

Areas to Consider in Reviewing Written Test Questions for Possible Cultural Bias

Wiesen, J.P. (2000) *Content-Oriented Cultural Bias Review of Test Questions*
24th Annual IPMAAC Conference; June, 2000; Arlington, Virginia.

1. Readability of Test Questions

If reading is not a pervasive aspect of the job, keep the reading level of most of the test below that needed on the job; otherwise your measure of various (non-reading) abilities will be confounded with (i.e., will reflect) reading ability. For most tests, only the reading questions should be at the reading level needed on the job.

Use readability formulas carefully. They are measurement tools, not guides to writing. Questions may be difficult to read despite low readability measures if you use hard words (including short hard words, such as acute), passive tense, or complex sentences. It is best to use clear, plain English.

Manually check the readability indices found in major word processors to be sure they are working as intended. For example, one major word processor takes every period to indicate the end of a sentence. So if the questions are numbered and each question number is followed by a period, each question number is counted as a sentence. That makes for artificially low readability statistics.

Use vocabulary of the type likely to be encountered on the job. Avoid phraseology with cultural or ethnic loading (such as, fair-haired child).

2. Reading Lists

Often reading lists are a good idea because they allow people the opportunity to prepare for the test. To allow for sufficient time to prepare, a reading list should best be published long before the examination date. Six months lead time is the most I've recommended. All people should have an equal length of calendar time to study.

3. Unnecessarily Academic Test Content

When textbooks are used as source material for jobs without an academic requirement, there is a serious risk that they introduce problems akin to those discussed above concerning readability. Often textbooks are written for a college audience, and the reading level may be too high, and the content may be overly academic for the job in question. Some texts are much better and others much worse in this regard. This should be a primary consideration in choosing texts to appear on a reading list.

It may be better to ask situational questions which tap understanding of application of principles and other knowledge, rather than more academic questions such as those which tap knowledge of definitions and memorization. If nothing else, candidates tend to perceive situational questions as more job-related and fair: as less academic, and as enabling application of acquired knowledge together with common sense gained through job/life experience.

4. Everyday, Practical Reference Material

Although authors in the last generation tended to try to make their textbooks readable, the content and presentation found in many books is quite academic, with much jargon and picayune material. Sometimes it is possible to base test questions on local SOPs, ISO9000 documentation,

local rules, laws or guidelines. These types of source material usually are more focused on practical aspects of the job than are textbooks. The use of local rules, regulations, operating procedures, and guidelines is generally desirable to the extent that these are current and reflect what is or should be done on the job.

5. Test Content Which Unnecessarily Evokes Emotional Responses

Some questions may arouse anxiety or other emotional responses from certain candidates. This may cause unequal testing conditions for various cultural groups. For example, it would be best to avoid tables depicting average family income or incarceration rates by ethnic or religious group in a test section designed to measure ability to read tables.

6. Representation Which is Balanced or Neutral as to Gender and Ethnic Group

One approach to this would have a balance of boys' and girls' names mentioned, rather than just one gender. If some people mentioned in the questions are in supervisory roles, they should not all be one gender or ethnic group (likewise for criminal roles, leadership roles, etc.) One should strive for test content which puts all test-takers at ease.

7. Test Content Which is Not Job Related

Test content which is not required for job performance sometimes has a differential impact on various groups. For example, a manufacturing company may say it needs employees with good basic math skills, given their emphasis on dimensions and tolerances of parts. But a basic math test might cover both decimal and common fractions whereas the plant might only use decimal fractions. These are different math abilities and, for many people, adding and subtracting common fractions is more difficult than adding and subtracting decimal fractions.

8. Irrelevant Test Content

Sometimes questions which are intended to measure one area are also measuring another. Reading level (discussed above) may be the most common example of this, but there are others. For example, math items which ask about import duties or increases or decreases in stock prices may confuse people who do not travel internationally or do not trade in stocks. Verbal analogy questions which use relatively obscure words will confound the measurement of reasoning ability with work knowledge. (It is easy to fall into this trap, perhaps because many common words have more than one meaning, and many more precise words are relatively obscure.)

9. Test Content Which is Equally Familiar to All Groups

Most test areas can be measured with questions in a variety of sub-areas. Often this is not mentioned in the test outline. Sometimes one or more areas are less familiar to some groups of candidates. For example, mechanical aptitude questions which focus on the working or repair of cars may be less familiar to women, and questions which focus on the working or repair of sewing machines may be less familiar to men. However, there are questions about common household objects which would test mechanical aptitude and be equally familiar to men and women.

10. Embarrassment/Liability

Would you be embarrassed by the content of this question if it should appear in the newspaper? How might an expert for the plaintiffs critique and attack this question? Is there a better way to ask this question? Drawings or pictures should not be likely to offend (not stereotypical, ugly). The keyed answer should be clearly correct, and the wrong answers clearly wrong. Avoid questions

which appear faulty or not job-related, even if they are perfectly sound and valid. Avoid outraging candidates.

11. Negative and Tricky Questions

Avoid questions which may be difficult due to their mode of presentation rather than the subject matter. Avoid questions with double or triple negatives, or even one negative (where it can be avoided). Capitalize the word “not” if it is likely to be overlooked. Avoid questions asking trivial rather than important information. Use words with their usual meanings, rather than with their relatively rare meanings. Avoid very fine distinctions among distractors. Avoid overwhelming extraneous information in the stem. Use clear and unambiguous wording. Avoid asking about unusual and trivial applications of an important principle. Do not put a very difficult item in the middle of very easy items. (Examinees get used to answering based on easy distinctions and so do not think about possible subtle issues.) Give candidates (especially internal candidates) an avenue for commenting on or protesting test questions.

12. Confusing Test Material

Avoid test-taking instructions or directions which are overly complicated or which contain unnecessarily difficult content. The first few questions on the test should not be the most difficult. There should be no typos or other errors in the test instructions, in sample test questions, or in the test itself. All these may be particularly troubling to people with less familiarity with tests.

13. Sensitivities of Various Cultural and Ethnic Groups

Various cultural and ethnic groups may have different sensitivities. Often members of a given group will be uniquely sensitive to certain material which may be offensive or confusing for their group. (However, detecting culturally biased test material is not easy to do. Having minority reviewers is not a guarantee of success.)