Investigation of Response Changes in the GRE Revised General Test

Educational and Psychological
Measurement

2015, Vol. 75(6) 1002–1020

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164415573988

epm.sagepub.com

**Ou Lydia Liu¹, Brent Bridgeman¹, Lixiong Gu¹,
Jun Xu¹, and Nan Kong¹**

Abstract

Research on examinees' response changes on multiple-choice tests over the past 80 years has yielded some consistent findings, including that most examinees make score gains by changing answers. This study expands the research on response changes by focusing on a high-stakes admissions test—the Verbal Reasoning and Quantitative Reasoning measures of the GRE revised General Test. We analyzed data from 8,538 examinees for Quantitative and 9,140 for Verbal sections who took the GRE revised General Test in 12 countries. The analyses yielded findings consistent with prior research. In addition, as examinees' ability increases, the benefit of response changing increases. The study yielded significant implications for both test agencies and test takers. Computer adaptive tests often do not allow the test takers to review and revise. Findings from this study confirm the benefit of such features.

Keywords

GRE, response change, computer adaptive test

Discussions on response changes on multiple-choice items date back to as early as the 1920s when Mathews (1929) reported that 86% of the students he surveyed believed that they did not profit from changing answers on multiple-choice examinations. Geiger (1996) found that about 53% to 88% of students believed that they would lose points if they changed answers on a test and more females tended to believe this than males. Kruger, Wirtz, and Miller (2005) also reported that 38% of students thought that their test scores would be hurt by switching answers.

¹Educational Testing Service, Princeton, NJ, USA

Corresponding Author:

Ou Lydia Liu, Educational Testing Service, 660 Rosedale Road, Mailstop R07, Princeton, NJ 08541, USA.

Email: lliu@ets.org

Students' beliefs reflect a common view that changing answers may lower one's test scores. Materials on study skills also suggest that students stick with their first hunch. For example, in the book *How to Prepare for the GRE: Graduate Record Examination* (Brownstein, Wolf, & Green, 2000), students were instructed to "exercise great caution if you decide to change an answer" (p. 6) as experience showed that students may change from right to wrong answers. Students are often told that their first response is usually correct or that they should not change answers on multiple-choice tests because their initial reactions to items could be more accurate than subsequent responses (Frender, 2004).

Many college instructors hold a similar view that changing one's initial response may lower students' test scores (Benjamin, Cavell, & Shallenberger, 1984). Through a search of popular websites in Australia that provide advice to students on taking a multiple-choice test, Milia (2007) found that a common suggestion for students was not to change their initial answers but instead to trust their original intuition.

Literature Review

Response Changes and Test Performance

Although popular beliefs promote the retention of original answers, research has consistently shown that judicious answer changing increases test takers' scores (Beck, 1978; Geiger, 1996; Harvil & Davis, 1997; Higham & Gerrard, 2005; Mathews, 1929; Vispoel, 1998, 2000). For example, Mathews (1929) reported that 53% of students were able to increase their scores by changing responses. Nieswiadomy, Arnold, and Garza (2001) reported that 87% of students benefited from answer changing on a nursing test, as opposed to 7% whose scores declined as a result of answer changing on the same test. Kruger et al. (2005) reported that among answers changed by undergraduate psychology students, 51% were changed from wrong to right, 25% from right to wrong, and 23% from one wrong answer to another. Kruger et al. (2005) also collected students' predictions of the effect of their changed answers. Results show that students significantly underestimated the benefit of answer changing.

Milia's (2007) study confirmed that although students only change a small portion of responses, changing answers is more likely to be beneficial than detrimental. Milia examined student responses on an undergraduate human resource management exam and on a postgraduate law exam. On both exams, there was higher percentage of test takers who gained scores as a result of answer changing than test takers who lost scores. Milia also found that high-performing test takers tended to make more wrong to right changes than low-performing test takers.

Al-Hamly and Coombe (2005) reported that 67% of students changed answers on a language test but the total percentage of answers changed was small (2.65%). Among the students who changed answers, 57% improved their test scores compared with the 19% who were disadvantaged by the changes. This study also found that, when compared with lower scorers, higher scorers were both less likely to change answers and made fewer wrong to wrong changes.

