

ИСПОЛЬЗОВАНИЕ ПРОТОКОЛОВ ОН-ЛАЙН КОНКУРСОВ ДЛЯ ОЦЕНКИ СЛОЖНОСТИ ЗАДАЧ И ПОВЫШЕНИЯ ВАЛИДНОСТИ ИЗМЕРИТЕЛЬНОЙ ПРОЦЕДУРЫ

Аннотация

На примере анализа протоколов Международного он-лайн конкурса по информатике «Бобёр-2012» рассмотрено несколько подходов к оценке сложности и трудности заданий. Показано, что во многих случаях априорная оценка сложности заданий составителями конкурса не соответствует объективной трудности заданий для участников. Показано, что школьники, особенно младшие, часто недооценивают сложность заданий. Предложен способ типологии задач конкурса на основании анализа их **трудности** и **сложности** для школьника, применение которого возможно и в других дисциплинах. Выделен кластер задач, **трудность** которых существенно недооценена организаторами конкурса. Показано, что в ряде случаев сложность задания для школьников может быть связана с особенностями интерфейса конкурса. Выявлен ряд недостатков в организации он-лайн протоколов, устранение которых позволит повысить валидность результатов конкурсов.

Ключевые слова: информатика, протоколы, on-line конкурсы, разработка тестов, ключ теста, сложность задач, трудность задач, образовательные тесты, типология задач, конкурс «Бобёр».

КОНКУРС КАК ИЗМЕРИТЕЛЬНАЯ ПРОЦЕДУРА

Любой конкурс в предметной области фактически является *тестом*, то есть стандартизированной методикой оценки. В он-лайн конкурсе «Бобёр» оцениваются знания и умения школьников в предметной области «информатика». Обязательными атрибутами теста являются: стандартизированный набор заданий, стандартизированное предъявление заданий, формализованное описание ответов и процедуры их обработки, адекватный ключ к тесту. Ключ к тесту – это алгоритм отображения протоколов ответов в точку на шкале тестовых баллов. Основное требование к ключу – его **соответствие измеряемому свойству** [1]. Только при наличии перечисленных атрибутов по результатам теста можно *объективно оценить испытуемого по определённой количественной*

шкале, указывающей на выраженность измеряемого свойства [10].

Простейший ключ к тесту предполагает суммирование «стоимостей» выполненных заданий теста. «Стоимости» заданий теста могут быть одинаковыми или иметь разные веса. В случае теста с дифференцированной стоимостью заданий возможны два принципиально различных подхода к назначению стоимости: априорное определение «весов» заданий организаторами (экспертами) или апостериорное приписывание «весов» с учётом результатов выполнения теста участниками. При апостериорном назначении «вес» задачи определяется количеством участников, которые с ней справились. Обычно большую стоимость получают задания, с которыми справилось меньшее количество участников¹. В случае априорного назначения

¹ Таким образом часто назначают веса заданий в конкурсах по программированию, например, <http://codeforces.ru/blog/entry/4172>, http://contest.vandex.ru/cpr_rules.html.

«весов» задач также предполагается, что у более трудных заданий – более высокие веса.

При простейшей организации ключа (участнику начисляют 1 балл за любую верно решённую задачу и 0 баллов за неверный ответ) принято, следуя Фредерику Лорду [10, с. 336], предъявлять задачи участникам в порядке увеличения сложности. Заметим, что для этого необходима априорная экспертная оценка сложности задач, хотя бы разбиение задач на несколько уровней сложности.

При априорном оценивании задач корректность результатов конкурса определяется квалификацией экспертов. Шмелёв подчёркивает, что «не стоит обольщаться тем, что эксперты умеют достоверно оценивать трудность заданий» [10]. По его мнению, *лучшая мера трудности – это та статистика реальных ответов, которую можно получить от реальных испытуемых*. Об этом же говорится в методических указаниях к олимпиаде школьников по информатике: «оценить сложность комплекта задач можно только по результатам выступления всех участников на основе распределения количества набранных баллов по участникам» [4].

СПОСОБЫ ОЦЕНКИ СЛОЖНОСТИ И ТРУДНОСТИ ЗАДАЧ

Задача, лёгкая для одного участника, может оказаться трудной для другого. Трудность задачи отражает **отношения между задачей и человеком**, который её решает. Чтобы подчеркнуть это, многие современные авторы разделяют понятия «сложность» и «трудность» [1, 2, 11, 13]. Под сложностью понимается некая объективная характеристика задания, а под трудностью – субъективная, то, как задания воспринимаются со стороны участника. Говоря о трудности, авторы фокусируют внимание на **деятельности** субъекта по решению задачи – анализе и переработке информации, планировании и принятии решений, прогнозировании последствий своих решений, формировании оперативных образов и концептуальных моделей [6, 9].

Сложность задачи можно измерить по результатам конкурса, подсчитав долю участников, правильно её решивших. Измерение, или хотя бы оценка, трудности задачи требует серьёзных усилий.

Трудность задачи для субъекта складывается из его «умственной рабочей нагрузки» (когнитивной, информационной, эмоциональной нагрузки, нагрузки на внимание) – *mental workload* и его затрат на регуляцию своего состояния [13].

