Running head: CHOICE EFFECT

Exploring the Choice Effect on Standardized Test Performance

Adrienne Sands

Washington University in St. Louis

Abstract

This study extends research on the framing and choice effects to determine whether framing a test as obligatory can impact performance on standardized testing. A within-subjects design controlled for academic ability. Participants completed verbal reasoning tests faster if researchers framed the test as an obligation rather than as a choice. These frames had no significant impact on accuracy rates. Proposed explanations for these results are discussed herein. These findings suggest that factors beyond ability impact test performance. To ensure that standardized tests provide a valid and reliable measure academic ability, replications of and extensions to this study are necessary.

*Keywords:* standardized testing*,* choice, choice effect, choice frame, framing effect

Exploring the Choice Effect on Test Performance

With the abundance of complaints against the education system, few top those aimed at standardized tests in terms of the number of complaints and diversity of protestors. Students don’t want to take them; parents don’t believe they reflect their child’s scholastic ability; professors (or, teaching assistants) don’t want to grade them; and administrators don’t want to be judged for low scores. But beneath these criticisms lies the belief that tests have some redeeming qualities: they provide a standardized measure of students’ aptitude and information necessary to develop better teaching methods. For these reasons (and despite our best attempts) standardized tests probably aren’t going anywhere.

To justify the use of standardized testing, it’s necessary to ensure that these qualities are maintained. In particular, it is necessary that tests are constructed to be a valid and reliable measure of student ability. For example, if we were able to determine that students of equal ability perform better on a test written in Times New Roman font we must always (or never) use this font; furthermore, if we are able to demonstrate that Times New Roman yields scores that accurately reflect academic ability, no other font should be used. Rather than analyzing the impact of font on performance, this study extends the research by Crookes (2007) on the impact of choice on assessments. In an effort to shape teaching methods, undergraduate students at the University of Leicester argued that they should be able to choose an essay topic which reflects their current level of understanding, thereby maximizing potential performance. Using a within-subjects design, Crookes sought to determine whether increased motivation and perception of ability (vis-à-vis topic selection) actually would impact performance. She found evidence of a “choice effect”—where choice of assessment topic improved assessment performance—in two out of the three cohorts that she studied.

But, essays and the assessment of performance in written assignments are very subjective; thus, it is rather difficult to apply Crookes’ results to standardized testing. Our study examines the impact of a “choice frame” on an objective multiple choice test. The classic study by Tverski & Kahneman (1981) demonstrated that the frame of an outcome as a gain or loss can drastically change decision making. It might be argued that multiple choice exams are a collection of decisions—which answers to choose and how fast to choose them—with correct answers as potential gains; thus (and as Crookes suggested), it is possible that the format of these exams has an impact on overall performance.

Consider the possible effects of a choice frame. Perceived freedom in a task elicits positive emotions connected to flexibility and control. It can also encourage a heightened sense of responsibility for the outcome by forcing the actor to ask why he/she would have chosen the task if they didn’t expect to perform well. But, choice isn’t always a good thing. Perceived freedom may decrease the sense of urgency needed to complete an important task. Thus, this study seeks to determine whether choice enhances or erodes performance on standardized tests. Given that mandatory tasks communicate a sense of urgency, we predict that participants will be faster on tests with an obligation frame; however, we also predict that the heightened sense of responsibility for the outcome of a chosen task will encourage participants to be more accurate under the choice frame.

For the use of standardized tests to continue, it is necessary to understand all factors that influence test performance. Many tests already allow students the freedom to choose which questions to complete; this study intends to justify that practice and move us one step closer to the ideal standardized test.

**Method**

**Participants**

Participants included twelve volunteers between the ages of 18 and 30 (nine female, three male). They were recruited individually and did not receive compensation for their efforts.

**Stimuli**

Stimuli consisted of two instruction pages and two tests, each described below. Instructions were modified for each condition. Tests were designed to be as similar as possible in terms of difficulty, content, and expected completion time.

*Instruction Pages*

For the obligation condition, instructions emphasized test procedures:

You have 10 minutes to complete this test. Mark your answers below. Raise your hand when you are finished.

For the choice condition, instructions emphasized the choice of the test to enhance beliefs of the participant’s control:

Please choose one of the following tests to complete. Raise your hand when you have chosen to receive your test. Once you’ve received your test, try to complete it in less than 10 minutes. Mark your answers on this page. Raise your hand when you are finished.

On the same page, participants were given three descriptions of “alternate” tests. Descriptions were designed to seem distinct but equally difficult:

Circle the number of the test you would like to take and show it to the proctor.

