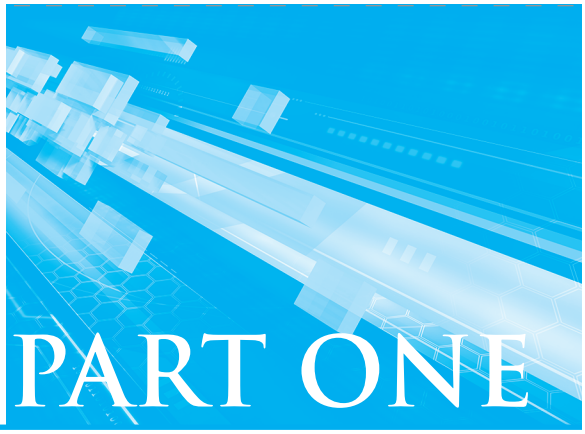
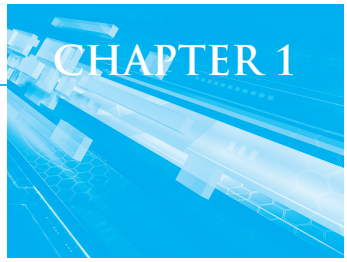


PART ONE

Introduction and Practical Aspects of Testing

CHAPTER 1



Introduction to Measurement and Evaluation

CHAPTER OUTLINE

KEY WORDS

age-appropriate physical activity
cardiovascular disease
coronary heart disease
criterion-referenced standard
evaluation
formative evaluation
health-related fitness
intermittent physical activity
measurement
moderate physical activity
norm-referenced standard
norms
obesity
objective
percentile rank
physical activity
prevalence
subjective
summative evaluation
vigorous physical activity

OBJECTIVES

Professions within the fields of kinesiology are constantly changing. Graduates of kinesiology programs are becoming not only teachers and coaches, but also exercise specialists, physical and occupational therapists, personal trainers, biomechanists, sport and exercise psychologists, and sport managers. Some are even starting or entering private business. The process of measurement and evaluation is an integral component of all of these professional efforts.

Obesity has reached epidemic proportions in the United States and worldwide. Regular physical activity can have a positive effect on obesity and other *cardiovascular disease* risk factors. Strong scientific evidence exists to demonstrate that even moderate increases in physical activity can substantially impact public health. Thus, it is becoming more and more important for students in kinesiology programs to understand the construct of physical activity and how to measure it. Kinesiology programs are adapting to societal needs by training students more thoroughly in the assessment and promotion of physical activity.

Although many factors influence professional preparation and many subfields exist within kinesiology (e.g., physical activity promotion, biomechanics, exercise physiology, physical education teaching, sport management, health and fitness, sport and exercise psychology, and motor learning, among others), certain highly influential research and historic reports should be understood by all students in the field. Forces in the dynamic environment in which we live (e.g., aging society, increasing numbers of overweight and obese people, more sedentary occupations) have led to the development of public health programs that improve health through promotion of physical activity and weight control. In 1996, the Surgeon General's Report on Physical Activity and Health was issued. This was a landmark document in the field of kinesiology. In 2008, the U.S. Department of Health and Human Services published the Physical Activity Guidelines for Americans. This was followed by the U.S. National Physical Activity Plan and the Toronto Charter for Physical Activity in 2010. These documents represent national and international commitments to a set of policies and programs designed to promote health-enhancing physical activity. These historic public health initiatives should be understood by kinesiology students because they document the evidence that physical activity improves many health outcomes, provide recommendations that multiple sectors of society can use to enable citizens to engage in sufficient physical activity, and highlight the issues that have made careers in kinesiology and physical activity promotion among the most important to the quality of human existence.

This chapter will help you understand the place of measurement and evaluation in our changing social and professional world.

After reading this chapter, you should be able to:

1. Define and differentiate between measurement and evaluation.
2. Define and differentiate between criterion- and norm-referenced standards.
3. Define and differentiate between formative and summative methods of evaluation.
4. Understand models of evaluation as they apply to teaching (K–12) and physical activity promotion settings.
5. Describe the role of public health initiatives in physical education and kinesiology.

6. Understand national and international physical activity guidelines.
7. Describe physical activity guidelines for children and adults.

Measurement and Evaluation

We tend to regard test results as a valid basis for decision making. Tests govern matters such as student promotion, college acceptance, and defining health-related levels of fitness. The terms *measurement* and *evaluation* are widely used, but often with little regard for their meanings. *Measurement* is the collection of information on which a decision is based; *evaluation* is the use of measurement in making decisions. This chapter clarifies these activities within the changing context of the field of kinesiology and introduces the procedures that have evolved to meet the challenges created by these changes.

Measurement and evaluation are interdependent concepts. Evaluation is a process that uses measurements, and the purpose of measurement is to collect information for evaluation. Tests are tools used to collect information and can be interpreted quite broadly. In the evaluation process, information is interpreted relative to established standards so that decisions can be made. Clearly, the success of evaluation depends on the quality of the data collected. If test results are not consistent (i.e., reliable) and truthful (i.e., valid), accurate evaluation is impossible. The measurement process is an important step in evaluation—improved measurement leads to more accurate evaluation.

People are different. They vary in behavior, size, shape, speed, strength, and many other respects. Measurement determines the degree to which an individual possesses a defined characteristic. It involves first defining the characteristic to be measured and then selecting the instrument with which to measure that characteristic (Ebel, 1973). Stopwatches, tape measures, written tests, skill tests, questionnaires, pedometers, activity monitors, skinfold calipers, treadmills, and cycle ergometers are common instruments used in the broad field of kinesiology to measure various constructs.

Test scores vary from highly *objective* to highly *subjective*. A test is objective when two or more people who score the same test assign similar scores. Tests that are highly objective are those that have a defined scoring system and are administered by trained testers. A multiple-choice written test, a stopwatch, skinfold calipers, and an ECG heart rate tracing all have a defined scoring system, but testers also need to be trained to secure objective measurements. For example, if percent body fat is to be measured by the skinfold method, the tester needs to be trained in the proper method of measuring a skinfold with a caliper. A highly subjective test lacks a standardized scoring system, which introduces a source of measurement error. We use objective measurements whenever possible because they are more reliable than subjective measurements.

