Position of Correct Option and Distractors Impacts Responses to Multiple-Choice Items: Evidence From a National Test

Séverin Lions[1] , **Pablo Dartnell[1], Gabriela Toledo[1], María Inés Godoy[1], Nora Córdova[1], Daniela Jiménez[1] and Julie Lemarié[2]**

Abstract

Even though the impact of the position of response options on answers to multiple-choice items has been investigated for decades, it remains debated. Research on this topic is inconclusive, perhaps because too few studies have obtained experimental data from large-sized samples in a real-world context and have manipulated the position of both correct response and distractors. Since multiple-choice tests' outcomes can be strikingly consequential and option position effects constitute a potential source of measurement error, these effects should be clarified. In this study, two experiments in which the position of correct response and distractors was carefully manipulated were performed within a Chilean national high-stakes standardized test, responded by 195,715 examinees. Results show small but clear and systematic effects of options position on examinees' responses in both experiments. They consistently indicate that a five-option item is slightly easier when the correct response is in A rather than E and when the most attractive distractor is after and far away from the correct response. They clarify and extend previous findings, showing that the appeal of all options is influenced by position. The existence and nature of a potential interference phenomenon between the options' processing are discussed, and implications for test development are considered.

[1]Universidad de Chile, Santiago de Chile, Chile
[2]University of Toulouse, France

Corresponding Author:
Séverin Lions, Center for Advanced Research in Education (FB0003), Institute of Education, Universidad de Chile, Periodista José Carrasco Tapia 75, Santiago de Chile, Chile.
Email: severin.lions@ciae.uchile.cl

Introduction

The multiple-choice format is widely used to assess learning in teacher-made tests and standardized tests (Gierl et al., 2017). Therefore, multiple-choice tests' results have major consequences for academic achievement and professional outcomes, which urges all potential sources of measurement error to be methodically investigated to ensure optimal validity (Haladyna & Downing, 2004). One of the items' features that can bias responses to multiple-choice items is the position of response options. Students do pay attention to options position (Carnegie, 2017) and avoid producing atypical answer sequences when choosing responses to consecutive items (Lee, 2019). Also, they have been shown to be unconsciously influenced by option position (Attali & Bar-Hillel, 2003).

However, studies measuring the impact of options position on individual multiple-choice items and overall test performance have produced conflicting results. Diverse option position effects have indeed been observed (middle bias, primacy. . .), those effects being reported to be either statistically significant, of consistent direction, and small; statistically significant but of unpredictable direction; or non-significant and negligible (see Hagenmüller, 2021; Holzknecht et al., 2020). These discrepancies were at first thought to be caused by methodological issues, such as the use of unbalanced conditions in study design or inappropriate data analysis (Fagley, 1987), but there might very well be other reasons.

One of them is that most studies focusing on options position are either experiments (conducted with a limited number of participants with no genuine motivation to complete the task) or incidental analyses of real tests responses (not adequately controlled, regardless of the numerous authentic observations that might have been collected). Conducting experiments using high-stakes standardized tests may sidestep this tradeoff by gathering data from large samples within a real-world context. In a systematic review of the literature (Lions et al., in preparation), just four of such studies were found. In two of them, items were found to be easier when the correct response (CR) was in the middle (Attali & Bar-Hillel, 2003) or when CR was positioned first (Hohensinn & Baghaei, 2017). In the other two, item performance was found to be hardly influenced by CR position (Sonnleitner et al., 2016) and options position (Wang, 2019). Further studies reporting clear-cut option position effects are thus still needed.

Option Position Effects: Middle Bias or Primacy Effect?

A response bias favoring middle options has been reported in several studies (Attali & Bar-Hillel, 2003; DeVore et al., 2016). Attali and Bar-Hillel (2003) showed this

middle bias to be systematically present in various settings. For instance, they conducted an experiment concealed within a standardized test (the Israeli Psychometric Entrance Test administered in 1997/1998) to evaluate whether switching the edge and middle options influenced performance when answering the same four-choice items. They observed that examinees (high-school graduates) selected middle options significantly more frequently than edge options and that items were more accurately responded to when CR was in the middle.

In contrast, other studies have reported a response bias favoring the first positions, known as the primacy effect (Hohensinn & Baghaei, 2017; Schroeder et al., 2012; Tellinghuisen & Sulikowski, 2008). For example, in a study analyzing almost one million examinees' responses to four-choice items of a standardized test (the Iranian University Entrance English Test administered in 2013), items were reported to be easier when CR was placed first and to be marginally more difficult when CR was moved toward the bottom of the options list (Hohensinn & Baghaei, 2017). Which conditions will lead to the presence of either middle bias or primacy effect remain unclear, but the selection of each option position (A, B, C. . .) and—especially—the selection of edge options should be analyzed separately to determine whether responses show a middle bias, a primacy effect, or both.

Item Difficulty When CR Is Positioned First Versus Last

In line with the primacy effect, several studies have claimed that items are easier when keyed on the first position rather than the last one (see Holzknecht et al., 2020). For instance, in a comparison between two randomly assigned test forms of a multiple-choice exam on economics responded by U.S. undergraduate students, the form which was more frequently keyed on A was linked to higher performance than the other form, which contained the same four-choice items but was more frequently keyed on D (Bresnock et al., 1989). Other studies reported that five-choice items were easier when keyed on A rather than E (Ace & Dawis, 1973; Hodson, 1984; but see also Taylor, 2005). However, the one study conducted in a real-world context with a large sample to explore this issue, which evaluated option position effects in Austrian tests for different tested domains, examinees' ages, and stakes (reading comprehension at 10, 14, and 19 years; vocabulary and mathematics at 10 and 19 years; science at 19 years; assessment of the 19-year-old students being high-stakes), concluded that CR position effects could be neglected, at least among adolescents (Sonnleitner et al., 2016). In sum, literature does not provide concurring results for the effects of shifting CR from one edge position to the other one.

Distractors' Position and Item Difficulty

The origin of the inconsistencies above reviewed might lie in the fact that many studies have examined CR position effects without controlling for distractors' position. Empirical research has shown that a strong distractor's position and its distance from

CR can affect item accuracy (Ambu-Saidi & Khamis, 2000; Cizek, 1994; Fagley, 1987; Friel & Johnstone, 1979; Kiat et al., 2018; Shin et al., 2020; but see also Wilbur, 1970). For instance, in a study in which the most attractive distractor's (MAD) relative position and distance from CR were experimentally manipulated throughout a Malaysian pilot test in statistics responded by 232 undergraduate students, MAD was found to be 1.67 times more likely to be selected when presented before CR than after it. Consistently, item accuracy decreased as CR moved toward the end of the options list (Kiat et al., 2018).

However, an experiment similarly designed but conducted with a real-world test (a test on educational assessment taken at a university in Canada) reported no effect of MAD position and found only a few items to be significantly easier when MAD was positioned far from CR (Shin et al., 2020). MAD position effects in multiple-choice tests were also found to be negligible in a study analyzing results to a political knowledge survey in Germany (Willing, 2013). Effects of MAD distance from CR remain notably unclear because other studies reported items to be more (not less) difficult when MAD was far from CR (Ambu-Saidi & Khamis, 2000; Friel & Johnstone, 1979). All of this reveals the need to conduct further experimental studies with large samples and in a real-world context to determine both CR position effects and distractors–positions effects.

