

# 11

## Developing Tests

*Preliminary Considerations*

*Objective Test Items*

*Essay Test Items*

*Item Analysis*

---

A test is an instrument which measures a subject's knowledge, understanding, or skill in a given content area, and produces a ratio score reflecting that measure. If the focus of a study is "testing subjects" on some variable (Bible knowledge, comprehension of various translations, current events), an appropriate test must be found, or one must be developed. This chapter introduces you to principles of developing tests.

### Preliminary considerations

You may be able to use existing tests for your study. Let's say the nature of your study is to identify a relationship between "job satisfaction" and "interpersonal dynamics among staff members". You may be able to find an existing test which will measure "job satisfaction." Check the *Mental Measurements Yearbook*, or *Tests In Print*, or other such resources for published tests in your area of interest.

Tests can also be found in research articles being gathered for the Related Literature section of your proposal. Study the **validity** and **reliability** scores on the test, the population(s) the test was designed for, and the conditions of test administration. If these factors fit your study, you're in business! Describe these characteristics in the "Instrument" section of your proposal.

You may need, however, to develop your own test, since there are many areas in the field of Christian Education that do not yet have tests. This chapter focuses on the procedure to use in developing such a test for use in a larger dissertation context.

Good tests gather good data. Good tests build good attitudes. Good tests can even produce a positive learning experience. The principles discussed here will help you in this task.

### The Emphases in the Material

A test should measure **important areas** of instruction, knowledge, understanding, or skill. The emphases of the test should parallel emphases in the material which subjects have learned. Avoid writing trivial, ambiguous, or simplistic questions.

### Nature of Group Being Tested

Study the group you intend to test. The level of difficulty of the test, the language

you use, the length of the test and other such variables depend a great deal on who your subjects are.

## The Purpose of the Test

What is the purpose of the test? What do you really want to know? Are you measuring knowledge, or comprehension, or the ability to solve problems, or to analyze new situations? Are you measuring simple recall, or mental reasoning? The purpose of the test provides a “North Star” to guide your developmental process.

## Writing items

Avoid ambiguous or meaningless test items. Use good grammar. Avoid rambling or confusing sentence structure. Use items that have a “definitely correct” answer. Avoid obscure language and “big words,” unless you are specifically testing for language usage. Be careful not to give the subject irrelevant clues to the right response. Using “a(n)” rather than “a” or “an” is an example of this.

In short, a test should not provide any barrier to subjects apart from demonstrating mastery over the test content. Otherwise, scores reflect more “noise” than “true measure.”

## Objective Tests

True-False  
Multiple Choice  
Matching

An “objective” test is a test made up of close-ended questions. Objective tests have several advantages over essay tests. Asking 100 objective questions over a given content field provides a much better sampling of examinee knowledge and understanding than asking three or four essay questions. With objective tests, grading is easier and the scores are a more reliable measure of what the examinee knows.

There are four common types of objective questions. These are the constant alternative (true-false) question; the changing alternative (multiple choice) question; the supply (or fill-in-the-blank) question; and the matching question.<sup>1</sup>

## The True-False Item

The true-false, or constant alternative item, presents the subject or student with a factual statement. The statement is judged to be either true or false.

### Advantages

The advantages of the true-false test item are efficiency and potency. It is efficient in that a large number of items can be answered in a short period of time. Scoring is fast and easy. It is potent because it can, in a direct way, reveal common misconceptions and fallacies.

---

<sup>1</sup>The material in this chapter is a synthesis of principles gleaned from Nunnally, “Chapter 6: Test Items,” 153-196; and Payne, “Chapter 5: Constructing Short Answer Achievement Test Items,” 95-136. These are excellent resources for those wanting to improve their test-writing ability. Another excellent (more recent) source is Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement: Classroom Application and Practice*, 2nd (Glenview, IL: Scott, Foresman and Company, 1987). Also more recent material can be found in my own *Created to Learn* (1996), Chapter 14 and *Called to Teach* (1999), Chapter 9, both from Broadman and Holman.

## Disadvantages

Good true-false items are **hard to write**. An item that makes sense to the writer may confuse even well-informed subjects. Statements require careful wording, evaluation and revision. Secondly, true-false items **encourage guessing**. An examinee can earn around 50% of the test score by mere chance simply by guessing at the right answer. If there are only two alternatives, then pure chance gives him 50% over the long run. This is if subjects know absolutely nothing about the subject. This is a lot of “noise” in the test scores. Thirdly, constant alternative items tend toward **response sets**. A response set is a repetitious pattern of answers, like the following 18-item test.

