Correlations Between Teacher Behaviors and Student Evaluations in College Level Physical Education Activity Courses

Del Engstrom

Abstract

The purpose of this investigation was to examine the correlations between teacher behaviors and student evaluations in college level physical activity classes. The physical activity courses were observed using the Physical Education Teacher Assessment Instrument (PETAI). At the conclusion of an observation, students were asked to complete a Student Response Sheet (SRS). The SRS detailed student demographic information, identified a student's primary reason for selecting the specific activity course and asked the students to evaluate the instructor's overall effectiveness for the particular lesson. In all, 349 students participated in the lesson observations. The responses indicated that the students enrolled in physical activity courses have a positive rationale for being in class beyond just meeting a requirement. Students overwhelmingly rated teachers with high marks. In this investigation there existed a positive correlation between student evaluations and the percentage of time teachers spent in instructional behaviors.

The practice of using student ratings to evaluate teachers dates back nearly 75 years to 1926 when the Purdue Rating Scale of Instruction was first introduced (VanArsdale & Hammons, 1995). Since that time there have been more than 1,300 published studies related to the issue of student ratings (Cashin, 1988). Yet even today the implementation of student evaluations continues to foster misuses and misconceptions concerning its convention (Simpson, 1995; Wagenaar, 1995; VanArsdale & Hammons, 1995). Many college professors lament the idea of college freshmen being placed in the role of an evaluator especially since many students are still adjusting to the rigors of a college education and are not acutely aware of the subtleties of effective teaching. Still other professors link grade inflation with the high priority placed on student evaluations in the pressure to achieve promotion and tenure. And finally there are individuals who disdain the process simply because college teaching is about a specific discipline, not the "mechanics" of the enlightenment of students.

Simpson (1995) highlighted this concern by describing zealous proponents of student evaluations as individuals who are willing to accept, "forcing every instructor, in every course, at every level, in every academic discipline to be judged by the same yardstick—a yardstick constructed by someone else and devised in some instances to report that half of the respondents are below average" (p. 3). Simpson (1995) further stated,

> To place a highly selective group of professionals on a bell-shaped curve is not only irrational, it is destructive. When the only way for a person to get 'better' is for someone else to get 'worse,' then we have a system that breeds an unhealthy brand of competition. (p. 5)

The purpose and the process of student evaluations must be kept in perspective when examining the teaching dynamic.

Despite the concerns about the practice of using student ratings, Marsh (1984) reported that,

> class-average student ratings are (a) multidimensional; (b) reliable and stable; (c) primarily a function of the instructor who teaches a course rather than the course that is taught; (d) relatively valid against a variety of indicators of effective teaching; (e) relatively unaffected by a variety of variables hypothesized as potential biases; and (f) seen to be useful by faculty as feedback about their teaching, by students for use in course selection, and by administrators for use in personnel decisions. (p. 707)

In summarizing the research on student ratings VanArsdale and Hammons (1995) stated that student evaluations are congruent with faculty peer evaluations; that a professor's age, gender, years of experience, and personality have minimal effects on student evaluations; and that the student variables of personality, year in school, gender, age and academic performance have negligible impact on student ratings.

In an attempt to determine the underlying factors in effective teaching, Benz and Blatt (1995) analyzed data from a spring term for an entire university covering 46 departments. The data were then aggregated to identify the factors which contributed to the one global item assessing the overall instructor effectiveness. Benz and Blatt (1995) stated,

> overall instructor effectiveness was predicted most strongly by three items: namely, students perceptions that the instructor was prepared, presented matter clearly, and was interesting. The prediction of students perceiving that they 'learned a lot' was the rating on three items: the instructor was interesting, the course met the objectives, and the instructor was well prepared. (p. 27)

The researchers concluded that behavioral factors rather than affective factors were the strongest predictors of an instructors overall rating. Cashin (1988) stated, "If one is willing to grant that the ratings of administrators, colleagues, and alumni have some validity, then student ratings share that validity" (p. 2).

Specifically in the area of physical education, Woods, Phillips and Carlisle (1997) investigated the characteristics of physical education teacher education (PETE) educators. They reported PETE faculty maintained a high degree of ambiguity in regards to the "weight assigned to evaluation from students, administrators, self and peers in the evaluation of their teaching" (p. 156). In their investigation, the section of the survey that focused on teacher evaluations had the least number of completed answers. Those PETE educators who responded to the particular question in the survey identified student evaluations as the primary method used to examine their teaching. PETE educators also expressed further confusion regarding the application of the various weights provided for tenure and promotion considerations among teaching, service and scholarship (Woods, et al., 1997). Too often this ambiguity in scrutinizing the effectiveness of instruction minimizes the impact of teaching in the overall evaluation of educators.