This Study

In this study, single-correct five-choice items embedded in a standardized test were used to run two experiments and obtain data from a large sample of examinees within a real-world context. In Experiment 1, the effects of CR position and MAD distance from CR were assessed by comparing item accuracy when CR was in A or E and MAD was close to or far from CR. Consistently with previous studies (Bresnock et al., 1989; Shin et al., 2020), it was predicted that items would be better responded when CR was in A and MAD was far from CR. In Experiment 2, the effects of MAD relative position and MAD location on item accuracy were measured when CR was held fixed in option B. Item accuracy was expected to be higher when MAD was located both after and far from CR, following Kiat et al. (2018). Given the scarcity of large-scale real-life experiments in the field, this study was expected to provide strong, replicable results that may help better understand the mechanisms involved in option position effects.

Method

Data and Participants

All examinees' responses to the language exam of the Chilean national examination called PSU (*Prueba de Selección Universitaria*) taken in 2020 were gathered (*n* = 195,715, 46.7% female). PSU is a paper-and-pencil, high-stakes, standardized examination that students need to take to enter most universities in Chile. PSU comprises

two mandatory exams that all students must take (one of them being the language exam) and several other optional exams belonging to different domains that students voluntarily take depending on the program they apply to (such as history or biology). Demographics such as socioeconomic status or age were diverse, but most examinees were Chilean fresh high-school graduates (that is, native monolingual Spanich-speaking adolescents). Examinees were not given any monetary incentives to take the test. On the contrary, they had to pay a ~ US$40 fee (a state-funded scholarship being available for low-income students). Since high performance on PSU maximizes the chances to be admitted into a Chilean university program, most students can be assumed to have been highly motivated to get the best possible score. All participants signed a written informed consent authorizing their data to be used for research purposes.

Instrument and Task

The language exam was administered in Chile on January 6, 2020, across the whole country, under standard conditions of high security. It was to be responded on a Scantron sheet within a maximum time of 2h30 (a no time pressure condition). It included 80 five-option items from which 75 were operational, that is, considered for students' final score computation, and five were pilot, that is, administered for research purposes only. All items were designed by the DEMRE (*Departamento de Evaluación, Medición y Registro Educacional*), the state institution in charge of developing and administering national university admission tests. Items assessed the achievement of the learning objectives advanced by the Chilean national high school's language and communication curriculum. They were organized in different blocks assembled in four different test forms (see Supplementary Table S1a). In all forms, the first block assessed knowledge about linking words and organizing sentences in text paragraphs. The remaining blocks assessed the reading comprehension of text passages of different text types and lengths.

Examinees were randomly assigned to take one of the four test forms. Between 48,000 and 49,000 examinees responded to each form. Forms had a varied distribution of answer keys (see Supplementary Table S1b). The four groups who received different test forms had an identical mean high-school grade point average (*M*: 5.7/7, *SD*: .5). Their mean scores in the test's operational part were very similar (mean between 40 and 41 CR, *SD* between 13 and 14 CR). Random assignment of test forms seemed thus to have led test forms to be responded to by equivalent groups of examinees.

Stimuli and Conditions

The manipulated five pilot items (hereinafter called Target Items 1–5) were embedded in the reading comprehension section of the test. When signing the informed consent, students were informed that some test items were pilot (and

therefore would not take part in their final score) but did not know which items were pilot and which ones were not. Across all test forms, pilot items were sequenced in the same order and located at the same place in the test, right after a block that was identical for all forms (called ''anchor,'' see Supplementary Table S1a). They pointed all to the same text passage (presented in Supplementary Figure S1). From the five target items, Experiment 1 involved Target Items 2, 3, 4, 5, and Experiment 2 involved Target Item 1. Target Items 2 to 5 had been field-tested during a prior pilot study before administration (with a fixed response order); Target Item 1 had not. Field-tested items presented acceptable psychometric characteristics (medium difficulty, acceptable discrimination index, no non-functional distractors).

All conditions consisted of specific arrangements for response options. Incorrect options for each item were labeled based on the frequency of responses selected by participants in the prior pilot study (for Experiment 1) or in this study (for Experiment 2). MAD label was assigned, for each item, to the incorrect option most frequently selected by examinees. The remaining distractors were numbered as 2, 3, and 4 following a simple criterion of decreasing frequency. Therefore, the final labels were MAD, Distractor 2, Distractor 3, and Distractor 4, hierarchically reflecting their efficiency (see Supplementary Figure S2).

In Experiment 1, items belonged to one of four conditions, labeled as ''*A_close*,'' ''*A_far*,'' ''*E_close*,'' and ''*E_far*'' to capture both CR position and MAD distance from CR. These conditions were obtained by crossing two levels of CR position (A and E) and two levels of MAD distance (MAD located either close to or far from CR). More precisely, conditions were configured by keying CR on A and placing distractors in a decreasingly efficient sequence first for *A_close* (MAD, Distractor 2, Distractor 3, and Distractor 4 as options B, C, D, and E, respectively) and then in a reversed sequence for *A_far* (which placed MAD as option E). The same logic was used to build the *E_close* and *E_far* conditions (see Table 1A). Conditions were counterbalanced across test forms generating a fully crossed design (see Tables 1B).

In Experiment 2, Target Item 1 was configured with CR always positioned at B (this was a requirement from DEMRE's test development unit). However, unlike Experiment 1, there was no prior information about distractors' efficiency. Because of this, a method was devised to make sure all four individual distractors were placed at options A, C, D, and E once (across forms). This was done by arranging distractors arbitrarily in test form 101. Then, form 102 was created by keeping CR at B and moving all distractors one position up in the list (the distractor placed at C was moved to A, the one placed at D was moved to C, the one placed at E was moved to D, and the one placed at A was moved to E). The same procedure was used to generate form 103 from form 102 and form 104 from form 103. The final arrangement of distractors across forms (showing distractors' distribution across forms after labeling each distractor based on examinees' actual selection) is presented in Table 1C.

Table 1. Experimental Conditions for Experiment 1 (A, B) and Experiment 2 (C).

A) Illustration of experimental conditions (Experiment 1).

	Conditions[a]			
Option position	A_close	A_far	E_close	E_far
A	CR	CR	Distractor 4	MAD
B	MAD	Distractor 4	Distractor 3	Distractor 2
C	Distractor 2	Distractor 3	Distractor 2	Distractor 3
D	Distractor 3	Distractor 2	MAD	Distractor 4
E	Distractor 4	MAD	CR	CR

B) Balancing of experimental conditions (Experiment 1).

Target item	Form 101	Form 102	Form 103	Form 104
2	A_close	A_far	E_far	E_close
3	E_close	A_close	A_far	E_far
4	A_far	E_far	E_close	A_close
5	E_far	E_close	A_close	A_far

C) Balancing of experimental conditions (Experiment 2).

	Target Item 1			
Option position	Form 101	Form 102	Form 103	Form 104
A	Distractor 2	Distractor 4	MAD	Distractor 3
B	CR	CR	CR	CR
C	Distractor 4	MAD	Distractor 3	Distractor 2
D	MAD	Distractor 3	Distractor 2	Distractor 4
E	Distractor 3	Distractor 2	Distractor 4	MAD

Note."A" and "E" refer to the position of the correct response; "close" and "far" to the distance from the correct response to the most attractive distractor.
[a]Conditions are labeled based on the position of the CR and the distance from CR to the MAD.

