

EIGHTH EDITION

# MEASUREMENT FOR EVALUATION

*IN PHYSICAL EDUCATION &  
EXERCISE SCIENCE*

Ted A. Baumgartner  
Andrew S. Jackson  
Matthew T. Mahar  
David A. Rowe

# 7

## AUTHENTIC AND ALTERNATIVE ASSESSMENT

*A. Barry Joyner Georgia Southern University*

*Steve Elliott University of North Carolina at Wilmington*

### C O N T E N T S

Key Words	140
Objectives	141
Introduction	141
Characteristics of Authentic Assessments	141
Rubrics	142
Types of Authentic and Alternative Assessments	146
Measurement Concerns with Authentic and Alternative Assessments	151
Summary	153
Formative Evaluation of Objectives	153
Additional Learning Activities	153
Bibliography	153

### K E Y W O R D S

alternative assessment  
authentic assessment  
rubric

## OBJECTIVES

The purpose of this chapter is to introduce the concepts of authentic and alternative assessments in physical education. Although these assessments are being used more often in physical education, reliability and validity issues have not always been addressed.

After reading Chapter 7, you should be able to

1. Differentiate between authentic and alternative assessments.
2. Identify the characteristics of authentic assessments.
3. Define the different types of scoring rubrics.
4. Identify the different types of authentic and alternative assessments.
5. Discuss measurement concerns with authentic and alternative assessments.

## INTRODUCTION

The educational reform movement of the 1990s and the passing of the No Child Left Behind Act of 2001 (NCLB; Public Law 107-110) have resulted in assessment and accountability becoming a hot topic in all educational subject areas, including physical education. Traditionally, physical educators have measured student learning through cognitive exams, skill tests, and fitness evaluations. Resulting from dissatisfaction with traditional testing methods and the need for accountability within education, newer **alternative assessment** methods have been incorporated into the teaching and learning process. Alternative assessments are “untraditional” and often involve allowing students to create a product that the teacher will have to grade (e.g., drawing a picture to show correct kicking technique, making a video of a particular skill). Another type of assessment that has become increasingly popular is **authentic assessment**. Authentic assessment can include traditional testing formats (i.e., cognitive tests and qualitative skill tests) and alternative assessments; however, authentic assessments differ in that they are conducted in a “real-life” or authentic context. Authentic assessment requires students to demonstrate meaningful comprehension and application

of essential knowledge and skills in a dynamic and realistic environment.

Proponents of alternative and authentic assessment methods have criticized traditional assessment methods in physical education that place students in a contrived setting for evaluation. For example, some skill tests place the student in situations that do not resemble those in which the skill would actually be used. How often in tennis are players responding to balls thrown to them rather than hit to them? Does hitting volleyball forearm passes against a wall as many times as possible in 30 seconds demonstrate an ability to use the forearm pass in a game? Are these authentic situations?

From a motor-learning perspective, authentic assessment addresses the issue of transfer of learning. Student learning of motor skills is best measured through a transfer test, or retention test (Magill 2004). One of the main goals of physical education is to help develop the “physically educated person” (NASPE 2004) who learns the skills necessary to participate in lifetime activities. Using the volleyball example presented, the physical education teacher should be concerned with whether students can transfer the skill of serving a volleyball to a game situation. An authentic assessment of the volleyball serve would allow the physical educator to assess this transfer of learning.

## Characteristics of Authentic Assessments

Lund (1997) has identified the following six characteristics that are present in most authentic assessments.

1. *Authentic assessments present challenges that are representative of real life.* How many times in a real-life situation are students going to be asked to take a multiple-choice test about tennis rules? Wouldn't it be more meaningful for the assessment to mirror what the students will do in real life? Assessing them within a game where they apply the rules, strategies, and skills needed may be more appropriate.
2. *Authentic assessments require students to demonstrate higher-level thinking.* Authentic assessments allow students the opportunity to

apply the concepts they have learned. After testing the students for their knowledge of the rules and strategies, authentic assessment could be used to determine whether the students understand how to apply them in a real-life situation.

3. *Students know the standards for assessment from the beginning.* Authentic assessment involves setting up scoring **rubrics** or standards that serve as the criteria for the assessment. By knowing the criteria in advance, students will constantly be receiving feedback about their progress (Lund 1994). This is not unique to authentic assessment. Many physical educators routinely do this with skill and fitness testing.
4. *Authentic assessments are part of the curriculum.* When using more authentic assessments, teachers essentially teach to the test (Lund 1994). Although some may consider this to be undesirable, a goal of testing should be that the test matches the real-life situation.
5. *Students often present the culmination of the authentic assessment publicly.* This may help to stimulate pride in their work and show the students that the material has meaning. The presentations might be at a PTA meeting, for other classes or other schools, or for the class itself.
6. *There is an emphasis on process and not just product.* How students arrive at the correct answer is just as important as the answer. By focusing solely on the product, students could be rewarded for using poor technique. In a basketball game, a student could have a terrible technique yet score a lot of points because he is taller than everyone else.