Livingston (1996) identified a crisis that is currently plaguing our basic instructional programs in physical education in higher education. Livingston proposed that "over the last few decades, undergraduate physical education programs have lost sight of the objectives and the values of physical activity courses in the preparation of future teaching professionals" (1996, p. 114). Siegal (1997) defined the problem as one of conflicting themes and grading practices which have contributed to confusion as to whether a program should exist to meet the recreational needs of students or to focus on developing student competencies. Additionally, Rink (1996) described the dilemma as a "multi-objective problem" wherein teachers struggle to balance effective teaching strategies that may produce short term student learning (i.e. an improved spe-

cific motor skill) which in the long run may diminish an individual's desire to be physically active and ultimately adversely affect the long term goal of the course. The purpose of this investigation was to examine the correlations between teacher behaviors and student evaluations in a college or university level physical activity course.

Method

College level physical activity classes were observed by a single researcher using the Physical Education Teacher Assessment Instrument (PETAI) (Phillips, Carlisle, Steffen, & Stroot, 1986). The PETAI is a computerized software program with an internal time clock that enabled the researcher to continuously record various teacher behaviors that occurred while an actual lesson was being observed. Using the PETAI enabled the researcher to account for every second of an actual lesson and to identify the percentage of time spent in a specific teaching behavior. The PETAI focuses on two general categories of teacher behaviors—teacher instructional behaviors and teachers management behaviors. Phillips (1986) stated that teachers maintaining higher percentages of instructional behaviors were more effective than teachers who maintained higher levels of management behaviors. Phillips and Carlisle (1983) described effective teachers as being better classroom managers. They spent less time beginning and ending class and less time involved in equipment management.

At the conclusion of an observation (lesson), students within the class were asked to complete a Student Response Sheet (SRS). The SRS contained three specific areas of information. The first area focused on student demographic information: gender, age, major, year in school and class activity. The second portion of the survey asked the students to identify the primary reason for selecting the particular class that they were enrolled in during the study. Nine options for responding to the question were provided: (a) required course, (b) I really enjoy this particular activity, (c) best available time slot, (d) I wanted to learn more about the activity, (e) I wanted to get a good workout, (f) I wanted to improve my skills in the activity, (g) I enjoy taking courses from this instructor, (h) I wanted an opportunity to meet people who enjoy a similar activity, and (i) other. The final section of the SRS asked the students to evaluate the instructor's overall effectiveness in the particular lesson observed. The students were provided with five possibilities: (a) excellent: purposeful, organized, motivating, provided specific feedback, (b) good: had an objective, organized, provided general feedback, (c) average: organized, had an objective, (d) weak: unorganized, had an objective, nonsupportive, and (e) poor: unorganized, lacked an objective, provided little incentive, univolved. Students participating in the study were guaranteed anonymity and the researcher individually collected the SRS.

Sample

The lesson observations occurred at two locations in Southeastern Pennsylvania during the fall semester of 1997. One of the locations was a small private liberal arts college with a student population of approximately 1200 students. Students in this college setting were not required to complete a physical activity course as a part of their general education requirements. However, students were allowed to count three units of physical activity courses towards graduation credit. The second location was a larger university in the Pennsylvania State system of higher education with an enrollment of approximately 7,900. Students at this university were required to take at least two physical activity courses to meet a graduation requirement. In all, 349 students participated in the investigation. Approximately one third, 32.4%, of the students (n=113) attended the liberal arts college while 67.6% (n=236) were enrolled at the state university. In some instances stu-

Table 1 Activity Profiles in the Student Evaluation Investigation

Activity	Number of Classes Observed	Number of Students in an Activity
1. Aerobics	3	62
2. Badminton	2	40
3. Body Recall	1	15
4. Field Hockey	1	20
5. Foil Fencing	1	17
6. Golf	1	20
7. Judo/Self-Defense	1	22
8. Personal Fitness	2	43
9. Racquetball	1	12
10. Self-Defense	1	19
11. Swimming	1	9
12. Table Tennis	2	22
13. Tennis	3	29
14. Weight Training	1	19
Totals	21	349

dents elected not to respond to a specific question on the SRS, as such, some totals in the discussion were reflective of a lower total.

