CHAPTER

9

MEASURING COMPLEX ACHIEVEMENT: THE INTERPRETIVE EXERCISE

Objectives

From this chapter you should be able to:

1. Describe the uses of the interpretive exercise.
2. Describe the advantages and limitations of the interpretive exercise.
3. Construct interpretive exercises.

Complex achievement includes those learning outcomes based on the higher mental processes, such as understanding, thinking skills, and various problem-solving abilities. Although some aspects of complex achievement require extended constructed responses and other types of performance assessment tasks, other aspects can be measured objectively.

We have already had some experience with measuring complex achievement, as the category encompasses all those learning outcomes requiring more than the mere retention of factual knowledge. The short-answer item to measure problem-solving abilities in mathematics and science, the true–false item to measure the ability to recognize cause-and-effect relationships, and the multiple-choice item to measure various aspects of understanding and application illustrate the measurement of complex achievement. These illustrations, however, were limited to the use of single, independent test items of the objective type. Greater range and flexibility in measuring complex achievement can be attained not only by moving to the extended response and other performance assessment tasks discussed in Chapters 10 and 11 but also by using more complex forms of objective test items.

A variety of learning outcomes are included in complex achievement. Following are some typical examples:

- Ability to apply a principle
- Ability to interpret relationships
- Ability to recognize and state inferences
- Ability to recognize the relevance of information
- Ability to develop and recognize tenable hypotheses
- Ability to formulate and recognize valid conclusions
- Ability to recognize assumptions that underlie conclusions
- Ability to recognize the limitations of data
- Ability to recognize and state significant problems
- Ability to design experimental procedures
- Ability to interpret charts, tables, and data
- Ability to evaluate arguments

These and similar learning outcomes have been classified under such categories as understanding, reasoning, critical thinking, scientific thinking, creative thinking, and problem solving. There is general agreement that learning outcomes based on higher-order thinking skills constitute some of the most significant outcomes of education. Given the importance of these complex learning outcomes, it is critical to use a full array of assessment techniques available for measuring those outcomes. The interpretive exercise provides one of those needed techniques. Used wisely, and supplemented by the techniques discussed in Chapters 10 and 11, the interpretive exercise can help ensure that complex learning outcomes are given adequate priority in classroom assessments.

NATURE OF THE INTERPRETIVE EXERCISE

An interpretive exercise (also called "classification exercise," "key-type item," or "master-list item") consists of a series of objective items based on a common set of data. The data may be in the form of written materials, tables, charts, graphs, maps, or pictures. The series of related test items may also take various forms but are most commonly multiple-choice or true–false items. Because all students are presented with a common set of data, it is possible to measure a variety of complex learning outcomes. Students can be asked to identify relationships in data, to recognize valid conclusions, to appraise assumptions and inferences, to detect proper applications of data, and the like.

The common set of materials used in interpretive exercises ensures that all students will be confronted with the same task. It also makes it possible to control the amount of factual information given to them. We can give them as much or as little information as we think desirable in measuring their achievement of a learning outcome. In measuring their ability to interpret mathematical data, for example, we can include the formulas needed or require the students to supply them. In other areas, we can supply definitions of terms, meanings of symbols, and other facts or expect students to supply them. This flexibility makes it possible to measure various degrees of proficiency in any particular area.

FORMS AND USES OF THE INTERPRETIVE EXERCISE

As with other objective items, there are so many forms and uses of the interpretive exercise that it is impossible to illustrate all of them. Here we present examples of this item type as applied to the measurement of complex learning outcomes in a variety of school subjects at the elementary and secondary levels. Different types of introductory material and different methods of responding also will be used to illustrate the great flexibility of the interpretive exercise. The references at the end of this chapter offer additional illustrative exercises.

