Assessment and Accountability in Secondary Physical Education

Jacalyn Lund

Accountability surrounds us in our daily lives. It is a key to keeping students on task and augmenting learning. Physical educators can use various types of accountability to increase the quantity and quality of student response rates. Accountability can take a variety of forms, with grading being used most often, along with public recognition, high-tech heart monitoring equipment, trophies, and other material rewards. Authentic assessments are becoming another way to hold students and teachers accountable for subject matter competency rather than completion of a class. When accountability and assessment are used in conjunction with goal setting and task-oriented teaching, instructional alignment occurs. Assessment and accountability need to be applied to the instructional task system as part of an effective physical education program.

We are surrounded by things that hold us accountable. The police hold us accountable for speed as we rush to our next appointment or to see a student teacher. Our bodies hold us accountable to sensible eating and exercise. Many of us are held accountable by tenure and promotion requirements to publish the results of our research projects. Almost everything we do has some type of accountability connected to it.

Teachers hold students accountable for their actions in the classroom as well. This can be done in many ways—for example, by giving positive feedback to students who follow the rules, sending tardy students to the office, or awarding “bonus bucks” that can be spent at the end of a month on T-shirts, sports equipment, or other commodities.

In a similar manner, grades can hold students accountable for meeting the goals of a physical education class. If a teacher has difficulty getting students to dress or participate, students can be held accountable for these tasks through a grade. Sometimes when this happens, physical performance requirements are decreased in exchange for student cooperation on managerial tasks. Although a grade is put on the report card, one cannot really say that the student has been evaluated on learning or performance.

Evaluation and grading are terms that are often used interchangeably, but I have come to see them as related but sometimes different entities. Evaluation

Jacalyn Lund is with the HPER Department, Crawford Gym, University of Louisville, Louisville, KY 40292.
is the process of determining if the objectives of a class have been met (Thomas, Lee, & Thomas, 1988). The teacher determines what students should know as a result of having participated in the class, and students are assessed to determine the degree to which they have met these goals. The grade that a teacher places on a report card should represent this.

Studies have found that many teachers use dressing for class, attendance, and participation/effort to determine students’ grades in physical education (Hensley, Lambert, Baumgartner, & Stillwell, 1987; Imwold, Rider, & Johnson, 1982). Applying the definition just described for evaluation to this, one could surmise that these managerial tasks represent educational goals for these teachers. For some noninstructional physical education classes, this may in fact be the case. Some teachers do plan just to keep students busy, happy, and good (Placek, 1983). In other instances, however, these managerial tasks are really not the only goals that teachers intend for students to accomplish. Teachers may use grades to hold students accountable for managerial tasks because student compliance is higher when these tasks are used as part of the grade.

The Importance of Accountability

Accountability is the consequence that a teacher uses to increase the likelihood of students’ completing a given task (Lund, 1990). A teacher may give a task to students, but if he or she fails to consequate it, chances are the task will not be done as intended. For example, a teacher may tell a student to serve the tennis ball over the net into the proper service court 10 times. Unless the teacher monitors the serves or assigns another form of accountability, completing the task becomes the choice of the student. The ideal situation, of course, is for the learner to be motivated enough to complete the task regardless of the teacher-imposed accountability. This, unfortunately, is not always the case. Therefore, to hold the student accountable, the teacher may assign a partner to monitor the serves or may require a written notation by the student of task completion.

Accountability drives the instructional task system (Doyle, 1980). Competent teachers “maintain consistent accountability procedures of all students’ progress with interventions to improve student learning” (Reynolds, 1992, p. 24). Task systems are eventually defined by what students are held accountable for by teachers (Siedentop, 1991). Thus, if teachers hold students accountable for only managerial tasks such as dressing and attendance, then the instructional task system directed by the teacher is suspended. In this situation, students do only as much in the instructional system as they choose to do, resulting in instructional pseudoaccountability. This occurs when teachers either fail to monitor or they rely on monitoring alone to hold students accountable; there is no follow-through to ensure that students complete a given task.

