для экспериментов в области отладки и сравнения различных тестовых технологий.

Хотелось бы отметить ещё один момент, относящийся к недостаткам тестовых и других видов заданий: может сложиться (и, по нашему опыту, часто складывается) ситуация, когда «слишком» хорошо подготовленные учащиеся в некоторых простых заданиях ошибаются чаще, чем менее подготовленные. Либо из-за невнимательности, либо из-за слишком глубокого понимания предмета и обнаружения неоднозначности в заданных вопросах, что, конечно, снижает валидность этих заданий, но очень незначительно, так как число очень хорошо подготовленных учащихся мало. И даже проверка заданий на пробной группе не помогает, так как для того чтобы отследить подобные проблемы у 1–5 % наиболее подготовленных участников, пробная группа должна быть очень большой, и должно быть проведено специальное исследование. Но, насколько нам известно, в научной литературе такая проблема до сих пор даже не поднималась, хотя про наличие заданий с низкой и отрицательной корреляцией хорошо известно, как и про проблему угадывания в заданиях с выбором варианта ответа.

2. ОСОБЕННОСТИ РАЗРАБОТКИ КОМПЬЮТЕРНЫХ ЗАДАЧ

Задания, которые мы в интернет-олимпиаде называем теоретическими задачами, на первый взгляд напоминают тестовые задания с вводом ответа в виде числа – как, например, в части В ЕГЭ по физике или математике. Имеется условие задачи и место ввода числовых ответов на заданные вопросы (рис. 1). Участник олимпиады решает задачу и вводит ответы, как и в задании теста. Но эти типы заданий принципиально различны по своей сути.

Задания в тестах имеют фиксированные числовые значения. Это связано с двумя причинами: во-первых, очень часто однотипные задания с разными числовыми значениями имеют сильно отличающуюся сложность.

Пример: Два автомобиля движутся перпендикулярно друг другу, один со скоростью $v_1 = 3$ км/ч относительно земли, другой со скоростью $v_2 = 4$ км/ч относительно земли. Чему равно значение скорости v_{12}, с которой первый автомобиль движется относительно второго?

Ответ: $v_{12} = \sqrt{v_1^2 + v_2^2} = \sqrt{3^2 + 4^2} =$ $= 5$ км/ч.

Такое задание многие способны решить в уме. А попробуйте-ка вычислить в уме ответ на тот же вопрос, но для случая $v_1 = 4$ км/ч, $v_2 = 5$ км/ч. Помимо прочего во втором случае возникает проблема точности, с которой должен быть введен ответ – ее обязательно необходимо оговаривать в условии.

В заданиях типа «задача» мы используем параметризацию условий, и каждый участник интернет-олимпиады получает вариант задания с условиями, сгенерированны-