Наиболее точные методы оценки рабочей нагрузки предполагают измерение разнообразных психофизиологических показателей. Процедура диагностики осуществима лишь в очных конкурсах с небольшим числом участников. Сама процедура может служить дополнительным стрессовым фактором для участников, увеличивая затраты на регуляцию состояния [3]. Оценить состояние участников в дистанционных конкурсах можно лишь при помощи опросников на оценку состояния в процессе решения задач. Результаты таких опросов будут адекватны лишь для школьников старших классов, поскольку младшим школьникам трудно объективно оценить своё состояние и свои возможности [8]. Согласно Пиаже, ученик начальной школы находится на стадии развития интеллекта «конкретных операций». Ограничения мышления, характерные для этого возраста, сказываются не только на познании окружающего мира, но и на том, как дети воспринимают самих себя. Говорить о полноценном формировании понятийного мышления можно только к 11–12 годам [7].

С возрастом связаны не только адекватность самооценки, но и уровень развития процессов мышления. Поэтому ученики начальных классов могут (и это закономерно!) плохо решать задачи, требующие оперирования абстрактными понятиями [7]. В период с 8 до 10 лет быстро увеличивается объём памяти, улучшается переключаемость внимания. Поэтому даже небольшая разница в возрасте в этот период может быть причиной существенных различий в результатах решения одних и тех же задач.

По мнению Навона и Гофера, результаты деятельности «человеческой системы» характеризуются соотношением качества рабочей информации (качество и количество стимулов, кодирование, размещение и т. п.) и величиной имеющихся ресурсов [11]. В предметных конкурсах нагрузку на внимание, процессы краткосрочной и оперативной памяти можно оценить через длину условия задачи (количество стимулов, которое нужно обработать для решения).

По мнению Рылея и соавторов, рабочая нагрузка может являться функцией от уровня трудности и количества задач, выполненных за единицу времени [12]. Используя протоколы он-лайн конкурсов, можно оценить рабочую нагрузку на участников, учитывая время, затрачиваемое ими на решение задач.

Правилами конкурса «Бобёр» допускается представление участниками решений не для всех задач. В этом случае отказ участника от решения задачи является его оценкой трудности задачи по бинарной шкале («трудная» – «не трудная»). Задачи, на которые участник дал решения, оценены им как «не трудные», задачи, в которых решения не представлены, оценены как «трудные». Доля участников, оценивших задачу как «трудную», характеризует трудность задачи для всей совокупности участников конкурса.

Априорное экспертное оценивание (взвешивание) задач будет корректным, если будет учитывать как сложность задачи, так и её трудность для участников каждой возрастной группы. Установить, действительно ли при взвешивании задач учтена трудность, согласуется ли априорная оценка с объективной сложностью, можно только проанализировав итоги конкурса.

Заметим, что ряд авторов использует понятия «сложность» и «трудность» как синонимы, поскольку для них принципиально не содержание психических процессов, а фактическое исполнение – сможет решить школьник задачу или не сможет [5]. Мы в дальнейшем будем, где возможно, различать термины «сложность» и «трудность», выделяя их жирным шрифтом. Если же речь

идёт о комплексной оценке, учитывающей оба параметра, будем использовать термин «сложность», не выделяя его.

ЦЕЛИ ИССЛЕДОВАНИЯ

1. Оценить валидность измерительной процедуры, осуществляемой при обработке результатов Международного конкурса по информатике «Бобёр». Оценить качество комплекта задач конкурса и способа подсчёта баллов.

2. Оценить адекватность экспертной оценки сложности задач. Сопоставить различные оценки **сложности** и **трудности** задач.

3. Классифицировать задачи конкурса по **сложности** и **трудности**.

4. Оценить возрастные различия в восприятии задач школьниками.

5. Сформулировать рекомендации организаторам конкурса, использование которых позволит повысить валидность измерительной процедуры. Определить дополнительные возможности использования протоколов он-лайн конкурсов для анализа результатов грядущих конкурсов.

КОНКУРС «БОБЁР»: ОРГАНИЗАЦИЯ, ПОДБОР ЗАДАНИЙ И УЧЁТ РЕЗУЛЬТАТОВ

Международный он-лайн конкурс по информатике «Бобёр» проводится с 2003 года. Россия в 2012 году участвовала в конкурсе впервые¹. Пул заданий для конкурса формируется представителями стран-участниц. Из этого пула заданий каждая страна-участница формирует собственные варианты.

В России конкурс проводится для шести возрастных групп школьников. Участникам каждой возрастной группы предлагаются задачи трёх уровней сложности: для учеников 1–2 класса – по 4 задачи (стоимостью 3, 6 и 9 баллов соответственно), для учеников 3–4, 5–6, 7–8, 9–10 классов – по 5 задач (стоимостью 6, 9 и 12 баллов), 11-классникам предлагаются 8 простых и 7 сложных задач (стоимостью 9 и 12 баллов). На решение задач школьникам даётся 40 минут. За непра-

¹ <http://bebras.ru/bebras13/main/go/p>.

вильный ответ участник штрафует, размер штрафа равен одной трети стоимости задания.

Задачи конкурса пронумерованы, простые задачи имеют меньшие номера, сложные – большие. Задачи одинаковой сложности пронумерованы в случайном порядке (отличающемся у разных участников). Участник может решать задачи в произвольном порядке, может возвращаться к уже решённым задачам.