## RUBRICS

Many physical education teachers have used skill tests to qualitatively assess student motor performance. This involves a student performing a skill a predetermined number of times while the teacher uses a checklist to record mastery or nonmastery of

critical elements of the skill. For example, the soccer kick may be broken down into four components (eyes on ball, nonkicking foot placed next to ball, contact with laces, and follow-through), which are assessed during each attempt in the skill test. When students are measured at the beginning and end of a unit, this skill test approach can provide a good indication of student learning that took place during an instructional unit. Teachers have favored this controlled skill test approach because it is relatively objective and can be standardized for all children. In contrast, one early criticism of authentic assessment measures was that motor skill tests were hard to administer and grade when performed in a dynamic, gamelike situation due to the lack of standardization. However, most authentic assessments now utilize grading rubrics that allow a teacher to qualitatively assess a student's performance in various contexts.

A rubric is a scoring guide that provides an outline of the guidelines for assessing student performance (Schiemer 2000). Rubrics allow the teacher to compare the student's performance to the criteria specified in various levels of the rubric. Rubrics have various levels for assessing performance, but may range from three (outstanding, acceptable, deficient) to five (excellent, good, satisfactory, fair, poor) levels.

Four main types of rubrics are used in physical education; checklists, point system rubric, holistic rubrics, and analytical rubrics (Lund & Kirk 2002). *Checklist rubrics* provide a list of characteristics or behaviors that the teacher records as being either present or not present (Lund & Kirk 2002). Checklists are useful for helping students understand the critical elements of a motor skill or for acknowledging completion of certain parts of performance or activity. Checklist rubrics provide a yes-no rating on a list of characteristics but fail to capture or describe the quality of the characteristics. An example of a checklist rubric designed to assess student performance within an indoor floor hockey unit is shown in Table 7.1.

*Point system rubrics* are very similar to checklist rubrics in that the teachers are recording the characteristics or behavior as being present or absent. However, they add a point value to each characteristic or behavior that is being assessed. For

**TABLE 7.1** A Checklist Rubric to Assess Student Performance within an Indoor Floor Hockey Unit

Yes	No	
X		Follows teacher's directions
X		Follows rules of the game
	X	Plays safely and never raises stick above waist height
	X	Cooperates with teammates
X		Uses correct technique during 5 v 5 games
X		Shows good sportsmanship at end of games

example, perhaps the teacher who designed the checklist for indoor floor hockey in Table 7.1 would like to place more of an assessment emphasis on the items that assess correct technique, following the rules of the game, and cooperation with teammates. To achieve this, the checklist could be converted to a point system rubric whereby the student is awarded a point value instead of a yes-no judgment. A point system rubric does not require the observer to address the quality of a certain aspect of a performance, but it does allow a greater emphasis to be put on certain characteristics. The addition of a point value allows the teacher to easily convert the point value to a grade for the assignment (see Table 7.2).

*Holistic rubrics* assess the student's motor performance as a whole without breaking the skill down into critical components. This type of rubric requires the teacher to make a judgment about the quality of all criteria simultaneously. The performance levels within a holistic rubric are usually presented in paragraph form. An example of a holistic rubric designed to assess student performance on the skill of ultimate frisbee game play is provided in Table 7.3. Once the holistic rubric has been constructed, the teacher can use a checklist to assess each student's performance in the authentic situation (see Table 7.4). By designing a rubric and a simple checklist, the physical educator can assess students while watching them perform ultimate

**TABLE 7.2** A Point Value Rubric to Assess Student Performance within an Indoor Floor Hockey Unit

1	Follows teacher's directions (1 point)
2	Follows rules of the game (2 points)
0	Plays safely and never raises stick above waist height (1 point)
0	Cooperates with teammates (2 points)
3	Uses correct technique during 5 v 5 games (3 points)
1	Shows good sportsmanship at end of games (1 point)
7	Total

## Grading Scale

10	A	Excellent
8-9	B	Good
6-7	C	Satisfactory
4-5	D	Unsatisfactory
0-3	F	Poor

frisbee skills in a game situation. This type of assessment is becoming increasingly popular in units taught through the sport education model approach (Siedentop, Hastie & van der Mars 2004).

In contrast to looking at the skill performance as a whole, *analytical rubrics* allow the teacher to assess the student's level of performance on each critical component of a motor skill performance. Although analytical rubrics require more detail and are generally more time-consuming to construct, they allow the teacher to pinpoint areas of weakness within the skill performance, allowing for individualized feedback to the student. Several physical education pedagogy textbooks (e.g., Graham 2001; Lund 2000; Schiemer 2000) and NASPE resources (e.g., *Authentic Assessment of Physical Activity for High School Students* 2002; *Creating Rubrics for Physical Education* 1999) specifically address the issue of rubric creation. For example, to help teachers develop analytical rubrics, Schiemer (2000) has provided specific guidelines that can streamline the rubric development process. First, she recommends creating a grid using either graph paper or a computer software program (e.g., Microsoft

