10-2003

Student Performance Data, School Attributes and Relationships

Murrary F. Mitchell
University of South Carolina - Columbia, mmitchel@mailbox.sc.edu

Darla M. Castelli
University of Illinois at Urbana-Champaign

Skip Strainer

Publication Info

Published in *Journal of Teaching in Physical Education*, Volume 22, Issue 5, 2003, pages 494-511.

JOURNAL OF TEACHING IN PHYSICAL EDUCATION, 2003, **22**, 494-511

Chapter 2: Student Performance Data, School Attributes, and Relationships

Murray Mitchell
University of South Carolina

Darla Castelli
University of Illinois

Skip Strainer
University of South Carolina

In a recent edition of the *Educational Researcher*, Robert Slavin (2002) strongly supported the notion of gathering evidence to support educational claims. He characterized the time we live in as the "age of accountability" and suggested that there would be no support for any school program area for which teachers were not willing to identify and assess outcomes. Physical educators have generally been unwilling to assess outcomes and in most cases to minimally define those outcomes.

Physical education as a school subject can be characterized as largely void of both consensus and accountability. Until the publication of the national content standards (NASPE, 1995) there was no explicit, shared vision regarding what is important to teach in school programs. There has been a sense that this lack of consensus somehow preserves the right of every teacher in every school to teach what they want to students—a right not granted to teachers in the core academic areas of the school curriculum. The rationale for this approach is that it provides opportunities for good teachers to be creative and to design appropriate curriculum. Instead, it could be argued that the lack of defined outcomes and accountability for their achievement has done nothing but ensure that many school programs accomplish very little.

Evidence-based educational policy has been embraced in South Carolina and stands as a prime example of what Kohn (2000) identified as "legislature centered" education. Physical education professionals in the state sought to be part of the standards, assessment, and accountability movement driving the distribution of resources in education (Rink & Mitchell, 2002; chapter 1). The reform movement started with legislative action to define outcomes for a one-year high school program. Seven years later, the first data on high school physical education programs were collected and the South Carolina Physical Education Assessment Program (SCPEAP) was initiated.

The purpose of this study is to report the results of the first statewide data collection on the effectiveness of physical education programs in the high schools of South Carolina. A related purpose is to investigate the relationships among the data that were collected and relationships among school and teacher performance and school and teacher characteristics.

Context for Assessment

With the publication of *Moving into the Future: National Standards for Physical Education* (NASPE, 1995) and *Physical Activity and Health: A Report of the Surgeon General* (U.S. Department of Health and Human Services, 1996), a great deal of professional consensus has been reached on the purpose and overriding goals of physical education programs. Most professionals would support the idea that physical education programs should be developing a physically active lifestyle.

Data on four performance indicators were collected as a part of the South Carolina Physical Education Assessment Program (SCPEAP). The four performance indicators for high school physical education (see Appendix A) were selected for two main reasons. First, these indicators are facets of the first four national standards (NASPE, 1995) and are consistent with established professional norms for what characterizes quality physical education programs. Second, these indicators are considered to be components of an education for a physically active lifestyle. Students need motor skills (Ernst & Pangrazi, 1999; Sallis & Patrick, 1994), they have to have some knowledge about health related fitness (Lindner, 1999), they need to be fit, and they need to participate in physical activity outside of physical education if they are to lead a physically active lifestyle and if they are to reap the health benefits from that lifestyle both now and in the future.

A substantial knowledge base is available to our profession on physical fitness and the role of physical activity in overall health (U.S. Department of Health and Human Services, 1996). There is support for the relationship between knowledge about how to be fit and being physically active (Dale & Corbin, 2000; Dale, Corbin, & Cuddihy, 1998; Pearman et al., 1997). Overweight children tend to grow into overweight adults (Guo, Roche, Chumlea, Gardner, & Siervogel, 1994; Must, Jacques, Dallal, Bajema, & Dietz, 1992), and active children tend to become active adults (Telama, Yang, Laakso, & Viikari, 1997). There is also some predictive value in childhood fitness testing where children who score low are likely to become sedentary adults (Dennison, Straus, Mellits, & Charney, 1988). The support for the relationship of motor skills to a physically active lifestyle is mostly indirect. We know that active children tend to be active adults. We also know that most of the vigorous physical activity of youth is primarily skillful activity. We assume that children who are not skillful tend not to be regular participants and that children who are not participants become inactive adults. We have little knowledge of the relationships among these components. As a profession we have little support that we can influence the lifestyle choices of adults. It is not only important that we try but it is important that we study efforts to exert such influence.

The South Carolina Physical Education Assessment Program (SCPEAP) is a long-term, statewide effort to evaluate specific program outcomes. The assessment program involves selective sampling to determine program effectiveness in developing student competence to meet explicit performance indicators. School scores for SCPEAP are reported by a school overall score, by indicator, and by teacher. Normally, a report of school scores in a subject area and their relationship to both school and teacher characteristics would not be publishable information. That kind of information appears on the front page of the newspapers in South Carolina on a regular basis for core academic areas in the school program. What is unique about SCPEAP is the large database on what students can do in a physical education program having multiple explicit objectives.

