

Дедикова Анна Олеговна

педагог дополнительного образования

МБОУ «Прогимназия №2»

г. Воронеж, Воронежская область

DOI 10.31483/r-106746

ЭКСПЕРТНОЕ МОДЕЛИРОВАНИЕ НАДЕЖНОСТИ ТЕСТОВЫХ ЗАДАНИЙ

Аннотация: в статье проанализированы результаты решения тестовых заданий шести видов, полученные методом экспертного моделирования. Модель тестируемого была представлена шестью уровнями обученности в зависимости от объемов изученного материала. Для каждого вида тестовых заданий определены вероятности угадывания верного ответа в зависимости от уровня обученности тестируемых. Проанализирована потенциальная устойчивость каждого вида тестовых заданий к ошибкам второго рода.

Ключевые слова: тестирование учебных достижений, единичные тестовые задания, достоверность оценивания знаний, ошибки второго рода, угадывание верных ответов, экспертное моделирование, вероятность.

Компьютерные тесты для оценивания учебных достижений обучающихся (КТОУДО) занимают ключевые позиции в инструментарии средств контроля текущей успеваемости, предварительной и промежуточной аттестации, контроля остаточных знаний, в том числе, в системах независимого мониторинга качества подготовки специалистов. Однако качество тестов зависит в большей степени от практического опыта их разработчика, нежели от полноты и глубины знания им содержания соответствующей предметной области. Данное обстоятельство обусловлено практическими трудностями интерпретации и анализа результатов тестирования в увязке с корректным априорным оцениванием уровня сложности тестов и достоверности результатов их применения. В основе решения таких проблем, как правило, лежит обеспечение некоторого требуемого уровня

надежности тестов, который в значительной мере определяет степень достоверности оценок тестирования.

Как надежность, так и достоверность оценок КТОУДО обеспечиваются надежностью и достоверностью оценок единичных тестовых заданий (ЕТЗ) в отдельности, поэтому очевидным решением обеспечения высокой достоверности оценок КТОУДО является увеличение количества ЕТЗ в выборке, что обеспечивает максимальный охват контролируемого содержания обучения и возможность варьирования пороговыми значениями оценочных критериев в широких пределах. Однако на практике увеличение количества ЕТЗ приводит к увеличению времени тестирования, что, в свою очередь, является фактором снижения надежности и достоверности результатов тестирования за счет повышения вероятности ошибок первого рода [1]. Лимитирование времени тестирования влечёт за собой ограничение количества ЕТЗ в их выборке, поэтому задача создания компактных и одновременно высоконадежных КТОУДО может быть решена за счет использования таких видов ЕТЗ, которые не только обладают повышенной устойчивостью к угадыванию верных ответов и временной деградации, но и характеризуются высоким дидактическим потенциалом.

В [2] приведены оценки надежности КТОУДО, составленных из различных видов ЕТЗ закрытого типа: с выбором единственного верного ответа, с множественным выбором, на установление соответствия, на установление последовательности, с градацией верных ответов, с упорядочиванием дистракторов. Модель испытуемого характеризуется несколькими уровнями обученности в соответствии с усвоенным объёмом контролируемого содержания обучения, что позволяет для каждого уровня вычислить вероятность получения положительной оценки с учетом угадывания неизвестных ответов. Каждый уровень обученности подразумевает не только знание определенного процента верных ответов, но и способность частичного исключения дистракторов (для ЕТЗ с выбором единственного верного ответа и множественным выбором), частичное знание верных ответов, например, одного из нескольких верных ответов в ЕТЗ с множественным

выбором или одного или двух парных соответствий в ЕТЗ на установление соответствий и т. п.

Для расширения верификационной базы исследований надежности тестов, составленных из однородных ЕТЗ, методов априорной оценки надежности и достоверности оценок КТОУДО, нами было проведено экспертное моделирование решения ЕТЗ различных видов по нескольким предметным областям.