That changing responses is a likely benefit to test performance seems to be a predominant finding across studies. Based on a review of 56 published studies that investigate the effect of response changes since 1929, Al-Hamly and Coombe (2005) reached the following conclusions about response changes: (a) only a small portion of answers are typically changed on a test (e.g., ranging from 2% to 6%, Benjamin et al., 1984), (b) most test takers change responses (e.g., 57% to 96%, Benjamin et al., 1984), (c) most of the changes made are from wrong to right responses, and most important (d) response changes are likely to improve test scores. In fact, these conclusions are true regardless of whether the test is multiple-choice or true-false, achievement, or aptitude, speeded or unspeeded, and computer-based or paper and pencil (Kruger et al., 2005).

The effect of response changes may interact with the nature of the mistakes students make on multiple-choice items. While response changes are able to correct errors made by careless or speeded responding, they do not affect errors due to misunderstanding or confusability among alternative options on multiple-choice items (Higham & Gerrard, 2005). If students do not understand the content being tested, changing answers is unlikely to increase their test scores.

Previous studies have also examined the relationship between response changes and test performance by a number of variables, such as gender (Al-Hamly & Coombe, 2005; Geiger, 1991a; Milia, 2007), test takers' ability or proficiency levels (Al-Hamly & Coombe, 2005; Benjamin et al., 1984; Mathews, 1929; Milia, 2007; Vispoel, 1998, 2000), testwiseness (Geiger, 1996), examinees' cognitive style (Friedman & Cook, 1995), item difficulty (Al-Hamly & Coombe, 2005; Green, 1981; Ramsey, Ramsey, & Barnes, 1987), item position (Johnston, 1978; Reille & Briggs, 1952), and item type (Al-Hamly & Coombe, 2005; Geiger, 1991b). The investigation also expands across a number of disciplines, including business (Geiger, 1996), English language (Al-Hamly & Coombe, 2005), psychology (Kruger et al. 2005; Mathews, 1929), law (Milia, 2007), and nursing (Gaskins, Dunn, Forte, Wood, & Riley, 1996; Jordan & Johnson, 1990; Nieswiadomy et al., 2001).

A recent study by van der Linden, Jeon, and Ferrara (2011), however, reported a contradictory (from previous studies) finding that the benefit of answer changing was *negative* for all ranges of ability groups, with lower ability groups showing the largest losses. The authors later discovered an error in their data and withdrew their publication. Bridgeman (2012) conducted a reanalysis using the correct data and reported conclusions consistent with prior research that a majority (76%) of the students received higher scores after revising their answers, only a small portion of students (6%) received lower scores after the answer changes, and the rest did not experience any score change.

Response Changes on Computer-Adaptive Tests

Although response changes on paper-and-pencil tests have been studied for more than 80 years and consistent conclusions have been reached regarding the benefits of

response changes, the research on response changes on computer-adaptive tests (CAT) has not been sufficient. This is because many adaptive tests do not allow item review and answer switching. As Vispoel (2000) describes, allowing examinees to change responses “can complicate item administration algorithms” (pp.328-329).

Limited studies have been conducted on response changing on conventional CATs with item-level branching decisions. These yield similar results to those on paper-and-pencil tests. Namely, the majority of the test takers change answers; only a small percentage of answers are typically changed; and those who change answers usually gain points on test scores (Ferrara et al., 1996; Lunz, Bergstrom, & Wright, 1992; Stone & Lunz, 1994; Vispoel, 1998; Vispoel, Wang, de la Torre, Bleiler, & Dings, 1992). For instance, Vispoel (1998) found that on a standardized vocabulary CAT, more than 60% of the test takers changed answers, about 4.4% of item answers were changed, and the item gain-to-loss ratio was 2.15. The effect of response changes also interacted with test takers’ ability levels. High-performing students were less likely to make changes but were more likely to achieve score gains when they did make changes.

For those CATs that allowed test review and response changes, studies found that examinee ability estimates before and after answer review were highly correlated (e.g., .98) and that pass/fail decisions remained the same for more than 95% of the examinees (Lunz et al., 1992; Stone & Lunz, 1994). The same authors also reported that student average ability estimates increased slightly after the test review and answer changes (Lunz et al., 1992; Stone & Lunz, 1994). The test information value and the relative test efficiency indicated by the difference in standard error of measurement were about the same before and after students’ answer switching. Another interesting finding from Stone and Lunz was that after the test review and response changes, examinees that passed the test originally had increased confidence that they would pass the test. Lunz et al. (1992) concluded that although test takers’ ability estimates only slightly improved after the review and response changing, just the ability to make changes to their answers may give test takers the feeling that they have more control over the testing procedure, and therefore, have greater confidence in testing.