T T F T T F T T F F T F T T F T T F  
 ^            ^            ^                            ^            ^

Notice that the pattern “T T F” repeats itself through the test. Test writers can produce these response sets **without being aware of it**. Subjects pick up these irrelevant clues, and score higher than their knowledge allows. The objective is not to insure high scores, but **to actually measure what subjects know and understand**.

## Writing True-False items

The following guidelines will help you avoid major pitfalls in writing true-false test items.

### Avoid specific determiners

Specific determiners, such as “only,” “all,” “always,” “none,” “no,” or “never,” give irrelevant clues to the correct answer. **When you find these terms in a true-false item, the answer is usually FALSE.** Terms like “might,” “can,” “may,” or “generally” are usually true. Write items without using these terms.

Determiners  
 Answers  
 Negatives  
 Language  
 Quotes  
 Item length  
 Sentences  
 False Items

### Absolute answer

Base true-false items on statements that are **absolutely true or absolutely false**. Avoid statements that are true under some conditions, but not others, unless the conditions are specifically stated. Well-informed subjects have greater difficulty answering ambiguous questions correctly, because they have more information to process in trying to understand the item.

### Avoid double negatives

A double negative is confusing. “T F It is not infrequently observed that three-year-olds play in groups.” **State the item positively:** “T F Three-year-olds play in groups.” The latter item tests knowledge of three-year-olds and social development. The former requires knowledge plus practice in “mental gymnastics.”

### Use precise language

Avoid using terms like “few,” “many,” “long,” “short,” “large,” “small,” or “important” in test items. These terms are **ambiguous**. How much is enough to determine the truth or falseness of a T-F question?

How big is “big”? How many is “many”?

### Avoid direct quotes

If the treatment has been a classroom situation or a series of directed readings, do not test over direct quotes from class notes or the readings. These, taken out of context, are usually **too ambiguous** to use as test items.

### Watch item length

Avoid making true statements longer than false items. This is easy to do because true statements often **need qualifications** to “make sure they are absolutely true.” The additional length is an irrelevant clue to the answer.

### Avoid complex sentences

Complex grammatical constructions and obscure language infuse questions with an **irrelevant level of difficulty**. Take a central idea and write two simple statements, one true and one false. Place these in your item pool.

### Use more false items

When developing a True-False test, make about 60% of the items false. False items **discriminate better** between examinees than true items.

**Yountian modification:** Improve the reliability of the test scores by having subjects **correct false items to make them read correctly**. Underline the most important concept in the statement, and have them change it to make the statement true. Score one point for the correct answer and one point for correcting the statement. This reduces guessing and increases reliability of scores.

## Multiple Choice Items

The multiple choice, or changing alternative, item consists of a **sentence stem and several responses**. One and only one of the responses is correct. All other responses are incorrect, but *plausible*. The most common form presents a stem and four or five responses.

### Advantages

The multiple choice question, with its multiple responses, can be written with less ambiguity and greater structure than the true-false question. **Guessing is reduced** since the probability of guessing the correct answer is 1 in 4 (25%) instead of 1 in 2 (50%) for true-false items. Multiple choice items can demand more **subtle discrimination** than other forms of objective questions. Lastly, one can write multiple choice items which test at **higher levels of learning**, such as application and analysis, than other question types.

### Disadvantages

Good multiple choice questions are **difficult to write**. Effective **detractors** — plausible wrong answers — are hard to create, particularly if you are providing a 5th or 6th alternative response. Secondly, multiple choice tests are **less efficient** because a subject can process fewer multiple choice items in a given time than other types.

## Writing Multiple Choice Items

The following guidelines will help you avoid major pitfalls in writing changing alternative items.

## Pose a singular problem

The **stem** of the question should **pose a clear, definite, singular problem**. A common mistake in multiple choice questions is the use of an incomplete stem. “In the continent of Africa...” could be followed with any number of responses that “fit.” Even better (though not a requirement) is to make the stem a **complete sentence** or a **direct question**, rather than an **sentence fragment**.

## Avoid repeating phrases in responses

Rather than putting the same phrase in every response, **include the phrase in the stem**. Keep the alternative responses as simple as possible.