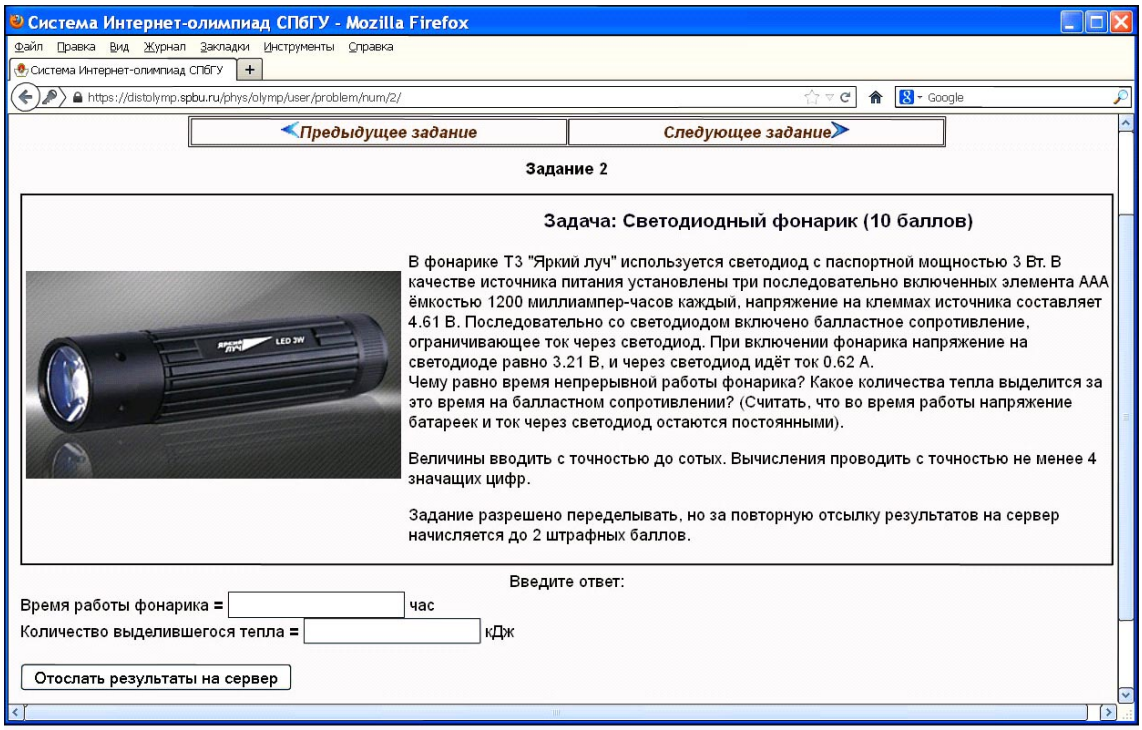

Рис. 1. Пример теоретической задачи

ми на основе псевдослучайной последовательности. В тестах так делать нельзя, а в задачах – можно. Это связано с тем, что в правильно составленном тестовом задании проверяется знание одного закона, или одного соотношения, или одной формулы, или конкретного факта. И принципиальное отличие задачи от тестового задания в том, что она проверяет способность анализировать и делать выводы, а также использовать знания для проведения расчетов на основе известных законов и формул. Это умение следующего уровня по сравнению с заучиванием конкретных утверждений. Поэтому время, которое требуется для решения задачи, обычно гораздо больше, чем для ответа на тестовое задание, и сложность таких заданий выше.

Конечно, проблема различия в сложности заданий с разными числовыми значениями параметров никуда не девается, однако в случаях, когда сложность установления взаимосвязей и нахождения решения намного выше, чем различие в сложности, порождённое различием в числовых значениях параметров, применение параметризованных заданий оправданно. При этом сложность вариантов оказывается фактически неразличимой. Но, конечно, нельзя допускать неэквивалентных по сложности вариантов решений, возникающих из-за различия параметров. Например, наличия в одном варианте двух различных корней квадратного уравнения, а в другом – совпадающих, и т. п.

Ещё одно важное отличие задания типа «задача» от тестового – наличие нескольких вопросов, на которые необходимо дать ответы в рамках одного задания. Необходимость такой структуры определяется назначением этого типа заданий. Конечно, возможна разработка тестовых заданий на проверку умения устанавливать связь между несколькими факторами и на наличие творческих способностей. Хорошим примером таких тестов являются задания по программированию на сертификацию Java SE Programmer Certified Professional Exam. Тем не менее, формулировка заданий в виде задачи гораздо ближе к форме заданий по физике и математике, принятой в российской школьной физико-математической системе обучения.

Возможность «угадайки» (случайного выбора вариантов ответа) заметно снижает ценность тестов при построении рейтинга учащихся. Для принятия решения типа «зачёт/не зачёт» (например, при упоминавшейся выше сертификации) этот фактор не имеет значения – он просто приводит к необходимости сдвига порога зачёта в сторону более высоких баллов. Но даже для зачёта крайне нежелательно использование сложных заданий с выбором одного варианта из нескольких (single choice) – в этом случае необходимо пользоваться заданиями с выборами нескольких вариантов (multiple choice), вставлением промежуточных слов и т. д., заметно уменьшающих возможность «угадайки». Как показывает наша практика, «угадайкой» занимаются вовсе не самые «слабые» учащиеся, как часто предполагают, а «сильные» учащиеся – и только для тех сложных заданий, с которыми они по каким-либо причинам не могут справиться. Поэтому задания с выбором одного варианта из нескольких целесообразно использовать только в области нижних значений сложности, то есть для самых легких заданий.