Other findings on CAT response changing concluded that allowing students to review significantly prolonged the testing time (Ferrera et al., 1996; Vispoel, 1998), and most test takers believe that it is important to have the item review option in a test (Vispoel, 1998; Vispoel, Rocklin, & Wang, 1994). Although much of the advice suggesting that students should stick with their original answers is targeted at high-stakes admissions tests, research on the possible benefits of changing answers on such examinations has been missing. This study addresses that gap.

Research Questions

This study addressed the following two research questions:

1. What is the general pattern and outcomes of response changes on the Verbal Reasoning and Quantitative Reasoning measures of the GRE revised General Test (rGRE), by subgroup characteristics?

2. What is the relationship between response changes and test takers' ability as indicated by their scores on the Verbal Reasoning and Quantitative Reasoning measures of the rGRE?

Method and Design

Instrument

Items from the Quantitative and Verbal Reasoning sections of the rGRE were used (Educational Testing Service, 2012). For both sections, the test follows a multistage design for which examinees first respond to a routing section containing 20 items and then are directed to one of three second-stage sections with 20 items based on their performance in the routing section.

Data

Six test forms each were selected for Quantitative and Verbal sections, balancing ability level (e.g., 33rd, 66th, and 99th percentile) and geographical location of the testing centers (e.g., North America, Asia). Each form has 80 items (20 in the routing panel and 20 in each of the three second-stage panels) which results in a total of 480 items for each of the Quantitative and Verbal sections. Final participants included 8,538 examinees for the Quantitative section and 9,140 examinees for the Verbal section, who took the rGRE in testing centers in 12 countries (Table 1).

Analyses

For the nonblank first responses, our analyses consisted of four components:

- (a) At the item level, the mean percentage of examinees changing their response to an item, the mean percentages of score gain, loss, or not changing for each item, and the mean item gain-to-loss ratio (i.e., percentage of examinees gaining scores divided by the percentage of examinees losing score)
- (b) At the person level, the percentage of examinees who changed at least one response, the percentages of examinees who gained, lost, or had unchanged scores as a result of response change, and the mean examinee gain-to-loss ratio
- (c) Response change patterns and outcomes examinee ability levels, including mean number of changed answers for each examinee, outcomes on score gain, loss, or not changing, and gain-to-loss ratio. Examinees were categorized into three groups of low, medium, or high ability based on their total Verbal or Quantitative Reasoning score
- (d) Response change behaviors and outcomes by gender, ethnicity, and test center country (i.e., North America vs. Asia)

Table 1. Participant Description.

	Quantitative			Verbal		
	<i>n</i>	%	Mean score ^a	<i>SD</i>	<i>n</i>	Mean score ^b
Gender						
Male	3,775	44.2	162	7.2	3,953	149
Female	4,297	50.3	157	9.1	4,621	150
Not answered	466	5.5	152	8.5	566	152
Ethnicity						
White (Non- Hispanic)	2,013	23.6	152	7.3	2,191	154
Black/African American	164	1.9	145	7.0	187	148
Mexican, Mexican American, Chicano	49	0.6	149	6.5	59	150
Asian, Asian American, or Pacific Islander	158	1.9	153	8.7	187	152
Other Hispanic or Latin American	110	1.3	149	8.2	130	151
Puerto Rican	18	0.2	145	8.1	23	151
American Indian, Alaskan Native, Other Native American Group	97	1.1	151	7.0	100	155
Not answered	5,929	69.4	162	6.8	6,263	148
Test center						
North America	3,729	43.7	152	8.2	4,076	153
Asia	4,802	56.2	164	4.7	5,057	147
Other	7	0.1	159	6.1	7	151
Total	8,538		159	8.8	9,140	150

^aMean Quantitative measure score. ^bMean Verbal measure scores.

For the blank first responses, we analyzed the percentages of blank to right, blank to wrong, and blank to blank responses by Quantitative and Verbal sections. We also analyzed the changing and outcome patterns from the examinee side (e.g., percentage of examinees making a change, percentage of examinees making score gains).

