

УДК 303.2

О. Б. Соболева¹, И. В. Антипкина², А. Е. Иванова³

¹e-mail: osoboleva@hse.ru; ²e-mail: iantipkina@hse.ru; ³e-mail: aeivanova@hse.ru
Национальный исследовательский университет «Высшая школа экономики»,
Москва, Россия

ВЕРИФИКАЦИЯ ПРАВИЛ СКОРИНГА В ТЕСТЕ СМЫСЛОВОГО ЧТЕНИЯ*

В исследовании проанализированы типичные паттерны ответов детей на вопросы в тесте смыслового чтения с форматом выделения ответа в тексте и проведена статистическая верификация определенных экспертами правил начисления баллов за выполнение компьютеризированных заданий. Показано, что паттерны ответов не организованы в соответствии с таксономией когнитивных процессов, а также что дети демонстрируют разнообразие паттернов.

Ключевые слова: смысловое чтение, оценивание, модели IRT.

Olga B. Soboleva¹, Inna V. Antipkina², Alina E. Ivanova³

¹e-mail: osoboleva@hse.ru; ²e-mail: iantipkina@hse.ru; ³e-mail: aeivanova@hse.ru
National Research University Higher School of Economics, Moscow, Russia

VERIFICATION OF SCORING RULES IN THE READING COMPREHENSION TEST

The study analyzed the typical patterns of children's answers to items in a computerized reading comprehension test and carried out a statistical verification of the expert-based rules for scoring. We demonstrate that the response patterns are not organized according to the Blooms taxonomy of cognitive processes, and that children exhibit a variety of response patterns.

Keywords: reading comprehension, assessment, IRT models.

Смысловое чтение – одна из основных технологий познания [1], в связи с чем важно, чтобы инструменты, измеряющие смысловое чтение, помогали сделать правильный вывод о навыках чтения у детей. Вклад в проблему валидности измерительного инструмента вносит способ скоринга задания [2]. В этом исследовании мы предприняли попытку обосновать

© Соболева О. Б., Антипкина И. В., Иванова А. Е., 2020

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-14110 «Использование контекстной информации и информации из цифровой среды оценивания при измерении индивидуального прогресса учащихся начальной школы с помощью цифровых технологий».

вать экспертные правила скоринга политомических заданий с опорой на статистические свидетельства.

Гипотезы исследования. В данном исследовании мы работали с разнообразными ответами, которые третьеклассники давали на открытые задания теста по чтению. Типичные способы выполнения заданий одного формата мы будем называть «паттернами». Разные паттерны могли привести к получению нулевого, частичного или полного балла. Для проверки было выдвинуто две гипотезы:

1) паттерны решения заданий одного типа организованы иерархически в соответствии с требующимися для их реализации когнитивными действиями обновленной таксономии Блума: вспоминание; понимание; применение; анализ; оценка; синтез [3]. Иными словами, дети с более высоким уровнем смыслового чтения используют паттерны, связанные с когнитивными действиями более высокого уровня;

2) выполняя задания одного типа, дети демонстрируют преимущественно одни и те же паттерны.

Выборка. Выборку исследования составили 1 369 третьеклассников из 57 классов десяти московских школ. Из них 50,3 % мальчики, 49,7 % девочки.

Инструмент. Настоящая работа была проведена на материале компьютеризированного теста смыслового чтения для третьих классов линейки ПРОГРЕСС. В фокусе этого исследования находятся восемь заданий, для ответа на которые необходимо было выделить ответ в тексте. Такие задания предполагают, что ребенок, пользуясь мышью, самостоятельно выделит отрывок в тексте. В начале теста присутствовало обучающее задание, которое знакомило детей с данным форматом. В таблице приведены типичные паттерны ответов детей на задания этого формата.

Экспертами было предложено ставить максимальный балл за паттерны 1 и 2, нулевой балл за паттерны 5 и 6 и частичный балл за паттерны 3 и 4. При этом эксперты предполагали, что паттерн 1 (лаконичное выделение только релевантной информации) отражает более высокие навыки анализа текста, чем, например, выделение полного предложения. В ходе анализа гипотез стояла задача верифицировать решения экспертов статистически.