Data Analysis

Data analyses were run independently by two authors to prevent inaccuracies. In Experiment 1, the influence of CR position (A, E) and MAD distance from CR (close, far) on selected responses was examined. This was done by conducting Pearson Chi-Square Tests of Independence, one for each of the five types of options, namely, CR, MAD, Distractor 2, Distractor 3, and Distractor 4. The five resulting chi-square tests were all based on $2\times2\times2$ contingency tables, with two levels for CR position (independent variable), two levels for MAD distance (independent variable), and two levels for option selection (dependent variable, either 1 or 0 depending on whether the option had been selected or not, see Carnegie, 2017 for an example of frequency analysis). Target Items 2 to 5 were entered as different layers in these analyses because we wanted to analyze option position effects for each item individually to evaluate

findings' replicability. Statistical results are presented as percentages built from raw frequencies (with 95% confidence intervals [CI] calculated thanks to a formula presented in Supplementary method), both for collapsed items and for each individual item. Effect sizes were calculated for CR position and MAD distance effects by extracting two-way contingency tables from the three-way ones and are reported as Cramers' V (see formula in Supplementary method and Cohen, 2013, for interpretation). Further analyses of effect sizes were obtained by calculating the odds ratios (ORs) of selecting each type of option relative to all others and by analyzing levels of significance, magnitudes of ORs, and associated 95% CI. The frequency of selection for correct and incorrect responses is next called item accuracy and distractor's efficiency, respectively. Note that omissions (i.e., absence of any selected option for an item) and null responses (i.e., response other than A, B, C, D, or E to an item) were excluded from the data used for these analyses.

In Experiment 2, four 2×2 contingency tables (one for each of the four types of distractors) were constructed and analyzed employing Pearson Chi-Square Tests of Independence (with two levels for relative position to CR as independent variable and two levels for distractor selection as dependent variable) to determine the effects of distractors' position relative to CR (before, after) on distractors' efficiency. Four 4×2 contingency tables were also constructed and analyzed to observe the effects of a second independent variable, distractors' location (A, C, D, E), on distractors' efficiency. These analyses were focused on the influence of MAD position. Pearson Chi-Square Tests of Independence were finally conducted to examine the effects of each distractor's position relative to CR and location on item accuracy, using four 2×2 and four 4×2 contingency tables, respectively (both with CR selection as dependent variable). ORs for Experiment 2 were also analyzed.

Since the design of Experiment 2 was not fully crossed and involved just one item (Target Item 1), further analyses were conducted at test level and examinee level to make sure that the differences between conditions were robust and not spuriously driven by test forms' and/or examinees' particularities. A comparative analysis of answer keys' distribution of test forms was first used to test whether the obtained differences between conditions could be related to this factor. Then, a comparative analysis of students' position-based response patterns and students' responses to the item just before Target Item 1 observed in the different test forms was done, to verify that responses to Target Item 1 was not linked to any particular response strategy. Association between responses to Target Item 1 and the item just before was evaluated by calculating mean square contingency coefficient.

Results

Experiment 1: Effects of CR Position When CR Is Located at A Versus E

The influence of CR position on selected responses across items and the accuracy obtained in the different experimental conditions for each individual item are presented in Table 2 and Figure 1, respectively (see also Supplementary Figure S2). As

Table 2. Response Frequency Observed in Experiment I.

	A	E	A or E	A or E	A	A	E	E
					Percentage of examinees selecting each option			
CR position	A	E	A or E	A or E	A	A	E	E
MAD position	After	Before	Close	Far	After/close	After/far	Before/close	Before/far
CR	52.3	48.1	49.5	50.8	51.6	52.9	47.4	48.7
MAD	16.7	18.7	18.9	16.5	17.9	15.5	19.8	17.6
Distractor 2	12.9	13.6	13.4	13.0	13.4	12.3	13.4	13.7
Distractor 3	10.0	10.8	10.2	10.6	9.9	10.2	10.6	11.1
Distractor 4	8.2	8.8	8.0	9.0	7.2	9.2	8.8	8.9

Note. CR = correct response; MAD = most attractive distractor.
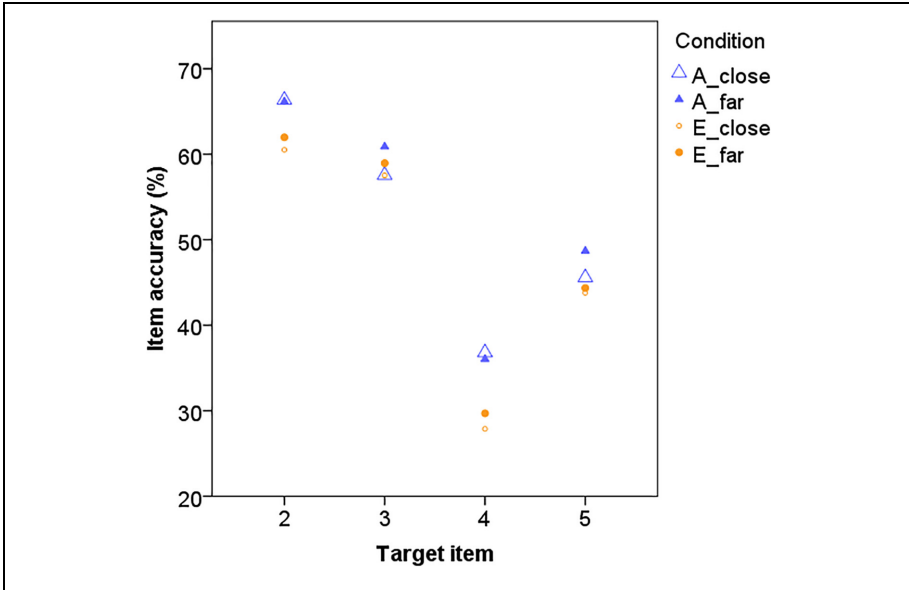
Figure 1. Item Accuracy Observed for the Four Items Manipulated in Experiment 1.
Note. The scale of Figure 1 Y axis starts at 20%. "A" and "E" refer to the position of the correct response; "close" and "far" to the distance from the correct response to the most attractive distractor.

predicted, overall item accuracy was higher ($+4.2\%$ in average) when CR was located at A (52.3%, 95% CI [52.1, 52.4]) than when it was located at E (48.1%, 95% CI [47.9, 48.2]). Consistently, a statistically significant effect of CR position on item accuracy was observed for all four collapsed target items, χ^2 (1) = 1,345.15, p < .001, V = .042. This effect was very small. It was also significant and very small for all individual items (all p_s < .001, V= .052, .010, .081, and .031, respectively, for Items 2–5), item accuracy being always higher when CR was located at A.

CR position also had a statistically significant effect on MAD efficiency, χ^2 (1) = 523.13, p < .001, V = .026, revealing that MAD efficiency was lower (-2.0% in average) when CR was in A (16.7%, 95% CI [16.6, 16.8]) than in E (18.7%, 95% CI [18.6, 18.8]). This effect was significant and very small for three of the four target items (all three p_s < .001, V = .043, .000, .046 and .013, respectively, for Items 2, 3, 4, 5). The effect of CR position on distractor's efficiency was also significant for all the other distractors. They were all less efficient when CR was located at A than when it was located at E, but effect sizes were even smaller for these weaker distractors (all p_s < .001, V = .011, .014, and .012 for Distractors 2–4, respectively).