Missing data were treated with pairwise deletion. Missing at random was assumed as the examinees were expected to finish most of the questions.

Results

Test Takers' Overall Response Change Patterns and Outcomes

The top of Table 2 presents results at the item level, averaged across items, while the bottom of Table 2 focuses on effects on the total test scores of examinees. On average about 15% of test takers made changes on any given individual Quantitative item and 36% on any individual Verbal item (Table 2, top). The item gain-to-loss ratio was 2.8 for Quantitative and 2.2 for Verbal. Most test takers made at least one change on the Quantitative (95.8%) or Verbal (98.4%) tests (Table 2, bottom). Most test takers gained scores from switching answers on the multiple-choice items (82.8% on Quantitative and 67.8% on Verbal). The examinee gain-to-loss ratio was 23.0 for Quantitative and 4.5 for Verbal.

Figure 1 provides a snapshot of examinees' gain and loss from changing answers. The 83% of examinees who gained scores on the Quantitative section had an average gain of 4.7 points, and the 68% who gained on the Verbal section had an average of 3.0 points. Four percent of the examinees who lost scores on the Quantitative section lost an average of 1.4 points, and the 15% of those who lost scores on the Verbal section lost an average of 2.1 points. Approximately, 9% and 16% of the test takers did not experience score change as a result of switching answers on the Quantitative and Verbal sections, respectively.

We also compared the percentages of examinees that experience score gain or loss on a particular item (Table 3). About 85% of the Quantitative and Verbal items showed more examinees experiencing score gains compared with score losses when changing responses.¹ For the Quantitative section, items for which examinees experienced more loss than gain appeared to be significantly more difficult than items that had equal gain and loss ($p < .001$). Similar findings apply to the Verbal section.

Response Change by Ability Groups

Examinees were categorized into three groups on the basis of their Quantitative or Verbal scores (Tables 4 and 5). As examinees' ability increased, the benefit of response changing increased in that both the item and examinee gain-to-loss ratio increased. Also, higher performing examinees tended to make fewer changes than lower performing examinees on the Verbal but not on the Quantitative section.

Table 2. Response Change Patterns and Outcomes for All Test Takers.

	Quantitative reasoning (%)	Verbal reasoning (%)
Item change (after-before)		
Mean percentage of test takers changing response on an item		
W-R	14.9	35.6
R-W	6.6	12.2
W-W	2.4	5.6
R-R	3.7	15.7
Item gain-to-loss ratio	2.2	2.1
	2.8	2.2
Examinees who changed at least one response		
Mean number of changes	8,178	8,992
Mean percentage	5.2	5.1
Examinees whose scores increased	95.8	98.4
Mean number of changes		
Mean percentage	5.4	5.6
Examinees whose scores decreased	82.8	67.8
Mean number of changes		
Mean percentage	8.2	6.6
Examinees whose score did not change	3.6	15.0
Mean number of changes		
Mean percentage	3.6	2.5
Examinees' gain-to-loss ratio	9.4	15.6
	23.0	4.5

Note. W-R = wrong to right changes; R-W = right to wrong changes; W-W = wrong to wrong changes; R-R = right to right changes (test takers chose the right answer initially, made interim changes, and chose the right answer again); gain-to-loss ratio = percentage of examinees with score gains/percentage of examinees with score losses.

Figure 1. Percentages of examinees with higher, lower, or unchanged scores as result of response changing. Number in brackets is the mean number of points gained or lost.

Table 3. Percentage of Score Gains and Losses on Quantitative and Verbal Reasoning Items.

	Quantitative		Verbal	
	Frequency	%	Frequency	%
Gain < loss	20 ^a	4.5	30	7.1
Gain = loss	47	10.7	34	7.9
Gain > loss	374	84.8	364	85.0
Total	441 ^b	100.0	428	100.0

^aOn these 20 items, the percentage of test takers who made score gains as a result of response change was smaller than the percentage of those who experienced score declines as a result of response change. Similar meaning applies to other categories in this table. ^bThere was a total number 480 items in the 6 Quant forms with duplicate items among them. The unique number of items was 441. Same note applies to the 428 unique number of items for the Verbal section.