В большинстве задач участникам предлагается 4 варианта ответа, из которых три неправильных и один правильный. Также у участников есть возможность выбрать вариант «нет ответа». В таком случае участник не получает ни баллов, ни штрафа за эту задачу. Некоторые задачи 1–2 класса являются динамическими – в них требуется произвести какие-то действия мышью. Начисление баллов в таких задачах происходит так же, как и в остальных. Результатом конкурса для каждого участника является сумма набранных баллов и штрафов. Победителями в пределах каждой из шести возрастных групп становятся участники, набравшие максимальное число баллов.

МАТЕРИАЛ И МЕТОДЫ

Исходным массивом данных для анализа является протокол конкурса, в котором фиксируются действия участников. Про каждого участника известно время, в которое он приступил к конкурсу, и время выбора каждого ответа. Если участник в одной задаче последовательно вводил несколько ответов, при подсчёте результатов учитывается ответ, введённый последним. Общее число участников конкурса в 2012 году составило 6588 школьников.

Оценка **сложности** и **трудности** заданий конкурса произведена по нескольким шкалам.

Шкала 1 – экспертная оценка. Производится по трёхбалльной шкале (1 – простые, 3 – сложные) на заседании международного оргкомитета представителями стран-участниц. Чтобы быть корректной, экспертная оценка должна учитывать как объективную **сложность**, так и **трудность** задач.

Шкала 2 – доля участников, которые в данной задаче выбрали вариант «нет ответа». При выборе ответа «наугад» из 4-х предложенных вероятность (p) выбрать правильный равна $1/4$. При этом математическое ожидание заработанных за эту задачу баллов равно

$$p \cdot x + (1 - p) \cdot (-x/3) = x(4p - 1)/3 = 0$$

и совпадает с количеством баллов, получаемых школьником при выборе варианта «нет ответа» (здесь x – стоимость задачи). Если же какой-либо ответ из предложенных можно отбросить как заведомо неверный, то вероятность случайного выбора правильного ответа из оставшихся становится больше $1/4$, и потому математическое ожидание заработанного за задачу балла становится положительным. Таким образом, у участника нет рациональных причин использовать кнопку «нет ответа». Использование этой кнопки может быть обусловлено только психологическими причинами – например, боязнью ошибки (и штрафа за неё) или чрезмерной неуверенностью в себе, испугом от вида задачи. Поскольку вычисление математического ожидания заработанного балла едва ли производится школьниками, выбор варианта «нет ответа» отражает их интуитивное представление о соотношении вероятностей выигрыша и проигрыша. Во всех случаях выбор «нет ответа» является итогом взаимодействия школьника с задачей, то есть характеризует **трудность** задачи для участника.

Шкала 3 – доля участников, давших правильный ответ, среди тех, кто взялся решать эту задачу. Определяется после завершения конкурса и фактически оценивает **сложность** задачи. Подчёркнём, что это именно **сложность** задачи, поскольку вычисляется только по тем учащимся, которые взялись за задачу, то есть по тем, кто по шкале 2 оценил её как «не трудную».

Шкала 4 – число символов в условии задачи. Является косвенной оценкой **трудности** задачи, поскольку связана с нагрузкой на память и внимание.

Поскольку шкалы 1–3 являются ранговыми, сопоставление шкалы 1 со шкалами 2 и 3 производили при помощи рангового коэффициента корреляции Спирмена.

Разбиение задач на кластеры проводилось методом Варда с использованием Евклидовой метрики.

Для оценки адекватности принятого в конкурсе способа подсчёта баллов и качества комплекта заданий конкурса при помощи критерия Лиллефорса проводили проверку на нормальность результатов конкурса. Как альтернатива реально использованному способу подсчёта баллов (задачи разной стоимости, установленной априорно, со штрафными баллами за неправильный ответ) рассмотрены следующие гипотетические способы:

- задачи разной стоимости, установленной априорно, без штрафных баллов за неправильный ответ;
- задачи разной стоимости, установленной апостериорно, со штрафными баллами за неправильный ответ;
- задачи разной стоимости, установленной апостериорно, без штрафных баллов за неправильный ответ;
- задачи равной стоимости, со штрафными баллами за неправильный ответ;
- задачи равной стоимости, без штрафных баллов за неправильный ответ.

РЕЗУЛЬТАТЫ

Итоги конкурса. Гипотеза о нормальности распределений результатов школьников каждого класса (набранного суммарного балла) отвергается на уровне значимости $p < 0.05$ для всех классов, кроме 6, 7 и 8 классов. Значения критерия Колмогорова-Смирнова приведены на рисунке 1 (ряд 1). Для младших (1–6) и старших (9–11) классов распределения результатов имеют достоверно положительную асимметрию, то есть относительно мало участников показали высокие результаты, большинство участников показали низкие результаты.

Анализ гипотетических вариантов оценивания результатов конкурса показал для всех классов, начиная с третьего, что используемый способ начисления баллов является лучшим из способов с априорным оцениванием задач (рис. 1). Исключение штрафов при разных весах задач или выравнивание весов задач при наличии штрафов несколько ухудшают реальное распределение результатов. Особенно неудачным является вариант задач с равными весами и отсутствием штрафов. При всех протестированных гипотетических способах начисления баллов участникам даже распределения результатов учеников 6–8 классов становятся далекими от нормальных. Для учеников 1–2 классов ситуация противоположная: реальный способ подсчёта баллов оказался наименее удачным, исключение штрафов или выравнивание весов задач улучшает распределение. Отметим еще раз, что лишь в трех случаях из всех рассмотренных отклонения распределений от нормальных можно считать случайными.