Evaluation is a dynamic decision-making process that involves (1) collecting suitable data (measurement), (2) judging the value of these data according to some standard, and (3) making decisions based on these data. The function of evaluation is to facilitate good decisions. For the teacher, this can mean facilitating student learning; for the exercise specialist, this could mean helping someone establish scientifically sound weight-reduction goals.

Functions of Measurement and Evaluation

Too often tests are administered with no definite purpose in mind. The ultimate purpose of testing is to enhance the decision-making process so that improvements can be made. Testing can have many purposes. Six general purposes of testing that facilitate the decision-making process are presented next.

PLACEMENT

Tests can be used to place students in classes or groups according to their abilities. Adult fitness tests are used to determine a person's current status so that an individualized program can be prescribed.

DIAGNOSIS

Tests can be used to diagnose weaknesses. While placement usually involves the status of the individual relative to others, diagnosis is used to isolate specific deficiencies that make for a low or undesirable status. In K–12 settings, tests can identify areas where students need to make improvements. In health promotion settings, test results are used to diagnose problems. For example, a treadmill stress test is used to screen for heart disease. An athletic trainer may use tests to identify symptoms of concussion.

EVALUATION OF ACHIEVEMENT

One goal of testing is to determine whether individuals have reached important objectives. Placement, diagnosis, and the evaluation of achievement together form the basis of individualized instruction. In K–12 settings, this can be the achievement of instructional objectives. In health promotion settings, this can be meeting important goals or showing progress; for example, documenting changes made during rehabilitation.

PREDICTION

Test results can be used to predict an individual's level of achievement in future activities or to predict one measure from another. Prediction often seeks information on future achievement from a measure of present status, and this information may help students to select activities they are most likely to master. For example, an individual found to have a high aerobic capacity may decide to engage in road racing or become a triathlete. The measurement and evaluation process is also used to predict variables that are difficult to measure from variables that are more easily measured. For example, aerobic capacity (VO_{2max}), which is assessed during a maximal treadmill test using a metabolic measurement system, can be estimated from performance on a distance run test. Percent body fat, which is measured in research settings using expensive equipment, can be estimated from skinfolds that can be easily measured in field-based settings.

PROGRAM EVALUATION

Test results of participants can be used as one piece of evidence to evaluate programs. By comparing the results of tests for a school district against national norms or standards, or by comparing the yearly changes made within a school district, important decisions can be made. Comparing changes in fitness between tests can provide evidence of the effectiveness of fitness training programs.

MOTIVATION

Test scores can be motivating. Achievement of important standards can encourage one to achieve higher levels of performance or to participate regularly in physical activity.

Formative and Summative Evaluation

Summative evaluation is the judgment of achievement at the end of an instructional unit and typically involves the administration of tests at the conclusion of an instructional unit or training period. *Formative evaluation* is the judgment of

achievement during the formative stages of learning. Motor-learning research shows that feedback is one of the most powerful variables in learning. Formative evaluation is used to provide feedback to learners throughout the instructional process.

Formative evaluation begins during the early stages and continues throughout instruction. It involves dividing instruction into smaller units of learning and then evaluating the student's mastery of these subunits during instruction. The main purpose of formative evaluation is to determine the level of mastery achieved and to identify any parts of the task that have not yet been mastered (Bloom, Hastings, & Madaus, 1971). The strength of formative evaluation is that it is used to provide feedback throughout the instructional unit.

Summative evaluation is used to decide whether broad objectives have been achieved. Certification examinations provide an example of summative evaluation. Certification examinations are often designed to determine whether a person is qualified to perform a job at some acceptable level. The similarities and differences between formative and summative evaluation are summarized in **Table 1.1**.

Formative and summative evaluation and mastery learning were developed for use by classroom teachers (Bloom et al., 1971). However, the logic of the system can be applied to fitness promotion programs. Information from periodic testing to determine if fitness goals have been achieved can be used to provide feedback that facilitates motivation. A key element of a successful self-supervised fitness program for NASA executives was periodic fitness testing (Owen, Beard, Jackson, & Prior, 1980). Periodic measurement of body weight is a behavioral strategy used for weight-reduction programs (deBakey, Gotto, Scott, & Foreyt, 1984; Wing & Hill, 2001).

Both formative and summative evaluation processes can be used to contribute to successful fitness training programs. Increases in intensity and/or duration of aerobic exercise can be used to provide feedback (i.e., formative evaluation) to the participant that he or she is improving and can become a source of motivation to continue in the exercise program. A fitness test after training can be used for summative evaluation to judge whether fitness goals have been met or whether the training program was successful.

TABLE 1.1 Similarities and Differences Between Formative and Summative Evaluation

	Formative Evaluation	Summative Evaluation
Purpose	Feedback to student and teacher on student progress throughout an instructional unit	Certification or grading at the end of a unit, semester, or course
Time	During instruction	At the end of a unit, semester, or course
Emphasis in evaluation	Explicitly defined behaviors	Broader categories of behaviors or combinations of several specific behaviors
Standard	Criterion referenced	Norm- or criterion-referenced

You are encouraged to use the formative evaluation exercises provided at the end of the chapter. After you have read the chapter, attempt to answer the questions. If you cannot answer a question or if you feel unsure of your answer, this is an indication that you need additional work. The key element of formative evaluation is the feedback it provides; it communicates to the participant what has been and what still needs to be achieved. For this course, your instructor probably will administer several major tests that evaluate your ability to integrate and apply the readings. These would be considered summative evaluations.

Standards for Evaluation

As previously explained, evaluation is the process of giving meaning to a measurement by judging it against a standard. The two most widely used types of standards are criterion- and norm-referenced standards. A *criterion-referenced standard* is used to determine if someone has attained a specified level. A *norm-referenced standard* is used to judge an individual's performance in relation to the performances of other members of a well-defined group; for example, 11-year-old boys. Criterion-referenced standards are useful for setting performance standards for all, whereas norm-referenced standards are valuable for comparisons among individuals when the situation requires a degree of selectivity.

