

Позиционный эффект в языковых тестах множественного выбора как фактор когнитивной нагрузки: теоретические основания и дизайн экспериментального исследования¹

М.О. Матвеев¹, Е.А. Матвеева², О.А. Ульянина^{3,4}

¹ФГАОУ ВО «Национальный исследовательский университет «Московский институт электронной техники», г. Зеленоград

²ФГБОУ ВО «Московский государственный лингвистический университет», г. Москва

³ФГБОУ ВО «Московский государственный психолого-педагогический университет», г. Москва

⁴ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)», г. Долгопрудный

В статье рассматривается позиционный эффект в языковых заданиях множественного выбора как потенциальный источник систематического смещения результатов тестирования. В отличие от содержательной трудности задания, позиционный эффект связан не с уровнем языковой компетенции тестируемого, а с расположением правильного ответа или дистракторов в списке вариантов. Теоретически обосновывается и описывается воспроизводимый дизайн экспериментального исследования, позволяющего проверить, влияет ли позиция правильного ответа на точность и скорость выполнения языковых заданий. Теоретическую основу составляют исследования серийного позиционного эффекта, стратегий снижения когнитивных усилий, response-order effects, а также работы по конструированию заданий множественного выбора, языковому тестированию и функционированию дистракторов. Экспериментальный план включает четыре формы теста, в которых правильный ответ перемещается по позициям 1–4 при сохранении состава опций. Материал тестирования – нейтральные языковые задания двух типов: задания на выбор синонима и задания на выбор нормативной коллокации или модели управления.

Ключевые слова: *психолингвистика, психометрия, языковое тестирование, когнитивная нагрузка, позиционный эффект, эффект первичности, дистрактор.*

¹ Финансирование: авторы выражают благодарность Министерству науки и высшего образования Российской Федерации за поддержку данного исследования в рамках дополнительного соглашения № 075-03-2026-305 (от 16 января 2026 г.), связанного с проектом «Прикладные исследования по внедрению технологий искусственного интеллекта в высшее образование» (шифр научной темы: FSMG-2025-0086).

Введение

Задания множественного выбора остаются одним из наиболее распространенных форматов языкового тестирования. Их популярность объясняется технологичностью проверки, возможностью массового применения, удобством автоматической обработки и оптимальной возможности стандартизации. В языковом тестировании задания множественного выбора используются в проверке чтения, аудирования, лексико-грамматических навыков и понимания текста; их преимущества и ограничения подробно обсуждаются в работах по языковому оцениванию (работы Дж.Ч. Олдерсона [4]; Л.Ф. Бахмана, А.С. Палмера [8]; Г. Фулчера, Ф. Дэвидсона [11]; А. Хьюза, Дж. Хьюза [15]). Вместе с тем сам формат множественного выбора содержит ряд факторов, которые могут влиять на результат независимо от проверяемой компетенции: качество дистракторов, длина опций, лексико-семантическая близость вариантов, наличие отрицаний и кванторов, а также позиция ответа в списке.

В отечественной тестологии подчеркивается, что качество тестового задания определяется не только содержательной корректностью, но и формой предъявления, структурой вариантов ответа и психометрическими характеристиками задания (работы В.С. Аванесова [1]; Н.Ф. Ефремовой [2]; М.Б. Чельшковой [3]). Следовательно, позиция варианта в списке может рассматриваться как один из формальных параметров задания, требующих контроля наряду с трудностью, дискриминативностью и функционированием дистракторов.

Последний фактор – позиция варианта, часто воспринимается как техническая деталь, однако в психологии памяти и принятия решений давно описаны эффекты края: эффект первичности и эффект недавности, или последней позиции [20]. В свою очередь данные обстоятельства могут влиять на ситуацию выбора ответов, занимающих более заметную позицию в ряду.

Некоторые исследования также отмечают влияние взгляда кандидата и трудность самого задания [14], положения правильной опции и дистракторов [18; 19]. В ситуации электронного тестирования эти факторы могут повлиять на порядок предъявлений, задаваемый оператором.