Распределения результатов участников при удачных вариантах апостериорного назначения весов задач могут быть близки к нормальным. Наличие штрафов за непра-

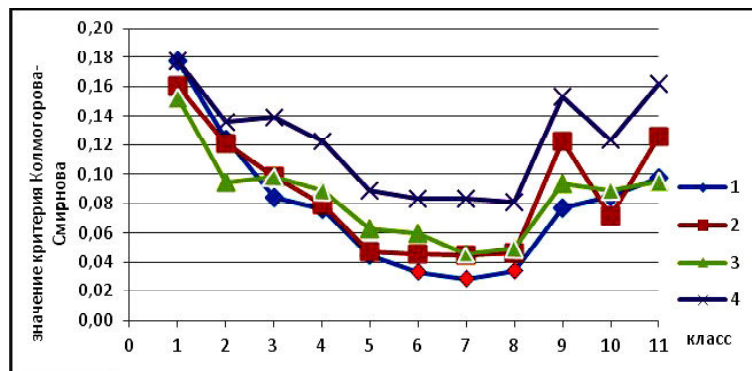


Рис. 1. Проверка на нормальность результатов конкурса при реальном и гипотетических способах подсчёта результатов. (приведены значения критерия Колмогорова-Смирнова для 1 – априорного назначения различных весов задачам и штрафов за неправильный ответ; 2 – априорного назначения различных весов задачам без штрафов за неправильный ответ; 3 – задач с одинаковыми весами со штрафами за неправильный ответ; 4 – задач с одинаковыми весами без штрафов за неправильный ответ. Красными маркерами отмечены те распределения, отклонения которых от нормальных можно считать случайными при $p < 0.05$).

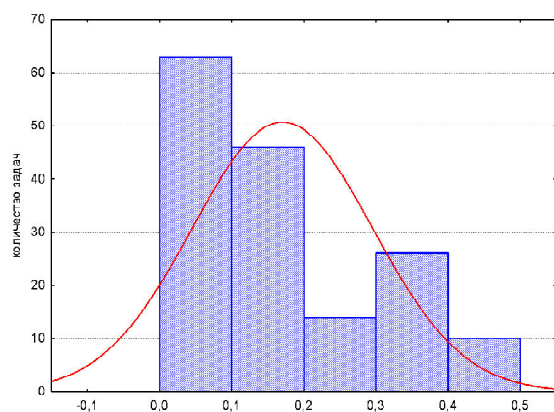


Рис. 2. Распределение задач конкурса по трудности

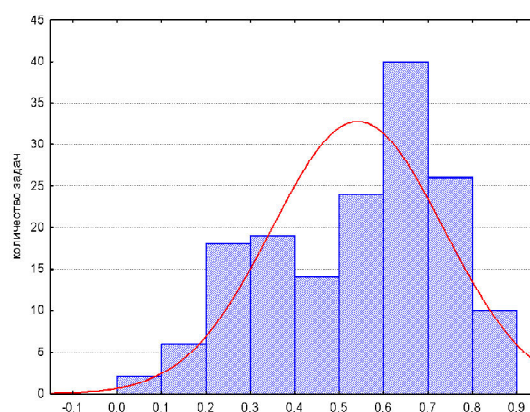


Рис. 3. Распределение задач конкурса по сложности

вильный ответ приближает распределение к нормальному.

Сложность и трудность задач. Распределения задач по **трудности** (шкала 2) и **сложности** (шкала 3) далеки от нормальных (рис. 2, 3). Коэффициент асимметрии распределения сложности задач положителен, а трудности задач – отрицателен, то есть в конкурсе преобладают задачи низкой **трудности**, при этом большая часть задач имеет высокую **сложность**.

В вариантах всех классов присутствуют задачи, **трудность** которых составляет менее 10 %. Задача самой низкой **трудности** обнаружена в варианте 8 класса (1, 1 %), самой высокой – в варианте 1 класса (47,6 %). **Сложность** задач в пределах вариантов каждого класса варьирует от 10–20 % до 80–90 %. Задача самой низкой **сложности** (7,7 %) обнаружена в варианте 2 класса, самой высокой – в варианте 7 класса (89,8 %). Примеры задач приведены на с. 3 обложки журнала.

Адекватность экспертной оценки сложности задач. Коэффициенты ранговой корреляции Спирмена экспертной оценки

сложности задач (шкалы 1) с их **сложностью** (шкалой 3) и **трудностью** (шкалой 2) для участников, вычисленные по всем задачам конкурса, достоверно положительны ($p < 0.01$), равны 0.56 и 0.60 соответственно. В таблице 1 приведены коэффициенты корреляции экспертной оценки сложности задач с их **трудностью** и **сложностью** для каждого класса. Для вариантов младших классов (1, 2, 3 классы) корреляции экспертной оценки сложности с **трудностью** задач не выявлено. Для классов, начиная с четвертого, имеется достоверная положительная корреляция. Наилучшее совпадение экспертной оценки со **сложностью** задач обнаружено в 1–2 классах. В вариантах 7–8 и 11 классов оценка сложности задач организаторами не соответствовала реальной **сложности** задач для школьников.

