

10-2003

Insights and Reflections on a State Assessment Program

Judith E. Rink

University of South Carolina - Columbia, jrink@mailbox.sc.edu

Susan Stewart

Duke University

Follow this and additional works at: https://scholarcommons.sc.edu/pedu_facpub



Part of the [Education Commons](#)

Publication Info

Published in *Journal of Teaching in Physical Education*, Volume 22, Issue 5, 2003, pages 573-588.

<http://journals.humankinetics.com/jtpe-contents>

© 2003 by Human Kinetics Publishers, Inc.

This Article is brought to you by the Physical Education, Department of at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact dillarda@mailbox.sc.edu.

Chapter 6: Insights and Reflections on a State Assessment Program

Judith Rink

University of South Carolina

Susan Stewart

Duke University

The purpose of this chapter is to synthesize the results of the studies presented in this monograph and to discuss the implications of those results. The chapter is divided into four sections. In the first section state level assessment as a reform mechanism for high school physical education is discussed in terms of the South Carolina experience. A second section discusses the implications of the identified relationships among the performance indicators and factors related to school performance. In a third section issues related to the appropriateness of the measures, materials, and the design of the assessment program are explored. The chapter concludes with comments on the generalizeability of both the process that was used to establish the assessment program and the results of the first data collection.

State Level Assessment as a Mechanism for Reform in Physical Education

The primary goal of the South Carolina initiative is to improve the ability of high school physical education programs to help students lead a physically active lifestyle. What is reported in this monograph are baseline data. It will be quite some time before this program is able to document longitudinal and meaningful change in student learning. More importantly, it will be quite some time before we will be able to document changes in adult activity patterns. The assumption is that good performance on the indicators assessed is related to students leading a physically active lifestyle. Ultimately, we will have to document change in performance on the indicators and the relationship of that change to a physically active lifestyle.

As daunting as the task of demonstrating program effectiveness may be, we are encouraged by the change that we have seen. The South Carolina experience with state level assessment has produced change in high school physical education programs and so far, more positive than negative change. Change has been produced not only as a result of the program itself but also as a result of the process of developing the program. The assessment program has created a shared vision of what physical education programs should accomplish and has played a strong advocacy role for physical education. There is evidence of many positive curriculum changes in high schools. As a result of developing and participating in the assessment program a professional energy was generated in the state. Teachers throughout the state have grown professionally and have learned to work together at all levels for common goals. All of this change is dependent upon establishing accountability for change. These ideas are explored below.

Shared Vision

One of the positive effects of high stakes assessment programs is the idea that high stakes assessment programs have the potential to establish a shared vision for what students should know and be able to do (DeStefano & Prestine, 1999; Fullan, 1991). In South Carolina the move to a state assessment program was first established with legislation describing what the content of the high school program should be. Later, policy to place physical education on the school report card was developed which created the need for the development of assessment materials. The assessment materials created a shared vision of what students should know and be able to do as a result of a one-year program. This step was a critical shift from abstract theoretical perspectives of what physical education should teach students to a concrete application of those ideas.

Although the national and state physical education standards represent somewhat of a consensus, when there is no accountability for using standards there is great diversity in the interpretation of their meaning. More specific performance indicators and/or benchmarks developed for broader standards are not enough. More often than not practitioners perceive these as flexible “suggestions” rather than mandates and they are either modified beyond recognition or ignored by most teachers. What is eventually delivered in a program may represent an adaptation of the standards ranging from a complete ignorance of the intended outcomes of the standards through full adoption.

The assessment materials that accompany more high stakes assessment define in measurable and therefore specific terms the intent of standards. When teachers are held accountable for meeting explicit goals, the goals are more firmly established as expectations rather than suggestions, and the probability increases that students will achieve prescribed goals, or at least experience instruction with that intent.

From a utopian perspective the flexibility given to teachers when they are not held accountable for achieving specific goals would seem to be desirable. In a flexible climate teachers can design instructional goals to meet the needs of the students. For physical education programs flexibility has resulted more often than not in few expectations for student learning rather than different expectations for student learning. For physical education a shared vision of appropriate expectations may be an important initial step in producing change.

Most teachers in this study appreciated the idea that what they were expected to do was made explicit. They were supportive of the indicators as being appropriate and reasonable expectations for students for the one-year program.

Assessment as Advocacy

An unanticipated result of the reform effort and perhaps the most significant result of these studies for others contemplating state level assessment was the finding that teachers perceived the policy initiative as something that was being done *for them* and *not to them* (Fleming, 1998). Teachers’ positive support for the indicators was first identified in the initial studies of the reform effort immediately following the legislation prescribing the indicators (Fleming, 1998; Wirszyła, 2002) and well before the assessment program was established. More critically, teachers who collected and submitted the first set of data reported in this monograph have expressed the same support (see chapter 3; Castelli, et al., 2001).