В рамках данного теоретико-методического исследования описывается разработанный экспериментальный протокол проверки позиционного эффекта в языковых заданиях множественного выбора.

Серийный позиционный эффект

Классические данные Б.Б. Мердока фиксируют преимущество краевых элементов последовательности при воспроизведении [20]. Хотя свободное припоминание и выбор ответа в тесте внешне различны, структурное сходство есть: альтернативы воспринимаются последовательно, поэтому эффекты первичности могут смещать фиксацию решения. Для языковых тестов с семантически близкими опциями это особенно критично, однако перенос требует эмпирической проверки.

Позиция ответа как психометрическая переменная

И. Аттали и М. Бар-Хиллель показали, что и составители, и тестируемые систематически ожидают определенного места правильного ответа [7]. Н. Бланч указывает: перестановка вариантов сама по себе не снимает позиционное смещение [10]. Более поздние работы подтверждают значимость позиции ключа для трудности задания [13]; С. Лионс и соавторы фиксируют систематическое влияние позиции дистракторов, причем сильные чаще располагаются в середине, слабые – в конце [18; 19]. Отсюда методический вывод: равномерного распределения ключей недостаточно, необходима контрбалансировка при неизменном содержании.

Влияние порядка ответов и стратегии минимизации когнитивных усилий

Дж. Кросник и Д. Олвин связывают эффект первичности в визуальных списках с высокой когнитивной нагрузкой [17]. Концепция satisficing [16] объясняет механизм: при недостатке ресурсов респондент выбирает первый приемлемый вариант, не завершая полного сравнения. Эта стратегия особенно вероятна при дефиците времени, усталости или близости опций, а этап сопоставления с альтернативами, по Р. Туранжо с соавторами, наиболее чувствителен к организации списка [24].

Задания множественного выбора и функционирование дистракторов

Качество задания определяется правдоподобностью дистракторов, которые не должны вводить в заблуждение компетентного испытуемого за счет формальных признаков [12]. Л. Бахман и А. Палмер включают в полезность теста несколько параметров – от конструктивной валидности до практичности [8]; А. Хьюз и Дж. Хьюз подчеркивают связь формата с интерпретируемостью результатов [15]. Надежность при этом зависит от структуры задания и локальной зависимости между элементами [25].

Работы по функционированию дистракторов показывают, что не все неверные опции одинаково полезны. М. Таррант, Дж. Уэр и А.М. Мохаммед различают функционирующие и нефункционирующие дистракторы: если неверный вариант почти никто не выбирает, он не вносит вклада в различение участников по уровню подготовки [23]. Метаанализ М. Родригеса показывает, что увеличение числа опций не всегда улучшает тест; часто три хорошо работающие опции могут быть эффективнее четырех или пяти формальных вариантов [22].

Современные исследования подтверждают эту логику. М. Аль-Лавама и Б. Кумвенда показали, что сокращение числа опций в заданиях множественного выбора может сохранять психометрическое качество теста и одновременно уменьшать долю нефункционирующих дистракторов [5]. Следовательно, для анализа позиционного эффекта важно контролировать не только расположение ключа, но и качество каждой неверной опции.

Для темы позиционного эффекта эти выводы имеют прямое значение. Чем слабее дистракторы, тем выше вероятность, что участник будет ориентироваться не на содержательное сравнение, а на поверхностные признаки – в том числе на позицию. Напротив, если дистракторы сбалансированы по длине, частотности, стилистике и семантической правдоподобности, позиционный эффект может быть слабее. Следовательно, эксперимент должен контролировать качество опций, иначе позиционный эффект будет трудно отделить от различий в привлекательности конкретных дистракторов.

Современные подходы к оценке качества дистракторов

В последние годы проблема качества дистракторов стала актуальной не только в классической тестологии, но и в области автоматической генерации заданий. Исследования автоматизированной генерации и оценки дистракторов важны для настоящей статьи, поскольку они предлагают операциональные критерии качества неверных опций: правдоподобие, неверность, разнообразие, семантическая близость и диагностичность.