На основании данной информации, в частности, можно сделать вывод о принципиальной ошибочности наличия в ЕГЭ по физике и по другим предметам сложных заданий с выбором одного варианта ответа из нескольких.

Опыт проведения интернет-олимпиады школьников по физике, охватывающей всю территорию России, показал наличие ещё одной проблемы использования тестов, которые всё чаще применяются в качестве основного средства проверки учебных успехов учащихся: с каждым годом увеличивается процент учащихся, отвечающих только на вопросы теста с выбором одного варианта ответа из нескольких. Эти учащиеся не могут решить даже простейшую задачу, но при этом простые задания на основе моделей виртуальных лабораторий они выполняют, и уровень их мыслительных способностей относительно высок. Их результаты по олимпиаде, если не учитывать задачи, мало отличаются от результатов остальных участников. Это означает, что у данных учащихся не сформированы не только умения анализировать условия задач и устанавливать взаимосвязи, но даже просто воспроизводить на чисто репродуктивном уровне решения стандартных задач. Об аналогичных проблемах, связанных с ориентацией на простейшие тестовые технологии при изучении физики, говорят и результаты других исследований [8, 9].

Но, конечно, не следует бросаться в другую крайность – отказываться от использования тестовых технологий. То, что любая сертификация в области компьютерных дисциплин осуществляется в форме тестирования, доказывает, что в некоторых областях тестовые технологии являются адекватным средством проверки знаний, умений и навыков.

Компьютерные задания типа «теоретическая задача» в интернет-олимпиаде по физике содержат от двух до четырех вопросов, на которые необходимо дать числовой ответ. Обычно первые части задания являются гораздо более простыми, чем последующие. Кроме того, в отличие от заданий тестов, компьютерные теоретические задания можно переделывать: сразу после отсылки отчета на сервер участнику интернет-олимпиады сообщается, какие части задания выполнены правильно, а какие – нет. И при наличии неверных ответов участник сам решает, переделывать ему задание или нет. При повторном выполнении задания с него снимается часть баллов. Подобная возможность отсутствует как в распространенных тестовых технологиях, так и в заданиях обычных олимпиад. Именно компьютерная обработка результатов дает возможность переделывать задания, а работа в серверном варианте в режиме онлайн позволяет использовать такие задания как для итоговой проверки (к которой можно отнести интернет-олимпиаду), так и для дистанционного обучения.

Необходимо отметить, что только в системе дистанционного обучения Moodle среди широко распространенных систем дистанционного обучения имеются средства для создания тестовых заданий с возможностью повторной отсылки ответа, при этом за повторную отсылку ответа на вопрос снима-

ется фиксированный процент баллов (по умолчанию 33,3%). Однако формирование задания типа «теоретическая задача» с единым условием и несколькими вопросами, на которые необходимо дать ответы, в Moodle не предусмотрено. Это связано с ориентацией Moodle на тестовые технологии и отсутствием в мире как практики, так и теории измерения результатов с помощью заданий типа «теоретическая задача».

3. ОСОБЕННОСТИ РАЗРАБОТКИ ЗАДАНИЙ НА ОСНОВЕ МОДЕЛЕЙ ВИРТУАЛЬНЫХ ЛАБОРАТОРИЙ

Задания на основе моделей виртуальных лабораторий [1, 4] не имеют мировых аналогов. В них участникам предоставляется набор инструментов и физических объектов. Несмотря на ряд принципиальных отличий от перечисленных ранее типов заданий, этот тип заданий может рассматриваться как более развитый вариант заданий типа «теоретическая задача». Ведь если в таком задании оставить только условие и форму ввода ответов, убрав из модели все инструменты и физические объекты, мы получим задание с функциональностью теоретической задачи, хотя и организованное чуть по-другому. Более того, в последнее время нами разработано несколько заданий на основе моделей виртуальных лабораторий, совмещающих особенности теоретических задач и виртуального эксперимента. На одни вопросы такого задания участник интернет-олимпиады должен найти ответы с помощью измерений, на другие – с помощью теоретических расчетов без использования измерений, на третьи – с помощью теоретически расчетов, использующих результаты измерений. При этом сам участник должен решить, какой путь должен быть выбран.