Experiment 1: Effects of MAD Distance When CR Is Located at A Versus E

The influence of MAD distance on selected responses across items and on accuracy for each individual item are presented in Table 2 and Figure 1, respectively (see also

Supplementary Figure S2). As predicted, MAD was less frequently selected (-2.4% in average) when far from (16.5%, 95% CI [16.4, 16.6]) than when close to CR (18.9%, 95% CI [18.7, 19.0]) and item accuracy was lower (-1.3% in average) when MAD was close to CR (49.5%, 95% CI [49.3, 49.6]) than far from it (50.8%, 95% CI [50.7, 51.0]). Consistently, MAD distance from CR had a statistically significant effect on MAD efficiency for the four target items, considered together, χ^2 (1) = 729.71, $p < .001$, V = .031 or separately (all $p_s < .001$, V = .015, .018, .055 and .030, respectively for Items 2–5), and MAD distance from CR significantly affected item accuracy for the four target items, considered together, χ^2 (1) = 133.78, $p < .001$, V = .013 or separately (all $p_s < .013$, V = .006, .024, .006 and .018, respectively for Items 2–5). All these effects were very small but revealed a systematic and consistent phenomenon: MAD efficiency was always higher and item accuracy was always lower when MAD was close to CR. This was true independently of CR position (A or E), indicating that a response option was not always more frequently selected when presented earlier in the options list.

The effect of MAD distance on the efficiency of other distractors was also significant in some comparisons. However, not all distractors were more frequently selected in the condition "close" because not all distractors were necessarily close to and far from CR in the conditions called "close" and "far," respectively. Interestingly, MAD distance effect on the efficiency of the weakest distractor was significant and inverted in direction (i.e., Distractor 4 efficiency was higher in "far" than in "close" condition) for three of the four target items (all three $p_s < .039$, V = .010, .000, .039 and .026, respectively for Items 2–5) as well as overall, χ^2 (1) = 246.773, $p < .001$, V = .018). This result indicated that even the weakest distractor was less frequently selected (-1.0% in average) when far away from CR (in "close" condition). This result must be interpreted with care, however, because the selection of MAD and Distractor 4 were interdependent phenomena.

Experiment 1: Further Analyses and Results Synthesis

OR analyses confirmed the results reported above. On average, examinees were 1.15 times more likely to select MAD when it preceded CR than when it was located after CR and 1.18 times more likely to select MAD when it was close to CR than when it was far away from CR. Consistently, examinees were on average 1.18 times more likely to select CR when it was located at A instead of E and were 1.05 times more likely to select CR when MAD was far away from it than when MAD was close to it. OR values reflected very small effects. These effects were even smaller for other distractors. However, these effects were almost always observed and statistically significant at a very low p level ($p < .001$), for each one of the four items considered in Experiment 1 (see Table 3).

Taken together, Experiment 1 results suggested that the efficiency of all distractors varied (although in a small amount) depending on both their relative position and distance to CR. It is reasonable to think that the decreases in item accuracy observed

Table 3. ORs Observed in Experiment 1.

Option	Option position	Reference category	OR [95% CI]				
			Item 2	Item 3	Item 4	Item 5	Overall
CR	A	E	1.24 [1.22, 1.27]***	1.04 [1.02, 1.06]***	1.42 [1.39, 1.44]***	1.13 [1.11, 1.15]***	1.18 [1.17, 1.19]***
	Far	Close	1.03 [1.01, 1.05]**	1.10 [1.08, 1.12]***	1.02 [1.00, 1.04]*	1.08 [1.06, 1.10]***	1.05 [1.04, 1.06]***
MAD	Before CR	After CR	1.30 [1.27, 1.34]***	1.00 [0.97, 1.03]	1.22 [1.20, 1.25]***	1.07 [1.04, 1.09]***	1.15 [1.13, 1.16]***
	Close	Far	1.10 [1.07, 1.13]***	1.12 [1.09, 1.15]***	1.27 [1.25, 1.30]***	1.17 [1.15, 1.20]***	1.18 [1.16, 1.19]***
Distractor 2	Before CR	After CR	1.02 [0.99, 1.06]	1.07 [1.04, 1.10]***	1.10 [1.08, 1.13]***	1.05 [1.02, 1.07]***	1.06 [1.05, 1.08]***
	Close (approximately)	Far (approximately)	1.12 [1.08, 1.16]***	1.09 [1.06, 1.12]***	0.96 [0.94, 0.98]***	1.04 [1.02, 1.06]***	1.03 [1.02, 1.05]***
Distractor 3	Before CR	After CR	1.22 [1.19, 1.26]***	1.02 [0.99, 1.05]	1.10 [1.07, 1.14]***	0.99 [0.96, 1.03]	1.09 [1.07, 1.10]***
	Close (approximately)[a]	Far (approximately)[a]	1.07 [1.03, 1.10]***	0.95 [0.92, 0.98]**	1.16 [1.13, 1.19]***	0.99 [0.96, 1.02]	1.04 [1.03, 1.06]***
Distractor 4	Before CR	After CR[a]	1.22 [1.19, 1.26]***	1.02 [0.99, 1.05]	1.10 [1.07, 1.14]***	0.99 [0.96, 1.03]	1.09 [1.07, 1.10]***
	Close[a]	Far[a]	1.07 [1.04, 1.11]***	1.00 [0.97, 1.03]	1.33 [1.29, 1.37]***	1.25 [1.20, 1.30]***	1.14 [1.12, 1.16]***

Note. OR = odds ratio; CI = confidence interval; CR = correct response; MAD = most attractive distractor.

[a]For Distractors 3 and 4, the "close" position corresponded to the experimental condition far.

Gray-colored text corresponds to results which were contrary to predictions.

*$p < .05$. **$p < .01$. ***$p < .001$.

when CR was switched from A to E or when the strongest distractor was closer to CR were mainly related to the fact that distractors had a higher power of distraction when explored before CR and in an adjacent time window. Consistently with this possible interpretation, omission and null responses' rates were quite similar in all conditions of CR position and MAD distance (from between 1.1% and 1.5% and between 0.0% and 0.1%, respectively), indicating that the observed changes in item difficulty were attributable to changes in distractors' appeal more than to accidental events. The fact that CR was systematically selected more often when it was located at A and when MAD was far from it, for all four items, indicated that the effects of CR position and MAD distance on performance were not due to an eventual bias in the test forms' assignment to participants.

Experiment 2: Effects of MAD Position When CR Is Held Fixed at B

Effects of distractors' position on selected responses for Target Item 1 are presented in Figure 2. As predicted, MAD was more frequently selected (between $+1.6\%$ and $+4.9\%$) when it was located at A (12.5%, 95% CI [12.2, 12.7]) than when it was located at any other position (10.9%, 95% CI [10.6, 11.2], 9.5%, 95% CI [9.3, 9.8], and 7.6% 95% CI [7.4, 7.9] in C, D, and E, respectively), being selected more frequently ($+3.1\%$ in average) when located before (12.5%, 95% CI [12.2, 12.7]) than after CR (9.4% on average, 95% CI [9.2, 9.5], see Figure 2A and 2B). Consistently, MAD position, either considered as relative to CR (before, after) or as location (A, C, D, E) had a statistically significant effect on its efficiency, χ^2 (1) = 383.388, $p < .001$, V = .044 and χ^2 (3) = 667.429, $p < .001$, V = .059, respectively. The effects of relative position and location on MAD efficiency were very small. Similar effects were observed for the four other distractors, although effect sizes were even smaller (all $p_s < .001$, V = .025 and .030 for Distractor 2, .019 and .028 for Distractor 3, and .023 and .032 for Distractor 4). These results indicated that all distractors were more frequently selected when they appeared before CR, and the earlier they appeared in the options list, the more frequently they were selected.