В работе В. Райны, А. Льюзи и М. Гейлса подчеркивается, что подбор хороших дистракторов является тонкой процедурой: неверная опция должна быть объективно неверной, не быть очевидно абсурдной и в то же время давать информацию о знаниях кандидата [21]. Обзор Л. Бенедетто, Ш. Таслимпур и П. Баттери систематизирует современные подходы к автоматизированной оценке дистракторов в заданиях множественного выбора [9]. Эти работы показывают, что проблема дистрактора сегодня рассматривается не только как вопрос ручного редактирования, но и как задача моделирования качества опции.

Работа Т. Андерссона и соавторов по генерации дистракторов для японских cloze-тестов демонстрирует: лингвистическая правдоподобность опции не менее важна, чем формальная неверность [6]. При неравноценных дистракторах позиция способна усиливать исходное различие в привлекательности.

Языковые опции не произвольны: они конкурируют внутри лексико-семантической или грамматической системы. При синонимии первый вариант может фиксировать начальную семантическую гипотезу, при коллокации – получать преимущество за счет легкости обработки. Позиция, таким образом, не технический параметр, а условие, определяющее порядок языкового сравнения. Эксперимент поэтому включает и синонимические, и коллокационные задания, чтобы проверить, усиливается ли эффект при высокой близости опций.

Статья представляет собой исследовательский протокол, а не отчет о собранных данных: описываются переменные, стимулы, контрбалансировка и план анализа до начала эксперимента. Такой подход критичен для цифровых систем тестирования: эффект можно

проверить на нейтральных заданиях, не раскрывая закрытый контент, и затем перенести протокол на реальные платформы.

Эксперимент предполагает четыре контрбалансированные формы с одинаковым набором из 24 заданий и четырьмя опциями; меняется лишь позиция ключа. Платформа должна фиксировать точность и, желательно, время ответа по каждому заданию. Материал включает два типа: синонимию и коллокацию/управление. Такое разделение позволяет проверить, одинаково ли проявляется позиционный эффект в задачах, основанных преимущественно на лексико-семантическом сопоставлении, и в задачах, связанных с языковой сочетаемостью.

Все стимулы должны быть созданы специально для исследования и не должны воспроизводить закрытые материалы каких-либо тестовых систем. Для снижения побочных факторов опции внутри одного задания необходимо уравнивать по части речи, длине, стилистической маркированности и степени очевидности.

Пример синтетической структуры задания

Тип 1. Синонимия. Стимул: выберите слово, наиболее близкое по значению к заданному слову. Опции: четыре слова одной части речи, близкие по общему семантическому полю, одно из которых является ключом.

Тип 2. Коллокация / модель управления. Стимул: выберите наиболее естественное словосочетание или грамматически корректное продолжение. Опции: четыре варианта сходной длины, один из которых соответствует языковой норме.

В публикации рекомендуется приводить только синтетические примеры, не связанные с реальными тестовыми базами. Если эксперимент проводится на авторских заданиях, их можно включить в приложение при наличии соответствующих прав.

Контрбалансировка

Для каждого задания создаются четыре версии, в которых правильный ответ занимает позиции 1, 2, 3 или 4. Остальные опции перемещаются вместе с ним таким образом, чтобы во всех формах сохранялся одинаковый состав вариантов (табл. 1).

Таблица 1

Схема контрбалансировки позиций ключа

Форма	Пункты 1–6	Пункты 7–12	Пункты 13–18	Пункты 19–24
А	позиция 1	позиция 2	позиция 3	позиция 4
В	позиция 2	позиция 3	позиция 4	позиция 1
С	позиция 3	позиция 4	позиция 1	позиция 2
Д	позиция 4	позиция 1	позиция 2	позиция 3

Участники

Минимально достаточный объем выборки – 120 участников, по 30 человек на каждую форму. Оптимальный объем – 200–240 участников, что обеспечивает более устойчивые оценки при анализе на

уровне отдельных ответов. Участниками могут быть студенты бакалавриата или магистратуры, владеющие языком теста на уровне, позволяющем выполнять задания без чрезмерных затруднений.

