

МЕТОДЫ МОДЕЛИРОВАНИЯ УСПЕШНОСТИ ВЫПОЛНЕНИЯ ТЕСТОВЫХ ЗАДАНИЙ ПРИ ОЦЕНКЕ НЕСКОЛЬКИХ КОМПЕТЕНЦИЙ

А. И. Безруков¹, В. Н. Гусятников¹, Л. В. Грахольская²

¹*Саратовский государственный технический университет, Россия*

²*Поволжский институт управления – филиал РАНХиГС, Саратов, Россия*

E-mail: bezr_Alex@mail.ru, victorgsar@rambler.ru, graholskayalv@yandex.ru

Рассматриваются методы моделирования успешности выполнения тестовых заданий в случае оценки уровня сформированности трех компетенций за один сеанс тестирования. Вероятность удачного выполнения задания зависит от трудности и характера задания, а также уровня подготовленности испытуемого. Оценка нескольких компетенций предполагает, что трудности задания с точки зрения каждой из компетенций различны. Смысловое наполнение компетенций не пересекается. Уровни сформированности различных компетенций, полученные с помощью имитационной модели.

METHODS FOR MODELING SUCCESS IN PERFORMING TESTS WHEN ASSESSING SEVERAL COMPETENCES

A. I. Bezrukov, V. N. Gusyatnikov, L. V. Graholskaya

Methods of modeling the success of test tasks in the case of three competencies are considered. The probability of successful completion of the task depends on the difficulty and nature of the task, as well as the level of preparedness. The assessment of several competencies suggests that the difficulties of the task from the point of view of each of the competencies are different. The semantic content of competencies does not overlap. The levels of formation of various competencies are obtained using a simulation model.

Компетентность – способность применения полученных знаний и умений для решения практических задач. В отечественных образовательных стандартах принято рассматривать компетентность как совокупность нескольких компетенций. В соответствии с компетентностным подходом, качество образования, полученное учащимся, оценивается по уровню сформированности у него требуемых компетенций. Уровень знаний, умений или навыков, в принципе, можно оценить, предложив учащемуся выполнить задание, условия которого специально подобраны для решения данным методом. Для оценки сформированности компетенций мы должны знать, насколько учащийся умеет:

- понять проблему, описанную в задании;
- выбрать метод, позволяющий решить эту проблему (т.е. превратить проблему в задачу);
- решить задачу выбранным методом;
- оценить, насколько полученное решение применимо для решения данной проблемы;
- сформулировать решение рассматриваемой проблемы.

Компетенции развиваются на протяжении всего процесса обучения и, как правило, при изучении каждой дисциплины формируются несколько компетенций. Поэтому инструменты контроля качества обучения (в том числе и компьютерное тестирование) должны обеспечивать оценку уровня сформированности сразу нескольких компетенций.

В работе [1] предложен алгоритм компьютерного тестирования, позволяющий оценивать уровень сформированности сразу трех компетенций за один сеанс тестирования. Для демонстрации возможности такой оценки и определения характеристик точности и достоверности результатов тестирования была разработана имитационная модель компьютерного тестирования [2].

В классической теории отклика (IRT) [3] вероятность успешного выполнения задания трудности δ учащимся, имеющим уровень подготовленности θ , оценивается по модели Раша:

$$P(\theta, \delta) = \frac{\exp(\alpha \cdot (\theta - \delta))}{1 + \exp(\alpha \cdot (\theta - \delta))} \quad (1)$$

Будем считать, что каждый учащийся имеет разные уровни сформированности разных компетенций. Вероятность удачного выполнения задания зависит от этих величин, а также от характера задания. Рассмотрим различные модели успешности выполнения задания в случае трех компетенций, понимая, что результаты могут быть обобщены на случай любого количества компетенций.

Величины θ и δ измеряются в логитах (логарифм отношения шансов успеха и неудачи). Диапазон возможных значений заключается в пределах $(-3; 3)$, где -3 означает для тестируемого минимальные шансы выполнить задание, а для задания – самый низкий уровень трудности. Соответственно, 3 – максимальные шансы для тестируемого и наибольшая трудность задания. Будем считать, что, с точки зрения разных компетенций, задание имеет разные трудности δ_1, δ_2 и δ_3 . Уровни подготовленности тестируемого студента по каждой компетенции обозначим как θ_1, θ_2 и θ_3 . Вероятность выполнения задания по каждой компетенции будем оценивать двухпараметрической моделью Раша:

$$P_i = P(\theta_i, \delta_i) \quad i = 1, \dots, 3 \quad (2)$$

Будем считать также, что компетенции сформулированы корректно и уровень овладения каждой из них не зависит от уровней по другим компетенциям. Смысловое наполнение различных компетенций не пересекается.

