AUDIO ACCOMMODATIONS IN TESTING: EXPERIMENTAL DIFFICULTIES, PROBLEMS WITH READERS, AND A SUGGESTED ROLE FOR AUDITORY DISPLAYS

Michael A. Nees, Lauren Berry & Charlotte Phillips
Lafayette College, Department of Psychology
Oechsle Hall, Easton, PA, USA
neesm@lafayette.edu; berryl@lafayette.edu; phillipsc@lafayette.edu

ABSTRACT

An experiment examined performance on a practice version of a high stakes standardized test under conditions that simulated both normal and accommodated (oral reader with extended time) testing conditions. Results did not show the anticipated negative effects of auditory test administration on scores, perhaps as a result of floor effects that we attributed to the demographics of the sample and the testing conditions. Difficulties with experimental approaches to the study of testing accommodations are discussed. Problems encountered during auditory administrations of the test are explored, and a potential role for auditory displays in accommodated standardized testing is suggested.

1. INTRODUCTION

Standardized, high-stakes tests attempt to measure an individual's aptitude for success in an important future situation such as school or work. People with disabilities, however, often cannot take tests under the usual (often pencil-and-paper) testing procedures, thus accommodations such as extended time, the use of assistive technologies, and oral readers are permitted as exceptions to standardized testing procedures. Oral readers—human administrators of tests that read questions and answers aloud to the test-taker—are a common accommodation for test-takers with visual or learning disabilities. This approach assumes that auditory and visual renderings of the test are equally valid for measuring the underlying construct (i.e., aptitude) being assessed. Concerns remain that accommodations such as oral readers may compromise the validity of tests [1]. For example, a U.S. test-taker recently claimed that the Law School Admissions Test (LSAT) violated the Americans with Disabilities Act, because he could not answer questions that required spatial reasoning and diagramming of visual concepts [2]. The test-taker claimed that his ability to practice law was not accurately represented by his performance on the LSAT.

We sampled sighted undergraduates under conditions that simulated how a person with a visual impairment would take the LSAT. The control was a pencil-and-paper administration. We also manipulated the availability of blank paper for spontaneous diagramming during the test. We predicted that auditory testing conditions would result in worse performance on the test due to increased demands on memory associated with the auditory processing of complex material, and we expected that a lack of paper for diagramming problems would also negatively affect scores.

2. METHOD

2.1. Participants

Participants (N = 44; 35 females; mean age = 19.43 years, *SD* = 2.94) were recruited from undergraduate psychology classes at Lafayette College and were compensated with course extra credit or $5 for every 30 minutes of participation. The sample included n = 5 first year, n = 18 second year, n = 9 third year, and n = 12 fourth year students. No students had previously taken an LSAT test and only one had previously taken an LSAT practice test. Students (n = 43) reported a mean grade point average (GPA) of 3.26 (*SD* = .37) on a 4-point scale.

2.2. Materials

Test questions came from LSAT practice tests in [3]. Participants answered 12 questions from each section of the test (Logical Reasoning, Analytical Reasoning, and Reading Comprehension). Each participant experienced approximately half of the scored portion of a 75-question LSAT exam.

2.3. Procedure

Participants were randomly assigned to conditions of the 2 (modality condition: auditory or visual) x 2 (paper condition: paper or no paper) design. In the auditory condition, the experimenter read aloud questions and answers and recorded participants' verbal responses. Experimenters adhered to the rules for LSAT readers [4] and did not define words, paraphrase, or otherwise deviate from the original text. In accordance with common practice for testing accommodations, participants were given 26 minutes (150% time in comparison visual testing condition) to complete each section. In the visual testing condition, participants were given 17 minutes to complete each of the three sections of the multiple-choice standardized test. The visual condition

simulated normal testing conditions; students read the questions and recorded answers themselves in writing. In the paper condition, participants were told that it may be useful to draw diagrams for portions of the test and were given a spare sheet of paper. In the no-paper condition, participants were asked to not make stray marks in the testing booklet and did not receive any spare paper to draw diagrams.

