

РЕГИСТРАЦИОННО-РАСЧЕТНЫЙ МЕТОД ИЗМЕРЕНИЯ УЧЕБНЫХ ДОСТИЖЕНИЙ ПО РЕЗУЛЬТАТАМ ТЕСТИРОВАНИЯ

Андрейченко В.А.

*Андрейченко Владимир Александрович – кандидат технических наук, доцент,
кафедра информатики, факультет пограничной деятельности,
Голицынский пограничный институт,
г. Голицыно*

Аннотация: в статье изложен метод реализации педагогических измерений, в котором используется циклический итерационный процесс тестирования учебных достижений. Для тестирования формируются батареи калиброванных по сложности тестовых заданий. Цикл тестирования заканчивается при достижении требуемой точности соответствия трудности тестовых заданий уровню обученности участвующих в тестировании обучающихся.

Ключевые слова: педагогические измерения, батарея тестов, шкалирование учебных достижений, педагогическая диагностика, статистическая теория тестов.

Введение

Не вызывает сомнений, что оценка учебных достижений обучающихся является важной и трудоемкой частью профессиональной деятельности преподавателей и образовательных учреждений в целом. В сложившейся практике с этой целью используют педагогические измерения и современный инструмент – компьютерное тестирование [1]. Одна из главных причин – повышение объективности оценки. Однако именно здесь возникает педагогическая проблема: как по результатам выполнения отдельных заданий получить объективную итоговую оценку? Проблема состоит в формулировании педагогически обоснованного правила подведения итогов и их сравнения. Решение этой проблемы частично облегчится, если решен вопрос о возможности работы с оценками, хотя бы по правилам арифметики (вычитать, складывать, делить для получения итоговых оценок и их сравнения). Обычно правила выставления итоговой оценки весьма субъективны и традиционно сводятся к суммированию баллов за правильно выполненные задания без учета или с учетом их трудности. Однако балльная оценка трудности заданий в плане объективности представляет отдельную проблему, а полученные в результате тестирования суммарные баллы, как с учетом трудности заданий, так и без него, можно только ранжировать, так как баллы относят к калибровке порядковых, а не метрических шкал. Сложение, вычитание или деление оценок (например усреднение) корректным признать нельзя, так как балльная оценка не имеет числовой природы, а является лишь обозначением того или иного качественного уровня учебных достижений. Поэтому возникает вопрос: чем можно заменить в тестировании учебных достижений балльное оценивание так, чтобы обработка этих оценок традиционными математико-статистическими методами стала корректной? Известно, что ответ на поставленный вопрос даст статистическая теория тестов, или IRT (Item Response Theory) в англоязычной терминологии [2].

Предлагаемый метод измерения учебных достижений основан на тестировании учебных достижений с использованием батареи калиброванных по возрастанию сложности наборов тестовых заданий и применении известной в IRT статистической модели Раша для интерпретации результатов тестирования с обеспечением гарантий валидности и достаточной надежности тестовых заданий. Интерпретация результатов тестирования с использованием модели Раша заключается в отождествлении уровня обученности тестируемого с такой трудностью заданий, для которой получены равные доли правильных и не правильных ответов. Представленное в статье описание метода содержит теоретическое основание, условия применения и спецификацию тестового процесса, включая подготовку тестовых заданий и обработку результатов их выполнения.

Теоретическое основание

Согласно однопараметрической модели Георга Раша, практикуемой в педагогической диагностике и измерениях, расчет вероятности правильного ответа тестируемого с определенным уровнем обученности на задание определенной трудности выполняют по формуле[2]:

$$P(\tau, \theta) = \frac{1}{1 + e^{(\tau - \theta)}} , \quad (1)$$

где θ - уровень обученности в логитах;

τ - трудность задания в логитах.

Формально логит обученности определяется натуральным логарифмом отношения доли правильных к доле неправильных ответов на тестовые задания, а логит трудности задания наоборот – логарифмом отношения доли неправильных к доле правильных ответов на задание.