Criterion- and norm-referenced standards have application in a wide variety of settings. They are used extensively in K–12 educational settings and in exercise and public health settings. Youth fitness tests have generally evolved from the use of norm-referenced standards to the use of criterion-referenced standards. The FITNESSGRAM® is the national youth fitness test, which is part of the Presidential Youth Fitness Program. The FITNESSGRAM® provides criterion-referenced standards that can be used to determine whether a participant has achieved a level of fitness associated with health benefits. The YMCA program (Golding, Meyers, & Sinning, 1989) is used to evaluate adult fitness with norm-referenced standards, but the trend for fitness testing is to use criterion-referenced standards. The criterion-referenced fitness standards evolved from medical research (Blair et al., 1989, 1995) showing that the relationship between health and aerobic fitness is not linear. Once a level of aerobic fitness is achieved, becoming more fit has little influence on health. FITNESSGRAM® criterion-referenced standards for aerobic fitness and body composition were established by documenting the relationship between these fitness measures and prevalence of metabolic syndrome in youth (Laurson, Eisenmann, & Welk, 2011; Welk, Laurson, Eisenmann, & Cureton, 2011).

Norm-Referenced Standards

Norm-referenced standards are developed by testing a large number of people of a defined group. Descriptive statistics are then used to develop standards. A common norming method is the use of *percentile ranks*. A percentile rank norm reflects the percentage of the group that can be expected to score below a given value. For example, a 1-mile run time of 11:31 for a boy 11 years of age is at the 25th percentile; only 25% ran slower, while 75% of the 11-year-old boys could be expected to run faster.

The characteristics of the group on which the standards were developed are important to consider when norm-referenced standards are to be used. The *norm* does not always translate to a desirable level. For example, if the average cholesterol level of men aged 40–49 years is 214 mg/dL, then this average does not represent a desirable level. A cholesterol level of less than 200 mg/dL is considered a desirable level for health. In this instance, average would not be desirable

because it has been shown that there is a relationship between high levels of cholesterol and risk of *coronary heart disease* mortality (Anderson, Castelli, & Levy, 1987; Castelli et al., 1977; Wood et al., 1988).

Criterion-Referenced Standards

A criterion-referenced standard is a predetermined standard that can be used to determine if an individual has achieved a desired level of performance. It is unlike a norm-referenced standard in that the performance of the individual is not compared with that of other individuals; instead, the performance is compared against the standard. A norm-referenced evaluation can be considered a relative evaluation—evaluation relative to norms developed on other people. A criterion-referenced evaluation can be considered an absolute evaluation—evaluation by comparison to an absolute criterion.

Many authors use the term *criterion-referenced test*, suggesting that the difference is not just with the standard, but also with the method used to develop the test (Glaser & Nitko, 1971; Safrit, 1989). Glaser and Nitko (1971) define a criterion-referenced test as one developed to provide measurements that are directly interpretable in terms of explicit performance standards. Although some tests used in education were constructed to be criterion-referenced tests, the more common practice is to apply a criterion-referenced standard to a test that was originally developed as a norm-referenced test. For example, the mile run is a youth fitness test item designed to assess aerobic fitness. The mile run was previously used in a norm-referenced fashion by interpreting how well the performer compared to others of the same age and sex. Currently, the mile run is used in the FITNESSGRAM® in a criterion-referenced fashion to determine whether the performance met or did not meet the standard. In this instance, the test itself has not changed, but the type of standard used to evaluate aerobic fitness has changed.

DETERMINING ACCURACY OF CRITERION-REFERENCED STANDARDS

Unlike norm-referenced standards, which use continuous variables, criterion-referenced standards create categories into which participants are classified. These categories are often dichotomies (a dichotomy is a division into two parts). Terms such as pass–fail, fit–not fit, healthy–needs improvement, mastery–nonmastery, or proficient–nonproficient are used to describe the dichotomy. The validity of the criterion-referenced standard can be examined by using a 2 × 2 contingency table. The accuracy of the criterion-referenced standard is analyzed by comparing how a participant scored relative to the criterion-referenced standard and a criterion measure that represents the person's true state. This creates four possible options, which are illustrated in **Figure 1.1**. Methods are available to determine the quality of a criterion-referenced standard by estimating its reliability and validity.

Criterion-referenced test reliability examines the consistency of classification—the percentage of people consistently classified as passing or failing a test that has been administered two times. Criterion-referenced test validity refers to the accuracy of classification—the percentage of participants correctly classified by the test as passing or failing compared to their true state.

When using the type of 2 × 2 table in Figure 1.1, a positive test indicates the presence of whatever is being tested. In a clinical setting, a positive test for heart disease indicates the presence of heart disease. In a fitness test setting, a positive test for fitness indicates the presence of adequate fitness levels.

A negative test indicates the absence of the entity being tested. In a clinical setting, a negative test for heart disease indicates the absence of heart disease. In a fitness test setting, a negative test for fitness indicates inadequate fitness levels.

For youth fitness testing, a *true positive* occurs when a participant's performance on the fitness test achieves the criterion-referenced standard and that participant's true fitness is considered

FIGURE 1.1 A 2 × 2 table is used to determine the accuracy of a criterion-referenced test. Categorization by a criterion-referenced test can be incorrect in two ways: false negative and false positive.

		True State: Criterion Test	
		Negative	**Positive**
Criterion-Referenced Test	Negative	True negative (TN)	False negative (FN)
	Positive	False positive (FP)	True positive (TP)

True Negative: The criterion-referenced test correctly indicates failure to reach the criterion.
True Positive: The criterion-referenced test correctly indicates reaching the criterion.
False Negative: The criterion-referenced test incorrectly indicates failure to reach the criterion.
False Positive: The criterion-referenced test incorrectly indicates reaching the criterion.

Courtesy of CSI Software Company, Houston, TX.

adequate for health purposes. A *true negative* results when a person's performance on the fitness test does not reach the criterion-referenced standard and his or her true fitness is below a level adequate for health promotion.

A *false negative* test results when a student's performance on the fitness test does not achieve the criterion-referenced standard, when, in fact, the student has a true fitness level adequate for health. A *false positive* test results when a student's performance achieves the criterion, when in reality he or she does not have an adequate level of fitness for health purposes.