Having baseline data on the motor skill, cognitive knowledge, outside participation, and health related fitness levels of large numbers of high school students permits the investigation of relationships among these indicators. Knowing the relationships among these indicators has the potential to be helpful in at least two important ways. First, these initial insights can facilitate better understanding of how these areas may contribute to a physically active lifestyle. Equally important is insight regarding the appropriateness of the level of expectations included in the four performance indicators for high school students.

This data set is context specific. Nonetheless, it represents a data set that can be used to ask important questions about the role of school context in teaching physical education and in choosing appropriate curriculum for high school physical education. We need to know what students know and are able to do in relation to defined outcomes. Knowing what students know and can do should help physical education teachers set more appropriate expectations for student achievement.

Perhaps the more important questions pertinent to this unique database are those related to understanding why some schools and some teachers are able to achieve very high student scores and some schools and some teachers are not. In chapter 3, Castelli and Rink have identified many department and teacher characteristics associated with both high performing and low performing schools. In the present study, the entire database was used to explore the relationship of several school characteristics and teacher characteristics to school and teacher performance.

School characteristics most associated with school academic performance have been identified in the literature. Profiles of schools completed since 1985 have contained many categories of school indicators with varying levels of relationships to student academic learning. Student attendance (Bobbett & French, 1993; Bobbett, French, Achilles, & Bobbett, 1995a; Bobbett, French, Achilles, & Bobbett, 1995b), percentage of teachers attaining higher levels on the career ladder (French & Bobbett, 1995), average teacher salary, and median income of the community have been the school characteristics most consistently related to academic performance in the effective schools literature. Negative relationships with school academic performance have been established for school drop out rate, student mobility rate, and the percentage students receiving free and reduced lunch.

The high relationship of socio-economic status (SES) with school academic performance is the relationship of most concern to educators. Many of the characteristics of high SES communities are the same as the factors associated with good school performance. Likewise, factors associated with low school performance are characteristics most associated with lower SES communities.

The established relationship between academic performance and the SES of a school presents problems for our society in general and for policy makers in particular. If poor school performance is related to a low SES community, then the teacher or school administration in a low SES school may not be responsible for, and cannot be held accountable for, poor student performance. Other negative consequences for citizens in lower SES communities include quality of life and healthcare concerns. The relationship between school performance in physical education and community SES, however, is unknown.

In addition to total school characteristics, another focus of this study was to look at the relationships among school and teacher performance and class size in physical education, teacher gender, and the amount of teacher training (in the standards and data collection) teachers received. In the classroom literature, small class

size has been shown to have a positive relationship to student achievement (cf. Achilles, Nye, Zaharias, Fulton, & Cain, 1996; Finn & Achilles, 1999). Intuitively, class size should be a factor associated with better student performance in physical education, too. A teacher with fewer students should be able to give more individual attention and use a greater variety of teaching methodologies (Rink, 2002). Physical educators often give large class size as a reason for not being able to run effective programs (Stroot, Collier, O'Sullivan, & England, 1994), but we do not have any data supporting the idea that students learn more in smaller physical education classes.

Class size has been a factor identified in the classroom literature related to school performance. Like other resource variables, however, the relationship to school performance is not as high as might be expected (Johnson, Barton, & Muldoon, 1998). In South Carolina, a strong negative relationship between class size and student performance led to state policy aimed at reducing class size in the core academic classrooms. As a result of the state initiative, some physical education classes in some schools were reduced and some were not.

The identification of teacher characteristics most associated with effectiveness is more complex. For this study, teacher characteristics were limited to data available on teacher gender and the training teachers received to implement the standards and assessment program. In this sense, the idea of teacher characteristics represents what Dunkin and Biddle (1974) refer to as presage variables. Teacher gender was a significant factor in the study done by Wirszyla (2002) investigating the efficacy of physical education reform in South Carolina. At the initial stages of the reform effort, female physical educators were found to be more supportive of the movement. More generally, there is an often insinuated bias in our field that females are more likely than males to run an instructional program.

The second variable investigated in this study was teacher training. The state professional association (South Carolina Alliance for Health, Physical Education, Recreation and Dance [SCAHPERD]) ran teacher development programs for seven years prior to the collection of data. The focus of these programs was to teach the state standards and to help teachers align their curriculums with the state standards. In addition, at least one representative from each school (not every teacher) attended a training session on how to collect assessment data. The variable of teacher training is discussed more specifically in chapter 5.

In summary, this chapter presents the results of a statewide program assessment effort in physical education. In particular, this study was designed to address four main research questions. First, what were the results of a program assessment surrounding explicit program goals? Second, to what extent are the results of program assessment data on the performance indicators related to each other? Third, what relationships can be identified among program assessment data and other school variables? Fourth, what relationships can be identified among program assessment data and selected teacher variables?