Consistently with our predictions, the lowest accuracy for the Target Item 1 was found when MAD was located at A and item accuracy was lower (-2.5% in average) when MAD was located before CR (78.3%, 95% CI [77.9, 78.7]) than after it (80.8%, 95% CI [80.6, 81.0]). MAD position had a very small but statistically significant effect on item accuracy when considered as MAD relative position (before, after) as well as MAD location (A, C, D, E), χ^2 (1) = 143.547, $p < .001$, V = .027 and χ^2 (3) = 252.602, $p < .001$, V = .036, respectively (see Figure 2C).

Experiment 2: Further Analyses and Results Synthesis

OR analyses confirmed the results just above reported. Distractors were more likely to be selected when they were located before CR than after CR (from 1.30 to 1.41 times, depending on the distractor). More precisely, the four distractors were more
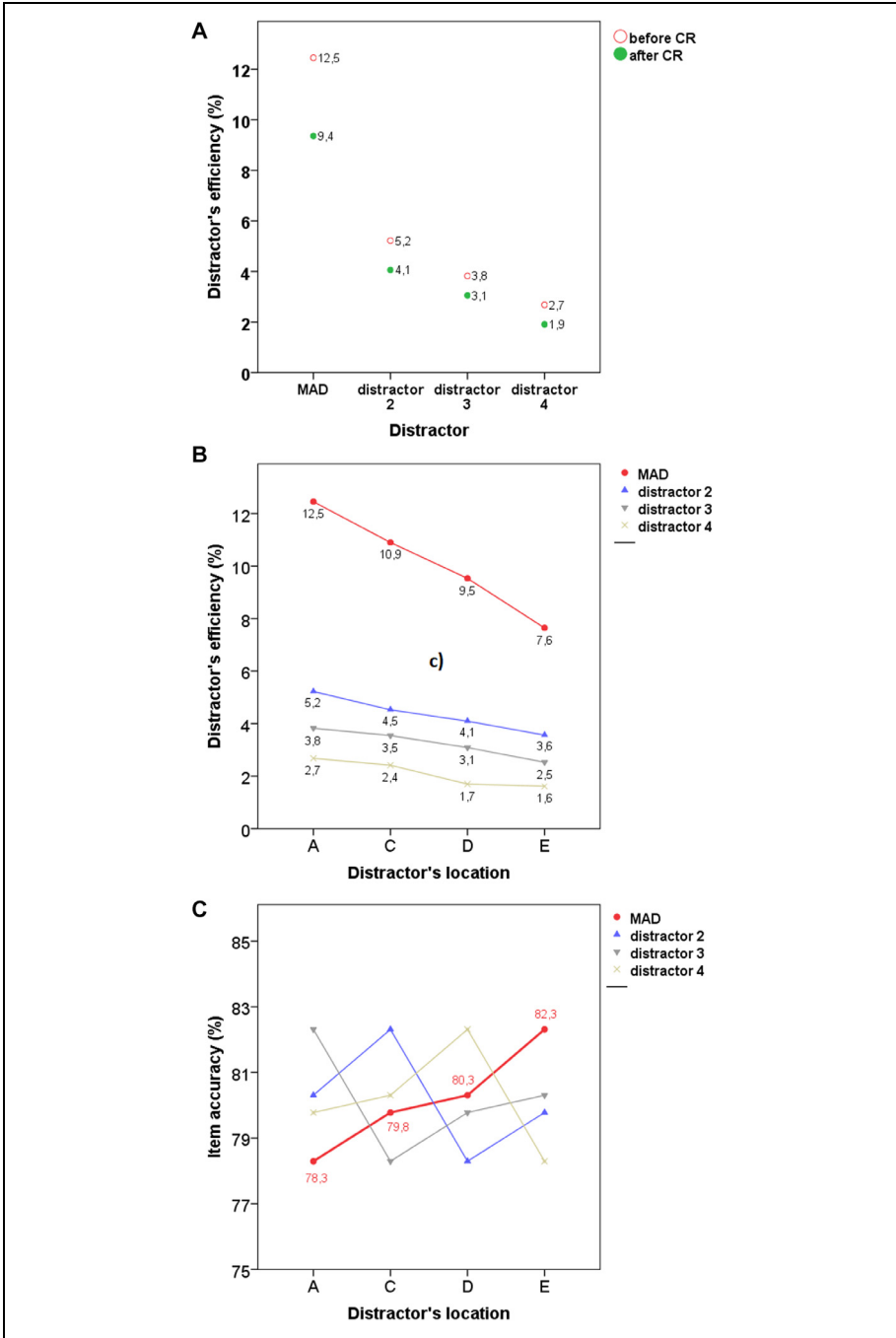
Figure 2. Distractors' Efficiency (A, B) and Item Accuracy (C) Observed in Experiment 2.
Note. The scale of Figure 2C Y axis starts at 75%. CR = correct response; MAD = most attractive distractor.

likely to be selected when they were located at A than at C (from 1.08 to 1.16 times), D (from 1.25 to 1.60 times), or E (from 1.49 to 1.72 times), when they were located at C than at D (from 1.11 to 1.43 times) or E (from 1.28 to 1.51 times) and when they were located at D than at E (from 1.05 to 1.27 times). The only distractor whose position seemed to affect on accuracy clearly and consistently was MAD: CR was 1.17 times more likely to be selected when MAD was located after CR than before CR and was more likely to be selected when MAD was in A than in any other location (from 1.09 to 1.29 times). The further away from CR was MAD, the more likely to be selected was CR. OR values in Experiment 2 reflected small or very small effects but, as in Experiment 1, effects of MAD position on MAD efficiency and on accuracy were always consistently observed and almost always statistically significant at a very low p level ($p < .001$, see Tables 4 and 5).

Supplementary analyses tested whether the obtained differences between conditions could be related to the answer keys' distribution of test forms. Answer keys' distribution was unbalanced differently in all test forms, with between 11 and 22 keys in each option position (see Supplementary Table S1b). There were never more than two keys at the same position in a row, key position repetition being very low in all forms (between 5.1% and 7.6%). The same block of items (anchor) preceded Target Item 1 in all forms, and the item just before Target Item 1, whose key was on C, always had the same arrangement of response options. Thus, the answer keys' distribution gave no clear incentive to select any specific position to answer Target Item 1 in any form.

Then, we tested whether students' position-based response pattern or students' responses to the item just before Target Item 1 could have affected the observed results for Target Item 1. Comparisons between the examinees' groups which received a different test form indicated that the group who solved Target Item 1 with MAD in A did not particularly over-select A (23.6 v/s 21.8, 23.2% and 20.7 %) or under-select B (18.9 vs. 16.9, 19.0% and 16.5 %) through the whole test. This group responded A and B to the item just before Target Item 1 faintly less frequently (30.1 vs. 31.7, 30.3% and 32.1 % of A; 9.8 vs. 10.8, 10.4% and 10.4 % of B), but neither selecting A nor selecting B in both items was correlated (mean square contingency coefficient for binary variables rφ < 0.01). In any case, examinees were not better at finding CR in B throughout the whole test in any test form (accuracy in B was 53.6 vs. 56.1, 59.0, and 53.3 %), and the percentage of items keyed in B was low in all forms (12.5, 15.0, 18.75, and 13.75 % of keys B), being inferior to 20% which would be expected with a balanced distribution of answer keys. It was thus reasonable to think that all examinees, including those who used balanced guessing as response strategies, probably considered B as a potential CR location for Target Item 1, independently of the test form they responded to.