When a criterion-referenced interpretation is applied to an exercise stress test that examines the electrical activity of the heart, a positive test indicates the presence of heart disease. The true coronary disease state is defined by a cardiac catheterization test, which involves passing a catheter into the coronary artery. The flow of a dye is traced to identify the extent of coronary blockage, and the presence of heart disease is defined by the degree of blockage. Because the catheterization test is dangerous, an exercise stress test is used first. A false positive test results when the stress test shows the person has heart disease but the cardiac catheterization test shows the person is healthy. A false negative results when the stress test suggests that the person does not have heart disease but the cardiac catheterization shows that heart disease is present.

DEVELOPMENT OF HEALTH-RELATED CRITERION-REFERENCED STANDARDS

The main challenge of the criterion-referenced approach is the setting of an appropriate standard. It is often not possible to find a criterion that explicitly defines mastery. Assume, for example, that a physical education teacher wants a criterion for the mastering of volleyball skills. Tests of mastery of complex motor skills are typically not readily available. However, other situations exist where criterion-referenced standards can be easily set. For example, some skills, such as beginning swimming, lend themselves to the criterion-referenced approach. The successful execution of these defined skills can be clearly determined and judged. Within the field of kinesiology, the setting of criterion-referenced standards for health-related youth fitness tests has received substantial attention.

The FITNESSGRAM® health-related youth fitness test provides criterion-referenced standards for aerobic fitness and body composition. These

standards were validated by comparison with the prevalence of metabolic syndrome. A health outcome–centered method was used to set the criterion-referenced standards (Zhu, Welk, Going, & Cureton, 2011). The key steps of the health outcome–centered method are:

1. Determine the components of health-related fitness.
2. Select field tests and criterion measures of the components of health-related fitness.
3. Determine the relationships between the criterion measure of health-related fitness and appropriate health outcomes.
4. Set criterion-referenced standards (also called *cut-off scores*) based on the relationship between the criterion measures of health-related fitness and the health outcomes.
5. Cross-validate the criterion-referenced standards.

The following section summarizes this approach applied to aerobic fitness tests of the FITNESSGRAM®.

DEVELOPMENT OF CRITERION-REFERENCED STANDARDS FOR AEROBIC FITNESS

Aerobic fitness tests for the FITNESSGRAM® include the PACER test, the 1-mile run, and the walk test (Meredith & Welk, 2010). Different criterion-referenced standards were developed for boys and girls for different ages. Setting criterion-referenced standards for youth fitness tests is difficult because of the complexity introduced by the effects of growth and maturation and because health outcomes are not as clear as in adults. The steps used to establish the criterion-referenced standards were as follows:

1. The two main tests of aerobic fitness in the FITNESSGRAM® were identified as the PACER test and the 1-mile run.
2. The criterion measure of aerobic fitness was defined as maximal oxygen consumption (VO_{2max}). Multiple regression was used to convert performance on the PACER test and the 1-mile run to estimates of VO_{2max}.
3. Metabolic syndrome was used as the outcome variable to indicate health risk (defined as the presence of three or more clinically poor values for the following five risk factors: waist circumference, blood pressure, HDL-cholesterol, triglycerides, and fasting glucose). Data for all variables were derived from the National Health and Nutrition Examination Survey (NHANES), which provided a nationally representative dataset of 12–18 year olds.
4. Receiver operating characteristic (ROC) curves were used to establish the criterion-referenced standards or cut-off scores for VO_{2max}. The ROC curves allowed examination of sensitivity and specificity (along with other statistics) values for various cut-off scores used to identify the presence of metabolic syndrome. LMS curves (L = skewness, M = median, S = coefficient of variation) were used to account for growth and maturation.
5. Three zones of aerobic capacity were developed using various values of sensitivity and specificity, separately for boys and girls, to allow categorization of participants into the Healthy Fitness Zone, the Needs Improvement Zone, or the Needs Improvement–Health Risk Zone.

Although some degree of misclassification will always be present when using criterion-referenced standards, the approach used to develop FITNESSGRAM® standards was developed on a nationally representative sample using an empirically sound approach.

Models of Evaluation

The professional environments of a K–12 physical education teacher and an exercise specialist are very different, but the evaluation processes used are quite similar. Illustrated next are evaluation models commonly used in K–12 educational and adult fitness settings. The main difference between the two models is test selection.

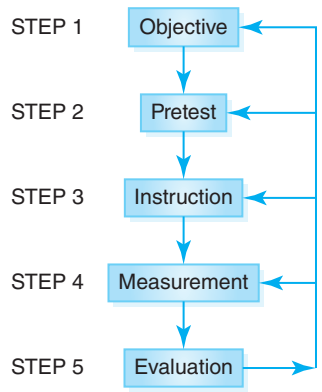
Educational Model

A primary purpose of teaching is to produce a measurable change in behavior. **Figure 1.2** is an evaluation model appropriate for use in K–12 educational settings. This dynamic model integrates measurement and evaluation with the instructional process. Each component of the model is briefly discussed next.

OBJECTIVE

Preparation of the objective is the first step in the evaluation process, because objectives determine what we will seek to achieve. The objective gives direction to instruction and defines the behaviors we want to change.

FIGURE 1.2 **A systematic model of evaluation suitable for K–12 educational settings.**



PRETEST

With some type of pretest, we can answer three questions: (1) How much has already been learned? (2) What are the individual's current capabilities and status? (3) What type of activity should be prescribed to help achieve the objectives? Pretesting does not necessarily involve administering the same test that will be given after instruction; it can include any form of measurement that helps to answer these questions. For example, observing students swim the width of a pool can be an effective pretest.

INSTRUCTION

Sound instructional methods are needed to achieve the agreed-upon objectives. Different instructional procedures may be needed to meet students' individual needs.

MEASUREMENT

This involves the selection or development of a test to gauge the achievement of the objectives. It is crucial that the test be designed to measure the behavior specified in the objectives. The objectives can be cognitive, affective, psychomotor, or fitness related. The key element is to select or develop a test that measures the objective. Often, teachers will need to develop their own tests, because standardized tests may not be consistent with their instructional objectives. Content evidence of validity is achieved when the test is congruent with the instructional objectives. This is a common method used to provide evidence of test validity in educational settings.