Under instructional pseudoaccountability, good behavior frequently becomes the task of the class, because this is the only task for which the teacher is holding students accountable. The instructional task system is suspended as the teacher fails to hold students accountable for instructional tasks, even though she or he may be telling students to do them. If the student elects not to do the task, it will not be done because the teacher has failed to designate any style of accountability to ensure completion.

Grading is the type of accountability used most frequently for instructional
tasks (Lund, 1990). Some research has found grading accountability to produce the most on-task behavior (Alexander, 1982). Unfortunately, as used by many teachers, grading typically affects response frequencies only on the actual day of testing. Grades are not the only way to hold students accountable for instructional goals. I have observed teachers who use other types of accountability. Response frequencies equal to those of grading accountability conditions were recorded for other forms of accountability. One teacher used grading in such a way that it affected response frequencies on several days. I’ll describe her accountability techniques in the next section.

**Multidimensional Accountability**

Mrs. Brown used grades to hold students accountable for skill through a very explicit system that incorporated additional types of accountability along with grading (Lund, 1990). On the first day of the unit, she told her students that they would be tested on three volleyball skills, representing one third of the volleyball unit grade. Mrs. Brown demonstrated the tests to students and explained the scoring rubric. She specified critical elements that needed to be present for an acceptable performance. She also gave a product criteria and the standards necessary for an A on each test. Mrs. Brown made it clear that a response would not be counted unless it was technically correct. She also told students that the reason she had shown the tests on the first day was that she expected students to practice the skills before class or when they had free time during class. Over the next 2 weeks Mrs. Brown reminded students that they would be held accountable for meeting the skill criteria she had outlined.

During the unit, Mrs. Brown used actual skill tests as some of her teaching tasks. She allowed students to practice the tests by giving refinements of these. She also gave extensions of the tests to challenge students with more difficult tasks as the unit progressed. Having students do the tests throughout the unit accomplished two things: students were able to practice for the test and improve their level of skill along with their grades; and students became familiar with the testing protocol so that only a single class period was needed for testing.

Although the tests and grades held students accountable for a long-range goal, Mrs. Brown incorporated additional accountability techniques to develop and maintain a true instructional task system. Public recognition, class competitions, bonus points, and competitions against the teacher’s performance all helped to hold students accountable on an intermediate basis. These accountability techniques produced rates of response equivalent to those under actual grading conditions (see Table 1).

When skill tests are shown to students on the day that they are to be given, response frequencies are influenced only on that day. By being integrated throughout the unit, tests can function as formative evaluations for both students and teachers as instruction progresses.

Mrs. Brown demonstrated instructional alignment by stating goals, giving tasks to help reach these goals, and holding students accountable for the goals on her assessments. Large gains in learning occur when instructional alignment exists (Cohen, 1987). By specifying criteria early in the unit and then teaching to these tests, Mrs. Brown kept response rates high on a daily basis. Research has shown that high rates of responding are essential to learning (Silverman, 1985).
Table 1
Frequency of and Mean Response Rate per Minute for More and Less Skilled Target Students for Mrs. Brown

<table>
<thead>
<tr>
<th>Type of accountability</th>
<th>No. of tasks</th>
<th>Mean response rate for</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>More skilled</td>
</tr>
<tr>
<td>Monitoring</td>
<td>13</td>
<td>3.91</td>
</tr>
<tr>
<td>Public recognition</td>
<td>9</td>
<td>3.45</td>
</tr>
<tr>
<td>Feedback on errors</td>
<td>7</td>
<td>.90</td>
</tr>
<tr>
<td>Aversive</td>
<td>1</td>
<td>4.00</td>
</tr>
<tr>
<td>Accountability check</td>
<td>17</td>
<td>26.00</td>
</tr>
<tr>
<td>Grade practice</td>
<td>5</td>
<td>5.55</td>
</tr>
<tr>
<td>Skill tests/grading</td>
<td>8</td>
<td>2.24</td>
</tr>
<tr>
<td>Combination of accountability</td>
<td>21</td>
<td>3.04</td>
</tr>
</tbody>
</table>