Рассмотрим несколько моделей оценки вероятности выполнения задания.

Модель максимальной вероятности предполагает, что для выполнения задания можно использовать любую из трех компетенций. Например, математическая задача может быть решена несколькими способами. Время и сил на решение задачи разными способами у испытуемого нет. Поэтому, студент выбирает наиболее знакомый ему способ. В этом случае, вероятность выполнения задания равна максимуму вероятностей P_i .

$$P_s = \max_i P_i \quad (3)$$

Если время тестирования позволяет, студент может попробовать решить задачу различными способами. В этом случае, вероятность успеха равна веро-

ятности того, что задание будет выполнено хотя бы по одной компетенции. Альтернативой этого события будет событие: «задание не выполнено ни по одной компетенции». Его вероятность $P_w = (1 - P_1) \cdot (1 - P_2) \cdot (1 - P_3)$. Тогда вероятность успеха равна вероятности того, что задача хоть каким-то способом будет решена.

$$P_s = 1 - P_w = 1 - (1 - P_1) \cdot (1 - P_2) \cdot (1 - P_3) \quad (4)$$

Отметим, что полученное значение всегда больше значения максимальной вероятности. Например, если все вероятности $P_i = 0,5$, то в случае максимума вероятности получим $P_s = 0,5$, а во втором случае: $P_s = 1 - 0,5^3 = 0,875$.

Усложним требования. Например, необходимо не только найти решение задачи, но и проверить его правильность, решив задачу другим способом. Для выполнения задания необходимо использовать, как минимум, две компетенции.

$$P_s = P_1 \cdot P_2 \cdot (1 - P_3) + P_1 \cdot P_3 \cdot (1 - P_2) + P_2 \cdot P_3 \cdot (1 - P_1) + P_1 \cdot P_2 \cdot P_3 \quad (5)$$

В этом случае, при всех $P_i = 0,5$, вероятность успеха $P_s = 4 \cdot 0,5^3 = 0,5$.

Пусть для выполнения требуются все три компетенции. Например, нужно правильно сформулировать задачу, найти ее решение и проинтерпретировать полученный результат. В этом случае вероятность успеха равна произведению вероятностей по каждой компетенции.

$$P_s = P_1 \cdot P_2 \cdot P_3 \quad (6)$$

Для $P_i = 0,5$, $P_s = 0,5^3 = 0,125$.

Рассмотрим, как влияет применение различных моделей успешного выполнения заданий на вероятность успеха.

Для проведения полнофакторного эксперимента был сформирован набор из 64 паттернов. Транзактами имитационной модели являются «студенты» (генерируются 64 транзакта этого типа, по одному для каждого паттерна) и «задания» (генерируется $64 \cdot 3 = 192$ «задания», каждое из которых принадлежит одному паттерну). Ключевым моментом моделирования являлось правило, по которому «разыгрывается» успешность выполнения задания. Разыгрывается случайная величина ξ , равномерно распределенная на интервале $(0, 1)$. Задание считается выполненным, если $P_s(\theta, \delta) > \xi$.

Выберем в имитационной модели «студентов» с одинаковыми уровнями развития всех компетенций и задания с одинаковыми уровнями сложности по всем компетенциям. Такой выбор позволяет сопоставить различные методы с одномерной моделью (одна компетенция), в которой вероятность успеха определяется формулой (1).

Таблица 1

Исходные данные для численного эксперимента

Студенты			Задания				
ID	θ_1	θ_2	θ_3	ID	δ_1	δ_2	δ_3
1	3	3	3	1	3	3	3
22	1	1	1	22	1	1	1
43	-1	-1	-1	43	-1	-1	-1

Также были вычислены вероятности успеха по каждой модели.