EVALUATION

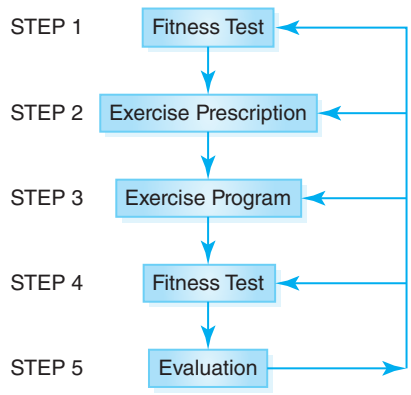
Once the instructional phase has been completed and achievement has been measured, test results are judged (i.e., evaluated) to determine whether the desired changes achieved the stated objective.

What happens if students do not achieve the desired objective? Figure 1.2 shows a feedback loop from evaluation back to each component of the model. Failure to achieve the stated objective may be due to any segment of the model. It may be discovered that the objectives are not appropriate and may need to be altered, or that the instruction may not have been suitable for the group, or that the selected test may not have been appropriate. The educational evaluation model is dynamic. The evaluation model provides information needed to alter any aspect of the educational process.

Fitness Assessment Model

The exercise specialist and the teacher have a common goal: to produce changes in fitness. This may occur in many different settings, such as hospitals, physical therapy clinics, or fitness centers. Personal trainers depend a great deal on the ability to demonstrate improvements in their clients. **Figure 1.3** shows how the evaluation model works in exercise settings.

FIGURE 1.3 **A systematic model of evaluation applied to fitness assessment.**



Courtesy of CSI Software Company, Houston, TX.

FITNESS TEST

The first step in the development of an individualized fitness program is a fitness test. This may consist of more than one component. For some, medical clearance may be needed before the person can start a fitness program. The American College of Sports Medicine has published guidelines (ACSM, 2014) for exercise testing and prescription. Once a person is cleared to engage in an exercise program, he or she is given a fitness assessment.

EXERCISE PRESCRIPTION

Once the person's fitness level and goals are known, an exercise prescription is developed. The purpose of the exercise prescription is to define the individual's fitness needs and to develop exercise parameters that are consistent with scientific research (ACSM, 2011, 2014). For example, the goal may be to exercise aerobically at 60% of heart rate reserve.

EXERCISE PROGRAM

Once the exercise parameters are known, the next step is to develop the person's exercise program. This is based on the initial fitness assessment, the exercise prescription, and the person's interests and goals. For example, a program for weight loss would be different from one designed to rehabilitate a knee injury.

FITNESS TEST

Once the program has been completed, a second fitness assessment is administered. The tests used will likely be the same as those used for the initial fitness assessment.

EVALUATION

Once the training program has been completed and the fitness parameters of interest have been measured, test results are judged (i.e., evaluated) to determine whether the desired changes have been made. Either norm- or criterion-referenced standards can be used to evaluate performance.

Like the educational model, the fitness evaluation model is dynamic. The model shows a loop from evaluation back to other components of the model. Once changes in fitness are made, the exercise prescription parameters also change. Other feedback loops in this evaluation model suggest that sometimes the evaluation process can lead an exercise specialist to judge that the fitness tests that were originally chosen are not sensitive enough to document changes in fitness. If this occurs, other, more sensitive, fitness tests can adopted.

Influential Research and Historic Reports

Influential research in kinesiology and public health set the stage for understanding the importance of physically active lifestyles for health promotion. Many careers in kinesiology involve helping people become more physically active or move more effectively. Physical educators try to teach students the knowledge and skills to lead physically active lifestyles. Health fitness specialists design exercise prescriptions to help clients adhere to exercise programs that increase fitness. Physical therapists evaluate components of movement and develop treatment programs to restore physical function. Public health specialists provide health education and promotion programs to help individuals and communities maintain healthy lifestyles. Promoting physical activity is one thing that all of these professions have in common. The following sections will describe landmark research and historic reports with which all professionals in kinesiology should be familiar.

Occupational Physical Activity and Coronary Heart Disease

In the 1950s, Dr. Jeremy Morris pioneered the study of physical activity and cardiovascular disease (Morris & Crawford, 1958; Morris et al., 1953). The original approach used to study the relationship between physical activity and health was to compare sedentary individuals with those who were more physically active. Physical activity was characterized through basic classifications related to occupation (e.g., active bus conductors compared to inactive bus drivers). Heart disease rates were compared between workers who held more physically active jobs and those who had sedentary jobs. The general conclusion from these studies was that individuals who had the most physically demanding jobs suffered fewer fatal heart attacks than their sedentary counterparts. For example, conductors who walked up and down the stairs of double-decker buses in London had fewer heart attacks than the more sedentary bus drivers (Morris et al., 1953).

Morris not only demonstrated the health benefits of movement (in the bus conductors), but also the health risks of sedentariness (in the bus drivers who sat for prolonged periods). Recent research on the negative health effects of sedentary behavior, independent of the health benefits of physical activity, suggests that public health initiatives need to focus on both promoting physical activity and limiting sedentary behavior.

Another classic study about the relationship between occupational physical activity and heart disease was conducted by researchers from the University of Minnesota (Taylor et al., 1962). They studied more than 191,000 American railroad workers. Because of union rules and benefits, railroad workers had excellent medical records, which provided the data for the study. In addition, union rules discouraged shifting from one occupation class to another. A 55-year-old person with 20 years of service was likely to have spent all 20 years at the same job.

The occupational groups studied were clerks, switchmen, and section men. The clerks represented men in jobs requiring little physical activity; the work of the section men was the most physically demanding; and the work demands of the switchmen were moderate. Death rates ascribed to coronary heart disease demonstrated that the most physically active workers (section men) had the

lowest heart disease rate, whereas the least physically active (clerks) had the highest heart disease rate. The heart disease rate for the switchmen fell between the two extremes. These data suggested that physical activity reduced the risk of heart disease.