I was observing two teachers at this school, who taught students from equivalent populations. Before Mr. Adam’s class, students dunked, kicked, and threw volleyballs—in short, they did anything with them except practice volleyball skills. In contrast, the preclass activities of Mrs. Brown’s students were practice and peer teaching. Her students helped each other as they prepared for their upcoming tests. Total in-class responses from the target students of these two teachers were also quite different (the responses before class were not counted in these totals, as I had failed to anticipate the use of this preclass time when I had set the parameters for data collection). Even without these additional numbers, Mrs. Brown’s students had as many as 4 times the total responses as Mr. Adam’s students (see Table 2). In addition, more of the responses from Mrs. Brown’s students were technically correct.

In a postclass conversation, Mrs. Brown said her administrators had not

Table 2
Response Frequency per Lesson and Percentage of Topographically Correct Responses by Target Student

<table>
<thead>
<tr>
<th>Target student</th>
<th>Adams</th>
<th>Brown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq. %</td>
<td>Freq. %</td>
</tr>
<tr>
<td>More skilled 1</td>
<td>45.44  85.33</td>
<td>60.75  90.53</td>
</tr>
<tr>
<td>More skilled 2</td>
<td>41.00  78.05</td>
<td>53.57  93.60</td>
</tr>
<tr>
<td>Less skilled 1</td>
<td>40.78  53.68</td>
<td>60.50  73.14</td>
</tr>
<tr>
<td>Less skilled 2</td>
<td>14.33  17.83</td>
<td>52.29  67.49</td>
</tr>
</tbody>
</table>
originally supported her grading techniques, although they now did. Whereas teachers and administrators may be reluctant to hold students accountable for psychomotor skills, this teacher found that student GPAs were not lower as a result of this practice.

**Nongrading Accountability Techniques**

I have also seen accountability techniques implemented without the benefit of grading. Physical education teachers at Branch High School were not allowed to grade on skill because of a Board of Education ruling. Students were held accountable daily for dressing and participation by a point system, with 80% of the student grade determined by points earned through dressing and participation. Substandard performance or inappropriate behavior resulted in deductions from the participation points.

This department used other accountability techniques to support their instructional system. Students were given trophies and ribbons rewarding performance during an end-of-the-semester physical education awards program. In bowling, for example, highest game, best series, and most improved player were among some of the achievements awarded mini–bowling pin trophies. One teacher observed that “nonathletes” tended to win bowling awards whereas track and field awards were won by those who were most athletic. In other words, all students, regardless of athletic ability, had a chance to be recognized in this physical education department. Competitions were arranged and winners received prizes. Also, class winners received additional recognition and incentive as they were encouraged by teachers to participate in interclass competitions against other class winners during the after-school activity periods.

Teachers at Branch also held students accountable by working one-on-one with students experiencing difficulty. One student was having problems with her volleyball serve. The teacher continued to work with the student even though the student wanted to give up. The teacher could have written the student off and deducted points for poor participation, but instead kept working with the student until she had successfully learned the serve. This is in contrast to a teacher who gives corrective feedback and then fails to remain with the student to ensure that the suggested changes in technique are implemented. When the teacher stays with a student until a task or skill is done correctly, the student is being held accountable for changing the motor response.

Public posting was another form of accountability used at Branch. Class leaders in the various fitness tests had their names posted on a chart that was displayed in one of the gymnasiums. Although students were not held accountable daily per se for the instructional system, the only way they could win an award at the conclusion of the unit was to have continuous achievement and improvement. Daily quality performance was necessary to receive end-of-the-semester awards for the various units.