Классификация задач. На основании сопоставления значений **сложности** и **трудности** задач при помощи кластерного анализа выделено 4 кластера (рис. 4). Два кластера включают задачи высокой **сложности**, два другие – задачи низкой и средней **сложности**. Среди задач высокой **сложности** выде-

Табл. 1. Коэффициенты ранговой корреляции Спирмена экспертной оценки сложности задач каждого класса с их **сложностью** и **трудностью**.

Коэффициенты корреляции, значимые при $p < 0.05$, выделены жирным шрифтом.

Класс	1	2	3	4	5	6	7	8	9	10	11
Трудность задач	0,41	0,46	0,51	0,67	0,53	0,76	0,76	0,76	0,63	0,79	0,84
Сложность задач	0,74	0,8	0,62	0,64	0,59	0,57	0,28	0,25	0,62	0,66	0,49

Табл. 2. Распределение задач каждого класса по выделенным кластерам (в клетках таблицы указано количество задач)

Класс	низкая трудность			высокая трудность
	низкая сложность	средняя сложность	высокая сложность	высокая сложность
1	3	1	1	7
2	3	2	3	4
3	1	4	10	0
4	4	2	9	0
5	1	6	5	3
6	4	4	4	3
7	3	4	4	4
8	4	3	4	4
9	1	2	5	7
10	1	3	3	8
11	1	1	4	9

лен кластер задач высокой **трудности** и кластер задач низкой **трудности**. Все задачи низкой и средней **сложности** имеют низкую **трудность**. Распределение по кластерам задач разных классов приведено в табл. 2.

Связь длины условия задачи с её сложностью и трудностью. В табл. 3 приведены коэффициенты корреляции числа символов в условии задачи (шкала 4) с её **сложностью** и **трудностью** (шкалы 2 и 3). В младших классах (с первого по четвёртый) выявлена достоверная ($p < 0.01$) положительная связь длины условия задачи и её **трудностью** (числом отказов от решения). При этом связь длины условия задачи с её объективной **сложностью** обнаружена лишь для задач 3–4 класса.

Возрастные различия в восприятии задач и в результатах конкурса. В каждом

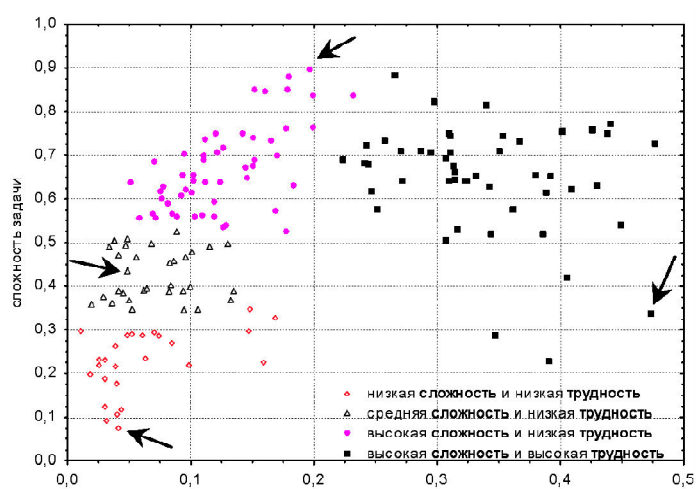


Рис. 4. Кластеры задач, выделенные на основании сопоставления их трудности и сложности (стрелками отмечены задачи, приведенные на с. 3 обложки журнала)

из первых пяти уровней конкурса участвовали школьники двух классов (табл. 4). На рис. 4 приведены результаты сравнения младших и старших школьников в каждой паре классов. На каждом уровне конкурса (то есть среди школьников, решавших одни и те же задачи) результаты младших участников ниже (рис. 5 а). При этом младшие школьники на первом-втором уровнях конкурса оценивают трудность задач (то есть выбирают вариант «нет ответа», рис. 5 б) как более высокую, а на четвёртом-пятом – как более низкую. Сложность задач (количество неправильных ответов, 5 в) оказывается выше для младших школьников на всех уровнях конкурса кроме первого. Отметим, что сравнивать результаты школьников разных уровней некорректно, поскольку и количество, и наборы задач на разных уровнях отличались.

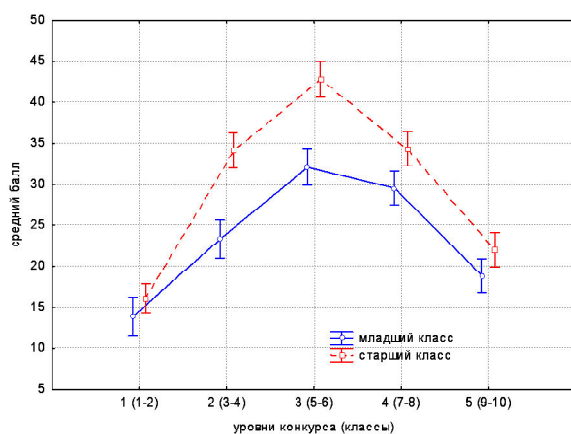
Табл. 3. Коэффициенты ранговой корреляции Спирмена длины условия задач с их **трудностью** и **сложностью**. Коэффициенты корреляции, значимые при $p < 0.05$, выделены жирным шрифтом