An accepted method of quantifying physical activity is by caloric expenditure (ACSM, 2014). Paffenbarger, Hale, Brand, and Hyde (1977) studied the role of caloric expenditure on fatal heart attacks of San Francisco longshoremen. This was an ideal group to study because most workers stayed in the same job throughout their working lives, and the amount of energy expended to perform a job could be determined. The study included a 22-year history of more than 3500 longshoremen. Workers were divided into three general categories based on the energy expenditure of their jobs. The job categories were as follows:

- High caloric expenditure: 5.0–7.5 kcal per minute
- Intermediate caloric expenditure: 2.4–5.0 kcal per minute
- Light caloric expenditure: 1.5–2.0 kcal per minute

The results of the study demonstrated that the longshoremen who expended the highest level of energy had the lowest rate of fatal heart attacks. Most impressive was the difference found between the active and inactive workers who had a history of diagnosed heart disease. The heart attack rate of the sedentary workers with a history of heart disease was about double the rate found for the physically active workers with previous heart problems. These data showed that energy expenditure provided a margin of protection from heart disease for these longshoremen.

Harvard Alumni Studies

The Harvard Alumni studies were surveys of the health and physical activity of nearly 17,000 Harvard alumni. Questionnaire data were used to quantify physical activity in terms of caloric expenditure. The forms of physical activity included various types of sports, stair climbing, and walking. The researchers showed that caloric expenditure was related to heart attack rate (Paffenbarger, Hyde, Wing, & Steinmetz, 1984) and all-cause mortality (Paffenbarger, Hyde, Wing, & Hsieh, 1986).

Harvard alumni who consistently exercised during their lifetimes had lower heart attack and mortality rates than their sedentary classmates. Walking regularly, climbing stairs, and playing either light or vigorous sports provided health benefits. The total amount of energy expended through all forms of exercise was highly related to heart disease and mortality rates. As total caloric expenditure increased, heart disease and mortality rates moved progressively lower. The highest heart disease and mortality rates were found in alumni who expended fewer than 500 kcal per week through physical activity. Heart disease and mortality rates dropped steadily as caloric expenditure increased up to about 2000 kcal per week expended through physical activity and then leveled off. Paffenbarger et al. (1984) also demonstrated that being a college athlete did not reduce risk of cardiovascular disease unless former athletes remained physically active after leaving college. The alumni at highest risk of a heart attack were the former athletes who were not physically active after college.

The key conclusion of the Harvard Alumni studies was that physical activity is a major determinant of public health. This was examined by calculating what public health researchers call community-attributable risk, an estimate of the potential reduction of heart attacks in the population if the risk factor were not present. This calculation considers the prevalence of the risk factor in the population. *Prevalence* in this context refers to the percentage of people in the group who have the risk factor. The higher the prevalence of the risk factor, the greater the potential improvement in public health if the risk factor is eliminated.

Surgeon General's Report on Physical Activity and Health

Regular physical activity and exercise are critical elements of health promotion. Increased physical activity is associated with a reduced incidence of coronary heart disease, hypertension, noninsulin-dependent diabetes mellitus, colon cancer, depression, and anxiety. *Physical activity* is a complex behavior, and its relationship with health is multifaceted. Physical activity is defined as any bodily movement produced by skeletal muscles that results in energy expenditure (Caspersen, Powell, & Christenson, 1985).

In 1996, the Surgeon General's Report on Physical Activity and Health was published (U.S. Department of Health and Human Services, 1996). The major purpose of the report was to summarize the existing literature on the role of physical activity in preventing disease and on the status of interventions to increase physical activity. This review led to eight major conclusions (U.S. Department of Health and Human Services, 1996, p. 4):

1. People of all ages, both male and female, benefit from regular physical activity.
2. Significant health benefits can be obtained by a moderate amount of physical activity on most, if not all, days of the week. Through a modest increase in daily activity, most Americans can improve their health and quality of life.
3. Additional health benefits can be gained through greater amounts of physical activity. People who can maintain a regular regimen of activity that is of longer duration or of more vigorous intensity are likely to derive greater benefit.
4. Physical activity reduces the risk of premature mortality in general, and of coronary heart disease, hypertension, colon cancer, and diabetes mellitus in particular. Physical activity also improves mental health and is important for the health of muscles, bones, and joints.
5. More than 60% of American adults are not regularly physically active. In fact, 25% of all adults are not active at all.
6. Nearly half of American youths 12–21 years of age are not vigorously active on a regular basis. Moreover, physical activity declines dramatically during adolescence.
7. Daily enrollment in physical education classes has declined among high school students from 42% in 1991 to 25% in 1995.
8. Research on understanding and promoting physical activity is at an early stage, but some interventions to promote physical activity through schools, worksites, and healthcare settings have been evaluated and found to be successful.

The Surgeon General's report gave a clear signal that physical inactivity is a major health risk and physical activity measurement research increased substantially after this historic publication.

Physical Activity Guidelines for Americans

Several national and international organizations have put forth physical activity guidelines, including the World Health Organization (WHO), the ACSM, the American Heart Association (AHA), and the U.S. Department for Health and Human Services. The physical activity guidelines from these organizations have many similarities. The physical activity guidelines from the U.S. Department of Health and Human Services (2008) for older adults, adults, and children follow.

PHYSICAL ACTIVITY RECOMMENDATIONS FOR HEALTH FOR ADULTS (AGED 18–64 YEARS)

- 2 hours and 30 minutes (150 minutes) a week of moderate-intensity aerobic

physical activity, or 1 hour and 15 minutes (75 minutes) a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic physical activity.
- Aerobic activity should be performed in episodes of at least 10 minutes, preferably spread throughout the week.
- Additional health benefits are provided by increasing to 5 hours (300 minutes) a week of moderate-intensity aerobic physical activity, or 2 hours and 30 minutes a week of vigorous-intensity physical activity, or an equivalent combination of both.
- Adults should also perform muscle-strengthening activities that involve all major muscle groups on 2 or more days per week.

PHYSICAL ACTIVITY RECOMMENDATIONS FOR HEALTH FOR OLDER ADULTS (AGED 65 YEARS AND OLDER)

- Older adults should follow the adult guidelines. If this is not possible due to limiting chronic conditions, older adults should be as physically active as their abilities allow. They should avoid inactivity. Older adults should do exercises that maintain or improve balance if they are at risk of falling.

PHYSICAL ACTIVITY RECOMMENDATIONS FOR HEALTH FOR CHILDREN AND ADOLESCENTS (AGED 6–17 YEARS)

- 1 hour (60 minutes) or more of physical activity every day. Most of the 1 hour or more a day should be either moderate- or vigorous-intensity aerobic physical activity.
- As part of their daily physical activity, children and adolescents should do vigorous-intensity activity on at least 3 days per week. They also should do muscle-strengthening and bone-strengthening activity on at least 3 days per week.