Разработка моделей виртуальных лабораторий, аналогичных разрабатываемым нами, является очень трудоемким и сложным процессом. На первый взгляд кажется, что, поскольку в мире существует огромное множество компьютерных моделей по физике, они легко могут быть доработаны до уровня виртуальных лабораторий. Это не так:

- Очень небольшое число моделей допускают параметризацию условий моделируемой системы. Как правило, это системы с жестко заданными параметрами.
- Ещё меньшее (намного меньшее) число моделей допускают недетерминированные действия пользователя: перемещение объектов в произвольные места сцены и их взаимодействие друг с другом, отличающееся в зависимости от внутреннего состояния объектов и их подсоединения друг к другу.
- Только некоторые модели имеют увеличительное стекло, позволяющее рассмотреть в увеличенном масштабе элементы модели и манипулировать ими. При увеличении в ряде случаев требуется особый режим для некоторых элементов управления установками – нам неизвестны модели других разработчиков, обеспечивающие такие возможности.
- Для многих моделей требуется построение графиков с использованием разнообразных возможностей научной графики (выделения участков графика для их просмотра в увеличенном масштабе, автоматическая оцифровка осей графика в широком изменении порядков показываемых величин, возможность фиксации десятичного порядка чисел на подписях к осям, и так далее).
- Модели должны иметь возможность показа условия параметризованного задания, получаемого с сервера.
- Модели должны иметь возможность получать с сервера в зашифрованном виде значения параметров объектов виртуальной лаборатории и самостоятельно устанавливать своим объектам эти значения. Шифрование требуется для того, чтобы имена и значения параметров не могли быть перехвачены недобросовестными участниками олимпиады.
- Модели должны иметь форму отсылки результатов на сервер с показом после отсылки информации о правильности или неправильности результатов по каждому из пунктов задания.

Остановимся на вопросе взаимодействия с сервером подробнее. В условии ряда заданий, как и в теоретических задачах, должны фигурировать значения, генерируемые псев-

дослучайным образом программным обеспечением серверной олимпиадной системы.

Пример текста задания: Найдите массу m_2 тележки, помеченной цифрой 2, если масса тележки, помеченной цифрой 1, равна m_1.

В коде генерации параметров системы **distolymp** в этом случае необходимо задать

```
m1=(50..100, 10);
```

Этот код означает, что задана переменная **m1**, значение которой выбирается псевдослучайным образом в интервале от 50 до 100 с шагом 10. Значения границ и шага могут быть любыми вещественными значениями. Вместо m_1 в тексте задания будет показано значение переменной **m1**, сгенерированное сервером для конкретного участника.

В системе **distolymp** возможен и другой вариант задания значений, выбираемых псевдослучайным образом, – перечислением возможных значений:

```
m1=[50, 70, 80, 100];
```

Мы считаем, что «вмуровывать» в исходные коды серверной системы тексты заданий и формы отчетов принципиально неправильно, так как в этом случае при составлении каждого нового задания меняется программное обеспечение всей системы. Поэтому в системе distolymp предусмотрено добавление для каждого задания следующих сценариев («скриптов»):

