International Journal of Advanced Educational Research
ISSN: 2455-6157
Impact Factor: RJIF 5.12
www.educationjournal.org
Volume 3; Issue 1; January 2018; Page No. 120-123

Improving SAT reading scores by mitigating the confirmation bias

[1] John Leddo, [2] Bhadra Kadangal, [3] Shyla Bisht, [4] Neha Rajeev
[1] Director of Research, My Ed Master, LLC, Virginia, USA
[2, 3, 4] Research Assistant, My Ed Master, LLC, Virginia, USA

Abstract
Standardized testing is pervasive in our education system and is often associated with high stakes as schools use test results in their admissions and graduation/certification decisions. One of the most common standardized test formats is multiple choice, where test takers are presented with a question and multiple answer choices and then asked to select "the best answer." This type of format and instructions could lead test takers to succumb to a form of the confirmation bias (Wason, 1960) where test takers look for evidence to support a preferred answer while ignoring evidence that may suggest the answer is incorrect. Because test makers often present multiple plausible answer choices, i.e., are at least part true, test takers can be induced to focusing on evidence that supports incorrect answers, thereby lowering their test scores. The present study investigated whether combatting the confirmation bias by instructing test takers to look for evidence that rejected answer choices were wrong as well as evidence supporting selected answers would lead to improved test scores. The testbed chosen was SAT reading because it is a universally known test used in college admissions. 30 students in Northern Virginia participated in a study where they were given two SAT reading passages containing 26 questions or one-half of the number of questions on a reading section of the SAT. Ten students were assigned to each of three conditions: a control group that received the standard instructions given to SAT students, a group that was given the standard instructions and asked to write down evidence to support their selected answer choice to each question, and a group that was given the standard instructions and asked to write down evidence to support their selected answer choice and evidence for why their rejected answer choices were false. Results showed that students receiving standard test instructions scored the same as those asked to provide reasons for their selected answer while those who were also asked to provided evidence why they rejected non-selected answer choices scored significantly higher than the other two groups. Results suggest that even without prior training or preparation, test takers can overcome the confirmation bias and dramatically increase SAT reading scores. Future research could investigate whether this applies to other types of reading.

Keywords: confirmation bias, SAT, standardized, whether

Introduction
Standardized testing has become a staple in public and private education. Each year, countless students are given high stake assessments that affect whether they graduate school, get accepted into advanced programs, and gain admission to their desired colleges. Perhaps the most common format used in these tests is multiple choice, where test takers are presented with a question and several answer choices, one of which is designated as the "correct answer." Students are then instructed to select "the best answer" to each question. Other answer choices vary in their plausibility in order to create variability in problem difficulty. This is done to create a distribution in performance across students such that most students can answer the easiest questions (those whose "incorrect" answer choices are highly implausible) while only a few can answer the most difficult questions (those whose "incorrect" answer choices seem plausible).

The coupling of multiple answer choices with the instruction to select "the best answer" choice as the single correct answer creates conditions where test takers may be vulnerable to the confirmation bias (Wason, 1960) [7] where people seek out and interpret information that is consistent with a preferred conclusion rather than also examining evidence that their preferred conclusions could be wrong (Koriat, Lichtenstein, & Fischhoff, 1980; Nickerson, 1998) [5, 6]. In cases where multiple plausible answer choices are given, the confirmation bias could lead to the selection of wrong answer choices as test takers may find themselves gravitating toward a plausible but incorrect answer choice and then focus on evidence that supports only that answer.

In spite of the fact that the confirmation bias could adversely affect the test scores of millions of test takers annually, with potentially negative consequences to their futures, there has been scant attention paid to whether standardized test scores can be affected by the confirmation bias. Although Garrison and Hoskisson (1989) [2] argued nearly 30 years ago that confirmation biases could affect predictive reading, there has been little research done on the extent to which confirmation bias could affect test scores and ways to combat its potential negative effects. More recently, Hernandez and Preston (2013) [13] found that confirmation bias effects could be mitigated through disfluency (presenting readers with degraded text). However, this technique is not likely to find its way into standardized testing as degraded text would undoubtedly initiate a wave of backlash from test takers who are unhappy with their scores, claiming that the degraded text was

confusing and difficult to read.