In sum, the observed effects of options position on selected responses for Target Item 1 did not seem to be related to differences between examinees or test's characteristics between the four test forms. General achievement, answer keys' distribution, examinees' position-based strategies, or response selection to the item just before the

Table 4. ORs Observed in Experiment 2 for Distractors' Selection.

Option position	Reference category	OR [95% CI]			
		MAD	Distractor 2	Distractor 3	Distractor 4
A	C	1.16 [1.12, 1.21]***	1.16 [1.10, 1.23]***	1.08 [1.01, 1.16]*	1.11 [1.03, 1.21]***
A	D	1.35 [1.30, 1.41]***	1.29 [1.22, 1.37]***	1.25 [1.16, 1.34]***	1.60 [1.46, 1.75]***
A	E	1.72 [1.65, 1.79]***	1.49 [1.40, 1.59]***	1.53 [1.42, 1.65]***	1.68 [1.53, 1.84]***
C	D	1.16 [1.11, 1.21]***	1.11 [1.04, 1.18]***	1.15 [1.07, 1.24]***	1.43 [1.31, 1.57]***
C	E	1.48 [1.41, 1.54]***	1.28 [1.20, 1.37]***	1.42 [1.32, 1.53]***	1.51 [1.37, 1.65]***
D	E	1.27 [1.22, 1.33]***	1.16 [1.08, 1.24]***	1.23 [1.14, 1.33]***	1.05 [0.95, 1.16]
Before CR	After CR	1.38 [1.33, 1.42]***	1.30 [1.24, 1.37]***	1.26 [1.19, 1.33]***	1.41 [1.32, 1.51]***

Note. Gray-colored text corresponds to results that were contrary to predictions. OR = odds ratio; CI = confidence interval; MAD = most attractive distractor; CR = correct response.

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5. ORs Observed in Experiment 2 for Correct Response's Selection.

Option position	Reference category	OR [95% CI]			
		MAD	Distractor 2	Distractor 3	Distractor 4
C	A	1.09 [1.06, 1.13]***	1.14 [1.10, 1.18]***	0.78 [0.75, 0.80]***	1.03 [1.00, 1.07]*
D	A	1.13 [1.10, 1.17]***	0.88 [0.86, 0.91]***	0.85 [0.82, 0.88]***	1.18 [1.14, 1.22]***
E	A	1.29 [1.25, 1.33]***	0.97 [0.94, 1.00]*	0.88 [0.85, 0.91]***	0.91 [0.89, 0.94]***
D	C	1.03 [1.00, 1.07]*	0.78 [0.75, 0.80]***	1.09 [1.06, 1.13]***	1.14 [1.10, 1.18]***
E	C	1.18 [1.14, 1.22]***	0.85 [0.82, 0.88]***	1.13 [1.10, 1.17]***	0.88 [0.86, 0.91]***
E	D	1.14 [1.10, 1.18]***	1.09 [1.06, 1.13]***	1.03 [1.00, 1.07]*	0.78 [0.75, 0.80]***
Before distractor	After distractor	1.17 [1.14, 1.20]***	0.99 [0.96, 1.01]	0.83 [0.81, 0.85]***	1.03 [1.01, 1.06]*

Note. Gray-colored text corresponds to results that were contrary to predictions. OR = odds ratio; CI = confidence interval; MAD = most attractive distractor; CR = correct response.

*$p < .05$. **$p < .01$. ***$p < .001$.

Target Item 1 did not clearly influence the observed results. The observed effects of options position in Experiment 2 seemed to be better explained by the fact that the earlier MAD appeared in Target Item 1, the higher its efficiency (and consequently item difficulty) was.

Discussion

In this study, data from a large-sized sample obtained in a real-world context showed that the same items were slightly more accurately responded to when (a) the CR was located at A instead of E, (b) the MAD was positioned after CR instead of before CR, and (c) the MAD was positioned far from CR instead of close to it. These results are in line with previous claims and findings that items are easier when CR is positioned first in the options list (Ace & Dawis, 1973; Bresnock et al., 1989; Hodson, 1984) and when strong distraction is located after CR (Fagley, 1987; Friel & Johnstone, 1979; Kiat et al., 2018) and far from it (Shin et al., 2020). They are consistent with the presence of an early answer advantage (Hohensinn & Baghaei, 2017; Holzknecht et al., 2020; Schroeder et al., 2012; Tellinghuisen & Sulikowski, 2008), but not with a middle bias (Attali & Bar-Hillel, 2003). They extend previous findings because consistent effects of CR position, distractors' relative position to CR, and distractors' distance to CR were observed repetitively and in a natural setting. They establish that the position of response options has a small but systematic impact on item solving even in a most consequential test. Moreover, the analysis of a very large sample has revealed that position influences even the weakest distractor's selection, which had not been observed in the past.

Theoretical Relevance: Psychological Principles Underlying Option Position Effects

It has been claimed that items are easier when keyed on A because reading CR first allows processing it as fast as possible (Bresnock et al., 1989). Eye-tracking studies have confirmed that Option A is the first students read (Holzknecht et al., 2020), and the economy of cognitive resources has been advanced as a general explanation for early-answer advantages (Schroeder et al., 2012). However, there is still no direct evidence that speediness is related to option position effects on item accuracy, and this mechanism does not explain distractors' distance effects.

It has also been claimed that presenting CR first allows processing it before considering any misleading alternative (Bresnock et al., 1989) and would prevent strong distractors from causing interference in CR memory retrieval (Kiat et al., 2018). The interference memory retrieval theory (Anderson et al., 1994) assumes that any information processing hinders the accessibility of closely related but non-retrieved memory traces. It would provide a more convincing explanation for option position and distance effects. The fact that MAD was more efficient when close to CR in this

study, even when located after CR, may suggest that interference can alter other stages of the CR recognition process, not only retrieval.

The satisficing theory (Simon, 1956) has also been invoked to potentially explain option position effects (Kiat et al., 2018). This theory predicts that options presented earlier have a better chance to be chosen because test takers often select the first acceptable option they read (Krosnick, 1991). Consistently, it has been noted that examinees do not always read all the alternatives before responding (Clark, 1956; Fagley, 1987; Willing, 2013). However, the high-stakes nature of the test observed in this study did not promote engaging in a satisficing strategy, and the effect of MAD distance observed when CR was in E seems to support interference theory. In fact, only a phenomenon of interference between options processing can explain why MAD was more efficient in D (condition "E_close") than in A (condition "E_far"). The fact that a strong distractor presented later could provoke higher distraction (if later means closer to CR) could reflect that interference decreases with time, any distractor's interference power being maximal when read just before CR. Presenting strong distractors close to CR, either before or after, seems in no case to help item solving by making the comparison between MAD and CR easier as proposed in other studies (Ambu-Saidi & Khamis, 2000; Friel & Johnstone, 1979).

Finally, the "middle bias" was probably not observed here because it is mainly detected when students answer multiple-choice items through guessing (Attali & Bar-Hillel, 2003; Bar-Hillel, 2015). Most examinees had received months of training before taking the national exam analyzed in this study. They probably had at least partial knowledge for most items and did not heavily rely on guessing strategies in such a high-stakes situation.