- ***Текст задания***.
- ***Декларация вспомогательных переменных*** с возможностью задания их значений псевдослучайным образом и через уже заданные переменные с использованием стандартных функций. Переменные могут быть числовыми и строковыми.
- ***Декларация переменных, соответствующих ответам***, и задание допустимой погрешности этих ответов. Число таких переменных задает число ответов.
- ***Сценарий задания баллов, начисляемых за каждый из ответов***. В системе distolymp возможно каждому ответу назначить независимый балл, например, в зависимости от его сложности. Однако анализ результатов прошедших олимпиад и теорий IRT [10,11] привел нас к выводу о том, что классическая для IRT схема назначения одинаковых баллов за задания разной сложности и линейного распределения заданий по сложности дает меньшую погрешность, чем при назначении тем больших баллов, чем больше сложность задания. На первый взгляд, кажется, что ценность сложных заданий должна быть больше. Однако этот подход, использовавшийся нами первоначально, не оправдал себя, и в настоящее время за каждую правильно решённую с первой попытки часть задания назначается 5 баллов.
- ***Сценарий назначения штрафных баллов за повторные попытки.*** За каждую повторную попытку выполнения задач и заданий на основе моделей назначаются штрафные баллы. Первоначально они назначались за задание целиком. После проведения первой же олимпиады пришлось принять меры, чтобы участники не отсылали одни и те же результаты повторно. Причём как с клиентской стороны (кнопка отсылки результатов была сделана недоступной для нажатия сразу после начала отсылки результатов на сервер, чтобы из-за дрожания руки при щелчке по кнопке результаты не отсылались повторно), так и со стороны сервера. Довольно часто участники после получения ответа со стороны сервера о частично правильном или даже полностью правильном ответе опять отсылали те же результаты! После теоретического анализа начисления баллов при наличии повторных попыток и компьютерного моделирования ситуации мы пришли к выводу, что необходимо начислять штрафные баллы только за те части задания, которые выполнены неверно. Если при этом за каждую неверную попытку снимать одну и ту же долю от максимального балла, назначаемого за данную часть задания, возможность повторных попыток с достаточно большой точностью в рамках теории IRT окажется эквивалентной понижению сложности данной части задания. Таким образом, мы пришли к системе назначения штрафов за повторные попытки, аналогичной использующейся в системе Moodle. Только вместо

33,3 % штрафа, по умолчанию выставляемого в Moodle, мы устанавливаем 1 штрафной балл за первую неверную попытку, то есть двадцатипроцентный штраф. Величина штрафа на настоящий момент назначена решением экспертов методической комиссии олимпиады и не имеет научного обоснования (как, впрочем, и в Moodle). Мы собираемся провести научные исследования по выбору оптимальной величины штрафа. Необходимо отметить, что участники олимпиады очень положительно отнеслись к наличию повторных попыток – в том числе показавшие в олимпиаде самые высокие результаты. Более того, они настояли, чтобы и в теоретических заданиях в ответах сервера показывалось, какие части задания выполнены правильно, а какие – неверно, и это было нами реализовано. Есть основания полагать, что возможность повторных попыток повышает точность измерения подготовленности и способностей учащихся, однако требуется специальное исследование для доказательства этого факта.

- ***Форма отчёта, показываемая участнику для отсылки результатов на сервер.*** Это HTML-текст со вставками на языке JavaScript. Подобный способ формирования задания отчёта и других сценариев позволяет на основе одной и той же модели формировать несколько различных совершенно независимых заданий, а также разделить работу программистов, разрабатывающих модели, программистов, разрабатывающих серверное ПО, и методистов, разрабатывающих конкретные задания и сценарии к ним.

Как можно заметить, разрабатываемые нами виртуальные лаборатории по физике неразрывно связаны с весьма сложным программным комплексом distolymp. Поэтому отсутствие аналогов таких виртуальных лабораторий не удивительно.

Следует отметить, что по функциональности distolymp мало уступает распространённым системам дистанционного обучения, а по ряду элементов (например, наличию поддержки виртуальных лабораторий по физике и возможностям системы статистического анализа результатов) заметно их превосходит.

4. ПРОБЛЕМЫ МУЛЬТИПЛАТФОРМЕННОСТИ, ВЗАИМОДЕЙСТВИЯ МОДЕЛИ ВИРТУАЛЬНОЙ ЛАБОРАТОРИИ С СЕРВЕРОМ И ШИФРОВАНИЯ ДАННЫХ

Разрабатываемые нами виртуальные лаборатории созданы на основе моделей BARSIC [1, 4]. Однако в настоящее время исполняющая среда («проигрыватель») BARSIC, разработанная в Delphi, работает только в операционной системе Windows, что препятствует использованию таких заданий в школах с операционными системами Linux или MacOS. В этих операционных системах имеется возможность запустить BARSIC под эмуляторами Windows API (Wine, Cedega или CrossOver), однако предварительно требуется установить под этими эмуляторами MS Internet Explorer, и в ряде случаев возникают проблемы с неадекватной работой программ.

В связи с этим нами ведется работа по переводу разработки программного комплекса BARSIC в мультиплатформенную среду Lazarus. Однако в Lazarus отсутствует полноценный компонент браузера, способный заменить MS Internet Explorer. Мы дорабатываем компонент GeckoPort [12], использующий мультиплатформенное ядро браузера XULRunner [13]. К сожалению, в нем пока отсутствует ряд необходимых обработчиков событий.