Thirteen teachers volunteered to be a part of this investigation. Seven of the instructors, four females and three males taught at the smaller liberal arts college. The remaining six teachers, five females and one male taught at the larger university. The greater number of female teachers involved in the study can in part be explained by the fact both departments had more female instructors than males. This profile was inconsistent with the overall findings of Woods, Phillips and Carlisle (1997) who reported that in "schools with less than 10K students, males accounted for over 62.5% of the PETE educators" (p. 152). Twenty-one separate lessons were observed to complete the investigation. The observations included 14 different activities ranging from team sports to individual and dual type activities to activities associated with personal fitness (Table 1).

Data Analysis

Descriptive statistics were applied to the data to determine the profiles of students involved in the study and to establish the primary reason for their course selection. The data was analyzed using SPSS for Windows, Release 6.1, 1994. Frequency distributions were computed for each variable to determine potential errors in data input and to analyze the score distributions. Pearson correlations were computed between variables to determine the direction of the relationships between the dependent variable and the independent variable. Correlation coefficients were computed to determine if any significant relationships (two-tailed) were associated with teacher behaviors and student evaluations.

Table 2 Student Demographic Profiles

	Frequency	Percentage
Gender		
Males	122	35
Females	205	59
Missing	22	6
Age		
17-18	12	3.4
19	63	18.1
20	101	28.9
21	96	27.5
22	42	12.0
23-29	27	7.7
30-40	8	2.3
Year in School		
Sophomores	70	20.1
Juniors	113	32.4
Seniors	152	43.6
Others	14	4.1
Academic Major		
Elementary Ed.	37	10.6
Exer. & Sport Sci.	33	9.5
Biology	28	8.0
Psychology	26	7.4
Others	225	64.5

Results and Discussion

The demographic profile of the students (n=349) involved in the study is described in Table 2. The students in the study were comprised of 122 males and 205 females (22 students chose not to respond to this question). While the ratio of males to females was not equal, the enrollment profiles of both colleges involved in the study contained a higher percentage of females than males. In addition some of the physical activity classes were more heavily populated by a single gender. In the case of this particular investigation, three classes were largely dominated by females: aerobics, field hockey and self defense.

The students ranged in ages from 17 to 40, however, the largest percentage of the students were represented by the ages 19

(18.1%), 20 (28.9%), and 21 (27.5%). While these figures are not surprising, it certainly should be apparent that a sizable amount of students fall outside the "traditional" age categories for higher education. Congruent with the information concerning age was the profile created for year in school. Seniors represented the largest group with 43.6% (n=152), with the junior class following with 32.4% (n=113) and the sophomore class represented by 20.1% (n=70).

Given the fact that students are generally required to complete student evaluations of faculty each semester, students in this study were experienced with the process of evaluating their instructors. Forty-three different majors were identified by the students as their major course of study. Elementary education had the largest representation of students with 10.6% (n=37) with Exercise and Sport Science 9.5% (n=33), Biology 8.0% (n=28), and Psychology 7.4% (n=26) completing the top four categories.

When asked to identify the primary reason for selecting the course four responses comprised the largest percentages of the survey. The largest response "required course," was identified by 22.6% (n=79) students. Considering that two physical activity courses were required by all of the students at the state university, and some of the Exercise and Sport Science majors at the liberal arts college, this is not surprising. Nevertheless, this figure represents less than 25% of the total number of responses in this investigation. It is interesting to report that the vast majority of students selected another rationale as the primary reason for selecting the course. Sixty (17.2%) students identified "I really enjoy this particular activity" followed by 56 (16.0%) responses for "I wanted to learn more about the activity," and 43 (12.3%) selections of "I wanted to get a good workout." The combination of these three responses indicated that students enrolled in physical activity courses have a positive rationale for being in the class beyond just meeting a requirement. The three categories that had the least number of responses were "I wanted an opportunity to meet people who enjoy a similar activity" (n= 1), "I enjoy taking courses from this instructor" (n=14), and "best available time slot" (n=18). For a complete list of frequencies see Table 3.

