Overview of Authentic Assessment

By Dr Jacalyn Lea Lund

Guiding Questions

1. What are authentic assessments?
2. When should authentic assessments be used to evaluate learning?
3. What are some types of authentic assessments?
4. What are some characteristics that describe authentic assessments?
5. What advantages do authentic assessments offer when used?
6. What are some limitations of using authentic assessments?

Introduction

In our world, we are often required to demonstrate some degree of competence. Every legal driver in the United States is required to pass a test and reach a criterion score on it. Next time you have your hair cut, look on the shop’s wall; your stylist should have a license to cut hair. In physical education, lifeguards are required to pass a test prior to supervising pools and potentially saving lives. Teachers must demonstrate competency during their student teaching and usually pass some type of licensing exam.

The assessments identified above all involved a written or cognitive assessment along with a practical assessment or internship. As you drive down the freeway at 65 mph, you probably feel a lot safer knowing that the drivers around you were required to demonstrate their ability to drive as well as an understanding of the rules of the road. The practical part of the assessment occurred on real roads, using a real car, with other vehicles moving in other lanes. The person who certifies people to drive has to be sure that the potential driver won’t be a safety hazard to him/herself or others. Although driving simulators are available which can closely approximate real driving conditions, novice drivers are given an authentic assessment that includes a “real world” performance. Increasingly, authentic assessments are becoming more popular in education. Given the performance nature of many activities, authentic assessments are also being used more often in physical education. The purpose of this chapter is to explain the characteristics of authentic assessment, describe the various forms of
authentic assessment, provide criteria for judging authentic assessment, and discuss the advantages and limitations of using authentic assessment.

What is Authentic Assessment?

“Authentic assessment is assessment that occurs continually in the context of a meaningful learning environment and reflects actual and worthwhile learning experiences that can be documented through observation, anecdotal records, journals, logs, work samples, conferences, portfolios, writing discussions, experiments, presentations, exhibits, projects and other methods.” (Winograd & Perkins, 1996, p. 2). Wiggins (1989a) first used the term authentic assessment when calling for a different type of evaluation – one that would measure the ability of students to apply information that they were learning in schools to meaningful tasks or projects rather than simply being tested on this knowledge with single dimension written tests. Because authentic assessments are usually performance-based, they tend to be ideal in many situations for evaluating learning in physical education.

Characteristics of Authentic Assessment

Authentic assessments are generally considered to be performance-based tasks done in an environment that actually is or closely approximates a real world setting. Physical education’s real world environment might include playing games, dancing, rock climbing, or inline skating. Danielson (1997) identifies a second type of authentic assessment that involves work that someone in the profession would actually do. For example, a dance critic would be required to write a review of a dance performance. A sports announcer would describe the play-by-play of a game. Either of these examples could demonstrate cognitive knowledge of the respective activity and thus measure student learning.

There are eight characteristics that can be used to describe authentic assessments: exhibit harmony with a shared aim, involve the presentation of meaningful or worthwhile tasks, require higher levels of thinking, be judged using criteria known by students in advance, be so firmly embedded in instruction that they are difficult to separate from instruction, give students multiple opportunities to demonstrate competence, involve the public presentation of student work, and assess process as well as the product. Each of these characteristics will be explained in the following section.

Exhibit harmony with a shared aim (e.g. content standard or a school-wide goal)

Authentic assessments are linked to some type of important knowledge such as the academic content standards (Danielson, 1997). A school might have additional knowledge or goals that it meets through the completion of authentic assessments. Most states and subject areas have developed standards to define what students should know and be able to do. In physical education, many states have adopted the NASPE
Physical Education Content Standards or have developed their own which are variations of the NASPE Standards. According to NASPE, a physically educated person:

1. Demonstrates competency in motor skills and movement patterns needed to perform a variety of physical activities.
2. Demonstrates understanding of movement concepts, principles, strategies, and tactics as they apply to the learning and performance of physical activities.
3. Participates regularly in physical activity.
4. Achieves and maintains a health-enhancing level of physical fitness.
5. Exhibits responsible personal and social behavior that respects self and others in physical activity settings.
6. Values physical activity for health, enjoyment, challenge, self-expression, and/or social interaction.

*Involve the presentation of meaningful or worthwhile tasks*

A good authentic assessment is engaging for students (Danielson, 1997; Wiggins, 1996). It can capture students’ attention and inspire them to put forth a lot of energy and time thus increasing the amount of interaction with the content to be learned. The assessments involve learning considered essential to the discipline and that students will need and use in the future. Since students are doing tasks or activities that they see adults in the real world doing, the assessments have greater significance and are more meaningful for students.