Очевидно, что функция вероятности правильного ответа (1) принимает значение 0,5 при равенстве уровня обученности трудности задания. Это равенство определяет граничное значение трудности менее и более сложных для тестируемого тестовых заданий. Именно это граничное значение принято считать уровнем обученности тестируемого в приложениях модели Раша к педагогическим измерениям учебных достижений. Найденное значение будет соответствовать точке перегиба графика функции вероятности успеха (1) при монотонном изменении трудности заданий в большую или меньшую сторону.

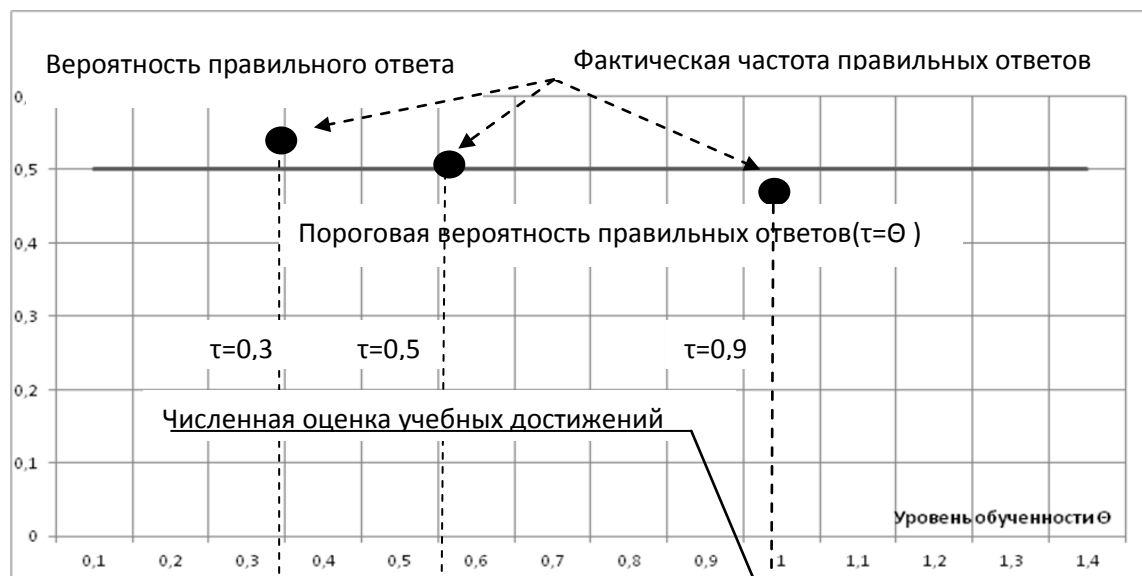


Рис. 1. Данные оценки учебных достижений

Исходя из того, что в модели Раша вероятность правильного ответа 0,5 означает равную возможность правильного и неправильного ответа, уровнем обученности тестируемого будем считать трудность задания, при многократном выполнении которого будут получены практически одинаковые доли правильных и неправильных ответов (рис. 1).

Таким образом, тестирование учебных достижений сводится в идеале к выявлению задания, на которое тестируемый дает одинаковое количество правильных и неправильных ответов при многократном повторении. Трудность этого задания в логитах фиксирует величину проявленного на тестах уровня обученности тестируемого.