The present study investigates whether confirmation bias appears to affect standardized test scores in reading assessments and whether this bias can be mitigated. SAT reading is chosen as the testbed. The SAT test is universal and taken by millions of high schoolers annually. It is a high stakes assessment in that colleges around the world use SAT test results in their admissions decisions. Moreover, the recent format change of the SAT reading section may very well induce greater risk of confirmation bias as there is a new type of reading question that asks students to indicate what lines in the passage provide support for the preceding question. This type of question asks students to support a single answer choice rather than provide evidence that other answers may be wrong.

Accordingly, the present study employs three in conditions. In the control condition, test takers are asked to read SAT passages and then answer the associated questions. In the confirmation bias inducing condition, test takers are asked to read the passages and then write down what evidence supports their selected answer choices. In the confirmation bias mitigating condition, test takers are asked to read the passages and then write down what evidence supports their selected answer choices and why the rejected the answer choices they did not select. If confirmation bias is normally present in SAT reading, we would expect that test takers in the control condition would perform the same as those in the confirmation bias inducing condition. If the confirmation bias can be mitigated by requiring test takers to consider disconfirming evidence for rejected answer choices, then test takers in the confirmation bias mitigating condition should score significantly higher than those in the other two conditions.

Methods

Participants

Participants were 30 high school students recruited from high schools in Fairfax and Loudoun counties in Virginia. High school students were chosen as those are the ones who take the SAT test for college. Because the SAT is a high stakes test, it was assumed that the participants would also be motivated to perform well on the materials they were provided.

Materials

The core materials used were two SAT reading passages. One was a fiction passage and the other was a science passage. These were chosen as they represent the range of reading passage types from fiction to non-fiction and would also serve as a test to see whether the reading techniques being tested depended upon the type of reading material used. Each passage had 13 associated questions, meaning that the 26 total questions represented half of a full SAT reading section. Each question had four possible answer choices, only one of which was deemed the "correct" answer.

The two core SAT passages constituted the reading passages for the control condition. There were two variations on the core passages that were used for the experimental conditions. One variation of the materials had space, after all the answer choices were listed, for participants to write what information from the passage they used to support the answer choice they selected as the correct answer. This version of the materials also had the instructions "After answering each question, write below the question why your answer is correct and what evidence is shown in the passage to prove it right."

The second variation of the materials had space after each answer choice in which participants could write what information from the passage they used to justify why they selected the answer as the correct one or rejected it as incorrect. This version of the materials also had the instructions "Choose the correct answer choice which applies to the passage. Write below each answer choice why it is/isn't the correct answer and use evidence within the given space."

Procedure

Participants were tested individually and randomly assigned to one of the three experimental conditions. They were given a copy of the materials that were associated with their respective conditions. The experimenter verbally reviewed the instructions for the condition with the participant. Participants then read the passages and answered the questions, including, as appropriate, giving their reasons for the answers they selected and the ones they turned down. To allow for the time taken to write their reasons, participants had no time limit, even though there normally would be one on that actual SAT test.

Results

Before the any analysis was done on whether participants correctly answered the questions associated with each passage, the answer sheets for participants in both the correct answer and correct and incorrect answers conditions were reviewed to determine whether the participants complied with the experimental instructions to provide justifications for why selected answers were chosen in the former condition and selected answers were chosen and non-selected answers were rejected in the latter condition. All participants had complied with the experimental instructions and, therefore, their responses were analyzed to see how many correct answers were chosen in each of the two passages (fiction and science).