**TABLE 7.3 A Holistic Rubric for Assessing Ultimate Frisbee Game Play**

3 - Outstanding	Student uses appropriate throws in different situations and consistently executes with proper technique. Throws are accurate and directed toward teammates in advantageous positions. Movement is consistently made to get open and provide passing outlets for teammates. Student provides strong defense within the defensive system of the team and is aware of teammates' positions, opposition runs, and the location of the disk. Student demonstrates a thorough understanding of the rules of the game and calls decisions fairly within the "spirit of ultimate Frisbee."
2 - Satisfactory	Although some incorrect-form performances are evident, throws are usually performed with correct form and technique, and appropriate decision making is usually demonstrated. Participation shows an understanding of rules and strategies of ultimate frisbee. Strong defensive skills are evident in individual situations although the student does not consistently engage in team-oriented defense.
1 - Unsatisfactory	Student performance is characterized by throws made with incorrect form. Decisions are poorly made in terms of shot selection, and minimal movement is made to provide passing opportunities for teammates. Consistent rule infringements demonstrate a lack of understanding of the rules of the game. Defensive skills are weak as demonstrated by poor footwork, anticipation, and individual and team defensive skills.

**TABLE 7.4 Checklist to Assess Ultimate Frisbee Skills in Game Play**

	Outstanding	Satisfactory	Unsatisfactory
John	X		
Thomas		X	
Jennifer	X		
Jim			X
Debbie	X		

Excel). Once this template has been constructed, it can be adjusted to score each skill the teacher wishes to assess. Second, the critical components that the teacher plans to assess should be listed down the left-hand column of the grid. Ideally, these critical components are directly related to the cues the teacher provided to the students during the instructional unit. Third, the different levels of performance should be listed across the top row of the grid. Schiemer (2000) recommends that the levels of performance match the reporting system that the teacher uses on the student reporting card, such as percentages, pass/fail, letter grades, or O, S, NI (outstanding, satisfactory, needs improvement). Fourth, each inner cell within the grid should con-

tain a qualitative statement of the criteria necessary to demonstrate competence at each of the designated performance levels. The constructed rubric should be made available to the students who are to be assessed. Finally, the teacher can create a checklist to record student performance while observing the students in a game situation. Tables 7.5 and 7.6 provide examples of an analytical rubric and a student checklist constructed to assess performance on the soccer kick.

Standard Five from the National Standards for Physical Education document urges teachers to design lessons that enable students to "exhibit responsible personal and social behavior that reflects self and others in physical activity settings"

**TABLE 7.5 Analytical Rubric for Assessing the Soccer Kick in 5-a-Side Game**

	Achieving (A)	Developing (D)	Not Yet (X)
Eyes on ball throughout kick	Eyes focused on ball throughout each kick attempt	Eyes focused on ball: (a) for most kicks or (b) for most of the time throughout each kick	Eyes never focused on ball throughout each kick
Nonkicking foot placed next to ball	Nonkicking foot placed next to ball during each kick attempt	Nonkicking foot usually placed beside ball during kicking attempts	Nonkicking foot never placed next to ball during kicking attempts
Contact with laces or instep	Contact always made with the laces or the instep of the foot	Contact sometimes made with the laces or the instep of the foot	Contact rarely or never made with the laces or the instep of the foot
Follow-through with kicking leg in direction of ball	Follow-through always made in the direction of the kick	Follow-through sometimes made in the direction of the kick	Follow-through never made in the direction of the kick

**TABLE 7.6 Checklist to Assess Soccer Kick in 5-a-Side Game**

	Eyes	Foot Placement	Contact	Follow-through
Jane	A	A	A	A
Steven	A	A	D	D
Alex	D	D	D	X
Jose	A	X	X	X
Crystal	D	A	D	D

**TABLE 7.7 Rubric for Assessing Student Sportsmanship during Basketball Game Play**

Cooperated with teammates	Always	Often	Sometimes	Rarely	Never
Followed referee's decisions	Always	Often	Sometimes	Rarely	Never
Played fairly	Always	Often	Sometimes	Rarely	Never
Reacted appropriately to adversity	Always	Often	Sometimes	Rarely	Never
Played with energy and enthusiasm	Always	Often	Sometimes	Rarely	Never

(NASPE 2004, p. 39). This standard is often reflected in affective objectives such as social behavior, sportsmanship, and cooperation skills. In addition to psychomotor objectives, affective domain objectives can also be evaluated in an authentic situation through the use of a rubric. Tables 7.7 and 7.8 provide examples of a grading rubric and student checklist that could be used either as a self-assessment for students to complete

or as an assessment for the teacher to complete while observing students in game play activities.