Ability to Recognize Warranted and Unwarranted Generalizations

The ability to recognize the validity of generalizations is of central importance in the interpretation of data. At minimum, students should be able to determine which conclusions the data support, which the data refute, and which the data neither support nor refute. The data may be in the form of tables, charts, graphs, maps, or pictures, and the test items may be true–false or multiple-choice items. An illustration of recognizing the validity of generalizations is shown in the following example:

EXAMPLE Percentage of population between the ages of 25 and 34 who have completed secondary and higher education, by gender, for large industrialized countries: 1995

	Males		Females	
Country	Secondary Education	Higher Education	Secondary Education	Higher Education
Canada	82.4	19.2	85.4	19.7
France	87.3	13.7	83.8	14.3
Germany	91.0	13.7	86.7	11.3
Italy	46.9	7.8	51.2	8.6
Japan	89.3	34.2	91.8	11.5
United Kingdom	87.5	16.3	84.7	13.1
United States	86.1	25.1	88.2	24.9

Source: Data from "The Condition of Education: 1998," Washington, DC: National Center for Education Statistics, U.S. Department of Education, 1998. Available at *http://nces.ed.gov/pubs98/98013.pdf*

Directions: The following statements refer to the data in the table above. Read each statement and mark your answer according to the following key.

Circle:

S if the statement is supported by the data in the table.
R if the statement is refuted by the data in the table.
N if the statement is neither supported nor refuted by the data.

Ⓢ R N 1. The United States has a smaller discrepancy in percentage completion of higher education for males and females between the ages of 25 and 34 than any of the other countries listed.

S	R	(N)	2. College admissions policies give preferential treatment to male applicants over female applicants.
S	R	(N)	3. It is more difficult to get into college in Germany than in Japan.
(S)	R	N	4. When males and females are combined, the United States has the highest secondary school completion percentage for young adults between the ages of 25 and 34.

Ability to Recognize Assumptions

Another learning outcome pertinent to the interpretation of various types of information is the ability to identify unstated assumptions that are necessary to a conclusion or course of action. The following item illustrates this type of interpretive exercise:

EXAMPLE Studies have shown that there is a relationship between vocabulary and crime. Crime rates are higher for people with poorly developed vocabularies, and crime rates are lower for people with well-developed vocabularies. Older studies have also shown that there is a positive relationship between the number of years of Latin studied and the size and preciseness of an individual's vocabulary. Conclusion: Crime rates can be lowered by reintroducing the study of Latin in the schools.

Which one of the following assumptions is necessary to reach such a conclusion?

A	Correlational methods were used to determine these relationships.
B	These reported relationships were statistically significant.
(C)	Relationships such as these imply causation.
D	Latin scholars have a low crime rate.

Ability to Recognize the Relevance of Information

A learning outcome important to all subject-matter areas and that can be measured at all levels of instruction is the ability to recognize the relevance of information. The exercise presented here was prepared for third-grade students. An example at the high school level may be found in Appendix D:

EXAMPLE Bill lost his boot on the way to school. He wanted to put a notice on the bulletin board so that the other children could help him find it. Which of the following sentences tell something that would help children find the boot?

Directions: Circle *yes* if it would help. Circle *no* if it would not help.

(yes)	no	1. The boot was black.
yes	(no)	2. It was very warm.
(yes)	no	3. It was for his right foot.
yes	(no)	4. It was a Christmas present.
yes	(no)	5. It was nice looking.
(yes)	no	6. It had a zipper.
(yes)	no	7. It had a gray lining.

Ability to Apply Principles

The application of principles may be shown in many different ways. In the following example, students are asked to identify principles that explain a situation and to recognize illustrations of a principle:

EXAMPLE **Mary Ann wanted her rose bush to grow faster, so she applied twice as much chemical fertilizer as was recommended and watered the bush every evening. About a month later she noticed that the rose bush was dying.**

Directions: **Which of the following principles is necessary in explaining why the rose bush is dying? If a principle is necessary, circle N; if unnecessary, circle U.**