Оценки вероятности успешного выполнения заданий, выполненные по различным моделям

Студент	1				22				43				64			
	1	22	43	64	1	22	43	64	1	22	43	64	1	22	43	64
Max	0,5	0,97	0,99	1	0,03	0,5	0,97	0,99	0	0,03	0,5	0,97	0	0	0,03	0,5
Anywise	0,88	1	1	1	0,09	0,88	1	1	0	0,09	0,88	1	0	0	0,09	0,88
2comp	0,5	0,98	1	1	0	0,5	0,98	1	0	0	0,5	0,98	0	0	0,1	0,5
3 comp	0,13	0,91	0,99	1		0,13	0,91	0,99	0	0	0,13	0,91	0	0	0	0,13

Сопоставим на графике зависимости вероятности успеха выполнения задания (для задания трудности 1) от подготовленности студента, рассчитанные по различным моделям.

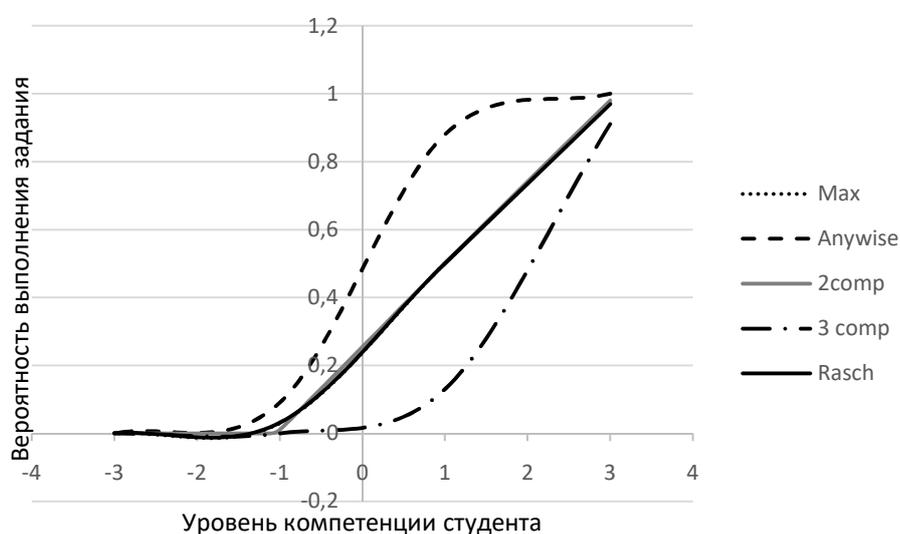


Рис. 1. Зависимости вероятностей выполнения задания от уровня подготовленности учащихся, оцененных различными методами

Обозначение линий: Rasch – модель с одной компетенцией; Max - модель максимальной вероятности; Anywise - решение задачи любым путем; 2comp – для решения необходимо использовать, как минимум, две компетенции; 3 comp - требуются все три компетенции.

На графике, хорошо видно, что только линии Max и 2 comp хорошо совпадают с линией Rasch. Это значит, что только эти модели хорошо «сшиваются» с классической одномерной моделью. Явное отличие результатов по остальным моделям показывает, что при составлении тестовых заданий от автора требуется в точности указывать выявляемые компетенции и характер выполняемого задания.

Проверка соответствия смоделированного и предсказанного уровня подготовленности по результатам множества прогонов имитационной модели, показала различную точность моделей. В идеале оценки должны совпадать с исходными данными (черная диагональ на рис. 2). Однако, например, метод мак-

симальной вероятности, столь хорошо зарекомендовавший себя на «типичных студентах» часто существенно ошибается в оценке уровня подготовленности «студентов», имеющих различную успешность по разным компетенциям. Наилучшими характеристиками в данном ключе обладает модель, учитывающая минимум две компетенции.