Response Change by Gender, Ethnicity, and Test Center

The response change patterns and outcomes are similar between males and females (Table 6). Among ethnic groups,² on the Quantitative test, the Puerto Rican group showed the highest percentage of test takers who experience score gain (94%, average score gain of 4.6 points). Black/American American examinees experienced the least amount of score gain in terms of percentage (74%, average score gain of 3.43 points). It is noteworthy that despite the variation in the gaining percentages and

Table 4. Response Change Patterns and Outcomes by Test Taker Ability for Quantitative Reasoning.

	Low ability (n = 918)		Medium ability (n = 1,372)		High ability (n = 6,248)	
	Mean	%	Mean	%	Mean	%
Score change (after – before)	1.3		1.5		1.6	
Number of changed answers	5.0	12.5	4.9	12.3	5.6	14.0
W-R	1.4	3.5	1.9	4.9	2.7	6.7
R-W	0.7	1.8	0.8	2.0	0.9	2.4
W-W	2.7	6.7	1.9	4.6	1.0	2.6
R-R	0.2	0.5	0.3	0.8	1.0	2.5
Item gain-to-loss ratio	2.0		2.4		3.0	
	n	%	n	%	n	%
Examinees with at least one changed answer	882	96.1	1,320	96.2	5,976	95.6
Examinees with score gains	730	79.5	1,113	81.1	5,226	83.6
Examinees with score unchanged	113	12.3	151	11.0	540	8.6
Examinees with score losses	39	4.2	56	4.1	210	3.4
Examinees gain-to-loss ratio	18.7		19.9		24.9	

Note. W-R = wrong to right changes; R-W = right to wrong changes; W-W = wrong to wrong changes; R-R = right to right changes. The ability ranges for the three groups are 130 to 148, 149 to 159, and 160 to 170, respectively.

Table 5. Response Change Patterns and Outcomes by Test Taker Ability for Verbal Reasoning.

	Low ability (<i>n</i> = 4,150)		Medium ability (<i>n</i> = 3,540)		High ability (<i>n</i> = 1,450)	
	Mean	%	Mean	%	Mean	%
Score change (after – before)	1.1		2.1		2.8	
Number of changed answers	5.6	14.0	5.1	12.8	3.7	9.3
W-R	1.5	3.8	2.0	4.9	1.3	3.2
R-W	0.9	2.3	0.8	2.1	0.3	0.8
W-W	2.9	7.2	2.0	5.0	0.8	2.0
R-R	0.3	0.6	0.3	0.8	1.3	3.2
Item gain-to-loss ratio	1.7		2.5		4.3	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Examinees with at least one changed answer	4,087	98.5	3,470	98.0	1,435	99.0
Examinees with score gains	2,427	58.5	2,556	72.2	1,211	83.5
Examinees with score unchanged	815	19.6	466	13.2	142	9.8
Examinees with score losses	845	20.4	448	12.7	82	5.7
Examinees gain-to-loss ratio	2.9		5.7		14.8	

Note. W-R = wrong to right changes; R-W = right to wrong changes; W-W = wrong to wrong changes; R-R = right to right changes. The ability ranges for the three groups are 130 to 149, 150 to 156, and 157 to 170, respectively.

Table 6. Responses Change Outcomes by Gender, Ethnicity, and Test Center Region.

	Percentage of items changed				Examinees gaining scores		Examinees losing scores	
	Total	W-R	R-W	W-W	R-R	%	Points	%
<i>Quantitative</i>								
Gender								
Male	15	7	2	3	3	82	4.84	4
Female	15	6	2	4	2	84	4.64	3
<i>Ethnicity</i>								
White (non-Hispanic)	13	6	2	5	1	81	3.98	4
Black/African American	12	4	2	6	1	74	3.45	5
Mexican, Mexican American, Chicano	12	5	2	5	—	84	3.63	8
Asian, Asian American, or Pacific Islander	14	6	2	5	1	84	4.36	5
Other Hispanic or Latin American	12	5	1	6	1	77	3.67	3
Puerto Rican	14	6	2	6	—	94	4.63	—
American Indian, Alaskan Native, other Native American group	14	6	2	6	1	87	4.08	3
<i>Test center</i>								
North America	12	5	1	5	1	81	3.90	4
Asia	17	8	3	3	3	84	5.29	3
Other	7	2	1	3	1	43	2.67	—
<i>Verbal</i>								
Gender								
Male	38	13	6	17	2	67	2.99	16
Female	34	12	5	15	2	69	3.03	14

(continued)

Table 6. (continued)