Условия применения

В тестовом процессе надо использовать батарею тестов из нескольких упорядоченных по возрастанию трудности предварительно калиброванных наборов тестовых заданий. Каждый набор должен содержать одинаковые по трудности задания, а задания разных наборов отличаться по трудности. Определение трудности и оценку надежности заданий следует провести путем предварительного тестирования (предтестирования) как минимум в двух итерациях. Калиброванные по результатам предтестирования задания следует подвергнуть педагогической обработке. В результате педагогической обработки трудность заданий батареи тестов должна покрывать шкалу логитов в достаточном для оценивания учебных достижений диапазоне. Сложность (в логитах) тестовых заданий того набора, на котором доли правильных и неправильных ответов одинаковы в пределах требуемой точности, будет соответствовать уровню обученности (в логитах) прошедшего тестирование обучающегося. Точность определяется исходя из количества правильных ответов, которое можно признать несущественным для оценки учебных достижений, например, в пределах 10% от общего количества вопросов (заданий). Если требуемая точность не достигнута на текущей батарее тестов, то уровень обученности приравнивается к трудностного набора тестовых заданий, доля правильных ответов на которые наиболее близка к 0,5 относительно других наборов.

Спецификация тестового процесса

Процесс: измерение учебных достижений.

Входные данные: требования к результатам обучения, точность измерения (ε).

Выходные данные: уровень учебных достижений (θ) в логитах.

1. Начало: сформировать наборы валидных тестовых заданий в соответствии с требованиями к результатам обучения.

2. Предтестирование: определить трудность и надежность тестовых заданий по результатам ответов как минимум двух групп обучающихся.

3. Провести компановку достаточно надежных тестовых заданий в наборы тестов из заданий одинаковой трудности.

4. Сформировать батарею тестов из наборов, упорядоченных по возрастанию трудности тестовых заданий.

5. Начать тестирование: выбрать набор заданий наименьшей трудности.

6. Установить начальное значение счетчика тестовых заданий в наборе $j=1$.

7. Установить начальное значение счетчика наборов в батарее $i=1$.

8. Установить начальное значение счетчика правильных ответов $r_i=0$ на наборе тестов i .

9. Выбрать задание j из набора тестов i с трудностью τ_j .

10. Получить результат выполнения задания.

11. Проверить правильность выполнения задания: если результат верный, увеличить счетчик правильных ответов $r_i=r_i+1$ на наборе тестов i ;

12. Проверить окончание набора: если выполненное задание последнее в наборе, перейти к п. 14, иначе увеличить счетчик заданий $j=j+1$.

13. Перейти к п.9.

14. Проверить: $\frac{1}{2}-\varepsilon \leq r_i/j \leq \frac{1}{2}+\varepsilon$: если да, запомнить уровень учебных достижений $\theta = \tau = \tau_i$ и перейти к п. 17, иначе – запомнить частоту правильных ответов на текущем наборе $\theta_i = r_i/j$.

15. Проверить окончание батареи тестовых заданий: если набор не последний, продолжить итерацию на следующем наборе: $i=i+1, j=1$, перейти к п.8.

16. Рассчитать $\theta = \text{Argmin}(0,5 - \theta_i)$.

17. Конец: показать результат θ .

Для верификации полученного результата предлагается следующая тестовая процедура:

1. Выбрать по одному заданию из каждого набора по возрастанию трудности.

2. Выполнять задания до первой ошибки.

3. Назначить уровню обученности трудность последнего в серии безошибочно выполненных заданий.

Если полученный в ходе фактического и проверочного тестирования результат будет отличаться более чем на 10%, достоверность результатов скорее низкая, чем высокая. В этом случае есть повод повторить тестирование на батарее более точно подобранных по трудности наборов тестов.

В случае если по п. 3 основного процесса не удастся отобрать равные по сложности задания или их количество слишком мало, то вместо нескольких наборов следует сформировать один набор. При наличии одного набора упорядоченных по трудности тестовых заданий в ходе тестирования надо установить трудность того из них, которое будет последним в серии безошибочно выполненных заданий.

Пример. Будем считать, что на основе изучения требований к освоению учебной дисциплины разработано 16 валидных тестовых заданий. Всем заданиям присвоен учетный номер и отобраны две группы из десяти и шести человек для предтестирования. Каждая группа получила по восемь разных заданий. По результатам предтестирования заполнены таблицы, куда для каждого задания внесены сведения о трудности и педагогическому отбору тестовых заданий (таб. 1, 2).