Methods

Participants and Variables

The school database for this study was the data collected on 62 schools in the first cycle of SCPEAP (see Rink & Williams, 2003). A plan was developed to

assess programs in one third of the districts in the state each year and these 62 schools represent the first cycle of data collection in this plan. Districts in each cycle were randomly distributed geographically and by district size. Data on four performance indicators were submitted to SCPEAP in the spring of 2001. One hundred sixty teachers from these schools submitted data.

The average enrollment of the participating schools was 943, ranging from 190–2533 students. Each of the schools received an academic rating on student performance the year prior to the collection of physical education data. This rating was reported on the first school report card sent to parents. School report cards in the state describe the performance of a school in the core academic areas and also include a variety of other indicators thought important to communicate to parents on a school. Scores for the overall academic rating of the 62 schools in this study were distributed as follows: excellent (n = 26), good (n = 12), average (n = 10), below average (n = 2), and unsatisfactory (n =10). Two schools had "not applicable" appear as the school's absolute rating.

Two critical school variables are the school poverty index and school scores on the 10th grade exit exam. The school poverty index is determined using both percent free and reduced lunch and median income of the community. The schools in this database represented the full range of school poverty index (11% of the school population receiving free or reduced lunch and a higher median income in the school, to 92% of the school population receiving free or reduced lunch and a lower median income in the school). High school students must pass an exit exam administered initially in the tenth grade. School scores for the exit exam represented the full range of exit exam scores (19–95)—this exam is scored on a 0 to 100 scale.

Of the teachers participating in this study (96 males and 64 females), 158 taught all four performance indicators. Two teachers only taught health-related fitness cognitive content. All teachers at schools included in this data cycle were required to submit assessment data to SCPEAP, however, not all teachers did.

At least 84% of the teachers in this study attended one or more Physical Education Institutes. Physical Education Institutes (PEIs) were teacher development programs offered in sessions across the year, designed to help teachers align their programs with state standards and teach more effectively to the standards. PEIs were sponsored by the State Department of Education from 1995 up to the year the data were collected for this study. The PEIs were conducted by SCAHPERD.

At least one representative from all schools assessed attended one or more data collection training sessions (41% of the teachers in this study). Data collection training sessions were designed to teach teachers how to collect data and how to use the assessment materials given to each school.

When schools submitted their data to SCPEAP in the spring of 2001, a peer monitoring committee assessed the accuracy of each teacher's data. A complete description of the process used to determine teacher compliance with protocols and teacher accuracy in collecting and analyzing data is provided in chapter 5. Of the 160 teachers required to submit assessment data, the monitoring committee accepted the data for at least one performance indicator from 146 teachers. For those teachers not submitting data, a score of zero was assigned. For those data sets submitted and not accepted by the monitoring committee because the data were inaccurate or because there were major violations in test protocols (e.g., the

assessment task protocols were not followed, student scores were not included with videos or scores but not videos were submitted, etc.), a zero was assigned.

Procedures

A SCPEAP assessment database listed the scores for every student in each class submitted for each performance indicator for each school. There were 20,975 student assessments administered by 160 teachers, across four performance indicators. Student scores were recorded as competent or not competent. School scores were determined by calculating the percentage of students achieving competence in each class and determining the average percent of students achieving competence for the school on each indicator. Teacher average scores for an indicator were determined in the same way. The overall score for each school was calculated in two steps. First, the school average percent of student competence was determined for each performance indicator. Second, scores for each indicator were weighted according to SCPEAP policy. Movement competence (PI-1) was weighted 50%, cognitive fitness (PI-2) was weighted 20%, outside activity (PI-3) was weighted 10% and fitness (PI-4) was weighted 20% to determine the overall school score.

School report card information on core subject areas was made public for all schools in the state during the fall of 2000 at the same time as schools were collecting physical education data. The report card indicators of (a) average academic class size, (b) school enrollment, (c) absolute school report card grade, (d) exit exam score for the 10th grade, and (e) exit exam scores for the 12th grade were obtained from the State Department of Education website for each school. The school poverty index (free and reduced lunch and eligibility for Medicaid benefits) was obtained from the Education Oversight Committee website.

Information on teacher gender, data collection training attendance, and Physical Education Institute attendance was obtained from SCPEAP records. Class size information was tabulated from class rosters submitted to SCPEAP as part of the assessment materials. Teachers were required to submit a recent computer printout of class enrollment with their data. Teachers were required to fill out special forms for any student for whom assessment data were not received.

Data Analysis

Descriptive statistics were used to determine state data and school data. State data included an overall average score, an average score for each performance indicator as well as an average score for each activity in PI-1 (motor competence) obtained by averaging all the class scores for that activity. School data included an overall school score based on the weighting for each indicator and school scores for each indicator. Teacher data included teacher scores for each class that was taught.