Today in many school districts, physical educators are responsible not only for assessing students, but also for reporting student scores to a district supervisor or central office. Teachers have had to collect student data using a pencil-and-paper approach in the gymnasium and then return to their office to manually input the scores into a

**TABLE 7.8** Checklist for Assessing Student Sportsmanship during Basketball Game Play

	Cooperated with teammates	Followed referee's decisions	Played fairly	Reacted appropriately to adversity	Played with energy and enthusiasm
Marissa	A	A	O	S	A
Juanita	A	A	A	A	A
Mario	O	O	O	O	O
Daniel	R	N	N	N	S
Will	S	S	O	A	A

A = Always

O = Often

S = Sometimes

R = Rarely

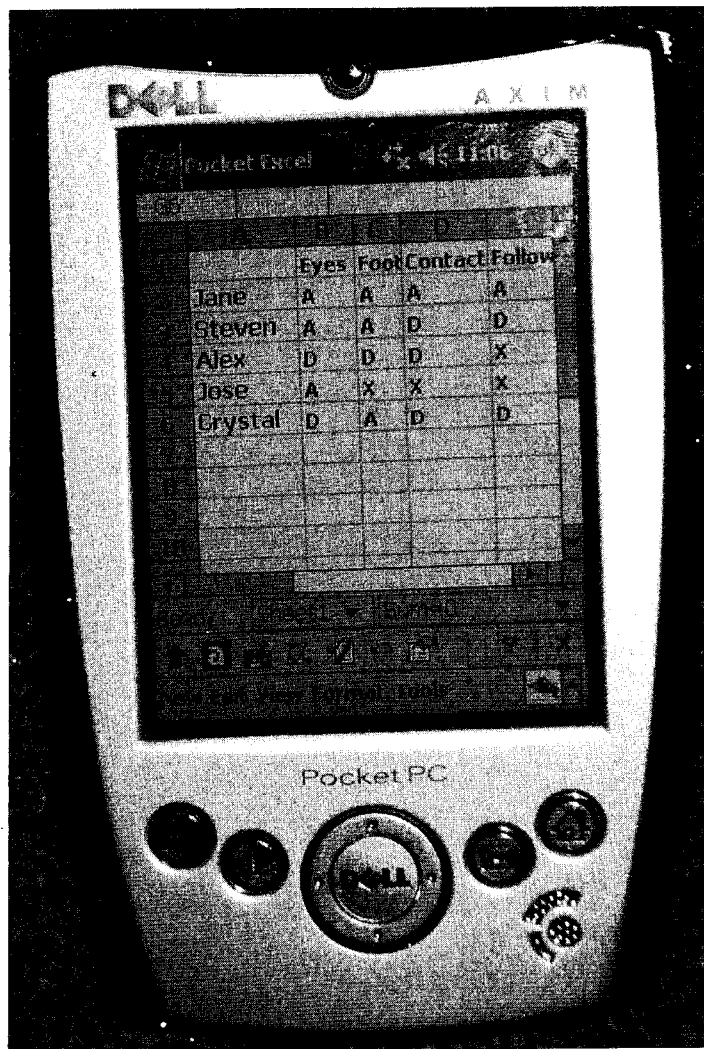
N = Never

spreadsheet on their desktop computer. This approach to assessment can be made more efficient through the use of technology, specifically a Pocket PC. A Pocket PC is a handheld computer that has many of the capabilities of modern desktop PCs. The common software bond between the desktop PC and the Pocket PC is the Excel spreadsheet software. Desktop Excel and Pocket Excel share many of the same functions, such as the sorting of data, summing functions, and the arrangement of data in columns. The obvious difference between Excel and Pocket Excel concerns the limited screen real estate on the Pocket PC. Rubrics and checklists are created using the Excel software program on the desktop computer and then saved to the Pocket PC. The teacher then uses the stylus to input data on the Pocket PC while circulating the class and watching students perform skills (see Figure 7-1). The final step in the process requires teachers to transfer their data back to the desktop computer to save in a folder or send to a central location (see Figure 7-2).

## TYPES OF AUTHENTIC AND ALTERNATIVE ASSESSMENTS

In 1995, NASPE published the first national document identifying what students should be learning in K–12 physical education classes. A newly revised second edition of *Moving into the Future:*

*National Standards for Physical Education* was released in 2004 and has been recognized as an essential tool for developing, implementing, and evaluating K–12 physical education programs. The 2004 National Standards clearly identify what students should know and be able to do as a result of a quality physical education program. The National Standards documents present assessment guidelines that encourage physical educators to instructionally align national content standards with clearly defined student expectations and performance outcomes. Furthermore, these documents present various alternative and authentic assessment options to measure student performance outcomes. Examples include student projects, student logs, student journals, peer observation, self-assessment, group projects, portfolios, event tasks, student artwork, video productions, and slide shows, as well as teacher observation (see Table 7.9 on page 149 for a list of alternative assessments). Although these examples are generally considered alternative assessments, they become authentic assessments when applied in real-life situations. Mintah (2003) surveyed 210 public school physical education teachers and found that over 75% of the teachers used authentic assessments. Some of the more commonly used authentic assessments were teacher observations, self-observations, checklists, and peer observations. Also, the teachers perceived a positive impact of the use of authentic



**Figure 7-1**

Checklist to assess soccer kick using a Pocket PC.

assessments on student self-concept, motivation, and skill achievement.