Limitations and Future Studies

The option position effect observed in this study can be deemed as clear, systematic, and consistent because predicted effects of CR position, MAD relative position, MAD distance, and MAD location were all reliably observed and significant at a very low p level ($p_s < .001$). However, all calculated effect sizes were small or even very small (all Vs $< .1$; all ORs < 2) and low p levels were probably in part due to the large sample size used in this study. As other authors claimed in the past, option position effects might thus be labeled "manifest but negligible" (Hohensinn & Baghaei, 2017), but this conclusion could underestimate the aggregate strength of option position effects at the test level. It seems more accurate to consider option position effect as "small but important," as advanced by Holzknecht et al. (2020). In this study, almost all predicted option position effects were visible even when considering a single item. One test form with a high frequency of items in "A_far" condition (i.e., keyed on A with MAD in E) would most probably be substantially easier to solve than one form with a high frequency of items in "E_close" condition (i.e., keyed on E with MAD in D).

In addition, option position effects could be stronger for some examinees' subgroups. Low achievers and young examinees were previously found to be particularly sensitive to these effects (Bolt et al., 2020; Shin et al., 2020; Sonnleitner et al., 2016), and even student personality could interact with option position effects (Kim et al., 2017). Supplementary analyses considering the influence of examinees' latent ability in language (estimated through item response theory) on item accuracy and MAD efficiency observed in the different conditions of Experiment 1 and Experiment 2 were run to challenge this possibility. These analyses indicated that option position effects observed in this study were instead found at all ability levels (see Supplementary Figure S3). They provide another confirmation that the effects of options position observed in this study follow a clear pattern; they are not noise in data and are not due to coincidence.

On a similar note, option position effects could be influenced by items' characteristics. These effects were previously found to be stronger when items are difficult (Bolt et al., 2020) and to potentially vary depending on the tested domain (Schroeder et al., 2012; Sonnleitner et al., 2016). In this study, these effects were observed for items of different difficulty levels. However, the difficulty level was not considered a study variable, so further studies should analyze the interaction between item difficulty level and option position effects more precisely. Moreover, this study's results should be replicated in testing fields other than language to inform which tested domains are most sensitive to such effects. Option position effects observed in different disciplines have never been compared through empirical studies or systematic reviews, and the possibility that option position affects option selection for options of some special form (e.g., numbers) more than for others might explain inconsistent results obtained in the past. Further research should also address how the phenomenon of interference between options' processing interacts with test-takers and items' characteristics.

Practical Implications

Multiple-choice items are not ''immune'' to positional response set, as was initially claimed by Cronbach (1950). Therefore, we believe that the position of response options must be considered in test development processes, at least for the construction and analysis of high-stakes tests administered in several test forms with different option position configurations. Consistently, existing item-writing guides have provided guidelines about response options placement. Most guides recommend that test developers arrange options in a ''logical'' order (e.g., numerically, alphabetically) and/or vary CR position arbitrarily, by key balancing, or through randomization (Haladyna & Rodriguez, 2013). The main principle behind these suggestions is that the position of item options should never lead examinees to detect the correct response. The fact that not all test developers adhere to this principle and that position biases are observable in the distribution of answer keys and strong distractors is an issue of concern (Attali & Bar-Hillel, 2003; Lions et al., 2021, 2022; Mentzer, 1982; Metfessel & Sax, 1958).

As far as we are concerned, any option placement strategy that does not provide any clue to the correct answer is acceptable. Arranging options in random order seems a good placement strategy, being an easy and quick procedure to apply using computer devices. Randomizing the order of response options rather than the key position seems preferable considering the influence of incorrect options' position observed in this study. Presenting options in a random-like order, such as alphabetically (or numerically if options are numbers), would be another good alternative, but it probably consumes more time. In any case, a computer device should be used if some randomization procedure is applied because humans cannot produce actual random sequences (Lee, 2019). It is important to note that in no case should key balancing be used, as balancing promotes balanced guessing, a successful position-based strategy that item designers should keep in check (Bar-Hillel & Attali, 2002; Bar-Hillel et al., 2005).

Moreover, it is cautious to present options in a fixed random order for all examinees: Although shuffling options differently for each student is potentially helpful to reduce cheating in tests administered online, it could lead to inequity situations and might be best not to be used, particularly when the test's stakes are high (Bolt et al., 2020). In any case, test instructions should include a warning against the potential misuse of options position to spot CR (Paul et al., 2014).

Test developers could increase the difficulty of a too-easy item or turn a non-functional distractor into a functional one by modifying response order. In standardized tests, options position should be considered when piloting items. When option position is scrambled for test security reasons, the impact should be evaluated on item and test analysis to make sure it does not generate any test inequities. No test developer should doubt that the options position influences the examinees' performance. It subtly does. Certainly, the impact of knowledge on performance is much higher, but still, options position is worth considering during development, application, and analysis of multiple-choice tests.

Author Contributions

S.L., P.D., and J.L. developed the study concept. S.L., M.I.G., N.C., and D.J. contributed to the study design. Testing and data collection were performed by the examiners of the *Departamento de Evaluación, Medición y Registro Educacional*, the state institution in charge of creating and administering national university admission tests. S.L., M.I.G., and G.T. performed the data analysis and interpretation. S.L. drafted the manuscript, and all the other authors provided critical revisions. All authors approved the final version of the manuscript before submission.

Declaration of Conflicting Interests

Funding

Open Practices Statement

Neither of the experiments reported in this article was formally preregistered. Neither the data nor the materials have been made available on a permanent third-party archive; requests for the data or materials can be sent via email to the lead author at severin.lions@ciae.uchile.cl.

Educational Impact and Implications Statement

Multiple-choice items are used to assess learning at almost every educational level. They can be found both in teacher-made tests and in standardized tests. Consequently, test developers should be aware that item-design features may impact test achievement, obscuring the assessment of examinees' knowledge. In this study, the position of both correct response and distractors are shown to be potential sources of measurement error in multiple-choice tests. Proposals to address this potential validity issue are discussed. Many test developers have probably asked themselves whether and how the position of response options influences test-taker's responses and items' psychometric properties so that results presented in this study may be of interest to a broad scientific and non-scientific audience.

ORCID iD

Séverin Lions https://orcid.org/0000-0003-0697-7974

Supplemental Material

Supplemental material for this article is available online.

References

Ace, M. C., & Dawis, R. V. (1973). Item structure as a determinant of item difficulty in verbal analogies. *Educational and Psychological Measurement*, *33*(1), 143–149. https://doi.org/10.1177/001316447303300115

Ambu-Saidi, A., & Khamis, A. (2000). *An investigation into fixed response questions in science at secondary and tertiary levels* [Doctoral dissertation, University of Glasgow].