*Require higher levels of thinking*

Authentic assessments are designed to measure complex thinking (Wiggins, 1996). With authentic assessment, students are required to take basic learned information and evaluate, analyze, or synthesize it while demonstrating the ability to use this knowledge. Schools want students to use higher levels of thinking; authentic assessments can create situations where critical thinking skills are used and assessed in physical education. Good authentic assessments will challenge students to build on prior knowledge and experiences as they demonstrate competence on the concepts evaluated by the assessment. They are designed to move students toward more sophisticated work, rather than something they can complete the night before the due date. When creating authentic assessments, teachers try to make student work relevant while creating complex intellectual challenges.

*Students understand the criteria by which authentic assessments will be judged*

When a teacher explains an authentic assessment to students, one of the key components is to also give them the criteria by which the work will ultimately be judged (Danielson, 1997; Wiggins, 1996). By having the criteria determined in advance, there is no confusion about the direction the authentic assessment should take. Identifying the criteria prior to beginning instruction also forces teachers to think through the assessment before giving it to students and may actually help clarify it. With a rubric,
when teachers explain the assessment to students, there should be no confusion about
the intent or focus of the assessment. Some teachers will have students discuss the
rubric or actually re-write it in the students' words to further eliminate confusion. Giving
students the criteria with the assessment helps remove some of the mystery that can
cloud the assessment process.

Assessments are firmly embedded in instruction

Frequently authentic assessments are very similar to tasks that teachers would
use for learning (Wiggins, 1989b). By adding criteria for performance or rubrics, the
task can become an assessment of learning. Physical education teachers sometimes
use “lack of time” as an excuse for not doing assessments. When assessments are
firmly embedded with instruction, they are difficult to separate from instruction.

Students have multiple opportunities to demonstrate competence

Authentic assessments are complex and may require several days or weeks to
comeplete (Danielson, 1997; Wiggins, 1996). Students are given criteria for the
assessment prior to beginning the project and thus have the opportunity to self-evaluate
their work as their skill and/or knowledge increases. Instead of giving the assessment
on a single day, teachers typically allow students multiple opportunities to demonstrate
competence on it. Just as people have more than one opportunity to pass the lifeguarding exam, students should have multiple chances to demonstrate competence on
authentic assessments.

Authentic assessments frequently involve work that is presented publicly

Presenting a final product or assessment publicly adds a form of accountability
to authentic assessments (Wiggins, 1996). Playing games in class tournaments,
presenting dance choreography to others, or at a parent open house handing out
brochures that were created in class, all add importance to the assessment. Additionally, students are given an audience for which to prepare. This audience helps
to focus the intent of the assessment. For example, a brochure on fitness for a family
open house would differ from one written for senior citizens.

If it is not possible to have a live audience, teachers should specify the audience
to give students a focus for their work. When there is not a live audience, a teacher
might have a contrived audience (e.g., videotape) or simply specify the audience for
students. This does not make the assessment less authentic nor decrease its worth as
an assessment (Wiggins, 1996). An audience provides a way to reduce the breadth of
an assessment as well as giving it another layer of accountability.

Assessment involves the examination of process as well as product

With authentic assessments, the process used to complete them is just as
important as the final product (Wiggins, 1989b). Much of the learning that can
accompany authentic assessments occurs while students are completing the
assessment. For this reason, the accompanying rubric should include elements of the
process that students should use to accomplish the assessment as well as a description
of the final product. Requiring students to explain the thoughts and decisions that went into arriving at the culminating assessment gives teachers greater awareness of student learning.

Types of authentic assessments

Authentic assessments usually fall into two categories, products and student behaviors (Danielson, 1997). This section will explain the various types of product and student behaviors associated with authentic assessments.

*Products*

When students are required to generate a concrete object that can be reviewed or evaluated at a later time, this is a student **product**. Student projects, journals, open response questions, and portfolios are all types of products used to measure student learning. Some examples of each of these assessments will follow.

*Student projects*

**Student projects** usually call for students to create something to demonstrate learning. Some projects may reflect individual work as well as group work. Instead of requiring students to write a report on a topic, an authentic assessment might require students to create a pamphlet or brochure for an upcoming sport camp to demonstrate their knowledge of basketball or tennis. A multimedia presentation on golf etiquette could also demonstrate student knowledge. Designing a piece of equipment to improve swimming effectiveness could demonstrate knowledge of swimming strokes and principles of physics. Students might develop an exercise video that others can use at unassigned time either during or after the school day. Because they find the final project engaging, students generally spend greater time completing it than they would spend studying for a written test. Projects usually require out of class time to prepare, which extends the amount of learning time in physical education.