## Minimize negative stems

Avoid negative stems **if possible**. “Which of the following is NOT a characteristic of...” This construction **can confuse** some subjects who might otherwise know the material.

## Make responses similar

Avoid making the correct response systematically different from the others (grammar, length, construction). Responses should be **written in parallel form** so that the form of the response is not a clue to the correct answer.

## Make responses mutually exclusive

Each response should be mutually exclusive of all others. **Avoid overlapping responses**.

## Make responses equally plausible

All responses in an item set should be **equally plausible and attractive** to the less knowledgeable subject.

## Randomly order responses

Responses (ABCD) should be randomly ordered for each question. Some test writers hesitate to place the proper answer first (A) because “subjects won’t read the others” or last (D) because “that’s an obvious place for the right answer.” That leaves “B” or “C” for the majority of correct responses. Use a random number table or a computer, or even a die, to assign the order of responses and **avoid unintentional response sets**.

## Avoid sources of irrelevant difficulty

Avoid irrelevant sources of difficulty in the statement of the problem or in the responses. Some test writers confuse subjects by using **complex vocabulary**, for instance. Do you want to know what the subject knows, or do you want to test his vocabulary?

## Eliminate extraneous material

Do not include extraneous material in a question. That is, do not attempt to mislead examinees by **including information not necessary for answering the question**.

Single Problem  
Repeats  
Negative Stems  
Responses Similar  
Responses Exclusive  
Responses Plausible  
Responses Random  
Irrelevance  
Extraneous  
“None of the Above”

## Avoid “None of the Above”

Alternative responses such as “none of the above,” “all of the above,” and “both b and d” should be eliminated if possible. The uninformed thinker can use simple reasoning skills to eliminate several alternatives. But the informed non-thinker may not be able to correctly manipulate the problem. **Unless the test is supposed to measure reasoning ability, do not make “ability to reason” part of the score.**

## Supply Items

Supply items, sometimes called “recall” or “fill in the blank” items, present a **statement with one or more blanks**. The task of the subject is to fill in the blank(s) with the most appropriate terms in order to correctly complete the statement.

### Advantages

Supply items are relatively **easy to construct**. Second, they are **efficient** in that a large number of statements can be processed in a given length of time. Third, remembering a term or phrase is **more difficult** than recognizing it in a list or response set. Therefore, supply items **discriminate better** between subjects’ knowledge of important definitions and concepts.

### Disadvantages

Supply items are notorious for being **ambiguous**. It is **difficult to write** a supply item that is clear and plainly stated. Supply items are unclear in the way they’re graded because usually more than one word will adequately fill the blank. **Grading can be arbitrary and unfair**, depending on how synonyms are handled.

## Writing Supply Items

The following guidelines will help you avoid major pitfalls in writing supply items.

When?

### When to use supply items

In general, only use supply items when the **correct response is a single word or brief phrase**.

Limit blanks

One Correct

Important Blank

End Blank

Irrelevant Clues

Text Quotes

### Limit blanks

Use only **one or two blanks** in a supply item. The greater the number of blanks, the greater the **item ambiguity** and the more difficult grading is.

### Only one correct answer

Write the item in such a way that **only one term or word** will correctly complete the statement. If there are equally acceptable terms for a given concept (i.e., “null” and “statistical” hypothesis), then credit should be given for either answer.

### Blank important terms

Leave only the most important word or term blank. Blanking out minor words makes the item trivial.

## Place blank at the end

In most cases, it is preferable to place the **blank at the end of the sentence**. This gives subjects the entire sentence to construct the basis for supplying the proper term. Placing the blank at the beginning reverses this natural process and causes confusion.

## Avoid irrelevant clues

An irrelevant clue is an element of the statement, unrelated to the conceptual focus of the question, which **hints at the correct answer**. An example is making the length of the blank equal to the number of characters in the word to be supplied. Another irrelevant clue is the use of “a” or “an” before a blank.

Avoid these two irrelevant clues by making **all blanks the same length**, and using the more general **“a(n)” before the blank**.

## Avoid text quotes

Do not use directly quoted sentences out of required reading as supply items. This is a easy temptation to fall into and seems to make sense: “If subjects have read the material, they should be able to supply the appropriate term.” But sentences taken out of context are usually **ambiguous**.

Write the supply item based on a clear concept, not a specific quote.