На этапе подготовки ЕТЗ для экспертного моделирования их выполнения была решена задача сопоставимости данных для анализа выполнения ЕТЗ одного вида для трех предметных областей (истории России, географии, физики (электричество и магнетизм)), а также обеспечен единый уровень содержательной сложности для ЕТЗ различных видов в каждой из предметных областей. Выполнение данных условий было подтверждено экспертами, участвующими в оценивании вероятности решения ЕТЗ тестируемыми в соответствии с моделью их подготовленности.

В состав экспертной группы входило пять преподавателей-предметников по каждой предметной области, а также один методист и один психолог. Экспертное решение утверждалось исключительно преподавателями-предметниками путем голосования, а задачей методиста и психолога было обеспечение единых алгоритмов выработки экспертных оценок и корректности их обсуждения и принятия. Процедура принятия экспертного решения включала в себя выдвижение рабочей гипотезы, её обсуждение, выдвижение альтернативной гипотезы, обсуждение альтернативной гипотезы, при необходимости корректировка рабочей или альтернативной гипотезы, утверждение одной из гипотез в качестве экспертного решения. Подавляющее большинство (около 90%) решений было принято единогласно, в остальных случаях – при одном воздержавшемся или одном «против», что явилось следствием достижения консенсуса при обсуждении гипотез и их корректировке.

Модель подготовленности тестируемых была представлена шестью уровнями обученности со следующими характеристиками уровней:

А – тестируемый не владеет учебным материалом, в ответе использует остаточные знания, в том числе, из смежных предметных областей, общую эрудицию;

Б – тестируемый имеет лишь общее представление о предметной области, оперирует обрывочными, бессистемными сведениями по изучаемому предмету, использует общую эрудицию;

В – тестируемый обладает фрагментарными знаниями в предметной области в объеме примерно половины дидактических единиц изученного содержания;

Г – тестируемый знает большую часть учебного материала, уверенно решает типовые качественные задачи (простые и среднего уровня сложности);

Д – тестируемый знает весь учебный материал, однако испытывает затруднения в решении сложных качественных задач, в том числе, ориентированных на междисциплинарное содержание, ответах на вопросы, требующих углубленных знаний или хороших аналитических навыков;

Е – идеальная модель тестируемого, дающего верные ответы на все рассматриваемые ЕТЗ.

Общей задачей экспертного моделирования было выявление видов ЕТЗ, наиболее устойчивых к угадыванию тестируемыми верного ответа при его реальном незнании.

Виды рассматриваемых ЕТЗ [3–7]:

- 1 – на установление парных соответствий;
- 2 – с выбором единственного верного ответа;
- 3 – с множественным выбором верных ответов;
- 4 – на упорядочивание дистракторов;
- 5 – на установление последовательности;
- 6 – с градацией верных ответов.

Рассчитанные вероятности угадывания правильного ответа [7] для каждого вида ЕТЗ по рассматриваемым предметам и уровня обученности тестируемых приведены в таблице 1.

Таблица 1

Вероятности угадывания правильных ответов на единичное тестовое задание

Уровень подготовленности	Предмет	Типы заданий					
		1	2	3	4	5	6
А	история	0,0128	0,25	0,0667	0,0104	0,0104	0,0104
	география	0,0128	0,25	0,0667	0,0139	0,0139	0,0417
	физика	0,0167	0,25	0,0625	0,0417	0,0139	0,00833
Б	история	0,1	0,333	0,143	0,0139	0,0294	0,0139
	география	0,0278	0,333	0,125	0,0208	0,0417	0,167
	физика	0,0833	0,333	0,125	0,0833	0,0333	0,0833
В	история	0,0333	0,5	0,25	0,167	0,0625	0,25
	география	0,167	0,5	0,25	0,0833	0,167	0,25
	физика	0,25	0,5	0,25	0,25	0,05	0,0833
Г	история	0,5	0,5	0,5	0,25	0,25	0,5
	география	0,25	1	0,5	0,25	0,25	0,5
	физика	0,5	0,5	0,5	0,5	0,125	0,25
Д	история	0,5	1	0,5	0,5	0,5	1
	география	0,5	1	1	0,5	0,5	1
	физика	1	1	1	0,5	0,25	0,25
Е	история	1	1	1	1	1	1
	география	1	1	1	1	1	1
	физика	1	1	1	1	1	1

Расчетные результаты хорошо согласуются с полученными ранее априорными оценками вероятности ошибок второго рода для тестов, составленных из однородных ЕТЗ различных видов [2].