In the final section of the SRS students were asked to evaluate their instructor's overall effectiveness in the particular lesson observed. Students were asked to rate their instructors on a five point scale selecting from excellent, good, average, weak, or poor. Students overwhelmingly rated the instructors with high marks. The vast majority of students 77.9% (n=279) rated the teachers as "excellent." Another 17.5% of the students (n=61) rated the teachers as "good." In all, 97% of the students provided the teachers with favorable ratings.

The PETAI contains two major categories for identifying teacher behaviors: teacher instructional time and teacher management time. Teacher instructional time was associated with various forms of presentation, monitoring activities and providing feedback. Teacher management behaviors included beginning/ending class activities, organizational and equipment management issues and the time spent dealing with student disruptions as well as outside interruptions. In concordance with the favorable student evaluations, teachers in this investigation spent a

Table 3 Primary Reason for Selecting Physical Activity Course

Reason	Frequency	Percentage
1. Required course	79	22.6
2. I really enjoy this particular activity	60	17.2
3. I wanted to learn more about the activity	56	16.0
4. I wanted to get a good workout	43	12.3
5. I wanted to improve my skills in the activity	29	8.3
6. Other*	19	5.4
7. Best time slot available	18	5.2
8. I enjoy taking courses from this instructor	14	4.0
9. I wanted an opportunity to meet people who enjoy a similar activity	1	0.3
10. Students selected two or more responses**	30	8.6

*Students wrote a number of responses in this category; the most common response dealt with the need to complete credits.

** Students in this category selected two or more responses but did not indicate which response was their primary reason for selecting the course.

high percentage of their time involved in instructional behaviors. Over the course of twenty-one individual lessons the instructors spent an average of 86.1% of their available time in instructional behaviors. The percentage of time spent in instructional behaviors ranged from 74.9% to 96.8%. This finding is in agreement with the work of Phillips and Carlisle (1983) where they reported more effective teachers spent a greater percentage of their time in instructional behaviors.

Specifically within this investigation there existed a positive correlation (r = .1317) between the student evaluations and the percentage of time teachers spent in instructional behaviors. The positive relationship seems to indicate that students were attentive to the behaviors of teachers as they completed their evaluations. This data was consistent with the findings of Cashin (1988), "student ratings tend to be statistically reliable, valid, and relatively free from bias" (p. 20). This finding is further substantiated by the fact that the primary reason for selecting a course did not prove to be a significant factor in the evaluation process. Additionally, the students' age, grade in college, or major did not prove to be a significant element in the evaluation process. Tatro (1995) reported that female students provided higher teacher evaluations than male students. However, this investigation failed to identify a significant relationship between the gender of the student and their overall evaluation of the teacher. These numbers further support the concept that student evaluations tend to be relatively free of various biases.

When identifying specific teacher behaviors that correlated with overall teacher evaluations, three behaviors achieved statistical significance (Table 4). Two of the behaviors were associated with teacher management time. There was a negative correlation between teachers who spent greater amounts of time involved in equipment management (r = −.1320) and in teachers who spent more time involved in what is described as "other" activities (r= −.1693). Other activities are described as those behaviors which are unrelated to the actual teaching of the class; outside interruptions, discussions with people outside of the class, etc. These negative correlations further suggested that students recognized when teachers were involved in management behaviors rather than instructional behaviors.

The third behavior, motivational feedback, was the singular teacher instructional behavior investigated that had a positive correlation (r = .1551) with an overall teacher evaluation. Phillips et al. (1986) stated that performance feedback is the information a learner receives from an instructor that pertains to specific aspects of a performance or skill, whereas motivational feedback provides general responses to a skill attempt. Stroot (1987) stated that teachers who spent higher percentages of their time providing feedback had more time allocated for skill learning and spent less time involved on management issues in the lesson. Graham (1992) further described the benefits of providing feedback as encouraging teachers to actively monitor class activity and providing opportunities for instructors to assess students' acquisition of skills. From this investigation it appeared that college students value motivational feedback as a positive measure of teacher effectiveness. Teachers who employed greater amounts of motivational feedback in their instructional methodologies were perceived as more effective instructors. The use of positive reinforcement and praise should not be trivialized as an important component of teaching. The students in this study appreci-

Table 4 Correlations Between Teacher Behaviors and Student Evaluations

Teacher Behaviors	Correlation Coefficient	p
Instructional Behaviors		
1. Planned Presentation	.0539	.316
2. Response Presentation	.0012	.982
3. Monitoring	−.0309	.565
4. Performance Feedback	.0051	.924
5. Motivation Feedback	.1551	.004
Management Behaviors		
6. Beginning/Ending Class	−.1040	.053
7. Equipment Management	−.1320	.014
8. Organization	.0542	.313
9. Behavior Management	−.0820	.127
10. Other	−.1693	.002

ated being affirmed and encouraged while participating in physical activity courses.