## Matching Items

Matching items presents subjects with two or three columns of items which relate to each other. An example of a matching question is one which provides a numbered list of authors with a parallel “lettered” list of the books they wrote. Match the books to their authors by writing the “letter” of each book in the space next to the numbered author. The list of authors is the **“item list”** and the list of books is the **“response option list.”**

## Advantages

The matching item can test a **large amount** of material **simply and efficiently**. Response pairs can be drawn from various texts, class notes, and additional readings to form a summary of facts. **Grading is easy**.

## Disadvantages

A good matching item is **difficult to construct**. As the number of response pairs in a given item increases, the more mental gymnastics is required to answer it. Matching items can present little more than a **confusing array of trivial terms** and sentence fragments.

## Writing Matching Items

The following guidelines will help you avoid major pitfalls in writing supply items.

### Limit number of pairs

Do not include too many pairs to be matched in a given item. The list should contain **no more than 8-10 pairs**.

Limit Pairs  
Option List  
One Correct  
Central Theme  
Responses  
Systematic

### Make option list longer

If each response should be used once and only once, then the **option list should be longer than the item list**. That is, it should contain more responses than are needed to match all the items. Then subjects cannot answer the last item by the simple process of elimination. However, **if responses can be used more than once, then both lists can be the same length**.

**Note:** There are times when a matching question can be used in place of several multiple choice questions. The response option list consists of only a few responses. Responses from the options list are used several times to match each of the items. This is not a true matching question, which consists of matched item-response pairs, but it is a common practice and eliminates the need for several repetitive multiple choice items. You can see an example of this kind of “matching” question at the end of the chapter.

### Only one correct match

It is important to insure that each term in the item list **matches only one term** in the option list. This becomes more difficult as the list grows larger. Response options may be used more than once, however.

### Maintain a central theme

A matching item should contain matched pairs that **all relate to one central theme**. Avoid mixing names, dates, events, and definitions in a single matching item. If this is not possible, **construct several matching items, each with a central theme**: dates-events, terms-definitions, and so forth.

### Keep responses simple

It is better to place longer statements in the item list and the shorter answers in the response option list. This helps subjects scan the response list for the correct match more efficiently.

### Make the response option list systematic

Arrange the answers in the response option column in some systematic way. This might be **alphabetical** or **chronological** order. This makes the task of searching through the list less taxing and allows subjects to concentrate on answering correctly.

### Specific instructions

Be sure to clearly instruct subjects on how the matching is to be done. Show an example, if necessary. This eliminates “test-wiseness” as an extraneous variable in the scoring.

## Essay Tests

Essay tests are constructed from **unstructured** or “open-ended” questions which require subjects to write out a response.

## Advantages

Essay test items allow **much greater flexibility and freedom** in answering. Grammar, structure, and content of the answer is left to subjects. Essay items permit testing at the **higher levels of learning** than most types of objective questions. Finally, essay questions permit a **greater range of answers** than objective items.

## Disadvantages

The greatest disadvantage of essay items is that they are **difficult to score consistently**. The answers are more **ambiguous and subjective** than objective responses. The **reliability of scores is lower** than those produced by objective tests over the same content because of the variability of response. Essay items test a **smaller sample of material** because of the amount of time required to analyze and understand the question, develop the answer, and write it out in complete sentences. They are **less efficient** than objective types.

## Writing essay items

The following guidelines will help you avoid major pitfalls in writing essay items.

### Use short-answer essays

It is much better to use **several short answer essay items** than one or two long ones. If the testing period is one hour, ask six ten-minute essays rather than two 30-minute essays. This **improves sampling** of material, **focuses the essays** sufficiently to increase reliability of grading, and produces a **better measure** of what subjects know.

[Short Answer](#)  
[Clear Question](#)  
[Grading Key](#)

### Write clear questions

Be sure that the question you ask gives **sufficient guidance** to examinees. The question, "Discuss sampling processes," is much too vague. Better essay questions **structure the thinking** of subjects: "Define four types of sampling and explain specifically how each is used in research."

### Develop a grading key

Outline a specific grading key for each essay item. Points may be awarded for each element in the key. Major elements should receive more points than minor ones. A point or two should be awarded for grammar, punctuation, organization, and the like.

This grading key provides a **systematic guide** for objectively grading an essay answer. Without such a key, the score is as much a result of the perception of the grader as it is a measure of the knowledge of the subject.