Класс	1	2	3	4	5	6	7	8	9	10	11
Трудность задач	0,90	0,83	0,77	0,78	0,17	0,13	–0,04	–0,06	0,10	0,04	–0,14
Сложность задач	0,37	0,36	0,77	0,78	0,29	0,32	–0,18	–0,20	–0,13	–0,01	–0,23

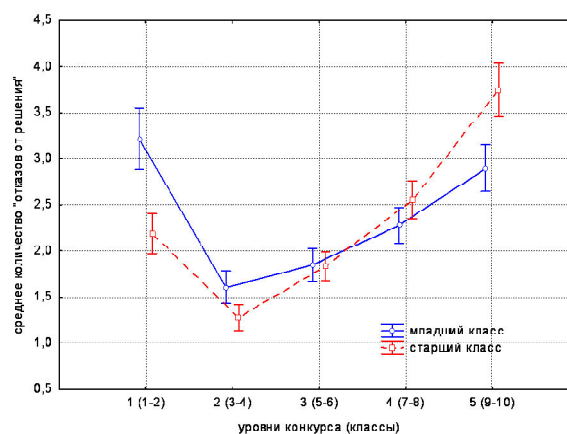
Табл. 4. Распределение задач каждого класса по кластерам.
В клетках таблицы указаны номера задач.
Изменения сложности или трудности задач показаны стрелками

Уровень	Класс	Низкая трудность			Высокая трудность
		Низкая сложность	Средняя сложность	Высокая сложность	Высокая сложность
1	1	3-4-7	6	1	2-5-8-11-9-10-12
	2	3-4-7	6-1	9-10-12	2-5-8-11
2	3	2	4-3-5-10	1-6-8-9-11-12-13-14-15-7	
	4	2-3-5-10	4	1-6-8-9-11-12-13-14-15	
3	5	10	2-7-15-3-4-8	1-6-9-14-5	11-12-13
	6	10-3-4-8	2-7-15-5	1-6-9-14	11-12-13
4	7	10-2-12	1-3-6-4	5-7-8-9	11-13-14-15
	8	10-2-12-4	1-3-6	5-7-8-9	11-13-14-15
5	9	1	5-7	2-6-8-3-9	4-10-11-12-13-14-5
	10	1	5-7-3	2-6-8	4-10-11-12-13-14-15-9

а)



б)



в)

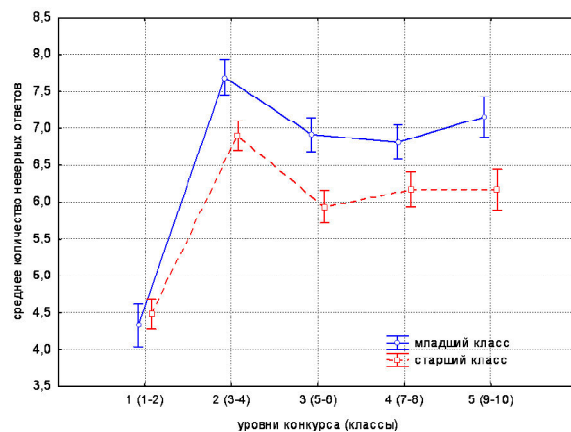


Рис. 5. Различия результатов младших и старших учеников на каждом уровне конкурса:

- а) различия среднего балла,
б) различия в оценке трудности задач,
в) различия в оценке сложности задач.
(уровень значимости – 5 %)

В табл. 4 показано распределение задач каждого класса по выделенным кластерам. Сопоставляя младший и старший класс в пределах уровней конкурса, замечаем, что все задачи (за единственным исключением) для старших школьников имеют равную или меньшую сложность и трудность. Только девятая задача пятого уровня конкурса для младших участников (девятиклассников) имеет низкую трудность, а для старших участников (десятиклассников) – высокую трудность, сложность этой задачи для всех участников одинаковая – высокая.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

О валидности измерительной процедуры. Как по распределению результатов конкурса, так и по **трудности** и **сложности** задач в вариантах конкурса можно заключить, что в конкурсе преобладают сложные для школьников задачи. Наиболее адекватен уровню участников вариант 7–8 класса – только в нём задачи высокой **сложности** и **трудности** составляют не более половины всех задач. Как следствие, именно в 7–8 классе распределение результатов конкурса наиболее близко к нормальному. В вариантах, предназначенных для всех остальных классов, сложность задач превышает возможности участников.

Среди разных вариантов подсчёта баллов, предполагающих априорное назначение весов заданий, выбранный организаторами способ (задания разной стоимости, штрафные баллы за неправильный ответ) оптимален для данного комплекта задач для всех классов, начиная с третьего. Для учеников самых младших классов введение штрафов и разбиение задач на группы сложности является ненужным усложнением, которое лишь ухудшает и без того плохое распределение результатов.