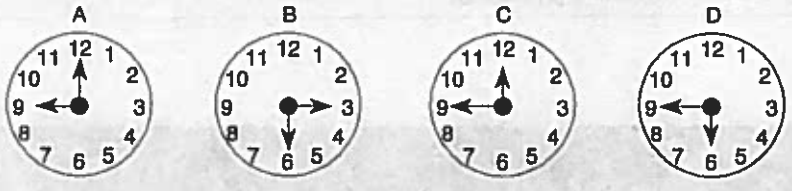
N	(U)	1. A chemical compound is changed into other compounds by taking up the elements of water.
(N)	U	2. Semipermeable membranes permit the passage of fluid.
N	(U)	3. Water condenses when cooled.
(N)	U	4. When two solutions of different concentration are separated by a porous partition, their concentration tends to equalize.

Use of Pictorial Materials

Pictorial materials can serve two useful purposes in interpretive exercises. First, they can help measure a variety of learning outcomes similar to those already discussed simply by replacing the written or tabular data with a pictorial presentation. This use is especially desirable with younger students and when ideas can be more clearly conveyed in pictorial form. Second, pictorial materials can also measure the ability to interpret graphs, cartoons, maps, and other pictorial materials. In many school subjects, these are important learning outcomes in their own right.

The following examples illustrate the use of pictorial materials:

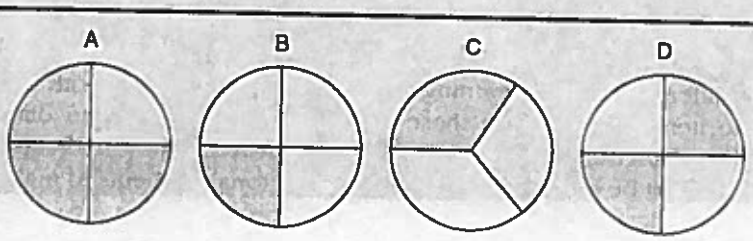
EXAMPLE I



USE ORAL QUESTIONS

What clock shows the time that school starts?	(A)	B	C	D
What clock shows the time closest to lunch time?	A	B	(C)	D
What clock shows half past the hour?	A	(B)	C	D

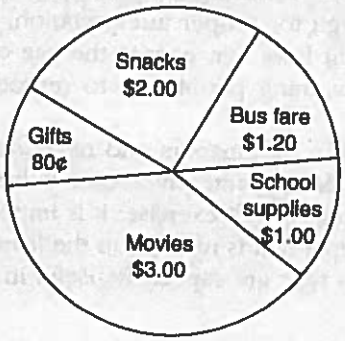
EXAMPLE II



USE ORAL QUESTIONS

What circle is 1/4 shaded?	A	(B)	C	D
What circle is 1/2 shaded?	A	B	C	(D)
What circle is most shaded?	(A)	B	C	D
What circle is least shaded?	A	(B)	C	D

EXAMPLE III



Above is a graph of Bill's weekly allowance distribution.

1. What is the ratio of the amount Bill spends for school supplies to the amount he spends for movies?
   - A 7:2
   - B 1:3
   - C 2:7
   - D 3:1
2. What would be the best title for this graph?
   - A Bill's weekly allowance
   - B Bill's money graph
   - C Bill's weekly expenditures
   - D Bill's money planning

These three examples were designed for use in lower grades. They illustrate the use of pictorial materials that can be drawn by the teacher and items that are useful for measuring rather simple interpretations of concepts and relationships.

Examples IV is an interpretive exercise designed for higher grade levels. They are included here to illustrate the use of various types of pictorial materials, the measurement of different types of learning outcomes, and the use of both multiple-choice and true–false items. As noted in these examples, the pictures and diagrams used in an interpretive exercise frequently can be obtained from published sources. When this is done, care must be taken in reproducing the pictorial elements to make certain that they are clear and detailed enough for proper interpretation. It is also important, of course, to be aware of the copyright laws that govern the use of the material. However, there is seldom a problem in obtaining permission to reproduce copyrighted materials for classroom use.