## Item analysis

Item analysis is a procedure for determining which items in an objective test **discriminate between informed and uninformed subjects**. If a test's purpose is to separate subjects along a scale of content mastery (and most tests have this purpose!), then it is important that this **separation be done fairly**. Every item in a test should contribute to this separation process. Those that do not should be revised or eliminated.

A popular method of item analysis is a procedure called the **Discrimination Index**. After administering and grading the exam, the procedure is applied as follows:

### Rank Order Subjects By Grade

Rank order subjects high to low by their grade on the exam. The rank position of each student is a reflection of their **overall preparation** for the examination.

### Categorize Subjects into Top and Bottom Groups

Identify top and bottom proportions of students to compare. You can choose a percentage ranging from 10- to 40-percent. Twenty-five percent is common, and gives you the **top and bottom quarters of the class**.

### Compute Discrimination Index

An example will illustrate this step better than a definition. Let's say you have a class of 40 students. You select top and bottom quarters (25%) for computation of the discrimination index. This means you have **10 students in the top group and 10 in the bottom**, as identified by test score rank. Count how many students in the groups answered question one correctly. Let's say that in our case, **8 of the top 10** subjects answered question #1 right, and **3 of the bottom 10** answered it right. The discrimination index is equal to **8 minus 3 divided by 10**, or **+0.500**.

### Revise Test Items

A discrimination index ranges from -1.00 to +1.00. **A negative index indicates a faulty question**: more "bottom" students answered it right than "top" students. This question should at least be rewritten and **may need to be eliminated** from the test.

Questions you expect everyone in the class to know, a so-called **barrier question**, will appear with a low discrimination index – often a "0.000" index.

Questions you design as discriminating questions – questions designed to separate students according to their mastery of the material – should have **moderate to high index** values (+0.500 and above).

A reasonable test should contain 60% barrier questions and 40% discriminating questions. A test can be more difficult (while remaining completely fair and unbiased) by including higher percentages of discriminating questions (say, 50-50), or by including questions with discrimination indexes of +0.750 or higher (or both!).

The use of the discrimination index by test designers solves one of the most frustrating aspects of education and research: **arbitrary testing**. The discrimination index provides a way to develop tests which contain questions that actually separate the prepared (knowledgable) from the unprepared, and yield test data which is more valid and reliable.

## Summary

In this chapter we have looked at procedures for developing various types of tests. We have considered four kinds of objective items: true-false, multiple choice, supply and matching. We have discussed the use of essay questions. Finally, we described item analysis, which allows test developers to determine whether objective test items are valid.

## Examples

In addition to the checklist in Chapter nine, **Dr. Mark Cook** also developed an objective test

...to measure the lesson objectives at three cognitive levels: knowledge, comprehension, and application. The process of development began by creating a thirty-item multiple-choice test to be used in the field test of the study (appendix D). The test was examined by three selected specialists. The specialists that were asked for validation of the test were as follows: [specialists listed]. These professors were provided complete lesson plans to use in evaluation.<sup>3</sup>

A copy of the test is located at the end of the chapter.

**Dr. Brad Waggoner** focused his entire 1991 dissertation on developing a standardized test to measure the "discipleship base" -- defined as 'that portion of a given church's membership that meets the criteria of a disciple'<sup>4</sup> -- of local Southern Baptist churches. He worked in conjunction with the International Mission Board of the Southern Baptist Convention to produce a valid and reliable instrument. **A final instrument of 136 items<sup>5</sup> produced a Cronbach's alpha reliability coefficient of 0.9618.<sup>6</sup>**

While we can certainly not replicate the fifty-eight pages<sup>7</sup> of his development procedure here, we will outline the procedure and focus on key aspects of test development.

### Phase One: Identification of Functional Characteristics<sup>8</sup>

Attitudes: A disciple is one who:

- Possesses a desire and willingness to learn
- Has conviction regarding the necessity of living in accordance to biblical principles and guidelines
- Evidences a repentant attitude when a violation of Scripture occurs
- Possesses a willingness to forfeit personal desires and conveniences, if necessary, in order to seek the interests of others
- Possesses and demonstrates the character trait of humility
- Possesses and demonstrates the character trait of integrity
- Is willing to be accountable to others

Conduct/Behavior: A disciple is one who:

- Manifests a lifestyle of utilizing time and talents for God's purposes
- Possesses a lifestyle depicted by intentional compliance with the with the moral teachings of the Bible. . .
- Maintains appropriate behavior toward those of the opposite sex
- Actively seeks to promote social justice and righteousness in society as well as to individuals