National Physical Activity Guidelines for Children

Charles B. (Chuck) Corbin has been one of the most influential people in the area of youth fitness and physical activity of the past half-century. His research in the late 1960s was among the first to assess fatness among youth using skinfold calipers (Corbin, 1969; Corbin & Pletcher, 1968). Additional work showed the capacity of youth to do vigorous physical activity in an era when distance runs for children were limited to 600 yards. Corbin and his colleagues were influential in changing the focus of youth fitness from skill-related to *health-related fitness* and provided the basis for many of the criterion-referenced health-related fitness standards in use today.

Corbin and Pangrazi (1998, 2004) authored the national physical activity guidelines for children. These guidelines are as follows:

> **Guideline 1:** Children should accumulate at least 60 minutes, and up to several hours, of *age-appropriate physical activity* on all, or most, days of the week. This daily accumulation should include *moderate* and *vigorous physical activity*, the majority of which is intermittent in nature. Age-appropriate physical activity is activity of a frequency, intensity, duration, and type that leads to optimal growth and development in children and contributes to development of a physically active lifestyle. *Intermittent physical activity* refers to relatively short bursts of movement (several seconds or minutes) interspersed with periods of rest. Moderate physical activity is defined as

an intensity equivalent to brisk walking. Moderate-intensity activities can be performed for relatively long periods. Vigorous physical activity is defined as movement that expends more energy than brisk walking. Figure 1.4 shows a computerized metabolic system assessing the exercise intensity for a youth.

Guideline 2: Children should participate in several bouts of physical activity lasting 15 minutes or more each day. It is typical for children that these bouts of activity will include both physical activity and time for rest and recovery. This recognizes that the physical activity needs of children are different from those of adults.

Guideline 3: Children should participate each day in a variety of age-appropriate physical activities designed to achieve optimal health, wellness, fitness, and performance benefits.

Guideline 4: Extended periods (periods of 2 hours or more) of inactivity are discouraged for children, especially during the daytime hours. This guideline recognizes that children should be active when opportunities to be active are available. Optimally, such opportunities would occur before school, after school, at appropriate times during school, and on weekends.

FIGURE 1.4 **Exercise intensity and oxygen uptake in youth can be assessed with computerized metabolic systems.**

Summary

Measurement uses tests with evidence of reliability and validity to secure the data essential to the evaluation process. Evaluation is a decision-making process with a goal of improvement. Tests can be used in six general ways: (1) placing students in homogeneous groups to facilitate instruction, (2) diagnosing weaknesses, (3) determining whether important objectives have been reached, (4) predicting performance levels in future activities, (5) comparing a program with others like it, and (6) motivating participants. The evaluation models appropriate for K–12 physical education teachers are quite similar to those used to evaluate adult fitness.

Evaluation is a means of determining whether objectives are being reached and of facilitating achievement. Formative evaluation clarifies what remains to be achieved; summative evaluation determines whether general objectives have been reached. Evaluation, then, is the feedback system by which the quality of the instructional or training process is monitored. Criterion-referenced standards specify the level of performance

necessary to achieve specific instructional objectives; norm-referenced standards identify a level of achievement within a group of individuals.

Although activity levels of youth and adults are less than desirable, the importance of physical activity, exercise, and fitness for health has been well documented. Physical inactivity increases the risk of premature mortality in general, and of coronary heart disease, high blood pressure, colon cancer, and type 2 diabetes. Significant health benefits can be obtained from a moderate amount of daily physical activity. Demographic trends in America are increasing the need for exercise and weight-control programs. Our population is getting older, and jobs are becoming more sedentary in nature.

Physical activity guidelines for adults generally suggest that adults should participate in moderate- and vigorous-intensity physical activity every week, with activity bouts of at least 10 minutes, preferably spread throughout the week. It is also noted that doing more physical activity than the amount that is minimally recommended will result in additional health benefits. Muscle-strengthening activity is also recommended.

Physical activity guidelines for children emphasize that children should accumulate 60 minutes or more of physical activity on all, or most, days of the week. The recognition that children are not little adults has led to physical activity recommendations for children that are different from those for adults. Children have unique needs and behavioral characteristics. Children will be physically active intermittently, interspersed with periods of rest. Children should participate in both moderate- and vigorous-intensity physical activity and should not be inactive for extended periods.

FORMATIVE EVALUATION OF OBJECTIVES

Objective 1 Define and differentiate between measurement and evaluation.

1. The terms *measurement* and *evaluation* often are used interchangeably. Define each term.
2. What are the key differences between measurement and evaluation?
3. Although measurement and evaluation are distinct functions, explain how they are interrelated.

Objective 2 Define and differentiate between criterion- and norm-referenced standards.

1. What are the key differences between criterion- and norm-referenced standards?
2. Explain how a physical education teacher or exercise specialist could use both types of standards.

Objective 3 Define and differentiate between formative and summative methods of evaluation.

1. Many believe that greater achievement is possible if both formative and summative evaluation procedures are used compared to the use of one type of evaluation by itself. Briefly describe formative and summative evaluation.
2. What are the key differences between formative and summative evaluation? Why could you expect to stimulate greater achievement by using both formative and summative evaluation?

(*continues*)

Objective 4 Understand models of evaluation as they apply to teaching (K–12) and physical activity promotion settings.

1. What are the steps of the K–12 evaluation model?
2. What are the steps of the fitness assessment evaluation model?

Objective 5 Describe the role of public health initiatives in physical education and kinesiology.

1. What are the major conclusions of the Surgeon General's Report on Physical Activity and Health?
2. Describe physical activity guidelines for adults and older adults.
3. Describe physical activity guidelines for children.
4. List characteristics of children's physical activity behavior.
5. How is children's physical activity different from physical activity of adults?

ADDITIONAL LEARNING ACTIVITIES

1. Visit a school physical education program and discover the types of fitness tests being used.
2. Visit a facility that conducts an adult fitness program. Identify the types of tests that are administered and the types of programs offered.
3. Visit a school physical education program. Analyze the program and determine the students' level of physical activity.
4. Develop a physical education unit of instruction and show how you could use both formative and summative evaluation procedures.
5. Develop a fitness program and show how you could use both formative and summative evaluation procedures.
6. Find or suggest criterion-referenced standards for the tests you would use for a physical education program, fitness training program, or any other type of program used to promote physical activity. Explain your logic and sources of data used to establish the standards.