Метод анализа данных. Для психометрического моделирования данных была выбрана Ordered Partition Model (OPM)[4]. Она применяется, если существуют: 1) предположение о прогрессе между уровнями, отражаемое в заданиях; 2) в рамках одного уровня существуют различные способы набрать один и тот же балл. OPM является расширением Partial Credit Model [5], но, в отличие от РСМ, позволяет оценить категории ответов, даже если они соответствуют одному и тому же баллу. Для применения OPM в каждом задании для каждого ребенка были размечены использованные паттерны (такие, как описаны в таблице выше). Эти паттерны оказались общими для разных заданий.

Типичные паттерны ответов учащихся
на задание теста смыслового чтения

	Паттерн	Типичные паттерны ответов на примере задания: «Выдели в тексте, где Сева нашел птичку»
1	Лаконичное выделение только релевантной информации	«за забором на куче сухой травы»
2	Выделение полного предложения, содержащего релевантную информацию	«И правда за забором на куче сухой травы Сева нашел маленькую птичку»
3	Недовыделение (ребенок не увидел часть информации)	«за забором» или «на куче сухой травы»
4	Перевыделение не более, чем на одно предложение	«Он побежал в ту сторону. И правда за забором на куче сухой травы Сева нашел маленькую птичку»
5	Очень сильное перевыделение	Например, выделен целый абзац
6	Выделена нерелевантная информация	«Сева жил у бабушки в деревне»

Результаты. Результаты показали значимые различия между всеми паттернами-категориями всех заданий, кроме одного, однако иерархия порогов не соблюдалась: не было постепенного уменьшения по трудности от «высокоуровневых паттернов» (точный лаконичный ответ) к низкоуровневым паттернам (сильное перевыделение, недовыделение).

Заметны перепады между трудностью порогов: к примеру, в двух заданиях трудность порогов срединных паттернов-категорий возрастает до экстремально высоких значений (6 и 8 логитов), в то время как крайние паттерны-категории обладают маленькой трудностью. Это означает, что в большинстве заданий от теоретически более трудной стратегии (как определено экспертами) легко перейти к менее трудной, но от стратегии, которая должна была обладать средней трудностью (по мнению экспертов), тяжело перейти к теоретически более легкой.

В целом заметно, что срединные паттерны-категории в целом функционируют плохо: у них нет места на оси, где вероятность их использования была наиболее высокой. Кроме того, пороги между паттернами-категориями не упорядочены по ожидаемой экспертно-определенной трудности.

Обсуждение. Исследование показало возможность применения Ordered Partition Model для верификации экспертных правил скоринга. Результаты свидетельствуют о том, учащиеся не соблюдают одни и те же паттерны при решении разных заданий, а сами паттерны плохо дифференцированы и не образуют иерархию по теоретически предполагаемой сложности. Таким образом, обе гипотезы не подтвердились.

К ограничениям интерпретации результатов относится, во-первых, влияние компьютерной грамотности на ответы: обучающего задания могло оказаться недостаточно и недовыделение или перевыделение могло стать результатом того, что некоторые дети плохо владели мышкой. Во-вторых, в данных встречаются пропуски ответов: некоторые дети пролистывали эти задания, что искажает полноту картины.

Практическим результатом исследования стали рекомендации по доработке заданий теста. Поскольку мы увидели, что дети, которые недовыделяли текст, имели в целом такой же уровень способностей по смысловому чтению, что и дети, выделившие всю релевантную информацию, текст и правила скоринга были переработаны. В тексте были сокращены предложения, убраны возможности сконструировать ответ из более чем одной единицы информации (например, вместо «за забором на куче сухой травы» осталось только «на куче сухой травы»).

Полученный результат ставит несколько вопросов для будущих исследований. Если показано, что паттерны не отражают когнитивную иерархию, необходимо понять, с чем тогда связаны способы решения заданий, а также почему дети не предпочитают один и тот же способ решения, а меняют их.

Список литературы

1. Мелентьева Ю. П. Общая теория чтения. М., 2015.
2. Tarasowa D., Auer S. Balanced Scoring Method for Multiple-mark Questions // CSEDU. 2013. С. 411–416.
3. Wilson L. O. Anderson and Krathwohl – Bloom’s taxonomy revised // Understanding the New Version of Bloom's Taxonomy. 2016.
4. Wilson M., Adams R. J. Marginal maximum likelihood estimation for the ordered partition model // Journal of Educational Statistics. 1993. Т. 18. №. 1. С. 69–90.
5. Wright B. D., Masters G. N. Rating scale analysis. MESA press, 1982.