*Journals*

**Journals** are excellent ways to assess students’ dispositions toward physical education. By giving students an opportunity to reflect on some aspect of the lesson, teachers gain insight on student perspectives toward the class. Instead of making the assignment a “free write,” teachers have better success when they create a specific question or **writing prompt** and ask students to respond. Although it is permissible to assess whether students have answered the question or prompt, teachers should avoid assessing student values or feelings toward a topic or activity, as this will decrease honesty when students write. Some teachers worry that writing in journals takes time away from activity in physical education. By making journal writing part of a class routine, students have a system for retrieving journals and returning them at the end of the writing episode. Journal writing can be done as classes conclude or at times when the whole class is not active (e.g., teams are playing in a tournament and there are not enough courts for every team to play simultaneously). Assessing the affective domain is
Open Response Questions

Open response questions give teachers an alternative method for checking cognitive knowledge. Open response questions provide students with a real world problem or scenario and give them an opportunity to solve it (Lund & Kirk, 2010). Students are expected to respond to the question by applying knowledge and information to address the situation presented. Since there are multiple ways to respond correctly to the question, the accompanying rubric must be open enough to allow for student choice in the response, while requiring the desired information that the question is designed to assess. Open response questions are often interdisciplinary and typically require higher order thinking skills to answer (Dick, Buecker, & Wilson, 1999).

There are five different types of open response prompts: single dimension, scaffolded, multiple independent component, student choice, and response to provided information (Kentucky Department of Education, 1997). Single dimension questions are the easiest to create and administer. Student responses usually are brief and little class time is required. Scaffolded questions contain a series of questions, with each response related to the previous response. Multiple independent components also have a series of questions but the response to one is independent of the response to the next. An example of a response to provided information would be when a teacher gave students a chart of fitness data and asked them questions requiring interpretation of that information while using knowledge from class that the teacher wished to assess. When teachers use student choice open response questions, they write questions that measure equivalent knowledge and then ask students to select one of the questions to answer. Each type of open response question offers unique ways to measure student learning.

Single dimension questions can require as little as 10 minutes to complete while some of the more complex questions will take over an hour to finish. In depth descriptions of open response questions are beyond the scope of this chapter. Lund and Kirk (2010) and the Kentucky Department of Education (1997) provide information explaining how and when to use open response questions.

Portfolios

Portfolios are collections of student work or artifacts that demonstrate student mastery or competence of some subject area (Lund & Kirk, 2010). They are an excellent way to demonstrate growth over time. There are two types of portfolios. Working portfolios contain multiple examples of student work and are places that students can store documents and artifacts. Evaluation portfolios contain selected items that demonstrate student competence on an area. Students select their “best piece” and then write a short explanation about why the artifact was selected and what type of learning it exemplifies. Chapter 2 in this book contains much information about portfolios.
Assessments of Student Behavior

Since physical education is very performance based, one of the best ways to determine learning is to look at student behaviors. Event tasks, role-plays, interviews, and observations are examples of authentic assessments that measure learning through student behaviors. The following provides a brief overview of these forms of authentic assessment.

Event Tasks

Performance tasks that can be completed in a single class period or less and which generally involve psychomotor activity are referred to as event tasks (NASPE 1995). Game play, choreography, and an inline skating obstacle course are all popular examples of event tasks. Teachers are able to see students in action as they apply the skills and knowledge learned from previous classes. Many of the adventure challenges designed to show teamwork or team building are also considered event tasks.

Role Plays

With role-plays, teachers provide students with a challenging, real-world problem or scenario that addresses a component of physical activity or physical education, and then ask students to demonstrate what they would do or how they would react. Measuring the affective domain is not an easy task as it is difficult to translate feelings and attitudes into observable behavior; role-plays can be an effective tool for doing just that. Teachers can also use role plays to assess cognitive knowledge. For example, a student might be asked to assume the role of a coach with 2 minutes left in the game. The prompt would give further details of the scenario and the student would be expected to come up with an appropriate offensive or defensive strategy that would be given during a time out.

Interviews

Interviews can assess student knowledge of many topics. Asking a student why he/she selected a certain play while coaching a team can reveal understanding of offensive or defensive strategies. Sometimes students make decisions that teachers question. By having the student give a rationale for the decision, teachers can detect instances of incomplete learning or even misinformation. Conversely, sometimes having students explain their actions will reveal new levels of understanding that teachers were unaware that they possessed. Interviews can be especially valuable for assessing learning for young children or students who do not have good writing skills. Students with English as the second language may also benefit from being assessed with interviews.

Observations

One of the most popular ways to assess student behaviors involves observations. Teachers, peers, or students themselves can do observations. When peers observe other students, they can provide valuable feedback and enhance student learning. Instead of having only the teacher to provide feedback, each student has his
or her own personal coach. Similarly, with self observations, students can assess their own learning or skill levels. Videotaping is useful with either peer or self observation as it gives students the opportunity to view performances multiple times. When using students to do observations (either self or peer), teachers must give clear guidelines for completing the task and specific criteria for the assessment. The purpose of the assessment must be understood and teachers must stress that honesty is the best policy. Students should not do assessments of others as part of the grading process.

Criteria for Judging the Value of Authentic Assessments

There are several ways to determine the value of authentic assessments (Lund, 1997). If teachers will ask these questions while creating authentic assessments they will be more likely to create worthwhile assessments of student learning.