Relational/Social: A disciple is one who:

- Values and accepts himself as created in the image of God
- Has an awareness of the reality and presence of God through the ministry of the Holy Spirit
- Experiences trust in God in times of adversity as well as in times of prosperity
- Seeks to commune with and learn about God through the means of meditation upon Scripture and prayer
- Is consistently involved in fellowship with other believers in the context of a local church
- Applies oneself to building meaningful relationships with other believers
- Maintains a forgiving spirit when wronged
- Confesses or seeks forgiveness when guilty of an offense

<sup>3</sup>Cook, 22-23

<sup>4</sup>Waggoner, 9

<sup>5</sup>*Ibid.*, 209

<sup>6</sup>*Ibid.*, 118

<sup>7</sup>Pages 65-118 of 233 pages

<sup>8</sup>*Ibid.*, Headings from 68-80

Ministry/Skills: A disciple is one who:

- Publicly identifies with Christ and the Church when provided an opportunity
- Seeks and takes advantage of opportunities to share the Gospel with others
- Is involved in ministering to other believers
- Seeks the good of all men with a willingness to meet practical social needs such as food, clothing, and the like

Doctrine/Beliefs:

- Eternal security
- Salvation
- The Holy Spirit (the nature and role of)
- The Eternal State (the literal existence of heaven and hell)
- Scripture (the authority and reliability of)

#### Phase Two: Testing of Content Validity<sup>9</sup>

The functional characteristics, categorized according to the five domains described above, were placed on a 9-point Likert rating scale, a value of "1" being "not valid," and a value of "9" being "very valid" with gradations of validity in between<sup>57</sup> (appendix B).

The purpose of the rating scale was for a panel of experts to determine the degree to which each characteristic was a valid and measureable function of a disciple.<sup>(58)</sup>

A list of names was compiled. . . the panel was to consist of five experts and two alternatives representing the academic, denominational, and local church levels (appendix C).<sup>(59)</sup>

A letter was constructed that explained the nature and purpose of the research and requested their participation on the panel (appendix D). . . . When the rating scales were returned, the mean scores were calculated for the characteristics (appendix F).

#### Phase Three: Revision of Characteristics<sup>10</sup>

Revisions to the list of characteristics were made based on the panel's scores, comments, and additions. It was predetermined that any item receiving a mean score of less than 7.0 would be considered for deletion.

#### Phase Four: Item Writing<sup>11</sup>

- Review Related Measures
- Construction of Questions
  - The Size of the Item Pool
  - The Issue of Relevance
  - The Issue of Clarity
  - The Issue of Simplicity
  - The Issue of Single Meaning
  - The Issue of Double Negatives
  - The Issue of Question Length
  - The Issue of Question Variety
  - The Issue of Response Categories
  - The Issue of Assuming
  - The Issue of "Leading" or "Loaded" Questions
  - The Issue of Grammar and Tone

#### Phase Five: Testing Content Validity of Questions<sup>12</sup>

- Selection of a Panel of Experts
- Development of a Validation Instrument
- Follow-Up of Validation Panel

<sup>9</sup>Ibid., 81-82

<sup>10</sup>Ibid., 82

<sup>11</sup>Ibid., 83-91

Calculation of Validity: "...questions receiving mean scores of less than 6.0 would be considered for deletion."

#### Phase Six: Questionnaire Design<sup>13</sup>

- Question Order and Flow
- Questionnaire Length
- Questionnaire Design and Layout
  - Size and Color of Paper
  - Layout
  - Instructions
  - Expression of Gratitude
  - Expression of Confidentiality
  - Identification of the Sponsor

#### Phase Seven: Refining the Pilot Test<sup>14</sup>

The process of refining the pilot test consisted of a small number of individuals evaluating the clarity of questions, word meanings, instructions, and procedure for completing the instrument. . . . Revisions were made to the instrument based upon the results. Subsequently, over 100 questionnaires were printed and put into booklet form (appendix M).

#### Phase Eight: Pilot Test #1<sup>15</sup>

- Selection of Sample Group [n=50 church members in two groups]
- Establish Time and Place of Pilot Test
- Letter of Invitation Constructed and Mailed
- Administering the Instrument

#### Phase Nine: Data Analysis<sup>16</sup>

[This is part of Chapter Four of the dissertation].