The first analysis done was to see whether the mean number of questions answered correctly differed as a function of type of passage, i.e., fiction vs. science. For all three conditions, the mean number of questions answered correctly for each type of passage was statistically identical. Therefore, the total number of correct answers given for each participant was combined across passages to form an overall number of questions answered. The mean total number of questions answered per participant was calculated based on condition. These means are shown in Table 1.

A one-way analysis of variance was performed on the data and the results of this test were statistically significant, $F(2, 27) = 9.65$, $p < .001$. To determine which means were statistically significantly different from each other, a Tukey Honest Significant Difference (HSD) test was conducted on the means. Results of this analysis revealed that the control and confirmation bias inducing condition means were not statistically significantly different from each other, $Q = .25$, NS. However, the confirmation bias mitigating condition mean was statistically different from both the control condition mean, $Q = 5.25$, $p < .01$ and the confirmation bias

inducing condition mean, Q = 5.5, p <. 01. This suggests that merely having students identify what information supports the answer choice they select for a question does not enhance performance, but asking them both to identify information that supports an answer choice and to identify information that disqualifies the other answer choices does enhance performance.

Table 1: Mean Number of Questions Answered Correctly Based on Condition

Control Condition	**13.5***
Confirmation Bias Inducing Condition	13.3*
Confirmation Bias Mitigating Condition	17.7

Numbers with an asterisk are not significantly different from each other.

Discussion

The results suggest that having students explicitly cite reasons why rejected answer choices are incorrect as well as citing reasons why chosen answer choices are correct significantly increases reading performance on the SAT reading section compared to having them cite reasons only why selected answer choices are correct or doing neither. Explicitly citing evidence for why a chosen answer is correct does not improve performance compared to not explicitly citing evidence. The fact that students who did not explicitly cite any evidence for or against answer choices performed virtually identically to those who cited evidence only for chosen answers and statistically significantly lower than those who cited evidence for both choosing and rejecting answers suggests that students who were citing no evidence were very likely looking, as is consistent with typical standardized test instructions, for evidence to support the "best answer" and were not focusing on evidence for eliminating rejected answers. Further research could explore whether students are, in fact, only looking for evidence to confirm their selected answer choices.

Given that the students completed the equivalent of one half of an SAT reading section (26 questions vs. 52 questions for a full reading section), it is possible to extrapolate how the increase in number of questions answered correctly would affect the scaled reading score. The overall SAT reading score is generated from the reading section score (Section 1 of the SAT) and the writing section score (Section 2 of the SAT). The total number of correct answers in each section is used to create a scaled subscore, which ranges from 10 to 40. The reading and writing subscores are added and multiplied by 10 to get an overall reading score. Extrapolating the data from the present experiment, one would project that a student who explicitly gives reasons why a selected answer is correct and rejected answers wrong would get approximately 35-36 out of 52 questions correct. This produces a reading subscore of about 30-31, which if duplicated in the writing section, suggests that a student would score at about the 610 level in the reading. This corresponds to about the 81st percentile according to College Board, the makers of the test (College Board, 2017) [1]. On the other hand, students who explicitly provide evidence only for why chosen answers are correct or do neither would be projected to correctly answer between 26 and 27 questions. This produces a reading subscore of about 25-26, which if duplicated in the writing section, suggests that a student would score at about the 510 level in the reading. This corresponds to about the 51st percentile, or about average, according to College Board.

Even though the overall reading score provided on the SAT results is a composite of both reading and writing, the results still suggest that that improvements in the reading alone would still boost the overall reading score by about 50 points, as indicated by an increase in reading subscore of five points that would be multiplied by 10 to get the overall score. Even so, for a single technique to boost a reading score by 30 percentile points from average to well above average is remarkable. We suspect that there are few, if any, such techniques that could have such a powerful effect without any training or preparation on the part of the student.