Нами также ведется альтернативный вариант создания мультиплатформенных моделей виртуальных лабораторий на основе классов Java в рамках изучения студентами кафедры вычислительной физики языка Java [14]. При наличии прототипов, созданных в среде BARSIC, создание их Java-аналогов заметно упрощается. На рис. 2 *а* показан пример модели BARSIC, а на рис. 2 *б* – модель на основе апплета Java.

Модель виртуальной лаборатории должна взаимодействовать с сервером, для того чтобы создать виртуальное окружение с нужными физическими параметрами. Тогда ответ, получаемый участником при правильном решении, будет совпадать с ожидаемым сервером. Однако возникает проблема – как

а) *б)*

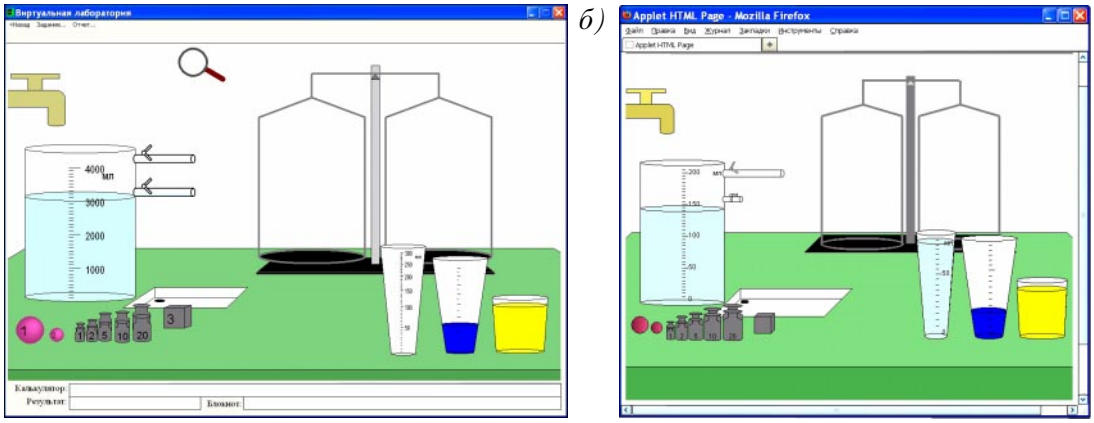

Рис. 2. Модели виртуальной лаборатории «Плотность тел и жидкостей»: *а)* модель BARSIC, *б)* модель на основе апплета Java

передать эти параметры таким образом, чтобы нельзя было их «подсмотреть» в коде страницы или отладочными средствами браузера при асинхронных запросах. Естественное решение проблемы – криптография. При этом использование Java дает заметные преимущества, так как вопросам надежности, совместимости и безопасности в Java уделяется особое внимание.

Ключ шифрования в таком случае скрыт внутри апплета, являющегося бинарным файлом, и извлечь значение ключа, а затам расшифровать данные, сложно.

Протокол передачи параметров апплетов выглядит следующим образом:

1. На стороне сервера PHP-сценарий записывает сгенерированные параметры задачи (например, массы гирь) в одну строку вида «ключ1=значение1 ключ2=значение2 ...».

2. Строка зашифровывается алгоритмом симметричного шифрования Blowfish в CBC режиме [15], используя в качестве ключа заданную строку. Для этого используется PHP-библиотека **`mcrypt`**.

3. Полученная строка передаётся в апплет стандартным для апплетов способом – через тег **`<param>`** внутри тега **`<applet>`**, имя такого объединённого параметра **`params`**. Пример:

```
<param name="params"
value="hwM+LFJSKuVnVJpLthMIgC81Bnq
Cammr43hggc7xcFEutsU6gjkcHZIJBX35H
sgImpv97Tnrpok=">
```

4. Апплет получает параметр params, расшифровывает его с помощью стандартной библиотеки Java Cryptography Extension (JCE) из пакета *javax.crypto*, анализирует строку, записывает параметры в хэш-таблицу, которая далее используется для работы модели.

Криптографический ключ дублируется в PHP-сценарии и Java апплете, необходимо следить за их согласованностью.