*Is the knowledge being assessed essential rather than tangential?*

Authentic assessments should be used to measure essential learning rather than tangential material. Teachers should use authentic assessment to measure learning about the main concepts of instruction. A good authentic assessment requires a great deal of time to create and a lot more time to evaluate thoroughly. For this reason, teachers should ask themselves whether the authentic assessment is truly measuring essential learning when determining that it is worth doing.

*Does the assessment measure rich or superficial learning?*

Students should not be able to complete an authentic assessment overnight or after a single day of instruction. Because of the complexity, a great deal of student learning must lay the groundwork for the authentic assessment. The more opportunity students are given to look in depth at a topic, sport, or activity the more likely they will be to develop a deeper understanding of the area of study. A good authentic assessment forces students to delve into the content and thoroughly understand the knowledge being assessed.

*Will the students find the assessments engaging or uninteresting?*

Some topics are of greater interest to students than others. When developing an authentic assessment teachers must give student interest special consideration. Students must sense that the assessment is relevant to them before they will get “hooked” and want to spend considerable time completing the project. If students have a topic that they can relate to or that they are interested in, they will do a much better job completing it. Giving students a choice in various aspects of the assessment also helps to increase student interest.

*Does the assessment require active or passive learning?*

Life is more interesting when one is a participant rather than a spectator. The same holds true for authentic assessments. A good authentic assessment should involve active learning. Too often students are passive learners and are expected to
soak up knowledge while in school. Authentic assessments should engage students, immerse them in learning, and let them play an active role in the assessment process.

Is it feasible or unfeasible to complete the assessment?

Although authentic assessments should be complex, they also should be doable. Although teachers will want to challenge students with authentic assessments, they should avoid developing an assessment so complex that students are overwhelmed and feel incapable of completing it in the time frame allotted. Teachers also must make sure that students have the necessary resources to complete the task. Sporting equipment, computers, books, or other equipment and resources must be available if needed for the assessment. Time can also be a limiting factor, especially if the assessment will require out of class time to complete. Any variable that may present a possible barrier to completion of the task must be considered prior to adding that component to the assessment.

Advantages of Using Authentic Assessments

Authentic assessments can provide new dimensions for teachers to evaluate student work (Wiggins, 1989a). The advantages of using authentic assessment will be explained in this next section.

Direct observations of student work

Many traditional assessments rely on proxies of learning to measure student competence. For example, skill tests, long used as indicators of student ability to play various sports, are actually proxies for game play ability. The assumption was made that if a student could perform well on the various skills, then he/she should be able to play the actual sport. In reality, although skill tests provide valuable ways to measure student skill in a closed environment, they are not always good indicators of playing ability. There is much more to playing a game of volleyball then being able to set, pass, serve, spike, and block the volleyball. The same is true for written tests. Questions on a written test are typically designed to sample student understanding of various rules, history, equipment, etc. Authentic assessments allow direct observations of student behavior and learning so that no assumptions need to be drawn between what the assessment is measuring and the intended knowledge. Authentic assessments allow teachers to see how well students can play a volleyball game and whether they know the rules that govern it. Content validity is very high with authentic assessments.

Chance for students to apply knowledge

Authentic assessments give students the opportunity to apply what they have learned to an actual task. It is one thing to assess knowledge apart from the activity but clearly another to apply it. Students can demonstrate their knowledge of pathways and force as they implement these concepts while choreographing a dance sequence. This gives teachers the opportunity to know whether students truly understand related cognitive knowledge. By requiring a student to analyze his/her level of fitness and then
develop an exercise regime designed to maintain or improve current levels, a teacher can determine the student's ability to apply fitness knowledge. Listening to a student announce the play-by-play of a tennis match or watching someone officiate a soccer game would also indicate student ability to apply rules for those activities.

Teachers know if students really understand the material

One of the most popular assessments for cognitive learning is a written test. Since it is difficult to write questions that measure higher levels of learning (especially given the large number of students for which most physical educators are responsible), most written tests contain questions that are at either a knowledge or comprehension level. It is difficult to determine through a written test if students truly understand some of the concepts important to physical education. When students are asked to take a body of knowledge and apply it to a different set of circumstances, teachers can observe the degree to which students comprehend the concepts. For example, creating a training video for future officials that shows how certain rules apply during gameplay is an excellent indicator of how well students know and understand rules.

Good instructional alignment

In 1987, Cohen published an article that talked about a two-sigma difference in achievement when instruction and assessment were aligned. With instructional alignment, teachers test what they teach. Authentic assessments have natural instructional alignment, as frequently the assessments actually are the learning tasks. Teachers become coaches for the assessment rather than assuming a gate-keeping role of merely measuring student learning. In today’s standards-based climate, some people worry that teaching to the test could limit the scope of what is being taught and narrow the curriculum. Wiggins (1989a) notes that if the test is worthwhile, there is no problem with teaching to it. In fact, Wiggins goes on to say that coaches teach to the test each week as they prepare for the upcoming game. In order to get ready for the next game, coaches scout the competition. The week prior to a game is spent preparing players for various situations that they are likely to encounter given the upcoming opponent’s strengths and weaknesses. In much the same way that a coach teaches for the contest, teachers who use authentic assessments teach to the test and prepare students for real world experiences that they will encounter as adults. As Cohen found out with his classroom research, learning is enhanced when there is strong instructional alignment. Given the nature of authentic assessment, if a good assessment is developed, instructional alignment will be very strong.