Cartoons can be found in newspapers and news magazines. Then simply prepare questions that require the desired interpretations. Either true–false or multiple-choice items might be used with this type of exercise. It is important to select a cartoon that illustrates a concept or principle that is relevant to the learning outcomes to be measured. Interpretive exercises of this type are especially useful in social studies.

EXAMPLE IV Average number of days of school per year for 13-year-olds, by country. School year 1990–91. (From *The Condition of Education*, 1993. National Center for Education Statistics.)

Larger Countries	Average Days per Year	Average Hours per Day	Average Hours per Year
Canada	188	5.1	953
England	192	5.0	960
France	174	6.2	1,073
Germany	210	4.6	966
Italy	204	4.8	983
Japan	220	4.0	875
Korea	222	4.4	977
Taiwan	222	5.3	1,177
United States	178	5.6	1,003

Directions: The following statements refer to the data in the chart above. Read each statement and mark your answer according to the following key.

Circle:

T—if the data in the chart are sufficient to make the statement true.
F—if the data in the chart are sufficient to make the statement false.
I—if the data in the chart are insufficient to determine whether the statement is true or false.

T (F) I 1. The number of days per year of school is lower in the United States than in any of the other countries shown.

(T) F I 2. The average number of days of school is higher in the three Asian countries than in the remaining six.

T F (I) 3. United States students spend fewer hours in school than students from Japan.

A chart like that in Example IV is typically easy to prepare but may sometimes be found in published sources. Maps and diagrams that contain data also make effective materials for interpretive exercises.

ADVANTAGES AND LIMITATIONS OF INTERPRETIVE EXERCISES

The interpretive exercise has several advantages. First, the introductory material makes it possible to measure the ability to interpret written materials, charts, graphs, maps, pictures, and other communication media encountered in everyday situations. The rapid expansion of knowledge in every subject-matter area has made it impossible to learn all the important factual information in a given field, which has led to greater dependence on libraries, reference materials, self-study techniques, and interpretive skills. Second, the interpretive exercise makes it possible to measure more complex learning outcomes than can be measured with the single objective item. Some data are usually necessary if students are to demonstrate thinking and problem-solving skills, and the inclusion of such data is awkward in most kinds of test items. Third, by having a series of related test items based on a common set of data, greater depth and breadth can be obtained in the measurement of intellectual skills. Fourth, the interpretive exercise minimizes the influence of irrelevant factual information on the measurement of complex learning outcomes. Students may be unable to demonstrate their understanding of a principle simply because they do not know some of the facts concerning the situation to which they are to be applied. This blocking of response, caused by a lack of detailed factual information not directly pertinent to the purpose of the measurement, can be largely eliminated with the interpretive exercise. In the introductory materials, we can give students the common background of information needed to demonstrate understanding, thinking skills, and problem-solving abilities.

The interpretive exercise is more structured than performance assessment tasks. Whether this is an advantage or disadvantage depends on the specific outcomes to be measured. Students are not free to redefine the problem or to demonstrate thinking skills at which they are most efficient on an interpretative exercise. The series of objective items forces them to use only the mental processes called for. This also makes it possible to measure separate aspects of problem-solving ability and to use objective scoring procedures. This structure has advantages for focusing the task on a specific outcome, such as the ability to identify assumptions underlying conclusions. For other types of outcomes, however, the structure may make it impossible to adequately assess the desired outcome. For example, an interpretive exercise would not be suitable for measuring whether students are able to generate the assumptions underlying a conclusion.

As with all forms of test items, the interpretive exercise does have some limitations. Probably the greatest limiting factor, and one that may have occurred to you as you reviewed the sample items, is the difficulty of construction. Selecting printed materials that are new to the students but that are relevant to the instructional outcomes requires considerable searching. When pertinent material is found, it usually must be edited and reworked to make it more suitable for testing purposes.