Portfolios, or collections of student work, have received much recent attention. The increased use of portfolios to assess student performance in physical education has been described by Lund and Kirk (2002) who stated, "With the current widespread emphasis on the use of performance-based, continuous, and authentic assessment of student work to determine their level of achievement of targeted goals and standards, the portfolio

has emerged as a new, exciting, and broadly used form of alternative assessment in the middle and secondary schools" (p. 97). NASPE (1995) includes hints for the development and use of portfolios. One suggestion is that the portfolio not include all examples of student work: the teacher specifies a certain number of pieces, and the student selects the pieces to include, submitting a cover letter explaining why those pieces were chosen. A scoring rubric should be included outlining the goals the teacher would like the student to meet

	A	B	C	D	E	F	G	H
1		Eyes	Foot	Contact	Follow			
2	Jane	A	A	A	A			
3	Steven	A	A	D	D			
4	Alex	D	D	D	X			
5	Jose	A	X	X	X			
6	Crystal	D	A	D	D			
7								
8								

**Figure 7-2**

Checklist to assess soccer kick saved from Pocket PC to desktop PC.

without limiting the creativity of the student. Kirk (1997) has developed sample portfolio tasks and a sample rubric for use in evaluating portfolios. Having the students include all possible information would make grading the portfolios impractical. Also, in some situations, physical educators may have too many students to make grading portfolios feasible.

Lund and Kirk (2002) and Melagrano (2000) provide examples of different types of portfolios, such as working portfolios (individual student's collections of projects, assignments, and class work), evaluation portfolios (finished portfolios that are submitted for evaluation), thematic portfolios (portfolios based on a specific unit of study), multiyear portfolios (collections of materials to show growth over a number of years), group portfolios (materials from groups within the class), and electronic portfolios (student work collected in an electronic format). Additionally, Lund and Kirk (2002) describe seven guidelines that the teacher should pay attention to when implementing a portfolio grading system in her physical education program (see Table 7.10).

Student portfolios have the advantage of allowing the teacher to triangulate numerous sources of information when engaging in the assessment process. One approach is to instruct the students on what items need to be placed into the

student portfolio. Possible items include student essays, stories, drawings, reflections, self-analysis of performance, peer analysis of performance, learning logs, personal fitness logs, activity plans, checklists of cues associated with skills, and opinion pieces (Hopple 1997). Providing a rubric and grade value for each assessment will make the assessing and grading procedure more transparent.

Two excellent examples of the use of portfolios in physical education can be found in *Moving into the Future: National Standards for Physical Education* by NASPE (1995). The first example is a fitness unit involving student assessment of the fitness levels of teachers and staff members at the school, development of individualized exercise programs, instruction for individuals on how to perform the activities, and monitoring of their progress (NASPE 1995). Students develop a portfolio of materials gathered throughout the project and are given feedback continuously. A scoring rubric would have been developed and distributed prior to the assignment so that students would know what criteria they must meet to succeed. The second example is an event task for gymnastics where the students have to develop a routine to perform at a halftime show for local basketball teams. The routine would be based on the gymnastics skills learned in class, and a scoring rubric has

**TABLE 7.9** Examples of Alternative and Authentic Assessments

Assessment	Description	Example
Slide Shows	Digital photos embedded within a PowerPoint presentation	Images are shown of students using correct technique to perform motor skills.
Video Production	Digital or analog video production of student motor performance	Students could be assigned to groups and asked to make an instructional video explaining and demonstrating the differences between man-to-man defense and zonal defense.
Poster / Banner Creation	An assignment that requires students to create a poster or banner that demonstrates their knowledge of content	Following a health-related physical fitness unit students could be instructed to create a poster illustrating the benefits of being physically fit.
Essays / Reports	A written summary or reflection on an event or concept	During a basketball unit students must watch a basketball game on television and describe and explain the fouls called during a 5-minute span.
Research Papers	An assignment that requires students to actively find information on an issue related to physical education	As part of a multicultural unit, students are assigned a country and are challenged to find one sport / dance / activity that originated in that country.
Personal Fitness Logs	A written personal summary of physical activity over a given time period	As part of a fitness unit, students must complete the Activity Log function of FITNESSGRAM® 8.0.
Opinion Pieces	A written reaction to an issue, concept, or other event	Students enrolled in a team sports class are required to write an opinion piece on different tactical approaches.
Dramatical Performances	Student-designed performances or skits	To develop skills in a cooperative learning unit, students are arranged into groups of five and instructed to create a situation in which a team handles losing well and a situation where the team does not handle losing well.
Officiating Games	Grading students on their knowledge of rules through their ability to officiate a game	In a sport education soccer unit, students take turns officiating the 5-a-side games. Officials are graded by the teacher and their peers through a specifically created rubric.
Oral Reports	Student presentations on an assigned topic	Students are assessed on the oral delivery of their research papers.
Worksheets	Assessing student knowledge of fundamental movement concepts	Images of a girl walking, jumping, and galloping are shown to assess whether a first-grade student can differentiate between these locomotor skills. The student circles the correct response.
Interviews / Focus Groups	Discussions with student(s) to evaluate knowledge	In a team sports unit, the teacher meets with each team to discuss why they implemented certain tactics and strategies.

**TABLE 7.10 Guidelines for Portfolio-Based Assessment in Physical Education**

1. Identify the learner outcomes that are to be demonstrated through the portfolio assessment.
2. Develop and communicate portfolio guidelines to students.
3. Build flexibility into the class schedule.
4. Provide a variety of performance-based learning and assessment opportunities for students.
5. Guide students in the creation of portfolio ideas.
6. Provide class time for students to work on their portfolios.
7. Provide opportunities for students to share or showcase their portfolios.