Процедура

Участник получает ссылку на одну из четырех форм случайным образом. Перед началом теста он видит нейтральную инструкцию: «В каждом вопросе выберите наиболее подходящий вариант ответа. Работайте в обычном темпе». Затем участник выполняет 24 задания множественного выбора. После теста можно предложить два дополнительных вопроса: оценить субъективную сложность задания по шкале 1–9 и уверенность в своих ответах по шкале 1–5.

Порядок заданий внутри формы желательно менять. Порядок опций внутри задания не следует дополнительно рандомизировать, поскольку позиция правильного ответа является экспериментальной переменной.

Переменные

Независимые переменные: позиция правильного ответа (1, 2, 3, 4); тип задания (синонимия/коллокация); форма теста (А, В, С, D). Зависимые переменные: правильность ответа (0/1); время ответа по пункту или время выполнения блока; выбранная позиция; субъективная сложность и уверенность (опционально). Контрольные переменные: длина опций, частотность слов, исходная трудность задания, порядковый номер пункта в тесте, уровень владения языком, если доступен внешний показатель.

План статистического анализа

Для первичной проверки гипотезы используется сравнение доли правильных ответов в зависимости от позиции ключа. Более строгий анализ предполагает смешанную логистическую модель:

$$\text{Correct} \sim \text{KeyPosition} + \text{TaskType} + \text{KeyPosition} \times \text{TaskType} + (1 | \text{Participant}) + (1 | \text{Item})$$

Если доступно время ответа по пунктам, анализируется логарифмированное время:

$$\log\text{RT} \sim \text{KeyPosition} + \text{TaskType} + \text{KeyPosition} \times \text{TaskType} + (1 | \text{Participant}) + (1 | \text{Item}).$$

Распределение выборов по позициям, рассматриваемое без учета корректности, выявляет общий позиционный сдвиг. Пороговые значения, заслуживающие внимания в тестовой практике, лежат в диапазоне 3–5 процентных пунктов для точности и 8–10% для времени ответа; эти цифры, впрочем, не универсальны и зависят от задач тестирования.

Если ключ на первой позиции систематически выбирается чаще или быстрее, мы сталкиваемся с эффектом первичности, который жестко ограничивает компоновку фиксированных форм: смещение ключа к началу недопустимо. Противоположный паттерн – преимущество четвертой позиции – указывает на стратегию

прочитывания всего списка с фиксацией последнего приемлемого варианта, и тогда опасность представляет сдвиг ключей к концу. Отсутствие значимых различий говорило бы о равномерной привлекательности дистракторов, однако балансировка позиций остается разумной предосторожностью, а не излишней формальностью.

Отдельного рассмотрения заслуживает взаимодействие позиции с характером языкового материала. В синонимических заданиях лексико-семантическая близость может усиливать закрепление первой гипотезы, тогда как в коллокационных тестируемый нередко полагается на субъективное ощущение естественности и выбирает первую опцию, которая «звучит» приемлемо, не завершая сравнения. Проверка этих предположений потребует отдельных контрастов.

С точки зрения психолингвистики позиционный эффект сводится к последовательности обработки: первая опция задает исходную гипотезу, последняя выигрывает за счет преимущества последних элементов списка. В рамках психометрии любое систематическое смещение вероятности правильного ответа в зависимости от позиции подрывает валидность, поэтому контроль позиций ключа и рандомизация опций – не техническая деталь, а часть обеспечения качества теста. Для цифровых платформ достаточно логировать код пункта, позицию выбранной опции, правильность и время; содержание заданий раскрывать не требуется.

Ограничения связаны с качеством стимульного набора: чрезмерно легкие или трудные пункты могут замаскировать позиционный сдвиг, поэтому пилотный отсев необходим. Если платформа не фиксирует время по каждому пункту, анализ латентности окажется огрубленным до уровня блока, что снизит точность выводов, хотя и не обесценит основной показатель – зависимость правильности от позиции. Наконец, индивидуальные стратегии (полное прочитывание опций против выбора первой приемлемой) вносят дополнительную дисперсию; косвенно их можно контролировать через время ответа или пост-тестовый опрос. Предложенный протокол прост в реализации, не требует доступа к закрытым материалам и применим в любой системе, поддерживающей контрбалансировку позиций и регистрацию поведенческих показателей.