3. RESULTS

The number of correct responses was analyzed with a series of two-way between subjects ANOVAs—one for each of the three sections of the test (see Figure 1). For the Logical Reasoning section, there was no significant effect of testing modality, $F(1, 40) = 1.01$, $p = .32$., or paper condition, $F(1, 40) = 1.58$, $p = .22$, and the interaction was not significant, $F(1, 40) = 0.25$, $p = .62$. For the Analytical Reasoning section, the main effects of testing modality, $F(1, 40) = 0.36$, $p = .55$, and paper condition were not significant, $F(1, 40) = 2.74$, $p = .11$, but there was a significant interaction of testing modality with paper condition, $F(1, 40) = 5.09$, $p = .03$, $\eta^2_p = .11$. Simple effects analyses at each level of modality showed that for the auditory condition, it did not matter whether participants received paper, $p = .67$, but in the visual condition participants who did not receive paper ($M = 5.27$, $SD = 2.57$) had a significantly higher mean score than participants who did receive paper ($M = 2.91$, $SD = 1.22$), $p = .009$. For the Reading Comprehension section, the effects of testing modality, $F(1, 40) = 2.24$, $p = .14$, and paper condition, $F(1, 40) = 1.03$, $p = .32$, and the interaction, $F(1, 40) = 0.09$, $p = .77$, were not significant. Results suggested a floor effect. Each section of the test had 12 multiple-choice questions with five possible responses, so chance performance was equivalent to 2.4 correct responses per section. Fourteen of the 44 participants performed below chance on at least one of the three test sections.

4. DISCUSSION

None of the predicted effects of modality or paper condition were confirmed. In fact, the only significant effect of the experiment was the curious finding that participants in the visual condition who were allowed to use paper actually did worse on the Analytical Reasoning ("logic games") section of the test than participants who were allowed to use paper.

Taken at face value, our findings suggested that an oral reader accommodation with extended time allowed for comparable performance on LSAT tests. These results should be interpreted with caution, however, as the presence of a floor effect in scores on the dependent variable suggested that our sample was not sufficiently representative of the type of person who might actually take an LSAT. Our sampling procedure assumed that undergraduates at a selective academic institution would be similar enough to the population of LSAT test-takers to draw generalizations, but our sample's performance on the test questions across experimental conditions suggested that research in this area should sample test-takers who are more rigorously selected for their similarity to LSAT test-takers (e.g., with respect to LSAT preparation courses, advanced or completed undergraduate coursework, etc.).

Another difficulty that potentially influenced our results was the extent to which our laboratory scenario accurately representated the conditions of high stakes testing. Most high stakes texts assume that the test-taker puts forth their maximum effort. Because our participants did not have "high stakes" consequences attached the outcome of their test scores, it is possible that the observed floor effect in some conditions reflects less than maximal effort on the part of our participants. This illustrates another important difficulty with simulating high stakes testing scenarios in experimental laboratory research.

5. CONCLUSIONS

5.1. Difficulties with Experimental Approaches

Our results illustrate the difficulties inherent in experimental approaches to assessing the validity of testing accommodations. Typically, post hoc correlational approaches [see, e.g., 5] are used to try to assess the validity of accommodated standardized tests after the accommodation has already been administered widely. Experimental research

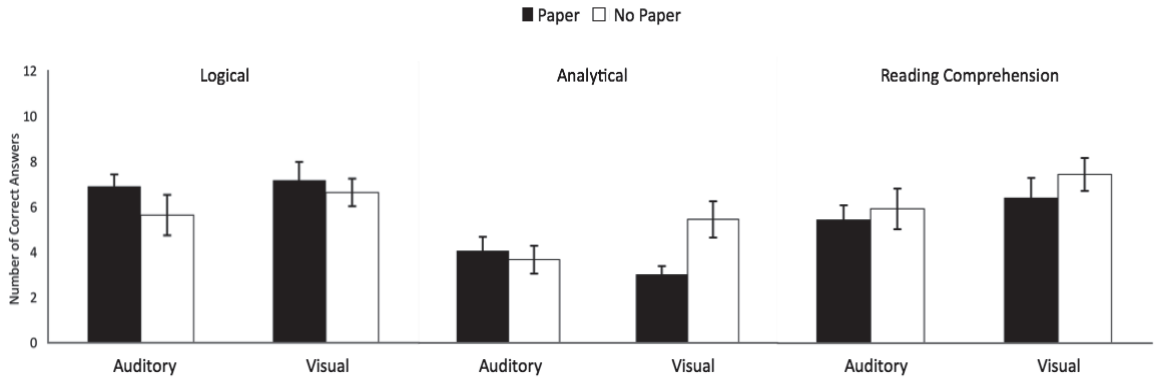

Figure 1: Mean number of correct answers on each section of the test for each condition of the experiment. Error bars represent standard error.