	Percentage of items changed					Examinees gaining scores		Examinees losing scores	
	Total	W-R	R-W	W-W	R-R	%	Points	%	Points
Ethnicity									
White (non-Hispanic)	26	11	3	10	2	75	2.97	9	-1.47
Black/African American	25	9	4	11	1	65	2.42	13	-1.52
Mexican, Mexican American, Chicano	28	9	4	13	1	66	2.59	12	-1.71
Asian, Asian American, or Pacific Islander	30	12	4	12	2	71	3.21	12	-1.18
Other Hispanic or Latin American	25	9	4	10	1	61	3.01	12	-1.40
Puerto Rican	27	9	4	11	3	61	3.14	22	-1.80
American Indian, Alaskan Native, other Native American group	28	12	4	10	2	77	3.32	12	-1.25
Test center									
North America	26	10	4	10	2	73	2.94	11	-1.45
Asia	44	14	7	20	2	64	3.04	19	-2.45
Other	30	13	2	14	1	86	3.20	—	—

score points, all racial/ethnic groups had more examinees showing score gain from response changes than those showing score loss, which also applies to the Verbal test.

Test takers from the North America and Asia test centers showed similar patterns and results of response changes.

Answer Changing of First Blank Responses

In this study, we differentiated between first blank response and first nonblank response. The average rate of first blank responses was about 12% on a Quantitative item and 15% on a Verbal item (Table 7). Most of the blank responses were later filled in with an answer. The rate of final blank answers was only 1.4% for the Quantitative and 0.3% for the Verbal section. Since a blank answer was scored the same as a wrong answer, there were only two possible outcomes for a response change with a first blank answer: score gain and score unchanged. The ratio of score gain to score unchanged was 1.2 for the Quantitative section and was 0.8 for the Verbal section (see top of Table 7). In terms of examinees, about 53% of examinees made score gains from changing a first blank response on the Quantitative section, but only 27% of the examinees gained scores on the Verbal section. The rest of the examinees had unchanged scores as they either made a blank-to-wrong change or left it blank in the end.

Discussion

Findings from this study support the 80-year literature on response change. Specifically (a) a majority ($> 96\%$) of test takers changed responses when taking the rGRE; (b) only a small number of responses were changed on average by each examinee (e.g., about 5 on Quantitative and Verbal sections, respectively); (c) at the item level, more answers were changed from wrong to right than from right to wrong (item gain-to-loss ratio was 2.8 for Quantitative and 2.2 for Verbal); and (d) most examinees made score gains by changing answers (examinee gain-to-loss ratio was 23.0 for Quantitative and 4.5 for Verbal). About 83% of the test takers experienced score gains on the Quantitative test and 68% on Verbal, and the average score gain was about 4.7 points for the Quantitative test and 3 points for Verbal on a 130-to-170 score scale. The standard deviation of this scale is about 9 for Quantitative and 8 for Verbal. Therefore, the effect size of score gain as a result of response change was .52 and .34 for Quantitative and Verbal, respectively. We consider these effect sizes to be substantially important given the high stakes nature of the rGRE.

We also examined response change patterns and outcomes by ability, gender, ethnicity, and test center groups. The results are consistent with the above findings from analyzing the entire group of test takers. Although in general all ability groups benefit from response changing, the higher performing examinees tend to benefit more

Table 7. Response Change Patterns for First Blank Responses.

Item change (after – before)	Quantitative reasoning (%)	Verbal reasoning (%)
Mean percentage of test takers having first blank responses		
B-R	11.9	14.9
B-W	6.4	6.7
B-B	4.1	7.9
	1.4	0.3
Ratio of correct to wrong and blank changes	1.2	0.8
Examinees who had at least one first blank response	7,777	7,804
Mean number of changes	4.6	7.0
Mean percentage	91.1	85.2
Examinees whose scores increased		
Mean number of changes	4.6	6.2
Mean percentage	52.5	27.2
Examinees whose score did not change		
Mean number of changes (B-W)	5.4	8.0
Mean percent (B-W)	26.1	47.3
Mean number of changes (B-B)	2.9	4.4
Mean percentage (B-B)	12.5	10.7
Ratio of correct to wrong and blank changes	1.4	0.5

Note. B-R = blank to right changes; B-W = blank to wrong changes; B-B = blank to blank changes.

than their lower performance peers, as indicated by the significantly higher gain-to-loss ratios.