Bibliography

American College of Sports Medicine. (2011). Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine and Science in Sports and Exercise*, *43*, 1334–1359.

American College of Sports Medicine. (2014). *ACSM's guidelines for exercise testing and prescription* (9th ed.). Philadelphia: Lippincott Williams & Wilkins.

Anderson, K. V., Castelli, W. P., & Levy, D. (1987). Cholesterol and mortality: 30 years of follow-up from the Framingham study. *Journal of the American Medical Association, 257*, 2176–2180.

Blair, S. N., Kohl, H. W., Barlow, C. E., Paffenbarger, R. S., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality: A prospective study of healthy and unhealthy men. *Journal of the American Medical Association*, *273*, 1093–1098.

Blair, S. N., Kohl, H. W., Paffenbarger, R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality: A prospective study of healthy men and women. *Journal of the American Medical Association*, *262*, 2395–2401.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports*, *100*, 126–131.

Castelli, W. P., Doyle, J. T. Gordon, T., Hames, C. G., Hjortland, M. C., Hulley, S. B., . . . Zukel, W. J. (1977). HDL-cholesterol and other lipids in coronary heart disease. The cooperative lipoprotein phenotyping study. *Circulation*, *55*, 767–772.

Corbin, C. B. (1969). Standards of subcutaneous fat applied to percentile norms for elementary school children. *American Journal of Clinical Nutrition*, *22*, 836–841.

Corbin C. B., & Pangrazi, R. P. (2004). *Guidelines for appropriate physical activity for elementary school children. 2003 update. A position statement. Council for Physical Education for Children (COPEC) of the National Association for Sport and Physical Education (NASPE)*. Reston, VA: NASPE.

Corbin, C., & Pangrazi, R. (1998). *Physical activity for children: A statement of guidelines for children ages 5–12. National Association for Sport and Physical Education (NASPE)*. Reston, VA: NASPE.

Corbin, C. B., & Pletcher, P. (1968). Diet and physical activity patterns of obese and nonobese elementary school children. *Research Quarterly*, *39*, 922–928.

deBakey, M. F., Gotto, A., Scott, L. W., & Foreyt, J. P. (1984). *The living heart diet*. New York: Simon and Schuster.

Ebel, R. L. (1973). *Measuring educational achievement*. Englewood Cliffs, NJ: Prentice-Hall.

Glaser, R., & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 625–670). Washington, DC: American Council on Education.

Golding, L. A., Meyers, C. R., & Sinning, W. E. (1989). *The Y's way to physical fitness* (3rd ed.). Chicago: National Board of YMCA.

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011). Development of youth percent body fat standards using receiver operating characteristic curves. *American Journal of Preventive Medicine*, *41*, S93–S99.

Meredith, M. D., & Welk, G. J. (2010). *FITNESSGRAM® & ACTIVITYGRAM® test administration manual* (updated 4th ed). Champaign, IL: Human Kinetics.

Morris, J. N., Heady, J. A., Raffle, P. A. B., Roberts, C. G., & Parks, J. W. (1953). Coronary heart-disease and physical activity of work. *Lancet*, *2*, 1053–1057, 1111–1120.

Morris, J. N., & Crawford, M. D. (1958). Coronary heart disease and physical activity of work. *British Medical Journal*, *2*(5111), 1485–1496.

Owen, C. A., Beard, E. F., Jackson, A. S., & Prior, B. W. (1980). Longitudinal evaluation of an exercise prescription intervention program with periodic ergometric testing: A ten-year appraisal. *Journal of Occupational Medicine*, *22*, 235–240.

Paffenbarger, R. S., Hale, W. E., Brand, R. J., & Hyde, R. T. (1977). Work-energy level, personal characteristics, and fatal heart attack: A birth cohort effect. *American Journal of Epidemiology*, *105*, 200–213.

Paffenbarger, R. S., Hyde, R. T., Wing, A. L., & Hsieh, C. C. (1986). Physical activity, all-cause mortality, and longevity of college alumni. *New England Journal of Medicine*, *314*, 605–613.

Paffenbarger, R. S., Hyde, R. T., Wing, A. L., & Steinmetz, C. H. (1984). A natural history of athleticism and cardiovascular health. *Journal of the American Medical Association*, *252*, 491–495.

Safrit, M. J. (1989). Criterion-referenced measurement: Validity. In M. J. Safrit and T. M. Wood (Eds.), *Measurement concepts in physical education and exercise science* (pp. 119–135). Champaign, IL: Human Kinetics.

Taylor, H. L., Klepetar, E., Keys, A., Parlin, W., Blackburn, H., & Puchner, T. (1962). Death rates among physically active and sedentary employees of the railroad

industry. *American Journal of Public Health*, *52*, 1692–1707.

U.S. Department of Health and Human Services. (1996). *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center of Chronic Disease Prevention and Health Promotion.

U.S. Department of Health and Human Services. (2008). *2008 physical activity guidelines for Americans*. Washington, DC: U.S. Department of Health and Human Services.

Welk, G. J., Laurson, K. R., Eisenmann, J. C., & Cureton, K. J. (2011). Development of youth aerobic-capacity standards using receiver operating characteristic curves. *American Journal of Preventive Medicine*, *41*, S111–S116.

Wing, R. R., & Hill, J. O. (2001). Successful weight loss maintenance. *Annual Review of Nutrition*, *21*, 323–341.

Wood, P. D., Stefanick, M. L., Dreon, D. M., Frey-Hewitt, B., Garay, S. C., Williams, P., . . . Haskell, W. L. (1988). Changes in plasma lipids and lipoproteins in overweight men during weight loss through dieting as compared with exercise. *New England Journal of Medicine*, *319*, 1173–1179.

Zhu, W., Mahar, M. T., Welk, G. J., Going, S. B., & Cureton, K. J. (2011). Approaches for development of criterion-referenced standards in health-related youth fitness tests. *American Journal of Preventive Medicine*, *41*(Suppl. 2), S68–S76.