Multiple chances to demonstrate competence or mastery

Formative assessments are frequently used with authentic assessment. With formative assessments, students are given multiple chances to gain competence and meet the criteria specified for the assessment. Some teachers do not like the idea of offering multiple opportunities to demonstrate learning. However, the idea of giving multiple chances to obtain mastery is very much attuned with real life. For example, students who fail a WSI test the first time are allowed additional chances to meet the requirements. Driver’s licenses work in the same manner. Not meeting the criteria on
the initial attempt does not diminish the privileges of those who ultimately achieve them. Skilled athletes rarely are able to learn a difficult skill by doing it once or twice. By allowing multiple chances to demonstrate competence or mastery, students learn the value of perseverance and working hard to achieve a final/ultimate goal.

Can provide feedback and learning for students

When students know the criteria for mastery in advance, they can self and peer assess as they strive for competence. These assessments provide valuable feedback to students helping them improve their skills and learning. Students can assume the role of coaches as they encourage peers to obtain mastery of a skill or concept and learn the material at a different level. Judging one’s performance against a standard also can increase motivation to learn. Students understand what they must do to reach the specified criteria and will work to achieve competence.

Students have more time to complete assessments

Different students require different amounts of time to learn skills and ideas. Since authentic assessments are complex, they are done over an extended period of time. When planning lessons, teachers usually allot a certain number of days to complete an assessment. Students can self-pace when completing the assessment, spending more time on the parts that are personally more challenging or difficult. The goal of performance assessment becomes student mastery, rather than putting in a certain amount of days on a task or concept. Advanced students might have the opportunity to learn additional skills if they are able to achieve mastery of an assessment early in a unit. Students understand that the assessment ends upon attaining a specified level of competency rather than after spending a given amount of minutes or days on the task. Additionally, given the authentic nature of the task, students frequently have the opportunity to work on assessments outside the school day. They can participate in games and activities with friends, or work on projects at home rather than just in school.

Tend to be more interesting; students spend more time completing them

Often authentic assessments have an “open” component, meaning that they are not as structured as more traditional assessment formats. As such, students can put a little of their own personality in the assessment allowing students to show hidden strengths and talents that might not be apparent with traditional forms of assessment. The final student work or product may resemble that of another, but still have unique differences. Different student learning styles, aptitudes, and interests are accommodated with authentic assessment while still addressing the desired content knowledge or standard. Since they are more meaningful to students, students tend to spend more time completing them.

Save instructional time

Even though authentic assessments are more complex and take longer to complete than most traditional tests, they tend to save instructional time. The simple reason for this is that with a good authentic assessment, it is difficult to separate
instruction from assessment. Teachers don’t need to set aside a day to do skill testing—they can evaluate the ability to use skills during game play or other types of culminating assessments planned for the unit. Some types of cognitive learning can also be assessed during game play. It is very easy to see knowledge of the rules of a game or activity by watching students participate. Affective domain traits can also be observed while students are engaged in psychomotor activity. Because assessments and instruction are not done separately, class time is used more efficiently and effectively by those engaged in learning.

Limitations of Using Authentic Assessment

As with any good thing, there also are disadvantages associated with using authentic assessment. Teachers must be careful when deciding to use authentic assessments as in some instances the disadvantages outweigh the advantages. The following section is intended to make the reader aware of some of the disadvantages associated with using authentic assessments.

*Much teacher time is required to prepare and grade authentic assessments*

Authentic assessments are very labor intensive for both teachers and students.

A good authentic assessment must be carefully planned and orchestrated to measure the desired student learning. Additionally, a rubric listing the characteristics by which the performance will be judged must be written prior to beginning the assessment. Since authentic assessments typically require several administrations to smooth out the rough spots, teachers will spend a lot of time developing the assessment as well as reflecting about and refining the assessment after it is initially created. This cycle is necessary to the process of developing quality assessments, but it does require a lot of teacher time.

It is far easier to grade a selected response exam (e.g. true/false, multiple choice) than it is to read 150 journals or portfolios. Teachers must be careful not to create an overwhelming burden for themselves when using authentic assessment.