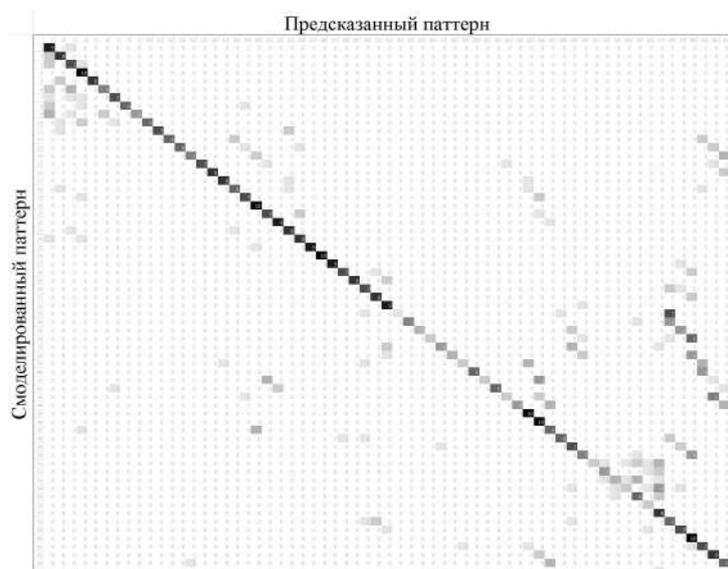


Рис. 2. Соответствие смоделированного на имитационной модели и предсказанного паттернов в модели максимальной вероятности

В реальных заданиях могут наблюдаться более сложные зависимости вероятности успеха от уровней различных компетенций испытуемого. Например, вероятность может зависеть от комбинации уровней компетенции и чувствительностей задания по каждой компетенции.

Будем считать, что каждое задание характеризуется различными чувствительностями по различным компетенциям: $\vec{a} = \{a_1, a_2, a_3\}$ и одной характеристикой трудности выполнения задания δ . Совокупный уровень компетенции $\hat{\theta}$ для данного задания будет функцией уровней развития студента и чувствительностей задания по каждой компетенции.

$$\hat{\theta} = F(\theta_1, \theta_2, \theta_3, a_1, a_2, a_3) \quad (7)$$

Вероятность успеха по-прежнему будем оценивать двухпараметрической моделью Раша. Выберем в качестве функции $F(\cdot)$ сумму произведений (линейную свертку) [4]:

$$\hat{\theta} = \sum_{i=1}^3 a_i \cdot \theta_i \quad (8)$$

В этих предположениях модель Раша можно записать так:

$$P_s = \frac{\exp(\alpha \cdot (\sum_{i=1}^3 a_i \cdot \theta_i - \delta))}{1 + \exp(\alpha \cdot (\sum_{i=1}^3 a_i \cdot \theta_i - \delta))} \quad (9)$$

Параметры a_i подбираются таким образом, чтобы в очевидных случаях параметр $\hat{\theta}$ принимал ожидаемые значения. Например, для «круглого отличника» (все θ равны 3) $\hat{\theta}$ также должна равняться трем, а «безнадежного двоечника»

ка» (все θ равны -3) $\hat{\theta} = -3$.

В более точной постановке, выбор a_i и δ можно рассматривать как задачу оптимизации: выделяются « типовые студенты » (паттерны), для которых ясно, как они должны выполнить каждое задание. Затем подбираются такие наборы α и δ , чтобы вероятности выбора нужного паттерна были максимальные.

Моделирование оценки вероятности выполнения задания существенно зависит от характеристик теста, а также особенностей проведения тестирования. При разработке тестов, позволяющих оценивать сразу несколько компетенций, мы должны уделять особенное внимание характеристикам тестового задания. Для каждого указать необходимо указывать не только уровень трудности по каждой измеряемой компетенции, но и предполагаемую модель решения.

Работа выполнена при финансовой поддержке гранта РФФИ (проект № 20-013-00783.).

СПИСОК ЛИТЕРАТУРЫ

1. Гусятников В. Н., Соколова Т. Н., Безруков А. И., Каюкова И. В. Адаптивная модель тестирования нескольких компетенций на основе алгоритма Байеса // Современные наукоемкие технологии. 2022. № 1. С. 40-46.

2. Гусятников В. Н., Безруков А. И., Соколова Т. Н., Каюкова И. В. Система оценки уровня нескольких компетенций в ходе одного сеанса тестирования на основе интеллектуального анализа данных. Свидетельство о регистрации программы для ЭВМ 2022664785, 04.08.2022. Заявка № 2022664231 от 28.07.2022.

3. *Partchev I.* A visual guide to item response theory. Jena: Friedrich Schiller Univer-sitat, 2004. 61 p.

4. *Mike Wu., Richard L. Davis, Benjamin W. Domingue, Chris Piech, Noah Goodman.* Variational Item Response Theory: Fast, Accurate, and Expressive. [Electronic resource]. <https://deepai.org/publication/variational-item-response-theory-fast-accurate-and-expressive>. (date of application: 22.09.2022).