Next, test items must be constructed that demand the specific behaviors indicated in the learning outcomes being measured. The construction process is often circular (i.e., it goes back and forth between revising the introductory material and revising the test items until a satisfactory product is obtained). This entire procedure is time consuming and requires much greater skill than that needed to construct single objective test items. But three positive comments can be made regarding the difficulty of constructing interpretive exercises. First, more and more items of this type now appear in various subject-matter fields. The references at the end of this chapter contain numerous examples that may serve as guides to test construction. Second, the greater instructional emphasis on complex learning outcomes resulting from the use of interpretive exercises offsets the additional effort required in test construction. Finally, the task becomes easier with practice and experience.

A second limitation, especially pertinent when the introductory material is in written form, is the heavy demand on reading skill. The poor reader is handicapped by both the difficulty of the reading material and the length of time it takes to read each test exercise. The first problem can be controlled somewhat by keeping the reading level low and the second by using brief passages. Both of these are only partial solutions, however, because the poor reader will still be at a decided disadvantage. In the primary grades and in classes that contain many poor readers, interpretive exercises might be better limited to the use of pictorial materials.

Compared to the extended essay questions and performance-based assessment tasks discussed in Chapters 10 and 11, the interpretive exercise has two shortcomings as a measure of complex achievement. First, it cannot measure a student's overall approach to problem solving. It is efficient for measuring specific aspects of the problem-solving process, but it does not indicate whether the student can integrate and use these skills when faced with a particular problem. Thus, it provides a diagnostic view of the students' problem-solving abilities in contrast with the holistic view of essay questions or other performance assessment tasks.

Second, because the interpretive exercise usually uses selection items, it is confined to learning outcomes at the recognition level. To measure the ability to define problems, to formulate hypotheses, to organize data, and to draw conclusions, performance assessment tasks must be used. Clearly, the interpretive exercise is not suitable for assessing a student's ability to communicate effectively in writing, perform an experiment, create a work of art, or make an oral presentation to a group. Essay questions and performance assessment tasks are needed for such outcomes. Nonetheless, the interpretive exercise is a valuable technique that can contribute to the valid measurement of complex outcomes.

SUGGESTIONS FOR CONSTRUCTING INTERPRETIVE EXERCISES

There are two main tasks in constructing interpretive exercises: (a) selecting appropriate introductory material and (b) constructing a series of dependent test items. In addition, care must be taken to construct test items that require analyzing the introductory material

in terms of complex learning outcomes. The following suggestions will aid in constructing high-quality interpretive exercises:

1. Select introductory material that is relevant to the objectives of the course. Interpretive exercises, like other testing procedures, should measure the achievement of specific instructional outcomes. Success in this regard depends to a large extent on the introductory material, as this provides the common basis for the test items. If the introductory material is too simple, the exercise may become a measure of general information or simple reading skill. On the other hand, if the material is too complex or unrelated to instructional goals, it may become a measure of general reasoning ability. Both extremes must be avoided. Ideally, the introductory material should be pertinent to the course content and complex enough to evoke the mental reactions specified in the course objectives.

The amount of emphasis given to the various interpretive skills in the course objectives is also important. Care must be taken not to overload the test with interpretive items in any particular area. The selection of introductory material should be guided by the emphasis to be given to the measurement of complex achievement and each type of interpretive skill.

2. Select introductory material that is appropriate to the students' curricular experience and reading level. Many complex learning outcomes can be measured with different types of introductory material. The ability to recognize the validity of conclusions, for example, can be measured with written materials, tables, charts, graphs, maps, or pictures. The type used should be familiar to the students so that the nature of the material does not prevent them from demonstrating their achievement of the complex learning outcomes. It would be unfair, for example, to ask students to recognize the validity of conclusions on the basis of data presented in graph form if they had not had experience in interpreting graphs similar to those used in the test. When various types of introductory material will serve a purpose equally well and all are familiar to the students, we favor material that places the least demand on reading skill. For elementary students, pictorial materials are definitely favored. For higher grade levels, pictorial materials and verbal materials with a low vocabulary load and simple sentences are preferred. Although general reading skill is necessary in all written tests, it can become prominent in interpretive exercises unless efforts are made to minimize its influence.