Таблица 1. Предтестирование: первая итерация

Учетный номер задания	Всего участников тестирования (n)	Количество правильных ответов (p)	Трудность задания $\ln(n-p)/p$	Педагогический отбор
1	10	0	очень трудное	исключить
2	10	3	0,84729786	
3	10	2	1,386294361	
4	10	6	-0,405465108	
5	6	6	очень легкое	исключить
6	6	4	-0,693147181	
7	6	2	0,693147181	
8	6	5	-1,609437912	
9	10	6	-0,405465108	
10	10	7	-0,84729786	

11	6	3	0	
12	6	3	0	
13	6	6	очень легкое	исключить
14	6	5	-1,609437912	
15	10	8	-1,386294361	
16	10	7	-0,84729786	

Таблица 2. Предтестирование: вторая итерация

Номер задания	Всего участников тестирования (n)	Количество правильных ответов (p)	Трудность задания $\ln(n-p)/p$	Педагогический отбор
1	6	0	очень трудное	исключить
2	6	2	0,693147181	надежно
3	6	1	1,609437912	надежно
4	6	3	0	исключить
5	10	9	очень легкое	исключить
6	10	4	0,405465108	исключить
7	10	3	0,84729786	надежно
8	10	8	-1,38629436	надежно
9	6	4	-0,69314718	исключить
10	6	5	-1,60943791	исключить
11	10	5	0	надежно
12	10	5	0	надежно
13	10	10	очень легкое	исключить
14	10	8	-1,38629436	надежно
15	6	5	-1,60943791	надежно
16	6	6	очень легкое	исключить

В результате педагогического отбора по двум итерациям предтестирования одно задание исключено как слишком трудное, два – как слишком легкие и пять заданий признаны малонадежными.

Таблица 3. Задания, прошедшие педагогический отбор

Номер задания	Номер задания до отбора	Трудность задания
1	15	-1,60943791
2	8	-1,38629436
3	14	-1,38629436
4	11	0
5	12	0
6	2	0,693147181

7	7	0,84729786
8	3	1,609437912

В данном случае к малонадежным отнесены те задания, трудность которых во второй итерации отличается более, чем на 25% от трудности в первой итерации. Величина порога может регулироваться.. Малонадежные задания подлежат исключению из списка. Оставшиеся после предтестирования, упорядоченные по трудности и заново пронумерованные тестовые задания показаны в таб. 3.

Из оставшихся заданий можно скомпоновать батарею только из двух наборов практически одинаковых по трудности заданий: в первый набор войдут задания № 2 и № 3, а во второй набор – задания № 4 и № 5.

Выборочные результаты тестирования учебных достижений одного из обучающихся на сформированной батарее тестов показаны в таб. 4.

Таблица 4. Результаты тестирования

Номер набора	Трудность задания (логит)	Доля правильных ответов
1	-1,386294361	1
2	0	0,5

Из таб. 4 видно, что учебные достижения соответствуют нулю логит. В отличие от балльной оценки, данный показатель имеет числовую природу и может корректно использоваться в вычислениях.

Таким образом, применение данного метода измерения учебных достижений основано на статистической модели Раша и опирается на организацию итерационного тестового процесса, включающего повторяющихся фазы предтестирования и тестирования учебных достижений. Целью предтестирования является отбор тестовых заданий: чем равномерней и чаще будут распределены по трудности задания, тем точнее можно будет измерить учебные достижения. Измерение учебных достижений сведено к отождествлению путем регистрации и расчетов уровня обученности с трудностью заданий, доля правильных ответов на которые как можно ближе к половине.

Список литературы

1. Звонников В.И., Челышкова М.Б. Современные средства оценивания результатов обучения : учеб. пособие для студ. вузов, обуч. по пед. спец. М.: Академия, 2008. 224 с.
2. Нейман Ю.М. Введение в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В.А. Хлебников. М., 2000. 168 с.