#### Phase Ten: Revision of the Instrument<sup>17</sup>

#### Phase Eleven: Second Pilot Test<sup>18</sup>

- Selection of [Three] Churches
- Procedure for Administering the Pilot Test
- Follow-Up Procedure

#### Phase Twelve: Data Analysis of the Second Pilot Test<sup>19</sup>

[This is part of Chapter Four of the dissertation].

As mentioned in Chapter One, this instrument -- with further revisions by Dr. Waggoner in conjunction with the IMB and LifeWay Christian Resources (SBC) -- is being **integrated into revised *MasterLife* materials produced by LifeWay.**

## Vocabulary

|                          |   |
|--------------------------|---|
| changing alternative     | synonym for a <b>multiple choice test item</b>                          |
| constant alternative     | synonym for a <b>true-false test item</b>                               |
| discrimination index     | procedure used to determine <b>quality of test items</b>                |
| distractors              | multiple choice options which <b>appear plausible</b> but are incorrect |
| multiple choice question | test item with one stem and 4 or 5 plausible options                    |

<sup>12</sup>*Ibid.*, 91-92

<sup>13</sup>*Ibid.*, 92-98

<sup>14</sup>*Ibid.*, 98-99

<sup>15</sup>*Ibid.*, 99-103

<sup>16</sup>*Ibid.*, 103

<sup>17</sup>*Ibid.*, 103-104

<sup>18</sup>*Ibid.*, 105-106

<sup>19</sup>*Ibid.*, 106

response set  
specific determiners  
supply question

**predictable pattern** in objective answers (e.g. T T T F T T T F T T T F)  
terms like `never' or `sometimes' that **give clues to correct answer**  
synonym for **fill-in-the-blank questions**

## Study Questions

1. Explain the four preliminary guidelines given for writing tests in your own words.
2. Explain why objective test items produce more reliable scores than essay test items.
3. Write out 3 TF, 3 MC, 2 supply and 2 essay questions relating to this material. Set them aside for a few days. Then go back and evaluate each of your questions according to the criteria given for each kind of question.

## Sample Test Questions

1. T F A true-false question which uses terms such as "only," "none," or "always" are usually true.
2. Choose the best true-false question below.
  - A. Disuse of double negatives does not impair item validity.
  - B. Payne writes, "Don't use direct quotes in t-f items."
  - C. Direct quotes, "fuzzy" language, double negatives, specific determiners and complex sentences should be avoided in t-f items; rather, focus on central concepts, precise language, and simple sentences.
  - D. Constant alternative items consist of a stem and several parallel responses.
3. Which of the following is an advantage of multiple choice items over true-false items?
  - A. Easy to write
  - B. Guessing is reduced
  - C. Less efficient
  - D. More open ended
4. Which of the following is a problem of matching questions?
  - A. The question contains less than 10 pairs
  - B. The response options list is systematically ordered
  - C. Each item matches one and only one response option
  - D. Response options cover multiple themes

## Sample Test

### APPENDIX B<sup>3</sup> PRE-SESSION TEST

Student Number \_\_\_\_\_ (see your name tag)

**Circle the letter of the phrase that best completes the sentence.**

1. The phrase "priesthood of believers" is found in the Bible
  - (a) in the New Testament,
  - (b) in the Old Testament,
  - (c) in both testaments,
  - (d) in neither testament.
  
2. The doctrine of the priesthood of the believer teaches that priests should
  - (a) be representative of all people,
  - (b) represent God to other persons
  - (c) be ordained by a church,
  - (d) remain completely separated from the world.
  
3. During the Reformation, the priesthood of all believers particularly emphasized
  - (a) infant baptism,
  - (b) personal witnessing,
  - (c) direct access to God,
  - (d) wrongs of the Catholic church.
  
4. The concept of priest in the Old Testament is most often associated with the priesthood of
  - (a) all Israelites,
  - (b) some Israelites,
  - (c) no Israelites,
  - (d) the special prophets of Israel.
  
5. The Old Testament covenant was designed by God
  - (a) to bless Israel as His people only,
  - (b) to assure that Israel worshipped only God,
  - (c) to help Israel conquer their world,
  - (d) to make Israel a blessing to all other nations.
  
6. Christians are referred to as a holy priesthood. This holiness is best reflected by Christians when they are
  - (a) motivated by love,
  - (b) pure in their thoughts,
  - (c) serving God at church,
  - (d) separated from the world....

---

<sup>3</sup>Cook, 62-63. The entire test runs 15 items.