1. This test will assess your ability to use English correctly
2. This psychometric test will assess your reasoning skills and your ability to understand word roots in a context
3. This test will assess linear reasoning and language functions (such as grammar and vocabulary) as lateralized by the right brain

*Test Material*

Two tests were constructed from the verbal reasoning sections of the 2011 Kaplan New Graduate Record Examination (GRE) practice book. Each test was designed to take approximately ten minutes and featured seven questions of three types: text completion, sentence equivalence, and reading comprehension. Samples of each question type are included below:

*Question Type*: TextCompletion *Number included*: 3

*Instructions*: For each blank select an answer choice from the corresponding column of choices that best completes the text.

*Sample Question*: Because the decision-making process was entirely \_\_\_\_\_\_\_\_\_\_, there was no way to predict its outcome. The process was \_\_\_\_\_\_\_\_\_\_ rolling dice, where there are a finite number of possibilities but no way to accurately predict which two numbers will come up.

Blank I:Blank II:

1. Arbitraryd. Likened to
2. Regimented e. Belittled by
3. Unilateral f. Dissimilar to

*Answer*: AD *Kaplan’s Expected Completion Time*: 1 Minute

Blank I:Blank II:

1. Arbitraryd. Likened to
2. Regimented e. Belittled by
3. Unilateral f. Dissimilar to

*Answer*: AD Kaplan’s Expected Completion Time: 1 Minute

*Question* *Type*: Sentence Equivalence *Number included*: 2

*Instructions*: Select the two answer choices that fit the meaning of the sentence and create two complete sentences that are similar in meaning.

*Sample Question*: W.C. Handy’s self-conferred sobriquet, “The Father of the Blues” is widely \_\_\_\_\_\_\_\_\_\_\_; although he has composed and published the first written blues song, other musicians had been playing the blues for years.

* 1. Professed d. Proven
  2. Deconstructed e. Contested
  3. Disputedf. Demonstrated

*Answer*: CE *Kaplan’s Expected Completion Time*: 1-1.5 Minutes

*Question* *Type*: Reading Comprehension *Number included*: 2

*Instructions*: The next two questions are based on the passage below.

*Sample Question:* Many Iranian Americans, whether they are immigrants or American-born, identify themselves as being of Persian heritage. This descriptor is a frequent cause for confusion among non-Persians who know the country as Iran and understand Persia as an antiquated name for the empire that encompassed part of Iran as well as parts of modern-day Pakistan and Afghanistan. Opponents of the term argue that because some Afghani and Pakistani groups refer to themselves as being of Persian heritage, the term loses meaning as a signifier of nationality. However, critics argue that just as the English language recognizes “Spain” rather than “Espana,” English speakers should refer to the country as Persia, and not as Iran, which is the Persian translation of the country’s name.

The author is primarily concerned with

1. Arguing that English usage of descriptors of nationality should reflect usage within the native languages of the countries in question.
2. Clarifying how the fall of the Persian Empire has influenced the terminology that modern citizens of Iran use to define their nationality.
3. Distinguishing among three groups that use the same term to describe their national identities.
4. Explaining two opposing positions in an argument about the use of a descriptor of national identity**.**
5. Persuading readers that in order for the term Persian to have a clear relationship to nationality, only Iranians, not Afghanis or Pakistanis, should use the term.

*Answer*: D *Kaplan’s Expected Completion Time:* 2-4 Minutes

The author is primarily concerned with

1. Arguing that English usage of descriptors of nationality should reflect usage within the native languages of the countries in question.
2. Clarifying how the fall of the Persian Empire has influenced the terminology that modern citizens of Iran use to define their nationality.
3. Distinguishing among three groups that use the same term to describe their national identities.
4. Explaining two opposing positions in an argument about the use of a descriptor of national identity**.**
5. Persuading readers that in order for the term Persian to have a clear relationship to nationality, only Iranians, not Afghanis or Pakistanis, should use the term.

*Answer*: D *Kaplan’s Expected Completion Time:* 2-4 Minutes

Participants received one point for each correct answer choice and lost one point for each incorrect answer choice.

**Procedure**

To begin, all participants were given an instruction page. Those assigned to the obligation condition were also given a test at this time. After reading the instructions, participants assigned to the choice condition revealed their choice to the experimenter and received their test. Each test was limited to ten minutes, and the experimenter gave a one-minute warning. After each test, each participant noted his/ her own time. When the last participant finished (or ten minutes had passed), participants were given a five minute break, and the experimenter distributed the second set of instruction pages. The second set of testing followed the same procedure as the first. Participants were debriefed at the end of the experiment and tests were subsequently graded.

**Design**

This study featured a within-subjects design. Initial conditions were randomly assigned to participants and orders counterbalanced to control for the practice effect. Test frame served as the independent variable with two levels: obligation and choice. The dependent variables were accuracy and completion time; they served as objective measures of overall test performance.