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087. https://doi.org/10.1037/0278-7393.20.5.1063

Attali, Y., & Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. *Journal of Educational Measurement*, *40*(2), 109–128. https://doi.org/10.1111/j.1745-3984.2003.tb01099.x

Bar-Hillel, M. (2015). Position effects in choice from simultaneous displays: A conundrum solved. *Perspectives on Psychological Science*, *10*(4), 419–433. https://doi.org/10.1177/1745691615588092

Bar-Hillel, M., & Attali, Y. (2002). Seek whence: Answer sequences and their consequences in key-balanced multiple-choice tests. *The American Statistician*, *56*(4), 299–303. https://doi.org/10.1198/000313002623

Bar-Hillel, M., Budescu, D., & Attali, Y. (2005). Scoring and keying multiple choice tests: A case study in irrationality. *Mind & Society*, *4*(1), 3–12. https://doi.org/10.1007/s11299-005-0001-z

Bolt, D. M., Kim, N., Wollack, J., Pan, Y., Eckerly, C., & Sowles, J. (2020). A psychometric model for discrete-option multiple-choice items. *Applied Psychological Measurement*, *44*(1), 33–48. https://doi.org/10.1177/0146621619835499

Bresnock, A. E., Graves, P. E., & White, N. (1989). Multiple-choice testing: Question and response position. *The Journal of Economic Education*, *20*(3), 239–245. https://doi.org/10.2307/1182299

Carnegie, J. A. (2017). Does correct answer distribution influence student choices when writing multiple choice examinations? *Canadian Journal for the Scholarship of Teaching and Learning*, *8*(1), Article 11. https://doi.org/10.5206/cjsotl-rcacea.2017.1.11

Cizek, G. J. (1994). The effect of altering the position of options in a multiple-choice examination. *Educational and Psychological Measurement*, *54*(1), 8–20. https://doi.org/10.1177/0013164494054001002

Clark, E. L. (1956). General response patterns to five-choice items. *Journal of Educational Psychology*, *47*(2), 110–117. https://doi.org/10.1037/h0043113

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, *10*(1), 3–31. https://doi.org/10.1177/001316445001000101

DeVore, S., Stewart, J., & Stewart, G. (2016). Examining the effects of testwiseness in conceptual physics evaluations. *Physical Review Physics Education Research*, *12*(2), Article 020138. https://doi.org/10.1103/PhysRevPhysEducRes.12.020138

Fagley, N. S. (1987). Positional response bias in multiple-choice tests of learning: Its relation to testwiseness and guessing strategy. *Journal of Educational Psychology*, *79*(1), 95–97. https://doi.org/10.1037/0022-0663.79.1.95

Friel, S., & Johnstone, A. H. (1979). Does the position matter? *Education in Chemistry*, *16*, 175. https://eric.ed.gov/?id=EJ213396

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, *87*(6), 1082–1116. https://doi.org/10.3102/0034654317726529

Hagenmüller, B. (2021). On the impact of the response options' position on item difficulty in multiple-choice-items. *European Journal of Psychological Assessment*, *37*(4), 290–299. https://doi.org/10.1027/1015-5759/a000615

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, *23*(1), 17–27. https://doi.org/10.1111/j.1745-3992.2004.tb00149.x

Haladyna, T. M., & Rodriguez, M. C. (2013). Guidelines for writing selected-response items. In T. M. Haladyna & M. C. Rodriguez (Eds.), *Developing and validating test items* (pp. 89–110). Routledge. https://doi.org/10.4324/9780203850381

Hodson, D. (1984). Some effects of changes in question structure and sequence on performance in a multiple choice chemistry test. *Research in Science & Technological Education*, *2*(2), 177–185. https://doi.org/10.1080/0263514840020209

Hohensinn, C., & Baghaei, P. (2017). Does the position of response options in multiple-choice tests matter? *Psicológica*, *38*(1), 93–109. https://files.eric.ed.gov/fulltext/EJ1125979.pdf

Holzknecht, F., McCray, G., Eberharter, K., Kremmel, B., Zehentner, M., Spiby, R., & Dunlea, J. (2020). The effect of response order on candidate viewing behaviour and item difficulty in a multiple-choice listening test. *Language Testing*, *38*, 41–61. https://doi.org/10.1177/0265532220917316

Kiat, J. E., Ong, A. R., & Ganesan, A. (2018). The influence of distractor strength and response order on MCQ responding. *Educational Psychology*, *38*(3), 368–380. https://doi.org/10.1080/01443410.2017.1349877

Kim, N., Bolt, D. M., Wollack, J., Pan, Y., Eckerly, C., & Sowles, J. (2017, July). Modeling examinee heterogeneity in discrete option multiple choice items. In *The annual meeting of the psychometric society* (pp. 383–392). Springer. https://doi.org/10.1007/978-3-030-01310-3_33

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Lee, C. J. (2019). The test taker's fallacy: How students guess answers on multiple-choice tests. *Journal of Behavioral Decision Making*, *32*(2), 140–151. https://doi.org/10.1002/bdm.2101

Lions, S., Monsalve, C., Dartnell, P., Godoy, M. I., Córdova, N., Jiménez, D., Blanco, M.P., Ortega, G., & Lemarié, J. (2021). The position of distractors in multiple-choice test items: The strongest precede the weakest. Frontiers in Education, 6, 731763. https://doi.org/10.3389/feduc.2021.731763

Lions, S., Monsalve, C., Dartnell, P., Blanco, M.P., Ortega, G., & Lemarié, J. (2022) Does the response options placement provide clues to the correct answers in multiple-choice tests? A systematic review, Applied Measurement in Education, 35:2, 133–152. https://doi.org/10.1080/08957347.2022.2067539

Mentzer, T. L. (1982). Response biases in multiple-choice test item files. *Educational and Psychological Measurement*, *42*(2), 437–448. https://doi.org/10.1177/001316448204200206

Metfessel, N. S., & Sax, G. (1958). Systematic biases in the keying of correct responses on certain standardized tests. *Educational and Psychological Measurement*, *18*(4), 787–790. https://doi.org/10.1177/001316445801800411

Paul, S. T., Monda, S., Olausson, S. M., & Reed-Daley, B. (2014). Effects of apophenia on multiple-choice exam performance. *SAGE Open*, *4*(4), 1–7. https://doi.org/10.1177/2158244014556628

Schroeder, J., Murphy, K. L., & Holme, T. A. (2012). Investigating factors that influence item performance on ACS exams. *Journal of Chemical Education*, *89*(3), 346–350. https://doi.org/10.1021/ed101175f

Shin, J., Bulut, O., & Gierl, M. J. (2020). The effect of the most-attractive-distractor location on multiple-choice item difficulty. *The Journal of Experimental Education*, *88*, 643–659. https://doi.org/10.1080/00220973.2019.1629577

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. https://doi.org/10.1037/h0042769

Sonnleitner, P., Guill, K., & Hohensinn, C. (2016, July). *Effects of correct answer position on multiple-choice item difficulty in educational settings: Where would you go?* [Paper presentation]. *The international test commission conference*, Vancouver, British Columbia, Canada. http://hdl.handle.net/10993/29469

Taylor, A. K. (2005). Violating conventional wisdom in multiple choice test construction. *College Student Journal*, *39*, 141–153. https://eric.ed.gov/?id=EJ711903

Tellinghuisen, J., & Sulikowski, M. M. (2008). Does the answer order matter on multiple-choice exams? *Journal of Chemical Education*, *85*(4), 572–575. https://doi.org/10.1021/ed085p572

Wang, L. (2019). Does rearranging multiple-choice item response options affect item and test performance? *ETS Research Report Series*, *2019*(1), 1–14. https://doi.org/10.1002/ets2.12238

Wilbur, P. H. (1970). Positional response set among high school students on multiple-choice tests. *Journal of Educational Measurement*, *7*(3), 161–163. https://www.jstor.org/stable/1434073

Willing, S. (2013). *Discrete-option multiple-choice: Evaluating the psychometric properties of a new method of knowledge assessment* [Doctoral dissertation, University of Düsseldorf]. https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=27633