I have had the opportunity to teach some of the students from this high school in a college class. They told me they “worked their tails off” to win those ribbons and awards. Even though the awards were inexpensive, students found them significant and worth working for. Public recognition is a powerful accountability technique. This program used it in a variety of instances.

A caveat must be offered here. Teachers need to be careful to avoid negative
effect with public recognition. Students with low scores can become the object of ridicule by other students or become overly self-critical, which can lead to lower self-esteem. With low-skilled students teachers might add group scores together rather than use scores from individual students, depending on the unit or task. All students need to have an opportunity to “win” or succeed or the negative results will far outweigh the positive.

High-Tech Accountability

I recently attended a workshop presented by Beth Kirkpatrick, a middle school physical educator who has received numerous awards for teaching excellence. Her students are assessed on four areas: effort, attitude, awareness, and expectations. Students are told that they will be observed and evaluated three times during the 9-week grading period. Because students don’t know when their evaluations will be, they are encouraged to participate maximally every day to ensure a high grade.

I would like to call attention to the component of effort. Teachers sometimes include this as a category for grading, raising or lowering grades subjectively. Not clearly specifying effort quantitatively, they encourage students to “try hard.” The problem with this is that the student’s perception of trying hard may be very different from the teacher’s. Tousignant (1982) noted that when “effort” was added as a dimension to the grading scheme, increased responses from students were likely. I suspect that teachers have discovered this, hence its popularity as a grading category (Hensley et al., 1987).

In Ms. Kirkpatrick’s classes, student effort is monitored in a unique manner. The teacher can tell when a student is putting forth adequate effort because every child wears a heart-monitoring device that records heart rate for the duration of the class. Beth can plug the monitors into a computer that gives a printout of cardiovascular activity after class. During class, she can “read” the heart monitoring devices with a wristwatch monitor and advise students to move faster or slower to maintain optimal target heart rates. Students are held accountable for highly specific fitness goals which they receive early in the grading period.

Although Mrs. Brown did not grade on effort per se, her students gave greater effort in terms of total responses than those of teachers who did. She indirectly held students accountable for effort because without effort, they could not have reached her criteria. Student responses were actually an objective measure of effort rather than a subjective one.

Authentic Assessments

Kentucky has recently adopted sweeping reforms (the Kentucky Educational Reform Act, or KERA) that will dramatically restructure schools. A key part of KERA is accountability. If a school fails to show improvement in student learning (status quo is not good enough), the state Department of Education has the authority to place sanctions on the school and take control of school governance and funding.

School improvement is measured in part by authentic assessments. These tasks are perceived as worthwhile to the learner. Authentic assessments simulate real-life situations. Completion of a class is no longer enough. A student must
demonstrate competence and mastery of a subject before the school is released from the responsibility to educate the child to a given performance outcome.

Sample assessments in physical education include such tasks as documenting participation in a sport activity. This means that students would be skilled enough to participate in an individual, dual, or team activity and then demonstrate that they are actually actively involved with the activity in some way. Fitness activity assessments may require students to explain and calculate a target heart rate and then write an exercise prescription demonstrating that they know how to keep themselves fit as adults.

Authentic assessments are gaining popularity in many states. Students are held accountable for knowing how the pieces fit together and for demonstrating command of a subject matter or competence in a skill. It is not enough to pass skills tests in physical education. Students must be able to integrate skills into games, dances, and so on to demonstrate mastery. Authentic assessments are very compatible with the sport education model described by Siedentop (in press). For example, in basketball students might record shots attempted, goals scored, free throws, rebounds, assists, turnovers, and steals during competition. Every student would get a performance record that could be used for assessment. Such data obtained during game play would be good indicators of how well students could apply their skills in an “authentic” situation.

Skills tests could also be used with authentic assessments. They would function as formative evaluations, to inform students and teachers about the degree to which the student had mastered the activity and where weaknesses still existed, rather than as summative evaluations.