on testing accommodations may be able to offer a priori evidence for or against a particular accommodation before it is implemented, but a known difficulty with all research (both experimental and correlational) on testing accommodations is small sample sizes. The pool of potential disabled test-takers is relatively small, and the low availability of representative samples for research becomes even more problematic as more criteria (e.g., interest in law school and previous LSAT test preparation) constrain the pool of representative subjects. For better or worse, researchers in many domains of science use samples of convenience. In studies of assistive technologies and testing accommodations, this often involves initial testing and prototyping of potential solutions with sighted participants, though ultimately testing with target visually-impaired populations becomes imperative [see 6].

5.2. Problems Encountered During Auditory Testing

The second and third authors administered the auditory conditions of the experiment to participants and anecdotally noted several potential difficulties with reading test items aloud to participants that were not explicitly addressed in [4], which stated the following regarding test administration:

> Readers shall be instructed to read the test verbatim, and will not be permitted to paraphrase, interpret, modify or otherwise vary from the text, except as set forth below. The reader shall not define words; however, the reader may, at the request of the test taker, spell words and re-read all or part of the text or questions within the section that is being tested. The test taker is permitted to direct the reader to specific portions of the test to be read or re-read, including the order in which the text is to be read. The test taker is also permitted to scan for particular words or phrases as specifically identified by the applicant to be read or re-read verbatim, and to identify the type and location of punctuation marks contained in the text. (pp. 2)

In reading the test verbatim, we noted that transitions (e.g., from questions to answers or from one answer to the next) are not always apparent to the test-taker. For example, when the reader announced the first potential multiple choice answer as "A," the test-takers were not always clear that the "A" designated the first in a series of possible answers (i.e., A, B, C, D, and E) rather than the start of a new sentence. The print version of the test is formatted to clearly indicate answers with indentations. Reading the text of the test aloud verbatim, however, does not explicitly indicate that "A" designates the first of several possible answers. Readers attempted to pause strategically during transitions in the test, but this did not entirely alleviate test-takers' confusion.

Further, some potentially informative notation in the text of test questions was not apparent when the reader followed the the rules for the accommodation. Quotation marks appear in certain test questions and passages. The rules for the accommodation do not specify whether or not the reader *must* indicate the start and end of quotations marks; the rules only state the the test-taker can ask for punctuation marks to be identified. If the test-taker does not specifically ask for punctuation to be noted, the reader might simply read the verbatim passage. Similarly, the reading comprehension portion of the test featured line numbering [e.g., the notation "(5)" in the margin of the fifth line of text in the passage]. From the rules for the reader, it was unclear whether the line numbering should be announced. In our experiment, the readers did not announce punctuation or line numbering. If the reader was to adopt a test-reading stratetgy that included announcing punctuation and line numbering, the flow and comprehensibility of passages could be negatively impacted. More research is needed to determine the best practices for maintaining equivalence across auditory and visual test versions. If necessary, accommodated tests should preserve notations that carry important information in print (e.g., punctuation, line numbering, etc.) during auditory testing.

The rules state that the test-taker may prompt a reader to scan the text for particular words or phrases. The experimenters (readers) in our study noted that many key phrases appeared more than once in the text of questions or passages. This presented a dilemma to the reader in determining which occurrence of the key phrase should be re-read. The reader must also determine the order in which to re-read queried phrases that have multiple occurrences. This likely presents a considerable disadvantage to the accommodated test-taker as compared to the visual test-taker who can strategically scan passages visually for key words or phrases by herself without relying on a reader's proclivities.

