AB-45 - Paper

Why Have Four-Option Multiple Choice Questions?

Grover E. Diehl
Robert Doucette

Navy Advancement Center
Pensacola, Florida

The Navy Advancement Center (NAC) insists that the multiple-choice items used in Navy Enlistment Advancement System (NEAS) examinations have four options. The correct answer to each question should have an ease of around 50 percent and the incorrect alternatives (off alts) should split the other 50% randomly. Among examination developers, the complaint is occasionally heard that the fourth option is hard to develop and is often not in play. Would it not be better to write better three-option items? The answer is "no" and this paper explains why.

Longer is Better

If two exams are identical in all respects except for length, the longer test will be more reliable.

It makes intuitive sense.

Let's say there are 105 pre-tested items, each with an ease of 50 percent and a discrimination index of .25. Five items are randomly assigned to exam A, the rest to exam B. Which would *you* rather take to demonstrate your proficiency on the subject? Obviously, sampling is "more likely" to be better with the longer test. On the short test, there is a huge penalty for not knowing an answer; much less so on the longer test. The longer test simply gives the respondent more "chances" to demonstrate knowledge. And, longer is longer, regardless whether the ratio is 5:100 or 135:150.

Here is another reason: The shorter, five-item test cannot fully differentiate among more than five individuals; there are only five intervals. The object of norm referenced testing is to separate each individual from another along a continuum. On short tests, candidates pile up on the same interval. The result is no discrimination among candidates at that interval. If nothing else, longer exams provide more intervals for differentiation.

And, better discrimination with more items happens to be the way the mathematics work.

Now, what does test-length have to do with four-option items? Guess! Or, rather, guessing. Other things being equal: an exam of 150 true/false (2 option) items has an effective length of 75 items, due to the effect of guessing. A 150-item exam of three option items has an effective length of 100 items. An exam of four-option

items has an effective length of 112.5 items. The number of questions that can be correctly "guessed" changes as the number of equally-plausible answer alternatives changes.

Score Inflation

Chance performance is a trigger for certain events and processes.

Since NEAS examinations are norm referenced, the actual location of a score along a continuum is not important so long as the overall distribution of scores is normal and the exam differentiates among individuals. In fact, the NAC's "compensative P-value" process is recognition of this fact and an attempt to bring the distribution of scores into line with model values. The promotion cutoff scores are independent of the score on the exam; the cutoff scores are billet (vacancy) driven. In a criterion referenced environment, of course, the effect of score inflation due to probability has to be incorporated into the cut score. For example, say you want students to demonstrate mastery of 50 percent of the material on a four-option multiple-choice test in order to be credited with having learned the minimum to pass. The pass score on the exam is not 50, however, but 63: 50 + (.25 x 50), because you have to add in the guessing factor on the portion of questions the student has to guess at.

While score inflation is not a factor on the promotion side, again it's relative, it is a factor when the individual scores at chance level or below. The exam raw-score pass cut is used to identify those candidates who will not be considered for advancement. (The candidates at or above the raw-score pass cut remain in the pool of candidates and compete for advancement.) The exam-pass cuts for each pay grade have been determined by the Bureau of Naval Personnel (BUPERS) advancement planners as follows: E4 - one chance raw score standard deviation above the chance raw score mean; E5 - two chance raw score standard deviations above the chance raw score mean; E6/7 - three chance raw score standard deviations above the chance raw score mean. That is:

- **For pay grade E-4: Raw score cut = (N / 4) + 1 x sqrt (Np (1 - p)) where N is the number of test items left in the exam and sqrt (Np (1 - p)) is the raw score chance distribution standard deviation.**
- **For pay grade E-5: Raw score cut = (N / 4) + 2 x sqrt (Np (1 - p))**
- **For pay grades E-6/7: Raw score cut = (N / 4) + 3 x sqrt (Np (1 - p))**

Let's look at an example. An ABE2 exam has 150 items. The raw score cut for this exam rate is calculated as follows: Raw score cut = (150 / 4) + 2 x sqrt (150 x .25 x .75) = 37.5 + 2 x 5.303 = 48.1 or 48. Any candidate scoring less than 48 on this ABE2 exam is dropped from consideration for advancement.

These are "confidence intervals" around the chance score estimate. Answer sheets reflecting better-than-chance performance are scored for credit; those scoring below chance are flagged and scrutinized. First, these sheets are scored using other

answer keys to determine whether the individual entered the wrong test code. The sheets may be visually inspected for clerical errors. Normally, no checking for answers 'out of sequence' is done for a candidate's answer sheet since answering the questions in order is considered part of the exam. However, all candidate answer sheets for an exam are checked for an "out of sequence" condition if the exam booklet has missing items, as sometimes happens since print on demand. Moving on, if it is determined that the individual deliberately failed or mistreated the exam, a stream of unpleasant personnel actions is initiated. If the score represents a true appraisal of the individual's knowledge, the local organizational processes that certified the candidate's fitness for promotion may be called into question. There is more, but the point is this: the whole process keys on 25 percent correct, chance performance. An exam with many three and two option items has a chance score above .25 and none of these great things protecting the integrity of the system and ensuring fairness to the candidates happen.

Cognitive Complexity

As the number of viable item alternatives increases, it becomes increasingly necessary for the candidate to know the answer to get the item correct.

True/false items are cognitively undemanding. It isn't necessary to know the answer to get the item correct, even above the chance level. All that is required is to know whether one of the alternatives is incorrect. NEAS exams are true/false when two off alts are obviously not plausible. A prime cause of this is noun/verb disagreement, which is why the NAC restricts the use of collective alts.

Three option items are better, unless one of the options is wrong on its face. The item then becomes a true/false question.

With four option items, even if one of the off alts is obviously wrong , the remainder still have the opportunity to present a viable three option item. At the least, it is less likely that two off alts will be wrong at first view.

All of this is, of course, standard test-taking strategy -- eliminate what is obviously wrong and guess at the rest. NEAS exams attempt to nullify this strategy by ensuring all four options are equally attractive by forcing the ease (P-value) of the correct alt to be 50% and each of the off alts to draw one-third of the rest. The NAC also balances the distribution of alts and examines dummy answer sheets to eliminate any inadvertent patterns.

Reality

Still, that fourth option is difficult. Would it not be better to just put that effort into better three option items?

No, it wouldn't. While the current exams may not have exactly .25 probability, they are a lot better than .33. Three options would guarantee that the probability could never be better than .33. Experience suggests that the probability of three option

items would probably be closer to .50, and it is likely to regress toward .50 faster than .25 regresses toward .33.

INTERNET DOCUMENT INFORMATION FORM

A . Report Title: Why Have Four-Option Multiple Choice Questions?

B. DATE Report Downloaded From the Internet 4/21/99

C. Report's Point of Contact: (Name, Organization, Address, Office Symbol, & Ph #): **Navy Education and Training Professional Development and Technology Center**
Navy Advancement Center Dept
Dr. Grover Diesel, (850) 452-1815
6490 Saufley Field Road
Pensacola, FL 32509-5237

D. Currently Applicable Classification Level: Unclassified

E. Distribution Statement A: Approved for Public Release

F. The foregoing information was compiled and provided by:
DTIC-OCA, Initials: VM_ **Preparation Date**: 4/22/99__

The foregoing information should exactly correspond to the Title, Report Number, and the Date on the accompanying report document. If there are mismatches, or other questions, contact the above OCA Representative for resolution.