Source: Lund and Kirk (2002, p. 105).

**TABLE 7.11 Assessment Rubric for Outdoor Education Assignment**

	Excellent (A)	Good (B)	Satisfactory (C)	Unsatisfactory (D)
Use of Compass	Always uses correct technique	Mostly uses correct technique	Sometimes uses correct technique	Rarely uses correct technique
Cooperation Skills	Cooperates and communicates effectively with group all of the time	Cooperates and communicates effectively with group most of the time	Cooperates and communicates effectively with group some of the time	Cooperates and communicates ineffectively with group
Tent Pitching Skills	Uses correct procedure throughout task	Uses correct procedure throughout most of task	Uses correct procedure throughout some of the task	Uses incorrect procedure throughout task
Fire Building Skills	Fire is built safely and efficiently	Fire is built safely and somewhat efficiently	Fire is built safely	Fire is built unsafely
Hot Chocolate Making Skills	Served at perfect temperature, contains no ashes, and is delicious	Served at acceptable temperature, contains no ashes, and is tasty	Served at acceptable temperature, contains some ashes, and is drinkable	Served at incorrect temperature, contains lots of ashes, and is undrinkable

been included for evaluation. This event task could be easily adapted for dance and other activities.

Smith (1997) advocates using a portfolio card for authentic assessment. He points out some of the drawbacks of portfolios previously mentioned and suggests using a portfolio card to document student achievement for each grade level. He has included an example portfolio card and scoring rubric and relates the assessment to national outcomes. This would seem to be a useful way to document student performance and could be adapted to fit different situations.

Event tasks are those the students could complete in one class period, are written so that more than one possible solution can be presented, and are tasks that simulate a real-world experience (NASPE 1995). A good example of an event task comes from Whelan (1997) as a culmination to an outdoor adventure unit. Students are given a compass and directions; using orienteering skills, they have to determine the correct place to pitch their tents. Students must find their way from a beginning point to a predetermined finishing point and pitch their tents

within a certain distance of that point. After successfully setting up the tent, the students have to build a fire and make the teacher a cup of hot chocolate without getting ashes in it. These tasks are to be completed within the class period. An example rubric for this assignment is provided in Table 7.11.

Scheimer (2000) published a book of alternative and authentic assessment strategies for elementary physical education. The book includes 58 ready-to-use, reproducible assessment sheets that physical educators can use to assess students' cognitive, psychomotor, and affective learning. All the sheets include teacher tips designed for implementation of the assessment strategies. Hopple (2005) wrote a textbook to help teachers understand, develop, and use alternative methods of assessment to assess program effectiveness and students' progress toward understanding and acquisition of movement and fitness concepts and skills. Additionally, this book addresses the design and use of rubrics for specific assessment tasks. Other examples of alternative assessment are available at the PE Central website ([www.pecentral.org/assessment/assessment.html](http://www.pecentral.org/assessment/assessment.html)). Alternative assessments for golf, tennis, motor skills, and fitness are available there. Also available are links to other websites containing assessments.

## MEASUREMENT CONCERNS WITH AUTHENTIC AND ALTERNATIVE ASSESSMENTS

Although authentic assessment seems attractive as a better way to evaluate students in physical education, it is not without potential problems. Within physical education, little attention has been paid to the psychometric properties (validity, reliability, and objectivity) of these assessments. Also, the question of how teachers can use authentic assessments for grading in physical education needs to be addressed.

Concern has been expressed about the quality (i.e., validity, reliability, and objectivity) of authentic assessments. Some authors have conveyed the idea that unless the assessments are used for high-stakes accountability, the quality of the assessments may not be as great a concern (Lund 1994). Grading would seem to involve high-stakes

accountability. Many teachers currently use teacher-made tests without regard for reliability and validity; therefore, many may use authentic assessments the same way. However, ignoring validity and reliability in one situation does not justify ignoring it in another situation.

The validity of authentic assessments can be defined as the accuracy of the interpretation of the assessment results. According to Safrit and Wood (1995), validation of authentic assessments has typically been demonstrated through a logical approach or face validity. In other words, does the assessment "look" like it is measuring the desired behavior? However, a logical approach does not provide enough evidence to say that an assessment is valid and would not be acceptable for high-stakes accountability (Burger & Burger 1994).

Elliot (1995) discusses three concerns when considering the validity of authentic assessment. The first deals with how well the test relates to other measures. This is synonymous with a criterion approach. An authentic assessment of volleyball skill should be highly correlated with another measure of volleyball skill. A second concern is the ability of the assessment to predict future performance, which is also a criterion approach for obtaining validity evidence. For example, can an authentic assessment dealing with fitness predict future fitness behaviors? The third issue concerns the assessment covering the content domain. Does the assessment cover all areas of the activity? If you are concerned with overall softball skill, is the assessment reflective of all the components of this domain?

