

*Национальный исследовательский университет «Высшая школа экономики»,
г. Москва, Россия*

***Аннотация.** Рассматриваются методы оценки заданий с множественным выбором ответов, широко применяемых в образовательных и психологических тестах. Описаны различные виды заданий с множественным выбором и подходы к скорингу таких заданий. Обсуждаются традиционные и альтернативные модели оценки, а также их преимущества и недостатки.*

Ключевые слова: психометрика; оценивание; вопросы с множественным выбором; скоринг заданий

Задания с множественным выбором ответов широко используются в образовательных и психологических тестах. Данный вид заданий обладает рядом преимуществ: объективность оценивания, простота проведения тестирования и обработки результатов, возможность оценить большое количество человек одновременно [1]. Скоринг модели таких заданий играют ключевую роль в обеспечении валидности и надежности теста. Однако, несмотря на большое количество существующих моделей, результаты исследований оптимальных методов оценки заданий с множественным выбором являются довольно противоречивыми [2]. Цель данного исследования – проанализировать существующие методы оценки заданий с множественным выбором и определить наиболее эффективный.

Задания с множественным выбором можно классифицировать по разным критериям. Наиболее распространенным является разделение на задания с одним правильным ответом и задания с несколькими правильными ответами [3]. Последние позволяют снизить вероятность случайного угадывания, а также оценить частичные знания испытуемых [4]. Кейс и Свонсон выделяют две основные группы заданий: на лучший ответ (например, "Какое из следующих утверждений наиболее соответствует...") и на все верные ответы (например, "Выберите все верные варианты...") [5]. Халадина и соавт. описывают семь форматов заданий, из которых наиболее распространены традиционные (с вопросом или утверждением со множественным выбором), альтернативного выбора (с двумя вариантами ответа) и на установление соответствия (между двумя множествами) [6].

Все методы оценки заданий с множественным выбором можно разделить на традиционные и альтернативные. В традиционных методах баллы начисляются за каждый правильный ответ. Например, в дихотомической оценке за все верно выбранные ответы дается 1 балл, в ином случае - 0 баллов. В политомической оценке за каждый верный ответ дается 1 балл, за неверный или пропущенный ответ – 0 баллов. Это позволяет оценить частичные знания тестируемых. Для снижения влияния угадывания применяются штрафы в виде отрицательных баллов за неверные ответы. Подобные скоринг-модели демонстрируют лучшие психометрические характеристики, чем модели без штрафов [7].

Альтернативные политомические методы позволяют оценить частичные знания испытуемых. К ним относятся метод исключения ответов и метод взвешивания вариантов ответа. В методе исключения ответов, или процедуре Кумбса, за каждый исключенный неверный вариант дается 1 балл, а за исключение верного ответа – штраф [8]. Например, в задании с 5 вариантами ответа исключение всех неверных вариантов дает 4 балла, а исключение верного ответа – штраф в 4 балла. Методы взвешивания вариантов ответа присваивают каждому варианту свой вес в зависимости от его правдоподобности или частоты выбора сильными испытуемыми. Например, в методе Гутмана вес варианта пропорционален среднему баллу испытуемых, выбравших этот вариант [9]. Однако эти методы редко используются из-за сложности расчетов.

Оптимальный метод оценки заданий множественного выбора играет ключевую роль в повышении надежности и валидности тестов, а также точности информации об уровне способностей испытуемых. Дальнейшее исследование скоринг-моделей с применением методов имитационного моделирования позволит получить более однозначные результаты и выработать практические рекомендации по выбору метода оценки в зависимости от цели и условий тестирования.

Список литературы:

1. Haladyna, T. M. (2012). *Developing and validating multiple-choice test items* (3rd ed.). New York, NY: Routledge.
2. Domnich, A., Panatto, D., Signori, A., Bragazzi, N. L., Cristina, M. L., Amicizia, D., & Gasparini, R. (2016). Uncontrolled web-based administration of surveys on factual health-related knowledge: A randomized study of untimed versus timed quizzing. *Journal of Medical Internet Research*, 18(4), e94.
3. Hsu, T.-C., Moss, P. A., & Khampalikit, C. (2015). The merits of multiple-answer items as evaluated by using six scoring formulas. *The Journal of Experimental Education*, 52(3), 152–158.
4. Dressel, P., & Schmid, J. (1953). Some modifications of the multiple-choice item. *Educational and Psychological Measurement*, 13(4), 574–595.
5. Case, S. M., & Swanson, D. B. (2002). *Constructing written test questions for the basic and clinical sciences* (3rd ed., rev.). Philadelphia, PA: National Board of Medical Examiners.
6. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333.
7. Abu-Sayf, F. K. (1979). The scoring of multiple-choice tests: A closer look. *Educational Technology*, 19(6), 5–15.
8. Ben-Simon, A., Budescu, D. V., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement*, 21(1), 65–88.
9. Frary, R. B. (1989). Partial-credit scoring methods for multiple-choice tests. *Applied Measurement in Education*, 2(1), 79–96.

E. D. Sidorkina

To penalize or not to penalize? Optimal scoring of multiple-choice tasks

National Research University Higher School of Economics, Moscow, Russia

Abstract. Multiple-choice tasks are widely used in educational and psychological tests due to their psychometric and administrative features. The study describes various types of multiple-choice tasks and approaches to scoring such tasks. There are discussed traditional and alternative scoring models, as well as their advantages and disadvantages.

Keywords: psychometrics; assessment; multiple-choice questions; scoring