Conclusion

To say that a good evaluation system coupled with accountability techniques can solve all the problems of physical education is a gross oversimplification. Evaluation must function in harmony with goal setting, quality lesson planning, good managerial systems, and all the other factors associated with effective teaching. The problem is that too frequently evaluation is tacked on at the end of a unit as an afterthought instead of being interwoven throughout.

Being able to document student learning becomes even more important given the current emphasis on accountability. Without some data, teachers have weak arguments to put forth about the value of their programs. This may place physical education in a precarious position in times of program review, budget cuts, or school restructuring. Accountability and authentic assessments are highly regarded topics in the educational literature. If physical education wants to assume a viable role in the education of children, change in evaluation procedures is essential. I feel the quality of physical education would improve dramatically if the following ideas were implemented:

1. Teachers need to determine criteria and curricular goals before beginning instruction and then to hold students accountable for reaching them. Evaluation systems should be designed to communicate the degree to which the students meet these goals. If the system is based on the instructional components of the class, grades communicate to others (such as parents and administrators) the achievements of students. These instructional goals can also provide targets for students to reach.
2. We must move beyond the practice of evaluating students on managerial categories and cooperation. These practices weaken our credibility and worth in the realm of education. Practices that do not hold students accountable for instructional goals do a disservice to them as well. We can hold students accountable for achieving skill in physical education instead of for only dress and attendance. We can give our physical education students “assignments” and require practice if we tell them in advance what we expect of them. Unfortunately, too many physical educators either have failed to envision just what they expect from students or they have never been taught to convey their expectations at the beginning of a unit.

3. If skills testing is used, it should be integrated throughout the unit. In this way it has a formative function, and it becomes a checkpoint or a means for giving feedback. It also can be an effective accountability technique. Evaluation should be done throughout the unit to ensure that students are progressing toward final goals. Although the ultimate goal is not really for a student to complete the skills on a test, this skills test can be one step in the learning progression toward competency. Teachers can also use this information to provide benchmarks for the quality of their teaching as students are held accountable for psychomotor goals. Teachers can give criterion tests that quickly and efficiently measure what they have taught. Although these tests do not have the formal standards of normative tests, they are probably more appropriate for the skill level of the students in that particular class. Teachers can set their own standards based on their instructional goals, the length of the unit, and initial student skill levels as instructional alignment strategies are implemented. These tests can be ecologically valid, as they assess teacher’s learning goals.

4. Accountability can take a variety of forms, with grading being the best known. Teachers need to be aware that other accountability techniques do exist and can produce high student response frequencies. Although grading and monitoring are important accountability techniques, they should not be the only ones teachers use.

5. Behavioral infractions need to be treated immediately rather than reflected in a grade at the end of the quarter, which dilutes the effect of punishment. Mrs. Brown gave rules about how students would act in class. They would not throw, kick, shoot baskets with, or sit on the volleyballs. Violations of these behavioral infractions were punished by having to spend part of the lunch hour with the teacher, rather than through a grade deduction. This manner of handling behavior gave immediate negative reinforcement and held students accountable. Ms. Kirkpatrick’s behavioral accountability system worked in much the same manner.

6. Preservice teachers need more experience in assessing students and holding them accountable for instructional tasks. They need to leave a teacher education program with an evaluation system (at least a skeleton) that works. Beginning teachers face a hectic year as they learn to negotiate the environments of their new profession and schools. They have little time to create an entirely new evaluation system. Research has shown that experienced teachers are more likely to test students (Kneer, 1986). If evaluation habits are to be cultivated, they must be started in preservice education and systematically added to the teaching repertoire at that level.

Grades can be a motivator for instruction as well as symbols of achievement.
when used in a program that promotes this. Instructional evaluation coupled with student accountability could significantly improve the quality of many physical education programs. Although the process requires some planning time up-front, the results from systems using assessment and accountability in harmony indicate that they are worth the effort.

References