Baker, O'Neill, and Linn (1993) identified several characteristics that good authentic assessments should possess. These include the following: having meaning for both students and teachers, serving as motivation for performance, evaluating attributes that are important to both the teachers and students, requiring demonstration of complex cognition, exemplifying current standards of content quality, minimizing the effects of irrelevant skills, and possessing explicit standards for rating or judgment.

The reliability and objectivity of the authentic assessment are dependent on the scoring rubric developed. A detailed rubric and practice scoring with that rubric can enhance reliability and objectivity by

increasing the chances that the scores a teacher assigns the students one day will be similar to the scores that would be assigned on another day or by another teacher. One issue that needs to be addressed within the scope of reliability of authentic assessments is the combination of the tasks. Reliability has usually been established for a single task that was independent of other tasks. With authentic assessments, teachers are usually concerned with a combination of tasks in a real-world setting where various single tasks are part of a set of tasks necessary to achieve a goal. Repeated evidence demonstrating that this combination of tasks is present would help increase the reliability of the assessment.

According to Frisbie (1988), reliability of scores from teacher-made knowledge tests, including alternative assessments, is usually lower than reliability of scores from standardized tests. Frisbie goes on to state that tests with low reliability can be acceptable if they are combined with other information for evaluation and that teachers should be more concerned with the reliability of scores that are a combination of all information rather than individual assessments. If the teacher has only a small number of items on which to evaluate, these must be of high quality. For example, if a teacher is assigning a grade solely on the basis of two skill tests, these must be of high quality. However, if a teacher were to base a grade on two skill tests, two written tests, and two event tasks, the reliability and validity for this combination of items would be of more concern.

Few studies have examined the reliability of authentic assessment scores and the validity of the interpretations of these scores in physical education. Grehaigne, Godbout, and Bouthier (1997) developed an authentic assessment technique for assessing individual performance in a team sport. They found acceptable objectivity, reliability, and criterion validity evidence. Oslin, Mitchell, and Griffin (1998) developed the Game Performance Assessment Instrument (GPAI) to examine tactical understanding and problem solving. The GPAI was field-tested across soccer, basketball, volleyball, and softball games. The results of their study showed that the stability reliability and the objectivity of the scores from the GPAI were acceptable. Also, there was acceptable evidence

for content and construct validity. Cucina (1999) found acceptable objectivity across multiple raters, acceptable stability reliability, and acceptable criterion validity evidence for a badminton game assessment rubric. She correlated the score from the rubric with subjective ratings of playing ability. In 2001, Kulinna and Zhu used Rasch modeling to support the construct validity of fitness portfolios. Williams and Rink (2003) found objectivity was higher when teachers used only a two-level rubric (competent or not competent) compared to four levels that discriminated among levels of competence.

A grade in physical education should reflect whether or not the student has met the objectives of the program. If an objective is for the student to develop skill in a particular activity, teacher observation, event tasks, and student performance logs could be used. To determine if students are meeting cognitive objectives, teacher observation, portfolios, event tasks, group projects, and other authentic assessments could be used. Many different authentic assessments could be used to determine whether affective objectives are being achieved. The use of authentic assessment should not mean the complete elimination of traditional means of evaluation. There may be times when a written test or skill test may be appropriate and should be utilized. Also, when used in conjunction with authentic assessment, more traditional assessments may help to improve the reliability and validity of the final grade given.

Many physical education teachers lament the amount of time needed to grade their students using skill tests or fitness tests. It is unclear whether the use of authentic assessments would result in more or less time spent on grading. Most physical educators spend at least part of each skill unit allowing the students to play the game. By making observations during this time with a rating scale that yields objective and reliable scores and for which the interpretations of the scores are valid, the teacher could grade most, if not all, students in one class period. On the other hand, if a physical educator has 40 students in a class and six classes a day, is it feasible to use portfolios for grading those students? This would seem to work better for classroom teachers who only have 25 to

30 students in a class. In physical education, the number of students may deter the use of portfo-

lios, which could be more time-consuming than traditional methods.

## SUMMARY

Authentic and alternative assessments are defined and characteristics of authentic assessments are discussed. Four different types of scoring rubrics—checklists, point system rubrics, holistic rubrics, and analytical

rubrics—are discussed in detail, and examples of each are given. Examples of the different types of authentic and alternative assessments are given. Measurement concerns with authentic and alternative assessments are examined.

## FORMATIVE EVALUATION OF OBJECTIVES

*Objective 1* Differentiate between authentic and alternative assessments.

1. What is the difference between authentic and alternative assessment?
2. When does an alternative assessment become authentic?

*Objective 2* Identify the characteristics of authentic assessments.

1. Traditionally, written and skill tests have been used in physical education. Discuss how each of these traditional methods would align with the characteristics in this chapter.

*Objective 3* Define the different types of scoring rubrics.

1. How do checklist rubrics and point system rubrics differ?
2. How do holistic rubrics and analytical rubrics differ?
3. How can technology be used to collect student assessment data?

*Objective 4* Identify the different types of authentic and alternative assessments.

1. List and describe four different types of alternative assessments in physical education.
2. Provide examples of how different types of alternative assessments can be utilized in authentic settings.