3. Select introductory material that is new to students. In order to measure complex learning outcomes, the introductory material must be new. Asking students to interpret materials identical to those used in instruction does not ensure that the exercise will measure anything other than rote memory. Too much novelty, however, must be avoided. Materials that are similar to those used in class but vary slightly in content or form are the most desirable. Such materials can be obtained by modifying selections from textbooks, newspapers, newsmagazines, and various reference materials pertinent to the course content.

4. Select introductory material that is brief but meaningful. Another method of minimizing the influence of general reading skill on the measurement of complex learning outcomes is to keep the introductory material as brief as possible. Digests of articles are frequently available and are good raw material for interpretive exercises. If digests are

unavailable, the summary of an article or a key passage may be sufficient. In some cases, the relevant information is summarized better in a table, diagram, or picture. In striving for brief introductory material, be careful not to omit elements that are crucial to the interpretive skills being measured. The material also should, of course, be complete enough to be meaningful and interesting to the students.

5. Revise introductory material for clarity, conciseness, and greater interpretive value. Although some materials (e.g., graphs) can be used without revision, most selections require some adaptation for testing purposes. Technical articles frequently contain long, detailed descriptions of events. On the other hand, news reports and digests of articles are brief but often present exaggerated reports of events to attract the reader's interest. Although such reports provide excellent material for measuring the ability to judge the relevance of arguments, the need for assumptions, the validity of conclusions, and the like, the material must usually be modified to be used effectively.

Revision of the introductory material and construction of the related test items tend to be interdependent procedures. Rewriting material often suggests questions to be used, and the construction of test questions often necessitates revisions of the material. In revising a description of an experiment, for example, assumptions, hypotheses, or conclusions explicitly stated in the description may be deleted and used as a basis for questions. Likewise, a question calling for application of the experimental findings may require the addition of new material to the selection. Thus, the revision of the introductory material and the construction of test items proceed in a circular fashion until a clear, concise interpretive exercise evolves.

6. Construct test items that require analysis and interpretation of the introductory material. There are two common errors in the construction of interpretive exercises that invalidate them as a measure of complex achievement. One is to include questions that are answered directly in the introductory material, that is, asking for factual information explicitly stated in the selection. Such questions measure simple reading skill. The second error is to include questions that can be answered correctly without reading the introductory material, that is, requiring answers based on general information. These questions measure simple knowledge outcomes.

If the interpretive exercise is to function as intended, it should include only those test items that require students to read the introductory material and to make the desired interpretations. In some instances, the interpretations will require students to supply knowledge beyond that presented in the exercise. In others, the interpretations will be limited to the factual information provided. The emphasis on knowledge and interpretive skill will be determined by the learning outcomes being measured. Regardless of the emphasis, however, the test items should be dependent on the introductory material while at the same time calling forth mental reactions of a higher order than those related to simple reading comprehension.

7. Make the number of test items roughly proportional to the length of the introductory material. It is inefficient to have students analyze a long, complex selection of material and then answer only one or two questions about it. Although it is impossible to specify the exact number of questions that should accompany a given amount of material, the items presented earlier in this chapter show a desirable balance. Other things being

equal, we always favor the interpretive exercise that has brief introductory material and a relatively large number of test items.

8. In constructing test items for an interpretive exercise, observe all pertinent suggestions for constructing objective items. The form of test item used in the interpretive exercise will determine the rules for construction. If multiple-choice or true–false items are used, the suggestions for constructing these item types should be followed. When modified forms are used, suggestions for constructing each of the various types of objective items should be reviewed for their applicability in construction. Freedom from irrelevant clues and technical defects is as important in interpretive exercises as it is in single, independent test items.