Список литературы

1. Аванесов В.С. Композиция тестовых заданий. М.: Центр тестирования, 2002. 240 с.
2. Ефремова Н.Ф. Тестовый контроль в образовании: учебное пособие. М.: Логос, 2005. 368 с.
3. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. 432 с.

4. Alderson J.C. Assessing Reading. Cambridge: Cambridge University Press, 2000. 398 p.
5. Al-lawama M., Kumwenda B. Decreasing the options' number in multiple choice questions in the assessment of senior medical students and its effect on exam psychometrics and distractors' function // BMC Medical Education. 2023. V. 23. Article 212. DOI: 10.1186/s12909-023-04206-3.
6. Andersson T., Nasukawa T., Miyazaki T., Kajino H., Takahashi K. Closing the Gap: Automated Distractor Generation in Japanese Language Testing // Education Sciences. 2023. V. 13. № 12. Article 1203. DOI: 10.3390/educsci13121203.
7. Attali Y., Bar-Hillel M. Guess Where: The Position of Correct Answers in Multiple-Choice Test Items as a Psychometric Variable // Journal of Educational Measurement. 2003. V. 40. № 2. P. 109–128.
8. Bachman L.F., Palmer A.S. Language Testing in Practice: Designing and Developing Useful Language Tests. Oxford: Oxford University Press, 1996. 377 p.
9. Benedetto L., Taslimipour S., Buttery P. A Survey on Automated Distractor Evaluation in Multiple-Choice Tasks // Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2025). Vienna: Association for Computational Linguistics, 2025. P. 55–69. DOI: 10.18653/v1/2025.bea-1.5.
10. Blunch N.J. Position Bias in Multiple-Choice Questions // Journal of Marketing Research. 1984. V. 21. № 2. P. 216–220.
11. Fulcher G., Davidson F. Language Testing and Assessment: An Advanced Resource Book. London; New York: Routledge, 2007. 416 p.
12. Haladyna T.M., Downing S.M., Rodriguez M.C. A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment // Applied Measurement in Education. 2002. V. 15. № 3. P. 309–334.
13. Hohensinn C., Baghaei P. Does the Position of Response Options in Multiple-Choice Tests Matter? // Psicológica. 2017. V. 38. № 1. P. 93–109.
14. Holzknecht F., McCray G., Eberharter K., Kremmel B., Zehentner M., Spiby R., Dunlea J. The Effect of Response Order on Candidate Viewing Behaviour and Item Difficulty in a Multiple-Choice Listening Test // Language Testing. 2021. V. 38. № 1. P. 41–61.
15. Hughes A., Hughes J. Testing for Language Teachers. 3rd ed. Cambridge: Cambridge University Press, 2020. 352 p.
16. Krosnick J.A. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys // Applied Cognitive Psychology. 1991. V. 5. № 3. P. 213–236.
17. Krosnick J.A., Alwin D.F. An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement // Public Opinion Quarterly. 1987. V. 51. № 2. P. 201–219.
18. Lions S., Monsalve C., Dartnell P., Godoy M.I., Córdova N., Jiménez D., Blanco M.P., Ortega G., Lemarié J. The Position of Distractors in Multiple-Choice Test Items: The Strongest Precede the Weakest // Frontiers in Education. 2021. Vol. 6. Article 731763. DOI: 10.3389/educ.2021.731763.
19. Lions S., Dartnell P., Toledo G., Godoy M.I., Córdova N., Jiménez D., Lemarié J. Position of Correct Option and Distractors Impacts Responses to Multiple-