*Objective 5* Discuss measurement concerns with authentic and alternative assessments.

1. How could you determine the objectivity of the rubric in Table 7.3?
2. Using the example from Whelan (1997), how could you determine the validity of this assessment?

## ADDITIONAL LEARNING ACTIVITIES

1. Using the information provided in this chapter, develop a holistic rubric for an activity of your choice. Have several of your classmates perform the activity and have three of your classmates rate their performance using the rubric. How good is the agreement among the raters? How could you improve the agreement?
2. From the list in Table 7.9, select three different types of alternative assessments and develop assessments of throwing ability.

## BIBLIOGRAPHY

Baker, E. L., H. F. O'Neill, Jr., and R. L. Linn. (1993). Policy and validity prospects for performance-based assessments. *American Psychologist* 48: 1210–1218.

Burger, S. E. and D. L. Burger. (1994). Determining the validity of performance-based assessment. *Educational Measurement: Issues and Practice* 13: 9–15.

- Cucina, I. M. (1999). *Specificity of feedback using alternative assessment techniques in a secondary physical education badminton class*. Unpublished doctoral dissertation, Springfield College, Springfield, MA.
- Elliot, S. N. (1995). *Creating meaningful performance assessments* (Report No. EDO-EC-94-2). Reston, VA: Council for Exceptional Children. (ERIC Document Reproduction Service No. ED 375 566)
- Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. *Educational Measurement: Issues and Practices* 79(1): 25-35.
- Graham, G. (2001). *Teaching children physical education: Becoming a master teacher*. 2nd ed. Champaign, IL.: Human Kinetics.
- Grehaigue, J. F., P. Godbout, and D. Bouthier. (1997). Performance assessment in team sports. *Journal of Teaching in Physical Education* 16: 500-516.
- Hopple, C. J. (1997). Authentic skills assessment in physical education. *Journal of Physical Education, Recreation, and Dance* 68(7): 19-24.
- Hopple, C. J. (2005). *Elementary Physical Education Teaching & Assessment*. Champaign, IL.: Human Kinetics.
- Kirk, M. F. (1997). Using portfolios to enhance student learning and assessment. *Journal of Physical Education, Recreation, and Dance* 68(7): 29-33.
- Kulinna, P. H. and W. Zhu. (2001). Fitness portfolio calibration for first- through sixth-grade children. *Research Quarterly for Exercise and Sport* 72(4): 324-334.
- Lund, J. (1994). Authentic assessment: Have we finally found user friendly assessment? *Proceedings of the World Congress for the Association Internationale des Ecoles Superieures d'Education Physique*. Berlin, Germany.
- Lund, J. (1997, January). What is authentic assessment? *Measurement News* 2:3.
- Lund, J. (2000). *Creating rubrics for physical education*. Reston, VA: American Alliance for Health, Physical Education, Recreation, and Dance.
- Lund, J. and M. F. Kirk. (2002). *Performance based assessment for middle and high school physical education*. Champaign, IL: Human Kinetics.
- Magill, R. A. (2004). *Motor learning and control: Concepts and applications*. 7th ed. New York: McGraw-Hill.
- Melograno, V. J. (2000). *Portfolio Assessment for K-12 Physical Education*. Reston, VA: National Association for Sport and Physical Education.
- Mintah, J. K. (2003). Authentic assessment in physical education: Prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement. *Measurement in Physical Education and Exercise Science* 7(3): 161-174.
- National Association for Sport and Physical Education. (1995). *Moving into the future: National standards for physical education*. Reston, VA: National Association for Sport and Physical Education.
- National Association for Sport and Physical Education. (1999). *Creating rubrics for physical education*. Reston, VA: National Association for Sport and Physical Education.
- National Association for Sport and Physical Education. (2002). *Authentic assessment of physical activity for high school students*. Reston, VA: National Association for Sport and Physical Education.
- National Association for Sport and Physical Education. (2004). *Moving into the future: National standards for physical education*. 2nd ed. Reston, VA: National Association for Sport and Physical Education.
- Oslin, J. L., S. A. Mitchell, and L. L. Griffin. (1998). The Game Performance Assessment Inventory (GPAI): Development and preliminary validation. *Journal of Teaching in Physical Education* 17: 231-243.
- Safrit, M. J. and T. M. Wood. (1995). *Introduction to measurement in physical education and exercise science*. 3rd ed. New York: Mosby.
- Schiemer, S. (2000). *Assessment strategies for elementary physical education*. Champaign, IL. Human Kinetics.
- Siedentop, D., P. A. Hastie, and H. van der Mars. (2004). *Complete guide to sport education*. Champaign, IL.: Human Kinetics.
- Smith, T. K. (1997). Authentic assessment: Using a portfolio card in physical education. *Journal of Physical Education, Recreation, and Dance* 68: 46-52.
- Whelan, J. (1997, March). *Authentic adventure*. Paper presented at the meeting of the American Alliance for Health, Physical Education, Recreation, and Dance, St. Louis, MO.
- Williams, L. and J. Rink. (2003). Teacher competency using observational scoring rubrics. *Journal of Teaching in Physical Education* 22: 552-572.