**Results**

Completion times are shown in Figure 1. On average, it took participants 5.23 minutes to complete tests framed as an obligation versus 5.48 minutes for tests framed as a choice; however, these times are not significantly different, *t*(11) = -0.31, n.s. Consider accuracy rates as listed in Table 1. On average, participants gave 73.08% correct responses under the obligation condition compared to 66.92% under the choice condition. These rates are not significantly different, *t*(11) =1.79, n.s. Overall, participants performed equally well (as accurately and as fast) under both conditions.  
 But as can be seen from the figure, there are two extreme values: participants 1 and 7 were particularly slow to complete their first tests. Excluding these results, the obligation condition yielded an average completion time of 4.46 minutes and an accuracy rate of 0.73; the choice condition yielded an average completion time of 5.75 minutes and an accuracy rate of 0.69. Excluding these participants, there was a main effect of test frame on completion time, *t*(9) = -3.528, *p* < 0.01. On average, participants were 1.29 minutes faster to complete tests framed as an obligation compared to tests framed as a choice. Despite the presence of a main effect on completion time, there was no main effect of test frame on accuracy, *t*(9) = 1.10, n.s. Even after data correction, participants were equally accurate under both frames.

**Discussion**

In this study, we found that participants were faster to complete tests if they believed they had no choice of which test to complete. To understand why this occurred, consider when frame could have impacted completion time. For example, participants with a choice might have been slow to start because they were distracted by thoughts of their motivations for making a certain choice or how the test matched their expectations and abilities. Throughout the test, these participants might continue to dwell on the consequences of their choice; however, if choice encourages responsibility for the outcome of that choice, perhaps these participants were giving more thoughtful (and time consuming) consideration of each question. Or perhaps, the more “responsible” participant finished the test just as quickly but kept it longer to review their answers. To determine how participants were slowed under the choice condition, it is necessary to determine when.

Perhaps choice doesn’t make people slower. It might also be the case that the obligation frame made participants faster. If an obligation endows a task with a sense of urgency, participants might speed through it as fast as possible or use heuristics over sustained thought. Or perhaps people are averse to situations where they sense a lack of control, and the need to avoid the negative emotions associated with this task would be enough to encourage a participant to rush. To determine whether obligation makes people faster or choice makes them slower, further study is necessary; regardless, this study suggests that ability alone can’t explain differences in completion time.

We might explain the absence of a main effect on accuracy to the type of task. Participants were given nine answer choices to questions with objective right and wrong answers; while the frame may have encouraged different strategies for test completion, the participants always had motivations to provide the right answers. For obligatory tasks, they were motivated by need to impress the enforcing authority; for tasks with a choice, the enhanced responsibility for the outcome would motivate a more careful, defendable (still accurate) approach. In both cases, the goal is always to get the right answer, but the time needed to do so is very different.

We might also attribute the absence of a main effect on accuracy to the numerous limitations of this study. With a larger sample size and a more convincing frame of obligation, one would gain clearer results and greater assurance of the experiment’s validity. Replications of this experiment might also include a neutral condition so researchers identify whether choice slows people down or obligation speeds them up. A more thorough accounting of time (i.e. total completion time, time until the first response, average time spent on each question, and time from last response until the participant raises his/ her hand) would clarify where the frame effect occurs.

This study extends Crookes’ results and Tverky’s frame effect to the realm of standardized testing. It gives further evidence that there’s more to test performance than a student’s knowledge or ability. As this study shows, it is not safe to assume that slower students have lower ability; perhaps perceive freedom and control over the tests they take. To justify the continued use of standardized testing, we must apply these results to develop tests that provide a sufficiently reliable and valid measure of overall performance. While an infinite number of factors might influence test performance, a more careful analysis of those most relevant to current test practices (i.e. test length, test frequency, content order, even font) is necessary.

# References

Crookes, A. (2007). The effects of topic choice on performance outcomes: an analysis of student selected third year essays. *Psychology Learning and Teaching* *, 6* (2), 85-90.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science* *, 211* (4481), 453-458.

Table 1

Accuracy rates for all participants by condition

|  |  |  |
| --- | --- | --- |
| Condition  Participant | Obligation | Choice |
| 1 | 0.77 | 0.58 |
| 2 | 0.67 | 0.77 |
| 3 | 0.75 | 0.54 |
| 4 | 0.58 | 0.69 |
| 5 | 0.92 | 0.92 |
| 6 | 0.92 | 0.83 |
| **7** | 0.67 | 0.54 |
| 8 | 0.62 | 0.58 |
| 9 | 0.77 | 0.67 |
| 10 | 0.50 | 0.62 |
| 11 | 0.85 | 0.67 |
| 12 | 0.75 | 0.62 |
| Average | 0.73 | 0.68 |

*Note.* Grey boxes indicate the condition of the subject’s first test.

Figure Captions

Figure 1*.* Completion times for each participant by condition.

Figure 1

Completion times for each participant by condition