The majority of test takers (53%) made score gains on changing a first blank response on the Quantitative section, but a relatively smaller percentage of test takers made score gains on the Verbal section (27%). Two interesting findings were noted from the above numbers: (a) compared with the gain score rate on nonblank responses (83% for Quantitative and 68% for Verbal; Table 2), such numbers for blank responses are substantially lower; and (b) the gain score rate was much lower for the Verbal section than for the Quantitative section. For (a), we speculate two factors as plausible explanations. One is that the fact that examinees left the answer blank suggested that the items appeared to be difficult to them. To confirm that speculation, we compared the average item difficulty of the items with more than 20% first blank responses with that of other items, and found that the items with higher rates of first blank responses were significantly more difficult than other items ($p < .01$) for both the Quantitative and Verbal sections. The other reason is that as they reach the end of the testing sections, examinees are more likely to leave items blank and then fill in an answer. We examined the last three items in both the first and second panels for both the Quantitative and Verbal sections, and found that the rate of first blank response was 19% on Quantitative and 20% on Verbal sections. Given the high stakes of the rGRE, if examinees were rushing to fill in the blank answers, it is likely that their performances were negatively affected.

The differences in the nature of underlying abilities assessed by the Quantitative and Verbal sections may explain the differential rates of gain scores on these two sections, not just for the first blank responses, but for the nonblank responses as well. It may be more direct for examinees to identify the right answer in revision on a Quantitative item than for them to evaluate among options for a Verbal item. On a Quantitative item, if through more careful recalculation a different answer occurs, it is likely the new answer could be the correct one. However, on a Verbal item, the change may not be as straightforward.

Implications

Findings from this study support response changing when there is a good reason for doing so. Also, test takers should try to avoid leaving items blank as there is no penalty of trying to fill in the answers: They either get a score gain for providing a right answer or their score remains unchanged for providing a wrong answer. The results of this study disprove the fallacy that the first instinct is always correct. As rGRE scores are associated with high-stakes admissions decisions, we urge test takers to be cautious when changing answers, but not to be afraid to change a response that, on review, they believe to be incorrect. Response change should not be driven by instinct, but rather by evaluation of the rationality of a switch. Despite the overwhelming benefits of response changing, the results are aggregated among all test takers and do not apply to every individual test taker. A test taker making changes to

rGRE responses may not necessarily gain score points from the changes, but gain is still a much more likely outcome than loss.

This study attests to the benefits of allowing test takers to change responses on computer-based tests. For measurement and technical considerations, many adaptive tests do not allow response review and change. A few studies (Vispoel, 1998, 2000; Vispoel et al., 1994) reported the advantages of providing opportunities for review and change on adaptive tests, and an important advantage of a multistage adaptive test design is that answer changing within a section is much more easily accommodated than on a CAT that branches at the individual item level. This study provided evidence that allowing examinees to change responses contributed to their improved performance on a high-stakes admissions test.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Note that among the total number of 480 items, some items are the same across second-stage panels, and therefore the unique number of items is less than 480.
2. The classification of race/ethnicity groups only applies to U.S. domestic test takers as the background question on ethnicity is only asked among U.S. test takers.

References

- Al-Hamly, M., & Coombe, C. (2005). To change or not to change: Investigating the value of MCQ answer changing for Gulf Arab students. *Language Testing*, 22, 509-531.
- Beck, M. D. (1978). The effect of item response changes on scores on an elementary reading achievement test. *Journal of Educational Research*, 71, 153-156.
- Benjamin, L., Cavell, T., & Shallenberger, W. (1984). Staying with initial answers on objective tests: Is it a myth? *Teaching of Psychology*, 11, 133-141.
- Bridgeman, B. (2012). A simple answer to a simple question on answer changing. *Journal of Educational Measurement*, 49, 467-468.
- Brownstein, S. C., Wolf, I. K., & Green, S. W. (2000). *Barron's how to prepare for the GRE: Graduate Record Examination*. Hauppauge, NY: Barron's Education Series.
- Educational Testing Service. (2012). *2012-2013 GRE® registration and information bulletin*. Princeton, NJ: Author.
- Ferrara, S., Frances, A., Gilmartin, D., Knott, T., Michaels, H., Pollack, J., & . . . Wise, S. (1996, April 6). *A qualitative study of the information examinees consider during item*