Similarly, the readers in our study noted that participants taking the auditory version of the test did not tend to employ strategic approaches to the test that could help to maximize their scores. Typically, participants in the auditory condition did not skip test questions to be revisited later for more thought. Even with 50% additional time, the readers noted that test-takers did not have time or inclination to review answers or return to difficult questions for a second consideration of responses. Instead, participants tended to answer the questions in the order they were presented and record their first response as their final answer. Skimming and scanning the test were generally not possible, although at least one participant did ask the reader to begin the reading comprehension portion of the test by reading the questions aloud before longer passage on which the questions were based.

5.3. Suggested Role for Auditory Displays in Accommodated Testing

Several of the problems presented in the previous section leave open a role for human error on the part of the reader to influence the scores of the test-taker and to thereby compromise the psychometric reliability and validity of the test. Auditory displays that use either pre-recorded human speech or synthetic text-to-speech [7] could ensure the uniform presentation of all auditory material to all test-takers. Further, a test presentation system with accelerated speech capabilities could likely offer the test-taker a more efficient presentation of test questions and answers that would save time. A well-designed auditory display for presenting the testing materials could employ "scrubbing" controls commonly found in personal audio devices (e.g., mp3 players, etc.) that allow the test-taker to flexibly advance and rewind

the digital audio. Such a system might permit the accommodated test-taker more flexibility in scanning the the test materials and using alternate test strategies such as skipping questions and returning to them later. Nonspeech auditory displays and sonification may even prove useful for presenting auditory versions of punctuation and other non-word printed cues to the test-taker.

5.4. Future Work

More research is needed to establish the reliability and validity (and thereby the fairness) of auditory versions of high stakes standardized tests. The current study used practice LSAT questions as a case study, but the underlying issues examined here apply to most all accommodated versions of tests that read questions aloud to the test-taker, including both aptitude tests and classroom achievement tests. Future work should continue to seek to overcome the difficulties with experimental approaches to studying testing accommodations (such as representative sampling), because experimental research may provide valuable evidence about difficulties with a particular accommodation before the accommodation has been administered in high stakes testing situations. Certainly post hoc correlational analyses provide useful insights about test validity, but these approaches examine the test scores of people who have already taken the test (and experienced the high stakes consequences of their test scores) under the accommodation being studied. Studies of representative samples of test-takers under various conditions of accommodated testing may be able to help to establish the comparability of accommodated and regular testing formats before the accommodations are widely administered and potentially used in college or graduate school admissions decisions, hiring or licensure decisions, etc.

Because of the inherent fallibility of human readers, speech auditory displays may be able to play an important role in reliable auditory testing. Although more research is needed, auditory displays that use speech and nonspeech sounds may be able to offer accommodated test-takers more efficient presentation of materials than an oral reader and also more independence in negotiating the test materials. For all auditory accommodations, researchers should provide evidence to ensure that accommodated and standard versions of tests are equivalently reliable and measure the same underlying aptitude or achievement construct.

6. REFERENCES

[1] S. E. Phillips, “High-Stakes Testing Accommodations: Validity Versus Disabled Rights,” *Applied Measurement in Education*, vol. 7, no. 2, p. 93, Apr. 1994.

[2] P. Rosenbaum, “Blind man files discrimination suit over law school admission test,” 2011. [Online]. Available: http://www.cnn.com/2011/US/05/24/michigan.aba.blind.lawsuit/index.html. [Accessed: 17-Feb-2013].

[3] Law School Admissions Council, *10 New Actual Official LSAT Prep Tests*. Newtown, PA: Law School Admissions Council, 2011.

[4] Law School Admissions Council, “Procedures and Policies on the Use of Readers on the LSAT By Blind and Severely Visually Impaired Test Takers,” Aug-2011.

[5] W. W. Willingham, M. Ragosta, R. E. Bennett, H. Braun, D. A. Rock, and D. E. Powers, *Testing Handicapped People*. Needham Heights, MA: Allyn and Bacon, 1988.

[6] B. N. Walker and J. Lindsay, “Using virtual environments to prototype auditory navigation displays,” *Assistive Technology*, vol. 17, pp. 72–81, 2005.

[7] M. A. Nees, “Correlations and scatterplots: A comparison of auditory and visual modes of learning and testing,” in *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, GA, USA, 2012, pp. 195–198.