

Benefits of Audio Exams

Wallace Judd, PhD

Benefits of Audio Exams

Wallace Judd, PhD

Performance Test Development Series

03

Benefits of Audio Exams

Industry Background

Testing has likely been going on in human societies forever.

In hunting societies there were tests of prowess like killing a male lion with a spear.

In the Middle Ages guilds assured the village that a practitioner was capable of producing a quality product.

In more modern times multiple choice testing arrived in time to induct recruits into World War I, and hands-on lab testing of software skills allowed large-scale performance testing.

And one future iteration of testing is audio testing . . . essentially a computerized interview.

Audio Performance?

It may not be clear how audio testing is or can be performance testing. It seems likely that it would be used for testing the articulation of knowledge, not the performance of a skill.

Admittedly, audio testing (or audio interviews) are appropriately performance tests only in specific contexts. A few examples are:

- Can a crane operator recognize signals given by a signalperson?
- Can a realtor identify a home type?
- Can a car salesman articulate the essential features of a car model?
- Can a police officer properly Mirandize an apprehended suspect?

In these and other contexts, audio responses constitute a performance which can be evaluated comprehensively and objectively.

What Is It?

Audio testing in this context is a test given by the computer speaking to the candidate, and the candidate responding by speaking to the computer.

It's like an automated conversation. There may or may not be any Artificial Intelligence involved; the computer is not pretending to be a human being.

The spoken responses of the candidate are transcribed, scored and reported. Scoring can be correct (the system recognizes the spoken reply and scores it as correct), incorrect (system recognizes an incorrect response) or indeterminate (computer can't recognize the response).

In cases where the computer can't recognize the response, the response is flagged and a human listens to the actual recording and determines whether the response was correct or incorrect. If it was unin-

telligible, it's scored incorrect.

No Guessing

Because candidates are not presented with choices, they must generate the response internally. Consequently the guessing parameter of items is essentially eliminated.

Because there is no guessing, the reliability of the exam (Alpha) is substantially higher than a multiple-choice exam of the same length would be.

For example, in developing a transactional database a reasonable multiple-choice item might be:

Assume you're developing a transactional database for an investment bank. How frequently should your transactions be backed up?

In a multiple choice exam, the choices might be

- A. Every day
- B. Every hour
- C. Every minute
- D. Every transaction

Most candidates will select D because it is prompted.

However, for an audio exam, the candidate must generate the response spontaneously.

Response Latency

The time from item presentation until speech response is an indicator of how frequently the candidate uses the information being elicited by the item. Long latency either means that the candidate is spending time interpreting the item or puzzling out the answer.

Response latency is insignificant for candidates who give a wrong response because their answer does not indicate competence at any level.

Alternative Correct Responses Accepted

In cases where there are synonyms for the correct response, alternatives don't need to be made explicit as choices. If a candidate knows the correct answer in one form that form doesn't have to be listed. Also, correct responses of different complexity or vocabulary level can be scored correct as well.

Brevity

Because the candidate doesn't need to read through four choices to select one, the test is substantially less time consuming to administer than multiple choice. Also, the candidate doesn't need to balance alternatives or try to tease out distinctions among choices.

Reading Skill Minimized

Reading skill level doesn't confound with stimulus difficulty because the candidate doesn't have to read the stimulus. Vocabulary level does interact with stimulus complexity, but this interaction is appropriate to the content being assessed. Irrelevant or unnecessary complexity should be eliminated at the item authoring level.

Authentication by Voice

Security is always a consideration in testing. In the case of audible response, the candidate's unique voice signature can be captured at regis-

tration. If someone else responds to the items in the test, the difference in voice signature can be immediately recognized.

Simplicity of Administration

Tests with little graphic or text content can be administered on a cell phone. Tests with more elaborate text or graphics can be administered on a conventional laptop computer or iPad. Virtually all portable computers have built-in speakers and microphone. Desktop computers will need access to a speaker and microphone.

Issues

Classroom settings require headsets to block out sounds from other test takers and the classroom.

Future: Stress Level

Methods of evaluating a speaker's emotional state are being researched and used in a variety of contexts. Whether these emotional indicators are stable enough to give indicators as to the candidate's comfort with the test content is a question for future research.

Actual implementations are just blossoming. Amazon Web Services has developed a sophisticated audio test enhanced by AI. Authentic Testing has a simplified audio testing system called DoTell that is available. And other implementations are around the corner.

Strategies for development, optimization and analysis will no doubt be enhanced in the near future.



3349 Monroe Avenue, Suite 336
Rochester, N.Y. 14618-5513
www.AuthenticTesting.com
+1-703-777-7321