9. In constructing key-type test items, make the categories homogeneous and mutually exclusive. The key-type item, which is used frequently in interpretive exercises, is a modified multiple-choice form that uses a common set of alternatives. In this regard, it is also similar to the matching item and so should be constructed in the same way, with special attention devoted to the categories used in the key. All the categories in any one key should be homogeneous; that is, they all should be concerned with similar types of judgment. At the same time, there should be no overlapping of categories. Each alternative should provide a separate category so that there is a clear-cut system of classification and each item has only one correct answer:

EXAMPLE The majority of medical researchers agree that exposure to secondhand cigarette smoke is detrimental to health. A number of cities have passed ordinances that prohibit smoking in public buildings. Despite an intensive educational campaign pointing out the dangers of secondhand smoke, many cities do not prohibit smoking in public buildings. Resolved: In the interests of national health, smoking should be prohibited in all public buildings in the United States.

Directions: Read each of the following statements carefully. In front of each statement mark

Key: **A** if the statement supports the resolution.
B if the statement contradicts the resolution.
C if the statement is a fact.
D if the statement is an opinion.

____1. The amount of reduction in exposure to secondhand smoke in cities with ordinances prohibiting smoking in public buildings has not been studied.

(Similar items complete the exercise.)

In this example, the key includes two overlapping categories, one concerned with the relationship of each statement to the resolution and the other with the nature of the statement itself. This makes it impossible to have only one correct answer for each statement. Item 1, for example, would have to be marked category B because it contradicts the resolution, and category C because it is a statement of fact.

The key could be improved by limiting the categories to the relevance of the statements to the resolutions, as illustrated in the following key:

EXAMPLE

Key:
A if the statement supports the resolution.
B if the statement contradicts the resolution.
C if the statement neither supports nor contradicts the resolution.

If judging both the factual nature of a statement and its relevance is important, these two elements can be combined to form discrete categories as follows:

EXAMPLE

Key:
A if it is a statement of fact that supports the resolution.
B if it is a statement of opinion that supports the resolution.
C if it is a statement of fact that contradicts the resolution.
D if it is a statement of opinion that contradicts the resolution.

The main drawback to combining two types of judgment in one category is the greater complexity of the key. This is especially undesirable with younger students.

In mathematics, a key-type item that has been found to be quite efficient is the quantitative comparison item. The use of a fixed-response format for quantitative comparison items reduces the reading required and makes it possible for students to respond to a larger number of items in a given period of time:

EXAMPLE

Directions: Each of the following questions consists of two quantities, one in Column A and one in Column B. You are to compare the two quantities and choose:

Key:
A if the quantity in Column A is greater.
B if the quantity in Column B is greater.
C if the two quantities are equal.
D if the relationship cannot be determined from the information given.

				Column A	Column B
(A)	B	C	D	7/8	0.75
A	(B)	C	D	3/7	7/3
A	B	(C)	D	1.25	5/4
A	B	C	(D)	X/2	X//3

10. In constructing key-type test items, develop standard key categories where applicable. Despite the usefulness of the interpretive exercise for measuring complex achievement, classroom teachers have not used it extensively, often because of the difficulty of construction. The popularity of the key-type item in interpretive exercises is probably because it uses a common set of alternatives. This makes it easier to construct than the regular multiple-choice form, which requires a different set of alternatives for each item.

It is often possible to simplify further the construction of key-type interpretive exercises by preparing key categories that can be reused with different content. For example, a learning outcome such as the ability to recognize assumptions might lead to the following key:

EXAMPLE Key: A an assumption that is necessary to make the conclusion valid.
B an assumption that would invalidate the conclusion.
C an assumption that has no bearing on the validity of the conclusion.

This key could be used with a brief description of a situation, a conclusion based on the situation, and a list of assumptions. Both the key and the form of the item could be used repeatedly, with only the content varying. Although selecting new content material is still a problem, the framework of the standard key categories simplifies the process.

Standard key categories, of course, cannot be used in all areas, and their use should not be permitted to determine which learning outcomes receive emphasis. Rather, the time and effort saved by such procedures should free the teacher to explore more creative applications of the interpretive exercise in other areas. See the "Checklist" box for reviewing interpretive exercises.