Pearson correlations were conducted to identify the relationships among all four performance indicators. Descriptive statistics were calculated for all school characteristic variables (average school physical education class size, average academic class size, school enrollment, poverty index, absolute report card grade, exit exam scores for the 10th and 12th grade). Analysis was organized by comparison of each variable to overall physical education score, and to each of the performance indicators.

A Pearson correlation with an alpha level of .05 was used to identify the strength of the relationships among overall performance in physical education and average physical education class size, school enrollment, poverty index, and the exit exam at the 10th and 12th grade. A Spearman correlation was used to identify the relationship between overall physical education performance and absolute report card grade. A Spearman correlation was selected because the variable of absolute school report card grade is ordinal. The same statistics were applied for comparison of overall physical education program by performance indicator.

An ANOVA was used to determine if a statistically significant difference existed between average academic class size and average physical education class size, within schools. Descriptive statistics were computed for all teacher characteristics (teacher gender, teacher attendance at data collection training, and teacher attendance at Physical Education Institutes). An ANOVA was used to determine the mean difference between physical education performance (by teacher) for gender and attendance. An alpha level of .05 was used to determine statistical significance. The analysis was conducted for overall physical education performance and for each performance indicator.

Results

Results are presented around the four research questions originally posed in this study. The results of the program assessment will be presented first as total school scores and for each of the performance indicators. Next, the relationships among performance indicators will be identified, followed by relationships to school variables and then to teacher variables.

Program Assessment Results

The school physical education program overall scores represent a weighted calculation of the percentage of students competent in each of the four performance indicators. Cycle 1 overall school scores ranged from 0% to 84% (M = 42.78, SD = 22.79). Scores on each of the four performance indicators are presented in Table 1. These scores are each represented by range, mean, median and standard deviation. There was a large standard deviation for each of the performance indicators (from a low of 22.79 to a high of 36.29). The highest scores appear for PI-2 (cognitive), and the weakest showing appears for PI-4, the Fitnessgram.

School scores for PI-1 (movement competence) are described by activity, in Table 2. Scores for this indicator represent a percentage of students competent in a motor activity. Average scores in this table represent the average class score for an activity out of all the classes in the state submitting data for this activity. The activities assessed are listed alphabetically with an indication of the number of different classes of data that were submitted for that activity.

Two separate scores for student competence in the activities are presented. First, the average scores for classes submitted for an activity that includes both compliant and noncompliant data are reported (third column). The term noncompliant is used to indicate the teacher did not follow specific testing protocols in the data collection and/or reporting process. Noncompliant data reported in this table received a score of zero for the class. A more thorough description and

Table 1 Percent Student Competency Overall School Physical Education and School Scores by Indicator

	Max.	Min.	*M*	Median	*SD*
Overall	84%	0%	42%	43%	22.79
Performance indicator 1 (Movement competence)	90%	0%	41%	41%	25.26
Performance indicator 2 (HRF cognitive test)	100%	0%	57%	68%	30.95
Performance indicator 3 (Outside activity)	100%	0%	41%	36%	36.29
Performance indicator 4 (Fitnessgram scores)	85%	0%	28%	28%	24.85

Table 2 Average Class Percent Student Competency by Activity

		Percentage Competence in the Activity	
Activity	Frequency	Compliant & noncompliant data	Fully compliant
Aerobic dance	7	76%	89%
Aquatics	1	100%	100%
Archery	4	48%	90%
Badminton	32	35%	60%
Ballroom dance	1	92%	92%
Basketball	47	50%	73%
Bowling	10	24%	58%
Flag football	11	53%	79%
Folk dance	1	91%	91%
Golf	9	37%	52%
Gymnastics	1	63%	63%
Line dance	3	89%	89%
Softball	13	31%	74%
Soccer	8	35%	70%
Square dance	3	100%	100%
Table tennis	8	68%	78%
Tennis	30	32%	55%
Track & field	5	43%	85%
Ultimate	5	41%	100%
Volleyball	31	48%	70%
Weight training	45	48%	90%

discussion of compliance issues is provided in the next chapter (Williams & Rink, 2003).

The fourth column in Table 2 reports the percentage of student competence based on data deemed in full compliance with assessment data collection and reporting protocols. For these data, teachers followed all procedures and protocols in the data collection, analysis and report. More than half of the students in all activities in full compliance reached competence. Student competence in the data reported for competence with non-compliance ranges from 24% to 100%. In the majority of activities, however, fewer than half of the students reach competence—defined as the ability to participate independently, safely and with sufficient continuity to make participation enjoyable.

Many schools submitted more than one class of data for popular activities such as basketball and weight training. Five activities were much more popular than other choices. The most popular activities were basketball (47), followed closely by weight training (45). Net activities were also popular, including badminton (32), volleyball (31) and tennis (30). Teachers in South Carolina chose other activities much less frequently for program assessment purposes.