*The assessment must measure what you intend to measure*

Measuring what you intend to measure is referred to as validity. Authentic assessments generally measure the actual, desired student performance and as such tend to have high face validity. Beyond the selection of the correct assessment, there are factors that influence validity. For example, if a teacher wants to know how well students can play tennis at the conclusion of a tennis unit, then the authentic assessment would logically be to have students play a game of tennis. However, if a team of good tennis players played a team composed of two beginners, the advanced players would probably not be able to demonstrate their higher level skills as rallies would probably occur infrequently during the games or set. Students should have multiple opportunities to demonstrate playing ability with opponents of various levels of skill. Additionally, the rubric would have to contain the correct indicators for the assessment to be a valid measure of student ability. If the tennis rubric failed to consider any items except for skill performance, then knowledge of strategies and rules
wouldn’t be evaluated. These as well as many other items must be carefully considered when planning authentic assessments and writing valid rubrics to use for assessment.

Low reliability

Measurement experts generally calculate reliability with statistical tests of high rigor. Calculating reliability with authentic assessments is done somewhat differently than for conventional assessments. Authentic assessments have been criticized for their subjectivity because of the reliance on teacher observation for scoring them. Leading advocates of authentic assessment take issue with those who consider the scoring subjective (Danielson, 1997). Since teachers often have many years of experience either playing or teaching the activity, they have a vast amount of knowledge. The professional judgment exercised by teachers when evaluating student performance should not be dismissed lightly. Teacher observation is a commonly used method for evaluating student achievement. In this instance, test and re-test reliability could determine reliability. With this, a teacher evaluates students and videotapes the performance. If the teacher goes back after a given period of time and gives the same student performance a score similar to that previously given, the assessment is said to have intra-rater reliability or objectivity. Inter-rater reliability is another type of reliability used with authentic assessment. With this, two different teachers score the same performance. The number of times that they have agreement on their ratings determines the degree of reliability. See more detailed information about reliability in Chapter 6.

If a teacher is the only person evaluating students during an authentic assessment, concerns about reliability are somewhat less, especially in situations of low accountability or low stakes assessment (the assessment does not have major implications for a student’s future). When multiple teachers are doing the same assessments for various groups of students, there must be consistency between raters. In other words, it should not matter who is doing the evaluation; the scores or results should be the same.

Observer drift, which refers to the consistency of the scoring, is another problem associated with reliability. It shouldn’t matter whether students are assessed at the beginning of the period or the end; the teacher should be consistent with his/her scoring. Observer drift can occur when teachers start focusing in on different aspects of performance over the course of the session. Fatigue can be a factor in observer drift when a teacher is not as alert as he/she was earlier in the day. Learning can be another factor. After a teacher has assessed several students, he/she learns to watch for certain specific elements as indicators of student success that may not have been apparent to him/her when the scoring began. A clearly written rubric that gives the scorer definite things to observe when evaluating students should have good reliability. Once the assessment begins, teachers must be sure that they assess students according to the rubric, even though problems may become apparent. Raters also must make sure that they are evaluating the items on the rubric and the performance they are witnessing, rather than what they think they should be seeing.
Tend to develop over time after multiple uses

Good authentic assessments typically are developed after several administrations, much revision, and lots of fine-tuning. For a teacher that uses the assessment once a year, it may take several years before the assessment really targets the exact knowledge and behaviors the teacher wishes to assess. Pilot administrations of the assessment are necessary for the development of a good rubric that focuses on the targeted knowledge. Teachers must be willing to commit the necessary time to this process, as authentic assessments usually are not perfect when first written. When the assessment is initially given, it should be more of an assignment than a high stakes assessment that is important to a student’s grade. As the assessment become more refined, higher stakes can be attached to it. In a related issue, because of the time lapse, teachers may forget the problems when the assessment was given and not fix them prior to administering the assessment again. It is important for a teacher to go back to the assessment as soon as it is completed and address any problems that arose. Only with this commitment to improvement will teachers develop the quality assessments that they desire.

Choosing the right type of assessment; traditional might be better

Since authentic assessments are time consuming for both teachers and students, teachers must be sensitive as to when to implement them. If the purpose of an assessment is to measure knowledge or comprehension, then traditional types of assessment are probably better choices than authentic assessments. Rules of a game, the knowledge about the history of an activity, and basic knowledge of terminology or correct form of skills (cues or critical elements that the teacher has taught) are all examples of information that can be assessed with selected response (e.g. multiple choice, true/false, matching) or short answer essay questions. Selected response items that measure higher levels of learning (analysis, evaluation, synthesis) are very difficult to write. If these levels of Bloom's Taxonomy are targeted, an authentic assessment will be appropriate. When choosing how to assess student learning, teachers should be very careful to select the most appropriate method of assessment. Given the large number of students that most physical education teachers have assigned to them and the student and teacher time that a quality authentic assessment requires, teachers should be selective when deciding to use authentic assessments with their students.

Rubrics are difficult to write

A good rubric takes time to develop and write. As stated earlier, if teachers do not look at the right criteria the assessment will not be valid. Rubrics should give students a good picture of what the final product will resemble. Since the easiest way to develop a rubric is to look at past student work, the rubric for first administration of an assessment is probably, at best, an educated guess by the teacher. As the assessment is developed over time, so is the rubric. Developing a good rubric is challenging. Chapter 3 will provide detailed information on how to write a quality rubric.