- Choice Items: Evidence From a National Test // Educational and Psychological Measurement. 2023. V. 83. № 5. DOI: 10.1177/00131644221132335.
20. Murdock B.B. The Serial Position Effect of Free Recall // Journal of Experimental Psychology. 1962. V. 64. № 5. P. 482–488.
21. Raina V., Liusie A., Gales M. Assessing Distractors in Multiple-Choice Tests // Proceedings of the 4th Workshop on Evaluation and Comparison of NLP Systems. Bali: Association for Computational Linguistics, 2023. P. 12–22. DOI: 10.18653/v1/2023.eval4nlp-1.2.
22. Rodriguez M.C. Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research // Educational Measurement: Issues and Practice. 2005. V. 24. № 2. P. 3–13.
23. Tarrant M., Ware J., Mohammed A.M. An Assessment of Functioning and Non-Functioning Distractors in Multiple-Choice Questions: A Descriptive Analysis // BMC Medical Education. 2009. V. 9. Article 40.
24. Tourangeau R., Rips L.J., Rasinski K. The Psychology of Survey Response. Cambridge: Cambridge University Press, 2000. 401 p.
25. Wainer H., Thissen D. How Is Reliability Related to the Quality of Test Scores? What Is the Effect of Local Dependence on Reliability? // Educational Measurement: Issues and Practice. 1996. V. 15. № 1. P. 22–29.

Об авторах:

МАТВЕЕВ Михаил Олегович – кандидат филологических наук, доцент, Институт лингвистического и педагогического образования, ФГАОУ ВО «Национальный исследовательский университет «Московский институт электронной техники» (124498, г. Зеленоград, пл. Шокина, д. 1), <http://orcid.org/0000-0003-3629-1332>, e-mail: mike-matveev@mail.ru

МАТВЕЕВА Елена Александровна – преподаватель, Переводческий факультет, ФГБОУ ВО «Московский государственный лингвистический университет» (119034, г. Москва, ул. Остоженка, д. 38), <http://orcid.org/0009-0007-3409-4200>, e-mail: e.a.shipova@mail.ru

УЛЬЯНИНА Ольга Александровна – доктор психологических наук, доцент, руководитель Федерального координационного центра по обеспечению развития психолого-педагогической помощи в системе образования Российской Федерации, ФГБОУ ВО «Московский государственный психолого-педагогический университет» (127051, г. Москва, ул. Сретенка, д. 29), главный научный сотрудник Центра прикладных лингвистических исследований и тестирования «ИСТОК», ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)» (141701, г. Долгопрудный, Институтский пер., д. 9, стр. 3), <http://orcid.org/0000-0001-9300-4825>, e-mail: ulyaninaoa@mgppu.ru

The position effect in multiple-choice language tests as a factor of cognitive load: theoretical foundations and an experimental study design*

M.O. Matveev¹, E.A. Matveeva², O.A. Ulyanina^{3, 4}

¹National Research University of Electronic Technology, Zelenograd

²Moscow State Linguistic University, Moscow

³Moscow State University of Psychology & Education, Moscow

⁴Moscow Institute of Physics and Technology (National Research University),
Dolgoprudny

The article examines the position effect in multiple-choice language tests as a potential source of systematic bias in test performance. Unlike item difficulty, the position effect is not determined by the test-taker's linguistic competence but by the placement of the correct answer or distractors within the response list. The aim of the article is to theoretically justify and describe a reproducible experimental design for testing whether the position of the correct answer affects accuracy and response time in language items. The theoretical framework draws on research into the serial position effect, response-order effects, satisficing strategies, multiple-choice item design, language testing and distractor functioning. The proposed design includes four counterbalanced test forms in which the same correct option is systematically rotated across positions 1–4 while the set of options remains constant. The experimental material may consist of neutral language items of two types: synonym selection and collocation/government selection. The expected contribution is to determine whether primacy or recency effects occur in language testing and to estimate the practical significance of this bias for test-form construction.

Keywords: *language testing, multiple choice, position effect, primacy effect, distractor, test item, psycholinguistics, psychometrics.*

Принято в редакцию: 15.05.2026 г.

Подписано в печать: 02.09.2026 г.

* Funding: The authors acknowledge the support of this research by the Ministry of Science and Higher Education of the Russian Federation under agreement № 075-03-2026-305 (January 16, 2026), associated with project «Applied Research on the Implementation of Artificial Intelligence Technologies in Higher Education» (project code: FSMG-2025-0086).