Relationships Among Performance Indicators

Relationships among the four performance indicators appear in Table 3. All of the Pearson correlations were significant at the .01 alpha level. The strongest correlation (r = .59) was between PI-2 (cognitive fitness) and the Fitnessgram performance indicator (PI-4; fitness). Understanding how to design a fitness program was related to meeting age and gender fitness levels. Demonstrating competence in movement forms was related to understanding how to design a personal fitness program, participation in a health related physical activity outside of class, and meeting age and gender levels of fitness on Fitnessgram.

Relationships to School Variables

A summary of relationships among performance indicators and school variables is presented in Table 4. The school variable of class size in physical education yielded statistically significant relationships to only the overall physical

Table 3 Pearson Product Relationships Among Performance Indicators (Teacher Data)

	PI 1 Motor skill	PI 2 Cognitive fitness	PI 3 Outside activity	PI 4 Fitness
PI 1	—	r = .47*	r = .42*	r = .55*
PI 2	—	—	r = .49*	r = .59*
PI 3	—	—	—	r = .45*

* Significance $p < .01$

Table 4 Performance Indicator Relationships to School Variables

School variables	Performance indicators PI 1 Motor skill	PI 2 Cognitive fit.	PI 3 Outside	PI 4 Fitness	Overall
Pearson product relationships					
Class size	–.26[2]	–.21	–.24	–.14	–.27[2]
School size	.13	.10	.03	.29[2]	.16
Poverty index	–.20	–.38[1]	–.01	–.44[1]	–.32
10th grade exit	.18	.53[1]	.18	.461	.36[1]
12th grade exit	.24	.50[1]	.15	.37[1]	.36[1]
Spearman relationships					
School report	.26[4]	.38[3]	.13	.46[3]	.39[4]

[1]Significance $p < .01$; [2]Significance $p < .05$; [3]Spearman Correlation, significance $p < .01$; [4]Spearman Correlation, significance $p < .05$

education score (r = -.27) and PI-1 (movement competence) (r = -.26). As class sizes increased, there was a decline in performance scores. A statistically significant relationship was also found for school size and PI-4 (fitness) (r = .29). As school size increased, so did performance on Fitnessgram. While statistically significant, these relationships account for very little actual variance.

The overall school report card grade showed more complex interactions. Statistically significant relationships were noted between the overall school report grade and the overall physical education score (r = .39); PI-1 (movement competence; r = .26); PI-2 (cognitive fitness; r = .38); and PI-4 (fitness; r = .46). The amount of variance accounted for is minimal, but as general indicators of academic performance increased, so did student performance in physical education in general.

The poverty index yielded a statistically significant relationship with performance indicators two (cognitive fitness; r = -.38) and four (fitness; r = -.44). Students in higher income communities performed better on the cognitive performance indicator and on Fitnessgram. The 10th grade and 12th grade exit exams showed statistically significant relationships with the overall physical education score and PI-4 "Fitnessgram" only.

Relationships to Teacher Variables

A summary of relationships among performance indicators and teacher variables with F scores and probability are presented in Table 5, with means and standard deviations in Table 6. There are more male teachers (n = 92) than female teachers (n = 62) in the sample of participants for this study. Overall, students with

Table 5 Performance Indicator Relationships to Teacher Variables (F Scores and Probability)

	Performance indicators									
	PI 1 Motor skill		PI 2 Cognitive fit.		PI 3 Outside		PI 4 Fitness		Overall	
Teacher variables	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Gender	8.785	.004	14.846	.000	2.094	.150	2.756	.099	11.089	.001
Data training	5.829	.017	6.056	.015	3.064	.082	1.530	.218	8.469	.004
PEI	9.010	.003	6.411	.012	4.406	.037	4.390	.038	12.612	.001

Table 6 Performance Indicator Relationships to Teacher Variables (Means and Standard Deviations)

	Performance indicators									
	PI 1 Motor skill		PI 2 Cognitive fit.		PI 3 Outside		PI 4 Fitnessgram		Overall	
Teacher variables	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Female	50.45	31.31	71.00	27.40	47.84	41.47	35.89	29.53	50.47	24.86
Male	35.42	31.02	49.75	37.19	38.22	39.77	27.73	30.24	36.52	26.66
Data training	46.50	30.22	64.06	31.44	46.89	39.97	33.53	29.97	47.34	25.13
No data training	34.28	32.98	50.22	38.46	35.34	40.84	27.48	30.22	35.19	27.47
PEI	52.08	27.53	68.27	30.51	51.75	39.30	38.16	30.16	52.71	23.18
No PEI	36.19	32.67	53.34	36.30	37.31	40.57	27.48	29.63	37.14	26.98

female teachers (M = 50.47%) scored significantly higher (p = .001) than those students who had male teachers (M = 36.52%). This relationship was also true for students of female teachers on both PI-1 (movement competence) (p = .004) and PI-2 (cognitive fitness) (p = .000). Gender was not a factor in student performances on the other two performance indicators.