Resources to administer might be problematic
Sometimes resources might be barriers to using authentic assessment. If an authentic assessment involved making a brochure or pamphlet so that students could demonstrate cognitive knowledge, teachers must be careful to not give students with home computers and color printers an advantage by adding those items/descriptors to the criteria. If the assignment could be completed at school and used in conjunction with a technology class, then these descriptors would be included in the rubric. Resources are an important consideration when designing an assessment and its rubric.

Video cameras are often used for evaluating student behavior assessments to allow teachers the opportunity to review student performance multiple times after the initial performance of the skill. When students create routines in gymnastics or perform dances as part of authentic assessments, the teacher does not need to watch the performance live—students could work on the tape and then turn it in for the teacher to evaluate at a later time. This format will give students the opportunity to do several “takes” (this actually increases student learning) before being satisfied with their product. The teacher has time to give feedback and assistance to other groups or performers as the taping is going on.

Problems can arise with videotaping. It is suggested that each performer or group have individual tapes so that other groups don’t record over a previous performance. Having students purchase their own tapes can solve this problem. If a teacher has several tapes that he/she is letting students use for videotaping their project, care should be taken to label these so that the contents are readily known. Tapes should be put in a secure place so that others will not accidentally reuse them. Additionally, if several groups are using the same camera, teachers should create a schedule to ensure adequate access to the camera. If some groups are unable to finish during class, video cameras can be available during lunch or before and/or after school.

If a teacher is using videotaping to show game play performance, it may be necessary for students to “script” the game. Watching a regular game and hoping to see students perform certain tasks is risky. By scripting a game, students can arrange to have certain plays occur. For instance, in a softball game students could put the ball into play off a tee, which would allow them to direct the ball and set up certain plays. Students could hit a fly ball to the outfield to demonstrate knowledge of how to tag up after a fly ball is caught. A situation could be set up to show how to turn a double play. This scripting allows teachers to see cognitive knowledge about game play. The described game would not be a good way to evaluate hitting but could demonstrate student knowledge of strategies and rules important to the game of softball. The second advantage of scripting the game is that the teacher does not have to watch hours of film to analyze the play of students. Students could have 3 innings where every player on a small-sided team has a chance to bat and play various positions. Students should show a variety of scenarios during this assessment so that the teacher has an idea of how well students can demonstrate skills important to the game. With this format it becomes the responsibility of the student to demonstrate the skills that the teacher has indicated he/she wanted to see.
If possible, have a student aide in charge of the camera/taping so that novices don’t break the camera or drop it off the tripod. Safety can also be an issue if the teacher is behind the lens of a camera rather than monitoring instruction. Remote mikes are available that will plug directly into the camera making it easier to get good sound in a noisy gym. If multiple mikes are used, teachers should be sure that they are set on different frequencies so that they don’t interfere with each other.

Care should be taken so that equipment or the lack of it is not a factor influencing a student’s performance of an authentic assessment. It is a teacher’s responsibility to ensure that lack of equipment or resources does not impede student learning and demonstration of competence on an authentic assessment.

Finding ways to translate them into grades is difficult

Many teachers rely on percentages or points for determining student grades. Since authentic assessments are typically evaluated with rubrics, translating these into a grade can be problematic. The easiest rubric to use for grading is a holistic rubric. A holistic rubric contains verbal descriptions of the expected performance for the various grades (A through F). The description that best matches student performance then becomes the grade for the authentic assessment. Rarely will a student performance fit completely at one level. Problems arise when a student’s performance is partially at one level while the other traits are at another level. A teacher must decide a priori whether a student needs to have every characteristic of the description in order to be at that level or if only a majority of descriptors must be at a given level. If the teacher uses the latter approach, then a decision must be made as to which items are the most critical to performance. These descriptors must be demonstrated at a given level of performance for a grade to be awarded, even though not every trait reaches that level. If holistic rubrics are used, it is unnecessary to convert student performance into percentages.

Scoring is more subjective and therefore more subject to criticism

Since authentic assessments are judged against a set of criteria, in some instances there is no clear-cut answer that is required. As such they have been criticized for their subjectivity. Danielson (1997) points out the teachers who are making these decisions are usually highly skilled, trained professionals with many years of experience. As such, rather than referring to the decisions they make about the assessments as subjective, she states that teachers doing the evaluation are exercising professional judgment.

There is a degree of professional judgment associated with any assessment. The person who writes the questions for a multiple choice test selects the content that will be tested, writes the questions that requires judgments about wording, and decides which is the correct answer. Danielson goes on to point out that errors have been made on tests as prestigious and scrutinized as the Scholastic Aptitude Test (SAT). Professional judgment also enters into whether students have reached the level of performance stated in the rubric of the authentic assessment. If rubrics are well written, high degrees of reliability are possible (Herman, Aschbacher, & Winters, 1992). The
The difference between subjectivity in authentic assessments and traditional selected response items is simply in the point at which the professional judgment is used.