Благодаря значительной правой асимметрии распределения суммарных баллов конкурса, определение победителей конкурса и ранжирование наиболее сильных участников происходит успешно. Для основной массы участников ранжирование является грубым. Комплект заданий, использованный

в конкурсе, был бы уместен, если бы целью конкурса был отбор наиболее сильных учеников. Поскольку конкурс «Бобёр» – массовый, нацеленный на повышение интереса к предмету, рассчитан, в первую очередь, на учеников общеобразовательных школ, комплект заданий следует признать не самым удачным.

Экспертная оценка задач. Как указано выше, большинство заданий оказались сложными для большей части участников конкурса. По результатам сопоставления экспертной оценки задач с их **сложностью** и **трудностью** можно сделать ряд предположений о причинах этого. Для задач, предложенных младшим школьникам, корреляция экспертной оценки со **сложностью** задач высока, а с **трудностью** – не значима. Чем старше школьники, тем точнее совпадала оценка задачи организаторами с установленной по протоколам конкурса её **трудностью**. Однако корреляции экспертной оценки со **сложностью** задач старших классов имеются не всегда. Это означает, что эксперты не достаточно точно оценивают **трудность** задач для младших школьников и **сложность** – для старших. Наши результаты отчасти подтверждают мнение Навона и Гофера [11] о том, что одним из компонентов трудности задачи является длина её условия. В младших классах трудность задачи существенно повышалась с длиной её условия. Возможно, именно этот фактор был не учтён экспертами при оценке задач младших классов, что и привело к составлению варианта завышенной сложности. Мы предполагаем, что в младших классах именно длина условия могла быть тем фактором (недооценённым организаторами), который обусловил большое число отказов от решения простых по сути задач, исказив тем самым результаты измерения знаний и умений младших школьников. Младшие школьники из-за более низкого уровня развития психических процессов плохо понимают длинный текст. Дополнительным усложняющим фактором могут быть особенности интерфейса конкурса. Задачи с длинным текстом могут не помещаться на одном экране компьютера. В этом слу-

чае для чтения с экрана требуется, во-первых, ряд навыков работы с компьютером (знать, что такое скроллер вертикальной прокрутки, и уметь им пользоваться), во-вторых, определённая сформированность мелкой моторики (работа мышью).

Для старших школьников длина текста уже не является существенным усложняющим фактором. К тому же в процессе обучения накапливается личный опыт по оценке трудности задач «на глаз» и вырабатываются стереотипы «трудной» и «простой» задачи. Поэтому задача, которую эксперты оценивали как трудную, таковой и оказывается для участников – старшеклассников.

Ошибившись с оценкой **трудности** задач младших школьников, эксперты именно для них наиболее точно оценили **сложность** задач. А вот в старших классах объективная **сложность** задач не совпала с априорным мнением о ней у организаторов. Объективная **сложность** задачи, вероятно, во многом определяется знаниями школьников. Знания старшеклассников-участников конкурса в области информатики были переоценены организаторами конкурса. Складывается впечатление, что эксперты при отборе и оценке задач конкурса ориентировались на условного ученика средней школы.

Классификация задач. В конкурсе преобладают задачи низкой **трудности**. Школьники более склонны давать ответ на задачи, чем выбирать «нет ответа». Как было показано выше, выбор варианта «нет ответа» не даёт преимущества в баллах перед простым угадыванием. Тем более интересны задачи, в которых от ответа всё-таки отказалось много школьников – это задачи второго кластера. Их оказалось больше всего в вариантах младших и старших классов. Мы полагаем, что этот кластер состоит из «нестандартных задач» – пугающих школьников своим видом. Вероятность дать правильный ответ в этих задачах интуитивно оценивается как чрезвычайно низкая (что не соответствует действительности – эта вероятность не менее j). Три других кластера – это задачи более понятные и/или привычные для школьников. Самые простые из них (задачи низ-

кой **трудности** и низкой **сложности**) – это задачи, называемые на олимпиадном жаргоне «утешительными», их решение доступно практически каждому участнику. Наличие в конкурсе таких задач обеспечивает позитивное ощущение от участия в конкурсе даже у тех, кто решил мало задач. Таких задач в конкурсе было предложено очень мало.

Возрастные различия в восприятии задач. Приведённые результаты показывают, что разница 1 год в возрасте участников вызывает существенные различия в восприятии одних и тех же задач. В младших возрастах более младшие школьники склонны оценивать задачи как более трудные и чаще, чем более старшие выбирать вариант «нет ответа». В старших возрастах младшие школьники, напротив, склонны оценивать задачи как менее трудные. Сложность задач для более младших школьников в каждой паре оказывается объективно выше, что проявляется как большее число неправильных ответов и, закономерно, приводит к более низким результатам.

ВЫВОДЫ

1. Задачи, предложенные на конкурсе «Бобёр»-2012, позволяют произвести отбор наиболее сильных школьников, но не позволяют дифференцировать основную массу участников. Разная стоимость задач и наличие штрафов за неправильный ответ повышают качество измерительной процедуры для 3–11 классов, но не делают её достаточно точной.

2. При подготовке задач конкурса организаторами не были учтены возрастные особенности школьников младших классов, в частности их низкая способность воспринимать длинные тексты. Также был переоценён уровень знаний школьников старших классов.

3. Предложенная процедура классификации задач позволяет выявлять нестандартные задачи.

4. Разница в один год обеспечивает существенные различия в восприятии задач школьниками и в их результатах.