Students of teachers who attended at least some of the data collection training scored significantly higher (M = 47.34%: p = .004) than students of teachers who did not attend any data collection training (M = 35.19%). This relationship held true with student performances on PI-1 (movement competence; p = .017) and PI-2 (cognitive fitness; p = .015). Teacher attendance at training for data collection did not have a statistically significant impact on student performance on either of the remaining performance indicators.

Students of teachers who attended at least one Physical Education Institute had significantly higher total scores (M = 52.71%; p = .001) than students of teachers who did not attend any Physical Education Institutes (M = 37.14%). This relationship held true overall and was statistically significant across all four performance indicators.

Discussion

Results have been presented on a statewide, program assessment effort in physical education. Answers are sought to questions about the extent to which program goals are achieved; the extent to which performance indicator results are related to one another; the relationships among program assessment data and other school variables; and, relationships among program assessment data and selected teacher variables.

Program Assessment Results

The highest program scores were achieved on PI 2—the cognitive indicator of student ability to develop a fitness program. This result may have been achieved for a variety of reasons. At least one explanation involves the availability of extensive instructional support materials. Students of women teaching in this area achieved very high scores, which clearly raised the overall average of all classes. The large standard deviation is accounted for, to some extent, by the large difference in means between students taught by males and females. These data may reflect some substantiation for the perception of a stereotypical gender difference in teaching physical education.

In many schools, students are still separated by gender for physical education classes with male students taught by male teachers and female students taught by female teachers. The discrepancy of scores across teacher gender is an area warranting closer study. It is not clear if classes began with different skill levels, if the females taught more effectively, or if female teachers had students more predisposed to learning—it is not uncommon for disruptive students to be assigned to male teachers. If more male students are taught by male teachers and do not achieve competence in understanding health related fitness concepts (including the ability to set appropriate goals and design appropriate programs), there may be long-term implications for our society. It is possible that, across time, if fewer males acquire the skills to live healthy lives, there may be a negative impact on the level of

quality and length of life for males. This is an intriguing web of relationships. Physical activity patterns drop dramatically for females when they hit adolescence and yet females still tend to have a longer life expectancy than males.

The lowest mean was achieved on PI-4, involving the Fitnessgram test. This program area was least well addressed by most teachers at most schools. The explanation for this low performance may be the result of low initial fitness levels of students. Poorly administered testing protocols are the cause for some of these scores. Teacher fitness data were not accepted and a class received a zero if a videotape of students doing the curl ups revealed any errors in protocol (e.g., all students were performing the curl up incorrectly by not following the cadence tape and/or by not being stopped when two errors occurred). Testing protocols have changed across time—especially for the curl up. It is possible that not all teachers have kept up with various revisions to testing procedures, preferring instead to adhere to a more custodial orientation of maintaining their original operational strategies of testing student performance (e.g., feet held, hands behind the head or crossed on the chest, etc.).

Poor performance on this indicator may have another explanation. It is possible that teachers remained confused about the difference between program evaluation and student evaluation—in spite of repeated efforts to distinguish between these two concepts. Student success on all parts of the Fitnessgram test battery is a measure of program effectiveness with respect to impacting health related fitness. We tried to emphasize how this test battery differs from the President's Council test battery, which is more skill-oriented, and from motivating and grading students. How teachers facilitate learning by differentiating among constructs measured in the test battery, provide feedback and assign grades to students are different decisions from assessing program effectiveness. Still, it is our sense that many teachers wanted their students to feel motivated to try hard and feared negative consequences of following protocols too closely, yielding low test scores. Clearly, this is an important finding warranting closer scrutiny and further study.

Another issue warranting consideration is the extent to which using Fitnessgram test scores is a reasonable way to assign grades for students or to evaluate programs. There are mixed messages in the literature on the topic. The Fitnessgram scientific advisory board identifies the use of these scores for assigning individual student grades, evaluating teacher effectiveness and evaluating overall programs as inappropriate uses for Fitnessgram (Cooper Institute for Aerobics Research, 2003). Yet, in an AAHPERD (1999) publication, using fitness test scores is described as not an appropriate basis for grading "unless you have provided sufficient class time for improving fitness" (p. 63). We believe that the time allocated to fitness improvement and the fact that these scores are only one component used to evaluate programs (worth 20% of the overall assessment), justify the inclusion. Further support for this position can be found in Keating (2003) where he contrasts arguments for and against the use of fitness scores. Keating concludes that "accountability is critical in creating a meaningful role of fitness test programs" (p. 154), and that "without accountability, the efficacy of fitness testing in promoting fitness is dubious at best" (p. 154).