Литература

1. *Монахов В.В., Стафеев С.К., Парфёнов В.Г. и др.* Проведение дистанционных экспериментальных туров олимпиад по физике с использованием программного комплекса BARSIC // Компьютерные инструменты в образовании, 2005.№ 2. С. 5–15.

2. *Монахов В.В., Стафеев С.К., Евстигнеев Л.А.* Интернет-олимпиады по физике. Опыт проведения и перспективы / Труды IX Междунар. конф. ФССО. СПб., 2007. С. 278–281.

3. *Монахов В.В., Стафеев С.К., Парфенов В.Г.* Развитие системы интернет-олимпиад СПбГУ и СПбГУИТМО / Труды X Междунар. конф. ФССО-09. СПб, 2009. Т. 2. С. 202–204.

4. *Монахов В.В., Ханнанов Н.К., Кожедуб А.В., Монахова С.В.* Интернет-олимпиады как способ развития творческих способностей школьников // Физика в школе, 2012. № 2. С. 27–40.

5. *Фер Р.М., Бакарак В.Р.* Психометрика: Введение / Пер. с англ. Челябинск: Изд. центр ЮУрГУ, 2010.

6. *L. Feldt, M. Steffen, N. Gupta.* A comparison of five methods for estimating the standard error of measurement at specific score levels // Applied Psychological Measurement, 1985. Vol. 9, № 4. P. 351–361.

7. *Аткинсон Р.Л, Аткинсон Р.С, Смит Э.Е, Бем Д. Дж. и др.* Введение в психологию / Пер. с англ. М.: Прайм-Еврознак, 2003.

8. *Серебрякова Д.В.* Корреляция навыка решения задач и результатов ЕГЭ по физике // Вестник НГУ: серия Физика, 2011. Т. 6. Вып. 3. С. 97–107.

9. *Зайцева Е.В., Лебедева О.В., Соколов В.М., Круглова С.С.* Результаты ЕГЭ и успехи обучения физико-математическим дисциплинам студентов первых курсов // Вестник Нижегородского университета им. Н.И. Лобачевского, 2011. № 3 (3). С. 47–54.

10. *De Ayala R. J.* The Theory and Practice of Item Response Theory. NY: Guilford Press, 2009.

11. *De Mars C.* Item response theory. Oxford, England: Oxford University Press, 2010.

12. http://wiki.lazarus.freepascal.org/GeckoPort (дата обращения 27.02.2013).

13. https://developer.mozilla.org/en-US/docs/XULRunner (дата обращения 27.02.2013).

14. *Монахов В.В.* Язык программирования Java и среда NetBeans / 3-е издание. СПб.: БХВ-Петербург, 2011.

15. *Брюс Шнайер, Нильс Фергюсон.* Практическая криптография. М.: «Вильямс», 2005.

Abstract

On the basis of experience of use of the program complex distolymp in the Online Competition in Physics the analysis of the general features of development of computer-aided tasks is carried out. Specifics and common features of three essentially different types of tasks are discussed: tests, theoretical tasks and tasks on the basis of virtual laboratories in Physics. The error of measurements, reliability and validity of tasks, influence of complexity of tasks on a correctness of measurements is discussed. The special attention is paid to discussion of features of tasks on the basis of models of virtual laboratories..

Keywords: Test, Assessment, Distance Education, Physics, Mathematical Simulation, Online Competition, BARSIC, distolymp, Score Level Errors, Validity, Reliability, IRT, Software.

Монахов Вадим Валериевич,
кандидат физ.-мат. наук,
доцент кафедры вычислительной физики СПбГУ,
v.v.monahov@mail.ru
Воропаев Роман Алексеевич,
студент 2 курса магистратуры физического факультета СПбГУ,
Бушманова Вера Алексеевна,
студентка 4 курса физического факультета СПбГУ,
Бушманова Елена Алексеевна,
студентка 4 курса физического факультета СПбГУ,
Фриш Владимир Сергеевич,
студент 4 курса физического факультета СПбГУ,
Васильева Анита Витальевна,
студентка 4 курса физического факультета СПбГУ.

© Наши авторы, 2013.
Our authors, 2013.