РЕКОМЕНДАЦИИ ОРГАНИЗАТОРАМ КОНКУРСА

- Конкурс в целом следует упростить: во всех классах увеличить количество «утешительных» задач. Нестандартные задачи в младших и средних классах следует упростить, к их формулировкам отнестись с большим вниманием. В младших классах следует избегать задач с длинными текстовыми условиями.
- Обдумать целесообразность наличия варианта «нет ответа». Возможно, стоит изменить соотношение начисляемых баллов и штрафов или количество предлагаемых ответов в задаче, чтобы выбор варианта «нет ответа» имел прагматический смысл.
- Обдумать возможность апостериорного определения стоимостей задач с сохране-

нием системы штрафов за неправильные ответы. Такой подход даст позитивный образовательный эффект – «дорогие» задачи не будут сразу отпугивать менее уверенных в своих силах участников. Одновременно, апостериорное определение стоимостей задач скомпенсирует ошибки, сделанные экспертами при их априорном назначении.

- Если школьники разных классов решают один и тот же комплект задач, то подведение итогов и выявление победителей проводить отдельно для каждого класса.

- В протоколах конкурса учитывать не только время ввода ответа, но и время, затраченное школьником на решение задачи (проведённое на странице с задачей). Это даст возможность альтернативным путём оценить **трудность** задач.

Литература

1. Балл Г.А. Теория учебных задач: Психолого-педагогический аспект. М., 1990.
2. Голиков Ю.Я., Костин А.Н. Психология автоматизации управления техникой. М.: Институт психологии РАН, 1996.
3. Кантовиц Б., Соркин Р. Человеческий фактор. М.: Мир, 1991. Т. 4. С. 85–113.
4. Кирюхин В.М. Методические рекомендации по разработке заданий для школьного и муниципального этапов всероссийской олимпиады школьников по информатике в 2012/2013 учебном году. М., 2012.
5. Кротов В.М. К вопросу о сложности (трудности) физических задач // Фізика: проблеми викладання, 1999. № 3. С. 69–74.
6. Леонтьев А.Н. Деятельность. Сознание. Личность. М.: Политиздат, 1975.
7. Пиаже Ж. Психология интеллекта. Пер.: А. М. Пятигорский. СПб.: 2003. [Электронный ресурс] // Центр гуманитарных технологий. URL: <http://gtmarket.ru/laboratory/basis/3252/3257> (дата обращения 20.12.2013).
8. Полозова Т.А. Развитие самооценки в детском возрасте // Советская Педагогика, 1973. № 11. С. 27–38.
9. Пономаренко В.А., Черняков Г.М., Кострица В.Г. Психические состояния оператора как предмет инженерно-психических исследований // Вопросы кибернетики, 1982. № 3. С. 131–137.
10. Шмелёв А.Г. Практическая тестология: тестирование в образовании, прикладной психологии и управлении персоналом. М.: 2013.
11. Navon D., Gopher D. On the economy of human information processing systems // Psychological Review, 1979. Vol. 86. P. 214–255.
12. Riley V., Lyall E. & Wiener E. Analytic workload model for flight deck design and evaluation // Proceedings of the Human Factors and Ergonomics Society, 1994. Vol. 38. P. 81–84.
13. Sammer G. Concepts of mental workload in psychophysiological research // IEA'97, Proceedings of the 13th Triennial Congress of the International Ergonomic Association, Tampere, Finland, 1997. Vol. 5. P. 368–370.

USE OF PROTOCOLS OF ONLINE COMPETITIONS FOR AN ASSESSMENT OF COMPLEXITY OF TASKS AND INCREASE OF A VALIDITY OF MEASURING PROCEDURE

Abstract

On the example of the analysis of protocols of the international online competition on informatics «Beaver-2012» some approaching to an assessment of complexity and difficulty of tasks is considered. It is shown that the assessment of complexity of tasks by organizers of competition the aprioristic doesn't correspond to objective difficulty of a task for the participant for a lot of tasks. It is shown that school students, especially younger, often underestimate complexity of tasks. The way of typology of problems of competition on the basis of the analysis of their difficulties and complexities for the school student is offered. Typology can be applied in other disciplines. The cluster of the tasks which complexity is significantly underestimated by organizers of competition is allocated. It is shown that in some cases complexity of a task for school students can be connected with features of the interface of competition. A number of shortcomings of the online protocols is revealed. Elimination of these shortcomings will allow to raise a validity of results of competitions.

Keywords: informatics, protocols, on-line competitions, development of tests, key to a test, complexity of tasks, difficulty of tasks, educational tests, typology of tasks, competition «Beaver».

*Ягунова Екатерина Борисовна,
старший научный сотрудник
СПбГЭТУ «ЛЭТИ»
им. В.И. Ульянова (Ленина),
преподаватель математики
ФНБУВПОиН Санкт-
Петербургского Академического
университета – научно-
образовательного центра
нанотехнологий Российской
академии наук, Лицей «Физико-
техническая школа»,
Katrin.home@mail.ru,*

*Рыжова Нина Евгеньевна,
ГК «Институт Тренинга – АРБ
Про», психолог, консультант
направления HR-консалтинга,
ryzhova-nina@yandex.ru.*



Наши авторы, 2013.
Our authors, 2013.