Benefits of Conversational Exams
Wallace Judd
3/14/2024

[A 'conversational' exam is one in which the device speaks a prompt to the candidate, and the candidate responds by speaking to the device.

We use 'conversational' because if we used 'audio tests', the title might indicate we were testing audio equipment.

Also, browser searches for 'audio tests' would confuse these tests for testing of decibel levels, noise, and frequency response.]

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 correct answer is, "Every transaction." , and 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.

No Distractor Authoring

Unless you want to score distractors differentially, there is no need to author distractors for each item. Anything other than a correct answer is incorrect.

Alternative Correct Responses Acceptance

When there are different ways to express a correct answer a number of answers can be accepted.   For example, if the question is, "How many planets are there?"
7 (they didn't count earth)
8 (they counted earth)
9 (they counted Neptune)

Alternative Responses Scored Differentially

Correct responses of different complexity or vocabulary level can each be awarded different positive scores.

7 (they didn't count earth)   4 points
8 (they counted earth)   5 points
9 (they counted Neptune)   6 points

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

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.

Security

Remote proctoring detects external prompting while the candidate is taking the test. During phone administration, the candidate can be asked to enable his camera and scan around the room. Testing with a laptop or with a desktop with a camera can also use the camera to detect a source of prompts. In addition, unusual response latency would be a cue that prompting may be going on.

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 computer, laptop or notebook. Virtually all portable computers have built-in speakers and microphones. Desktop computers will need access to a speaker and microphone.

Issues:

Classroom settings require microphones for audio input and headsets to block out sounds from other test takers in 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.