Conclusions

The teachers involved in this study did an excellent job of employing a high percentage of their allotted class time for instructional purposes as opposed to managerial aspects of teaching. The teachers were organized and prepared. Students spent little of their class time waiting for the teacher to initiate the activity of the day or to organize the equipment necessary for the lesson. There were virtually no behavior management issues evident during the lessons. The students cooperated with the instructors in completing the instructional components of the lessons. As was the case, the students responded by evaluating the individual lessons of their instructors with very favorable marks. The results from this investigation confirm the work of previous researchers in finding that student evaluations are a reliable and valid means of measuring an instructor's effectiveness. The inclusion of student evaluations in physical activity classes appears to be an appropriate and important component in the overall evaluation of physical education teachers.

Furthermore, the importance of providing motivational feedback cannot be understated. Students at the college level believed it was important for their instructors to affirm their efforts and to be recognized in class. Even as adults students valued the encouragement of their instructor. Too often this simple instructional behavior which personalizes teaching is neglected and trivialized in the learning environment.

In addition, even in situations where physical activity courses were required, students identified alternative positive rationales for selecting their particular activity course. Key items in this selection process included; enjoyment of the activity, a desire to learn more about the activity or to improve skill, and an interest in achieving a good workout. As physical educators construct syllabi and implement their lesson plans, the interests and desires of students should be highly considered if we want to develop individuals who value physical activity and choose to pursue it as a lifelong endeavor.

REFERENECS

Benz, C., & Blatt, S.J. (1995). Factors underlying effective college teaching: What students tell us. *Mid-Western Educational Researcher, 8* (1), 27-31.

Cashin, W.E. (1988). *Student ratings of teaching: a summary of research.* IDEA paper No. 20. Manhattan: Kansas State University, Center for Faculty Evaluation and Development.

Graham, G. (1992). *Teaching children physical education: Becoming a master teacher.* Champaign, IL: Human Kenetics.

Livingston, L.A. (1996). Re-defining the role of physical activity courses in the preparation of physical education teaching professionals. *The Physical Educator, 53* (3), 114-121.

Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76,* 707-754.

Phillips D.A. (1986). *Intervention as a tool in improving teaching.* Paper presented at the annual meeting of the Central Association for Physical Education in Higher Education, Estes Park, Colorado.

Phillips, D.A., & Carlisle, C. (1983). The physical education teacher assessment instrument. *Journal of Teaching in Physical Education, 2* (2), 63-67.

Phillips, D.A., Carlisle, C., Steffen, J., & Stroot, S. (1986). *The computerized version of the physical education teacher assessment instrument.* Unpublished manu-

script, University of Northern Colorado, Greeley, CO.

Rink, J.E. (1996). Effective instruction in physical education. In S. J. Silverman, & C. D. Ennis, (Eds.), *Student Learning in Physical Education: Applying Research to Enhance Instruction* (pp. 171-198). Champaign, IL: Human Kinetics.

Siegal, D. (1997). What grading tells us about college physical education. *The Physical Educator, 54* (4), 208-215.

Simpson, R.D. (1995). Uses and misuses of student evaluations of teaching effectiveness. *Innovative Higher Education, 20* (1), 3-5.

Stroot, S.A. (1987). The relationship between selected teacher variables and selected feedback variables and student engaged skill learning time (Doctoral dissertation, University of Northern Colorado, 1987).

Tatro, C.N. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education, 28* (3), 169-173.

VanArsdale, S.K. & Hammons, J.O. (1995). Myths and Misconceptions about student ratings of college faculty: separating fact from fiction. *Nursing Outlook, 43* (1), 33-36.

Wagenaar, T.C. (1995). Student evaluations of teaching: some cautions and suggestions. *Teaching Sociology, 23* (1), 64-68.

Woods, M.L., Phillips, D.A. & Carlisle, C. (1997). Characteristics of physical education teacher educators. *The Physical Educator, 54* (3), 150-159.

Dr. Del Engstrom is a member of the Exercise and Sport Science Department at Ursinus College, Collegeville, PA.