

Administrative Issues for Computerized Tests and Surveys

Mary E. Lunz, Ph.D., DM
Measurement Research Associates, Inc.

Introduction

The flexibility of computerized testing makes it necessary for administrators to make a number of decisions about the test prior to administration. The following are some of the administrative issues to be considered if computerized tests or surveys are to be administered.

The first issue is to determine the basic principles of computerized test administration. *Fixed computer based testing* is like a written test or survey presented on the computer screen. It has all of the assets and liabilities of fixed form tests or surveys. *Computerized adaptive testing* uses the power of the computer to select the best test or survey for each person. The disadvantage is that people may not trust tests or surveys that are composed of different items. The advantage is that the persons can take shorter tests that are challenging for them, without compromising reliability.

Test or Survey Format

When the computer is used to administer the test or survey, it may be fixed or variable length.

The **fixed format** requires that all persons take the same set of preselected items.

The **adaptive fixed format** requires all persons to answer the same *number* of items. However, these items are targeted to the person's current ability estimate or some other preselected point.

The **variable length format** presents persons with the number of items necessary to determine whether or not they meet the criterion for success. This means that the person is presented with the items which are the best measures of his/her ability, attitude or feeling. Very few items are needed for persons who are clearly able or persons who are clearly unable. However, the person whose ability is close to the criterion take more items.

Stopping Rules

Prior to starting a computerized program the standards for performance must be defined on a Benchmark Scale. This standard becomes the pass point or point of acceptability for the test or survey. The stopping rule usually revolves around the pass point. The following are some stopping rules that have been used.

A **fixed number** of items may be the criteria for success. For example, when the person answers 100 items, the test or survey stops, regardless of the person's position on the scale. There may be 90% confidence in the accuracy of the pass or fail outcomes, or there may be more or less, depending upon the ability of the person.

A **confidence interval** stopping rule may be used for a computerized adaptive test. A specified level of confidence in the accuracy of the decision must be achieved before the test terminates. Thus, the Board determines that they want 90% confidence in pass or fail decisions about persons. The CAT program then is programmed to continue presenting items until this criteria are met. An upper limit on the number of items presented may also be a factor in this stopping rule.

These stopping rules may have serious consequences for the persons. Therefore, it is imperative that the Board understand the implications of the stopping rule that they adopt.

Starting Difficulty

Selecting the difficulty of the first few test items is very important. When using computerized adaptive test administration, there are three possibilities.

1. Target to the pass point. Since we have no idea of the person's ability at the beginning of the test, it may be useful to target to the criterion. After a given number of items, it is possible to begin targeting to the current estimated ability of the person.

2. Present items that will be easy for the person, targeted at 80% probability of a correct response. This is in line with the old theory of a few 'easy warm up' items for the persons. However, this will yield less information about the ability of the person, making it more difficult to determine the current estimated ability of the person early in the test.

3. Present difficult items. Difficult items may be a 'wake up call' for persons at the beginning of the test. However, this will yield less information about the ability of the person, making it more difficult to determine the current estimated ability of the person early in the administration.

With fixed computer based testing, the items may be administered in a designated difficulty order, random order, or in the same order for all candidates. When the items are preselected for the fixed form, the order of item administration is less important.

Targeted Level of Difficulty

Computerized adaptive testing is often targeted to the ability of each individual person.

The best information about person ability is gained when the test or survey is targeted to the person's current ability estimate. This means the person has a 50% probability of a correct response. It is possible to target the test to any level of difficulty such as 60%, 70% or 80% probability. This may be important for surveys to keep the respondents interested. However, as the test is made easier, it becomes more off target, and more items may be needed to make a confidence pass or fail decision about the person.

Fixed form tests or surveys are often selected to include items close to the criterion. This reduces the error of measurement at the point where most of the persons are measured. Targeting to the criterion will be more accurate and appropriate for some candidates than for others. This is the nature of a fixed test or survey.

Minimum Test or Survey Length

A properly targeted computerized adaptive test, can often determine a pass or fail decision about a person with less than 10 items. This may be psychometrically valid, but not always content valid. Often, examinations must cover content guidelines. Thus, it is necessary to present a certain number of items so that a representative sample is presented to each person. Consequently, minimum test or survey length may be set at 25, 50, 60, or 100 items, even though the particular items and their difficulty varies among persons. A success or acceptability decision can be made about the persons as soon as the minimum number of items have been presented, or more items can be presented to cover the content areas.

With fixed length tests or survey, all persons take the same number of items. This could be five or 100 items depending on the design of the test or survey.

Opportunity to Control the Test or Survey

Computerized adaptive testing offers the opportunity to allow the person to control the test to the level the Board is comfortable. It may be necessary to make the test longer, if the person has more control; however, this may be worth it in face validity. The following are options for person control of the test.

Review allows the person to go back and review and alter their responses after they have completed the test. This is more or less comparable to the way persons can take written tests.

Skip allows the person to decide not to respond to an item. Persons would likely choose to skip and items to which they do not know the answer. An item of comparable difficulty is then presented for them to answer. Persons cannot return to items they choose to skip.

Defer allows the person to defer answering an items until the end of the test. After they complete all of the test items, the deferred items are presented to them for response, and response is required at this point.

Fixed length paper and pencil tests and surveys, by their nature, allow persons to review, skip and defer answering items at their convenience. However, when fixed tests or surveys are presented on the computer, some of these conveniences may not be allowed. There is really little reason not to allow control of the test, since the items are fixed. It depends on the computer programming. Administrators must think very carefully about allowable behaviors.

Conclusion

Computer based administration is a powerful education tool which could benefit Boards and persons simultaneously. However, the administrative issues must be carefully considered before beginning the program. The implications of these decisions can be to the advantage or disadvantage of the persons. It is incumbent upon the organization to understand these implications before beginning the program.