- review on a computer-adaptive test. Paper presented at the annual meeting of the National Council on Measurement in Education, New York City, NY.
- Freder, G. (2004). *Learning to learn: Strengthening study skills and brain power* (Rev. ed.). Nashville, TN: Incentive.
- Friedman, S. J., & Cook, G. L. (1995). Is an examinee's cognitive style related to the impact of answer changing on multiple-choice tests? *Journal of Experimental Education*, 63, 199-214.
- Gaskins, S., Dunn, L., Forte, L., Wood, F., & Riley, P. (1996). Student perceptions of changing answers on multiple-choice questions. *Journal of Nursing Education*, 35, 88-90.
- Geiger, M. A. (1991a). Changing multiple-choice answers: Do students accurately perceive their performance? *Journal of Experimental Education*, 59, 250-257.
- Geiger, M. A. (1991b). Changing multiple-choice answers: A validation and extension. *College Student Journal*, 25, 181-186.
- Geiger, M. A. (1996). On the benefit of changing multiple-choice answers: Student perception and performance. *Education*, 117, 108-117.
- Green, K. (1981). Item-response changes on multiple-choice tests as a function of test anxiety. *Journal of Experimental Education*, 49, 225-228.
- Harvil, L. M., & Davis, G. (1997). Medical students' reasons for changing answers on multiple-choice tests. *Academic Medicine*, 72, S97-S99.
- Higham, P. A., & Gerrard, C. (2005). Not all errors are created equal: Metacognition and changing answers on multiple-choice tests. *Canadian Journal of Experimental Psychology*, 59, 28-34.
- Johnston, J. J. (1978). Answer-changing behavior and grades. *Teaching of Psychology*, 5, 44-45.
- Jordan, L., & Johnson, D. (1990). The relationship between changing answers and performance on multiple-choice nursing examinations. *Journal of Nursing Education*, 29, 337-340.
- Kruger, J., Wirtz, D., & Miller, D. T. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology*, 88, 725-735.
- Lunz, M. E., Bergstrom, B. A., & Wright, B. D. (1992). The effect of review on student ability and test efficiency for computerized adaptive tests. *Applied Psychological Measurement*, 16, 33-40.
- Mathews, C. O. (1929). Erroneous first impressions on objective tests. *Journal of Educational Measurement*, 20, 280-286.
- Milia, L. D. (2007). Benefiting from multiple-choice exams: The positive impact of answer switching. *Educational Psychology*, 27, 607-615.
- Nieswiadomy, R. N., Arnold, W. K., & Garza, C. (2001). Changing answers on multiple-choice examinations taken by baccalaureate nursing students. *Journal of Nursing Education*, 40, 142-144.
- Ramsey, P. H., Ramsey, P. P., & Barnes, M. J. (1987). Effects of student confidence and item difficulty on test score gains due to answer changing. *Teaching of Psychology*, 14, 206-210.
- Reille, P. J., & Briggs, L. J. (1952). Should students change their initial answers on objective-type tests? More evidence regarding an old problem. *Journal of Educational Psychology*, 43, 110-115.
- Stone, G. E., & Lunz, M. E. (1994). The effect of review on the psychometric characteristics of computerized adaptive tests. *Applied Measurement in Education*, 7, 211-222.
- Van der Linden, J. W., Jeon, M., & Ferrara, S. (2011). A paradox in the study of the benefits of test-item review. *Journal of Educational Measurement*, 48, 380-398.

- Vispoel, W. P. (1998). Reviewing and changing answers on computer-adaptive and self-adaptive vocabulary tests. *Journal of Educational Measurement*, 35, 329-346.
- Vispoel, W. P. (2000). Reviewing and changing answers on computerized fixed-item vocabulary tests. *Educational and Psychological Measurement*, 60, 371-384.
- Vispoel, W. P., Rocklin, T. R., & Wang, T. (1994). Individual differences and test administration procedures: A comparison of fixed-item, computerized-adaptive, and self-adapted testing. *Applied Measurement in Education*, 53, 53-79.
- Vispoel, W. P., Wang, T., de la Torre, R., Bleiler, T., & Dings, J. (1992, April 20-24). *How review options, administration mode, and anxiety influence scores on computerized vocabulary tests*. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, CA. Retrieved from ERIC database. (No. TM018547)