Анализ данных, представленных в таблице 1, показывает, что для задания в виде вопроса с выбором единственного верного ответа уровень обученности «Д» практически гарантирует верное решение. Для данного вида ЕТЗ по географии верный ответ, по мнению экспертов, обеспечивался уровнем «Г». Несмотря на достаточное количество дистракторов (четыре) ЕТЗ с выбором единственного верного ответа оказались подвержены угадыванию верного ответа обучающимися, чьи модели обученности характеризуются неполнотой и даже фрагментарностью усвоения тестируемого содержания. Усложнение ЕТЗ с множественным выбором за счет заранее не известного тестируемым количества верных ответов не привело к существенному снижению вероятности угадывания верных ответов. По мнению экспертов, ЕТЗ на установление последовательности и с упорядочиванием дистракторов ни по одной из предметных областей не могли быть

корректно решены обучающимися, чей уровень подготовки позволял бы им получать преимущественно отличные оценки по соответствующим учебным дисциплинам. Данный результат объяснен экспертами исключительно специфической структурой данных ЕТЗ, требующих от обучающихся одновременно конкретных знаний и понимания их содержательной и логической взаимосвязи. Обеспечение трудности ЕТЗ с выбором единственного ответа (вид 2), с множественным выбором (вид 3) и на установление парных соответствий (вид 1), адекватной трудности ЕТЗ с упорядочиванием дистракторов (вид 4) и на установление последовательности (вид 5), оказалось возможным лишь за счёт повышения содержательной сложности заданий. Трудность ЕТЗ с градацией верных ответов (вид 6) и с упорядочиванием дистракторов (вид 4) оказалась зависящей от предметной области (в естественно-научной предметной области трудность данного вида ЕТЗ признана наивысшей). По мнению разработчиков тестов и экспертов, имеет место корреляция между сложностью разработки ЕТЗ и сложностью их решения, что особенно характерно для ЕТЗ видов 5 и 6.

Результаты экспертного моделирования могут быть полезны для специалистов – практиков разработки КТОУДО различного целевого назначения.

Список литературы

1. Булгаков О.М. О применимости методологического аппарата теории надежности к оценке качества тестов для проверки знаний / О.М. Булгаков, А.О. Дедикова // Вестник Воронежского института ФСИН России. – 2017. – №4. – С. 214–221.

2. Булгаков О.М. Модели оценки качества тестов для контроля знаний / О.М. Булгаков, И.Н. Старостенко, А.А. Хромых [и др.]. – Краснодар: Краснодарский университет МВД России, 2021. – 138 с.

3. Булгаков О.М. Тестирование остаточных знаний обучающихся / О.М. Булгаков, А.И. Ладыга. – Краснодар: Краснодарский университет МВД России, 2021. – 146 с.

4. Аванесов В.С. Теория и практика педагогических измерений (материалы публикаций в открытых источниках и Интернет) / В.С. Аванесов. – Екатеринбург: ЦТ и МКО УГТУ-УПИ, 2005. – 98 с.

5. Ефремова Н.Ф. Тестовый контроль в образовании: учебное пособие / Н.Ф. Ефремова. – М.: Логос, 2005. – 368 с.

6. Ким В.С. Тестирование учебных достижений / В.С. Ким. – Уссурийск: Издательство УГПИ, 2007. – 214 с.

7. Булгаков О.М. Тестирование учебных достижений: от проверки знаний к проверке понимания / О.М. Булгаков, А.О. Дедикова // Вестник Санкт-Петербургского университета МВД России. – 2020. – №2 (86). – С. 183–190.

8. Гмурман В.Е. Теория вероятностей и математическая статистика: учебник для прикладного бакалавриата / В.Е. Гмурман. – 12-е изд. – М.: Юрайт, 2017. – 479 с.