CHECKLIST

Reviewing Interpretive Exercises

	Yes	No
1. Is this the most appropriate item format to use?	___	___
2. Is the material to be interpreted relevant to the intended learning outcomes?	___	___
3. Is the material to be interpreted appropriate to the students' curricular experience and reading level?	___	___
4. Have pictorial materials been used whenever appropriate?	___	___
5. Does the material to be interpreted contain some novelty (to require interpretation)?	___	___
6. Is the material to be interpreted brief, clear, and meaningful?	___	___
7. Are the test items based directly onthe introductory material (cannot be answered without it), and do they call for interpretation (not just recall or simple reading skills)?	___	___
8. Has a reasonable number of test items been used in each interpretive exercise?	___	___
9. Do the test items meet the relevant criteria of effective item writing?	___	___
10. When key-type items are used, are the categories homogeneous and mutually exclusive?	___	___
11. If revised, are the interpretive exercises still relevant to the intended learning outcomes?	___	___
12. Have the interpretive exercises been set aside for a time before reviewing them?	___	___

SUMMARY

Complex achievement refers to those learning outcomes based on higher mental processes. Such outcomes are classified under various general headings, including understanding, reasoning, thinking, and problem solving. The attainment of goals in these areas can be measured by both objective and subjective means. The most commonly used objective item is the interpretive exercise.

The interpretive exercise consists of a series of objective questions based on written materials, tables, charts, graphs, maps, or pictures. The questions require students to demonstrate the specific interpretive skill being measured. For example, students might be asked to recognize assumptions, inferences, conclusions, relationships, applications, and the like. The structure of the interpretive exercise makes it possible to obtain independent measures of each aspect of thinking and problem solving. Although it is efficient for measuring such learning outcomes, it does not measure a student's ability to integrate and use these skills in a global attack on a problem. Thus, it is limited to a diagnostic analysis of problem-solving skills.

Probably the main reason for not using the interpretive exercise is the difficulty of construction. This process involves (a) selecting appropriate introductory material, (b) revising the material to fit the outcomes to be measured, and (c) constructing a series of dependent test items that call forth the desired behavior. Although these steps are admittedly time consuming, the rewards in improved teaching–learning practices seem to justify the time and effort.

LEARNING EXERCISES

1. What are the advantages of the interpretive exercise over the performance-based assessment for measuring complex achievement? What are the disadvantages?
2. For which types of learning outcomes is the interpretive exercise most likely to be appropriate? Why?
3. Discuss the relative merits of the interpretive exercise and the single-item multiple-choice question. For which situation would each be most useful? What are the limitations of each?
4. Construct one interpretive exercise for each of the following:
   a. A paragraph of written material
   b. A picture or cartoon
   c. A chart or graph
5. What steps would you follow in examining an interpretive exercise to determine whether it had been properly constructed?
6. What are some of the factors to consider when you are deciding whether to use interpretive exercise in a classroom test?

FURTHER READING

Educational Testing Service. (1973). *Multiple-choice questions: A close look.* Princeton, NJ: Author. Illustrates the use of the multiple-choice item for measuring complex achievement in a variety of fields. Maps, graphs, pictures; diagrams, and written materials are used. Each item is followed by a statistical and logical analysis of its effectiveness.

Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of student achievement* (9th ed.). Upper Saddle River, NJ: Pearson. See Chapter 7, "Writing Selection Items: True-False, Matching, and Interpretive Exercises," for examples of interpretive exercises and their uses.

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). New York: Holt, Rinehart & Winston. See Chapter 7, "Writing Objective Test Items: Multiple-Choice and Context-Dependent," for sample interpretive exercises and suggestions for construction.

Wesman, A. G. (1971). Writing the test item. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 81–129). Washington, DC: American Council on Education. An extended treatment of the topic of item writing. See pages 120–128 for the construction of interpretive exercises.