It should be noted that the process of writing evidence to support both chosen and rejected answer choices is not practical for a timed test such as the SAT (although many standardized tests such as state reading assessments are not timed). Therefore, further research is needed to determine whether students could effectively employ the technique of looking for disconfirming evidence without actually writing the evidence down. Our experience working with students suggests that this would require some training, not because the process is difficult, but rather the process requires a certain mental discipline to sustain. Moreover, it is probably the case that searching for disconfirming evidence is not required for every question. SAT questions typically range in difficulty from easy to medium to hard, so there may be several questions within the test that are relatively easy to answer. We imagine that the procedure for searching for disconfirming evidence will be most useful for difficult questions where more than one answer seems potentially correct.

Conclusion

As noted in the Introduction, people all over the world take high stakes assessments, many of which include reading. Often instructions to those tests are worded to induce a confirmation bias, which could potentially lower test takers' scores. The present research suggests a simple, but powerful technique for overcoming the confirmation bias and one that leads to significantly improved reading scores with no training or preparation on the part of the test taker. However, the process of looking for disconfirming evidence is not one that students generally appear to use naturally. Given that there is an abundance of training materials and programs available for students preparing for standardized tests, the present findings suggest that those materials and programs could be enhanced by including practice in overcoming the confirmation bias.

The present research raises other important questions and may have implications for how education should proceed. First, the present research used SAT reading as a testbed. The SAT is generally taken by high school students. It is worth investigating whether the confirmation bias affects other types of reading and readers of different ages. For example, the multiple choice format of standardized testing lends itself to confirmation biases as test takers search for correct answers. However, in standard classroom instruction, students are taught to be "active readers" where they make predictions about what will happen in the texts they are reading and then search for evidence that their predictions are true. In Loudoun

and Fairfax counties in Virginia, where we live, high school students are taught to write "claim, evidence and commentary essays" where they make assertions about text they are reading and then cite evidence from the text that supports their assertions. Arguably, both active reading techniques and claim, evidence, commentary essays induce students to succumb to the confirmation bias as both direct students to look for supporting, rather than disconfirming, evidence.

Second, the goal of education is to prepare students for succeeding in their adult lives. Instruction teaches the knowledge and skills that students need and assessment measures the degree to which students have learned what they have been taught. Part of this preparation for adulthood involves teaching reasoning, judgment and decision making skills. The psychological literature is replete with studies that demonstrate that people exhibit a wide range of reasoning and decision making biases (cf., Kahneman, Slovic and Tversky, 1982) [4] including showing that professional decision making is also subject to the same biases (Dawes, 1982) [4]. Although such biases are well documented, formal training in overcoming such biases is generally lacking in mainstream education nor does standardized testing focus on measuring them.

Beyond the educational implications, the present study ties to a recent and perhaps growing societal phenomenon. Sophisticated computing technology now makes it possible to target content to individual users' interests. Some of this is self-selection as people may seek out websites or publications that feature content that match their points of view. However, technology also permits content providers to tailor news items and other content to the interests and political leanings of their audience. This phenomenon has the potential to "societalize" the confirmation bias by creating a cycle whereby people seek out and are presented which information that confirms existing points of view and does not challenge those views with disconfirming evidence. In the US, we have seen the consequences of this as political parties have become more polarized in recent years. An important question is whether training in overcoming the confirmation bias can overcome this polarization.

References

1. College Board. SAT: Understanding Scores, 2017.
2. Garrison J, Hoskisson K. Confirmation Bias in Predictive Reading. The Reading Teacher. 1989; 42(7):482-486.
3. Hernandez I, Preston JL. Disfluency Disrupts the Confirmation Bias. Journal of Experimental Social Psychology. 2013; 49:178-182.
4. Kahneman D, Slovic P, Tversky A. Judgment under uncertainty: Heuristics and Biases. Cambridge University Press, New York, 1982.
5. Koriat A, Lichtenstein S, Fischhoff B. Reasons for confidence. Journal of Experimental Psychology: Human Learning & Memory. 1980; 6:107-118.
6. Nickerson R. Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology. 1998; 2:175-220.
7. Wason PC. On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology. 1960; 12:129-140.