Conclusion

Authentic assessments are excellent ways to measure student learning in physical education. They allow teachers to measure the degree to which students have met various standards set by states or local districts. Additionally, they have excellent instructional alignment. Authentic assessments tend to be more meaningful for students as they involve the completion of meaningful and worthwhile tasks. Students also like knowing the criteria by which the assessment will be evaluated in advance, when beginning the assessment task. Since authentic assessments are firmly embedded in instruction, they are often difficult to separate from instructional tasks. Teachers must remember that authentic assessments are time consuming to develop and can be labor intensive to complete and evaluate. As such, they should use them for assessing essential knowledge, which requires higher level thinking skills. This chapter has provided a brief introduction to authentic assessments. A list of resources has been provided at the end of this chapter to readers who wish to seek additional information.
References


Exercises

1. Develop a culminating assessment that you feel is authentic. Using the eight characteristics of authentic assessment found in this chapter, is the assessment really authentic? Explain your answer.

2. Think of some type of student learning you wish to assess.
   a. How would you assess it using traditional forms of assessment
   b. Develop an authentic assessment for the same learning
   c. Identify ways that assessment and learning could be linked using authentic assessment.

3. Create a list of possible artifacts for a portfolio designed to evaluate diversity.

4. Develop an authentic assessment project designed to measure biomechanical concepts important to dance (or any activity or sport). Identify the higher level thinking skills the project involves and explain how that lets the teacher know whether the student truly grasps/understands the material.

5. Describe procedures a teacher could implement when planning an authentic assessment that would circumvent the disadvantages described in this chapter.
Additional Readings


A collection of articles that discuss many aspects of assessment and authentic assessment. Covers relevant topics. Articles are done by leaders in the field of assessment.


A collection of articles from the magazine *Educational Leadership*. Has several articles by Grant Wiggins and other assessment gurus.


Gives some good perspectives on using authentic assessment. The study described in the article contains teacher interviews that are insightful while looking at authentic assessment.


Stiggins is one of the leaders in the field of educational assessment. He covers the nuts and bolts of authentic assessment.


This article raises some interesting issues regarding authentic assessment. It urges people to consider some issues before jumping on the authentic assessment bandwagon.

Wiggins first introduced the concept of authentic assessment in 1989. This book will give many helpful ideas to people designing authentic assessments. The book covers all the major topics on authentic assessment.
Glossary

**Authentic assessment**: Assessment that occurs continually in the context of a meaningful learning environment and reflects actual and worthwhile learning experiences that can be documented through observation, anecdotal records, journals, logs, work samples, conferences, portfolios, writing discussions, experiments, presentations, exhibits, projects, and other methods (Winograd & Perkins, 1996).

**Bloom’s Taxonomy**: A hierarchical way of looking at student learning. Knowledge and comprehension are considered lower levels of learning according to this classification while synthesis and evaluation are considered the highest forms of cognitive learning.

**Evaluation portfolio**: A document containing selected materials which students submit to show competence or mastery in/of a subject area. A reflection explaining why the document or artifact was selected and the learning it represents is included with each portfolio entry.

**Event task**: A performance task which can be completed in a single class period or less and which generally involves psychomotor activity (NASPE, 1995).

**Formative assessment**: Assessments given to students multiple times throughout the instructional process. They can provide information to teachers about the instructional needs of a student before mastery is achieved.

**Instructional alignment**: An instructional strategy where teachers test what they teach.

**Interview**: A type of assessment involving teachers asking questions of students to determine knowledge on a topic.

**Journal**: A writing assessment where students are asked to respond to a question or writing prompt to demonstrate learning. Journals can be used to assess cognitive or affective domain learning.

**Observer drift**: When an evaluator unintentionally changes the criteria used for scoring an assessment.

**Open response question**: A writing prompt that describes some type of real world problem or situation and requires students to apply knowledge learned in class to solve or explain it. Responses typically require interdisciplinary learning and higher order thinking skills to answer.

**Portfolio**: A collection of student artifacts used to demonstrate student mastery or competence on some subject area. Portfolios are often used to demonstrate growth or learning over time.

**Product**: A type of assessment containing student work that is tangible and can be evaluated at a later time.
**Project**: A type of product assessment that requires student to create something such as a brochure, piece of equipment, display, etc. to demonstrate student knowledge on a topic.

**Reliability**: The degree of consistency with which an assessments measures content.

**Role-play**: A method of measuring student learning in which students are presented with a problem and then demonstrate how they would act if they found themselves in this situation.

**Rubric**: A document that contains the criteria by which an assessment will be judged.

**Skill test**: A method of measuring student performance in a closed environment for a sport skill.

**Working portfolio**: Places where students can store documents and artifacts prior to deciding which they will include in their evaluation portfolio.

**Writing prompt**: A problem, scenario, or question given to students to stimulate writing on a given topic.