The fitness results for South Carolina are better than those achieved in California. This comparison is not offered as a direct ranking challenge—there are too many differences in the ways in which these data were collected for this to be a balanced comparison. Still, the lack of similar efforts at meaningful, related

databases make this an intriguing comparison. In California, there were 19.4% in 1999, 22.6% in 2001 and 22.7% in 2002 of ninth grade students in the healthy fitness zone in all components measured (California Department of Education, 2002). In South Carolina, 28% of the students measured were in the healthy fitness zone in all components measured. Certainly, 28% is nothing to be proud of; but, it may indicate that the effort to improve the fitness level of students in South Carolina is on the right track, especially since the scores in California have continued to increase on successive data collections.

Even when scores attributed to procedural errors are removed, student fitness levels are still low. Future studies appear warranted to discover if the low scores recorded in this study are indicative of low levels of fitness in children, low levels of student motivation to perform, unrealistic fitness standards, or some other explanation.

The fact that the lowest program performance was achieved on performance indicator 4 (fitness) rather than PI-3 (outside activity) comes as something of a surprise. The majority of high school teachers expressed resistance to PI-3 (outside activity), in that they did not support tracking student performance outside of the classroom. Concerns included fear of student cheating, being held accountable for student performance that was perceived to be out of their control (i.e., not in their instructional space and time), and the time commitment involved with tracking student performance on this indicator (Castelli, Hawkins, Koutseorgous, Rairigh, & Strainer, 2001; Fleming, 1998). It is unclear why the scores were higher for this performance indicator than for the fitness test scores. It is possible that teachers found this requirement to be more valuable and manageable than they originally feared. It is also possible that some misrepresentation of student performances went undetected or that the fitness standards could not be achieved with the time and techniques used.

An examination of the activities taught reveals distinct differences across movement forms. The various dances showed the highest proportions of student competence. If helping students reach competence is a key feature in students choosing to stay involved in activity, then these results support the inclusion of dance in the high school physical education curriculum. All other movement forms (e.g., team sports, individual, dual, aquatics, and outdoor activities) taught by more than a handful of teachers, reflect a substantive range of student competence scores. The wide range is encouraging because it means that it is possible for high percentages of high school students to achieve competence across a range of activities. A critical question warranting further study is whether students bring competence to class and score well because they can already do the activity, or is achieving competence the result of good teaching?

An attractive nuisance illustrated in Table 2 surrounds the discrepancy between student competency scores of compliant and noncompliant data sets. Competency scores were markedly higher when data sets were fully compliant with data collection and reporting protocols. Was this relationship evident because it is easier to concentrate on following directions with talented students? Or, are teachers capable of following directions more effective teachers? These relationships are undoubtedly more complex than the simple generalizations offered in this false dichotomy. Some type of qualitative follow-up with these cases would be informative.

A more thorough analysis of differences in teaching movement for competence than was undertaken for this study is needed. For example, what sorts of curricular and instructional approaches will yield the highest rates of achievement in movement forms? Is the level of competence sought, higher or lower than the threshold necessary for individuals to continue in activity? Is there a relationship between competence in selected movements and choices for an active lifestyle?

One interesting additional relationship is the high correlation across performance indicators—the level of performance in movement competence was strongly related to student performance on each of the other targets of program assessment. There are several possible explanations for these relationships. First, this correlation may suggest that there is some consistency across all areas of program delivery. Put differently, it is possible to deliver quality programs across areas; exerting effort and using resources with one part of the program need not be at the expense of other parts of the program. Alternatively, when programs are poorly delivered, they are delivered poorly across program components. The image of teachers being very strong with only one part of the program (i.e., good only in the classroom, or good only on the field), is not supported by these data. Strong teachers achieve results across a variety of goals and poor teachers are uniformly weak.

A second possibility for the high relationship among all four indicators is that the correlation is due to an overlap of student characteristics and content—both may contribute to an active lifestyle. It is possible that when students are competent movers, they are competent in more than one movement form, they are fit , they know the cognitive material and are participants in physical activity outside of school. A growing literature on the connection between a physically active lifestyle and cognitive performance would support this interpretation (cf. Cotman & Engesser-Cesar, 2002; Lindner, 1999; Symons, Cinelli, James, & Groff, 1997).

Variables Related to Performance

Overall physical education scores showed a significant negative correlation to class size. In spite of being significant, very little of the variance is explained through the relationship between these two variables. Class size would not appear to be the most significant factor in determining student performance scores. Class sizes ranged from small (9 to 20 students; n = 48), through medium (21 to 29 students; n = 79), to large (30 to 53 students; n = 28). Class size could not be determined for 6 classes with missing data. Very few of the schools in this data base had unmanageable class sizes.

The size of schools yielded no statistical relationship and poverty index had a negative but low relationship to the overall physical education score. The most significant correlations appeared to be between the overall physical education score and other academic indicators. Quite simply, the level of achievement demonstrated in other subject areas is mirrored in student achievement in physical education—effective schools are likely to have effective physical education programs.

The relationship of school physical education scores to the poverty index is negative. Unlike academic performance at a school that has a high relationship to the poverty index, however, the relationship to physical education performance is low. Most of the relationship can be attributed to PI-2 (cognitive fitness) and PI-4 (fitness). Low fitness levels and low cognitive scores have been identified as related

to SES previously in the literature (King et al., 1992). One reasonable explanation for the fitness level differences between SES groups may also involve nutritional variables. The quality of eating habits and general diet differences across groups has also been well documented in the literature (Centers for Disease Control and Prevention, 2001).

It is unclear if students perform well as a result of expert instruction or if students enter classes as high performers and remain essentially unchanged. Likewise the low relationship between SES and physical education performance leads to questions about what variables are most influential over student performance in physical education. Understanding what factors influence high performances is perhaps one of the most important questions addressed in this project and the following chapter will discuss the topic in greater detail.

Relationships Among Performance Indicators

The data from this study suggest that training for assessment and professional development can help teachers do a better job of teaching and assessing student performance. There is a positive and significant relationship between program performance and teacher attendance at these two types of sessions. Unfortunately, it is unclear if these teachers were effective already and wanted to capitalize on professional development opportunities available to them, or if the sessions effectively enhanced their abilities. As usual in exploratory studies of this nature, more questions are generated than answers. New cycles of state-wide assessment will add to the database and may shed more light on some of these complex relationships.

More insight into differences across teachers is available through the gender comparison. Students of female teachers fared better than students of male teachers overall, and significantly so with regard to movement competence and cognitive ability with fitness program design. The teacher gender differences in student performance are a cause for concern. An attractive explanation for such differences would be that male teachers spend more time coaching and therefore do not have the time to give to their classes that female teachers have. Yet, it is unlikely that this explanation alone can account for large differences in performance. Are more misbehaving, unmotivated students scheduled into classes with male teachers, yielding substantive competency score differences? Are females simply more meticulous about following protocol directions for testing and data collection? Are male teachers attempting to accomplish goals other than what was measured?

There is good news related to the big differences discovered between student performances with male and female teachers. These data suggest that the extent to which students achieve competency in physical education is teacher-related. Whether a student does well or does not do well is more strongly correlated to the quality of teaching than it is to factors beyond the teacher's control. We have also identified in this study that professional development experiences for teachers are correlated with student performance. Neither of these relationships can be described as causal; these are correlations only. Still, efforts to provide in-service educational opportunities for teachers would seem to be investments worthy of the time and efforts involved and there are grounds for the belief that teachers make a difference.

Conclusions

While many conclusions could be drawn from these data, several will be highlighted. First, students in South Carolina demonstrate the greatest levels of competence in PI-2 (cognitive fitness)—valued as 20% of program assessment. This accomplishment is consistent with teacher espoused beliefs valuing student understanding of fitness concepts. An added benefit is the potential link between demonstrating this understanding and increasing activity levels (Dale & Corbin, 2000; Dale, Corbin, & Cuddihy, 1998). If physical education programs can have a positive impact on student understanding, there is a possibility of changing the unhealthy patterns of adults in South Carolina.

A second conclusion to be drawn from these data is that the most significant program assessment area (worth 50% of program scores), PI-1 (movement competence), is achieved by fewer than half of the students in this data base. It is troubling to note that so few students achieve such an important goal.

Third, students in South Carolina are not meeting the minimal standards for health related fitness. Nearly three-quarters of the students assessed fell below this standard and this is cause for concern when children who score low on fitness tests are likely to become sedentary adults (Dennison, et al., 1988). Whether or not the gains achieved on impacting student understanding of fitness concepts can overcome low current levels of fitness and lead to an active lifestyle in adulthood is a significant question worthy of longitudinal study.

Fourth, the four performance indicators selected for program assessment in South Carolina appear related. That is, success on each performance indicator is statistically related to success on the others. Good programs can be good in all areas and do not appear to have to sacrifice achievement in one area to succeed in another. In fact, it is more likely that poor performance in one program area will be related to poor performance in other areas.

As a result of the combined efforts of committed professionals, quality programs exist and can be identified, when defined by student performance on explicit program goals. Still, high percentages of students are leaving physical education programs without achieving movement competence; without being involved in physical activity outside of class; and, unable to achieve healthy fitness scores on a health related fitness test. The ability to demonstrate an understanding of how to design fitness programs—a skill that may be of use later in life when values may change—is mastered by only slightly more than half of the students assessed. Arguably, the main targets of physical education instruction are not being achieved.

If there is one thing that must change about required physical education programs, it is this: The majority of students must leave our programs physically educated. Ultimately, the true test of the value of the South Carolina Assessment Program will be the extent to which we are able to move beyond simply identifying student and school performance levels to improving performance. In South Carolina, we have defined very specifically what it means to be physically educated. Program viability may depend upon achieving this basic goal. More importantly, and without intending to appear melodramatic, we believe that quite literally, lives depend upon our achieving this goal.