Although the level of teacher support for the indicators and assessment program could not have been predicted, a large part of the success of the program can be attributed to that support. Teachers who thought they were doing a good job welcomed the opportunity to demonstrate that they were doing a good job. Teachers were also supportive of the assessment program because the assessment program played a strong advocacy role in their schools. The assessment program was seen by teachers and acted as a mechanism to support a marginalized program within the school (Castelli et al., 2001). As a result of the state level policy many programs were able to reduce class sizes, acquire equipment they needed, and obtain scheduling concessions from high school guidance counselors. Having expectations for what students should learn put physical education on a more equal footing with the academic areas of the school program. Few of the schools in this first data collection had unreasonably large class sizes. At the very least, program resources were preserved in a climate when all but academics was being reduced. Although the low performing schools did contend with more difficult inhibitors, the major problems of these programs seemed to lie more with factors that were in control of the teachers at a school.

A vision shared by the profession is very helpful in communicating with legislators, administrators, parents, and students. The professional association of the state was able to derail an early effort by music educators to substitute marching band for physical education. A group of music educators placed a heart rate monitor on a tuba player to demonstrate how much fitness and physical activity was achieved in marching band. Clearly, carrying a tuba in marching band is physical activity and the physical exertion was thoughtfully documented on videotape. Because the performance indicators were low, physical educators were better able to describe a quality program as more than being physically active and developing fitness.

Curriculum Changes

The studies of this monograph did not focus on curriculum, but changes in curriculum were clearly an outcome of the state assessment program documented in several of the studies. Curriculum changes were identified in both increased time spent planning curriculum and changes in the content of the high school physical education curriculum.

Increased planning time. As a result of the assessment program schools had to plan a year's program based on intended outcomes. They had to make the decision of what would be taught before the school year started as part of submitting an assessment plan. The assessment plan asked teachers to designate the classes that would be used to collect data on each of the indicators. Because programs had to designate ahead of time the classes that would be used to assess each indicator, the assessment plan as much as the evaluation itself forced programs to focus their teaching on intended outcomes in their classes. For many schools planning a year's program was a new experience. Targeting instruction to intended outcomes was even a newer experience. The more successful schools found ways to work together as a department to consider how best to do this.

Ninety-six percent of the high school teachers surveyed in this monograph reported an increase in the extent to which they planned (Castelli et al., 2001). Planning by itself may not lead to better programs. Nonetheless, it is unlikely that

most programs will get better at helping students become competent without thinking through and planning how to help students become competent.

Changes in content. Linn (2000) was correct in stating that schools would do whatever they had to in order to do well on the test. Usually the change in curriculum as a result of high stakes assessment is curriculum compression that results from teaching to the test. The opponents of high stakes assessment cite the “narrowing of the curriculum” as one of the major problems of high stakes assessment as a reform effort (AERA, 2000; Amrein & Berliner, 2002). What is identified as a minimal expectation becomes the maximum expectation.

We are not unaware of the potential dangers of the direction we are taking. Our experience so far is that as programs considered how they would address the performance indicators, curriculums have been expanded from what they were prior to the assessment program. The changes in curriculum we have been able to identify so far have all been positive ones.

Support for the idea that schools have actually improved their curriculum is based on the idea that many high school physical education programs in the state were primarily team sport dominated programs prior to the assessment. Short units of many different kinds of team sports were offered in an exposure curriculum. The performance indicators directed schools to include fitness concepts, fitness development, content helping students to make the transition to being physically active outside the physical education class, and movement forms other than team sports. Although we have no data to demonstrate a direct effect, anecdotal evidence supports the idea that changing the state computer scheduling software that is used by all of the schools helped to institutionalize a choice program and longer units of instruction for many schools.

All schools in the state have been encouraged to change their curriculums and move toward longer units and a choice curriculum. The high performing schools actually began to make these changes shortly following the 1994 state law defining the performance indicators.

Many programs had to begin to teach fitness concepts for the first time. In fact, 68% of the teachers who were surveyed indicated that they had increased the amount of fitness they were teaching (Castelli et. al., 2001). The scores for PI -2 (cognitive fitness) were the highest of all the performance indicators and teachers who taught the fitness actually enjoyed it (see chapter 4).

Teachers had many choices to make in terms of what they would teach. At this time we do not have a good sense of the basis upon which teachers chose what to teach. However, we have no evidence that schools chose what to teach based on their perceptions of how difficult or easy it would be to get students to a level of competence. Teachers are changing their curriculums so that they can develop competence in the indicators. In this sense teachers are teaching to the test. However, authentic outcomes and program flexibility are integral parts of the overall program accountability design. It is our sense at this point, that teaching to the test has not involved substantive changes for quality programs, and has been a dramatic improvement for low quality programs.

One of the most encouraging findings of this monograph is the idea that the more effective high school programs were able to integrate the content of all four performance indicators (see chapter 3). The programs conducted in these schools did not seem dominated by a narrow search for competence in the indicators. Teach-

ers in these schools established goals for their programs that were more reflective of the broader standards rather than the more narrow indicators.

Clearly the effect of the assessment program on school curriculum is one of the most important issues related to high stakes assessment. This initial investigation only gives us a glimpse of the curricular orientations of the schools in this database. The more important issues are likely to develop over time. Although the research is mixed, longitudinal research in the academic areas has produced evidence that high stakes assessment and the curriculum compression it causes may not lead to increased student learning (Amrein & Berliner, 2002).

Although the South Carolina program is not a high stakes assessment program it has created a shared vision and has created more accountability than physical education has had in the past. Physical educators need to be aware of research that takes place in the academic areas. On the other hand, they need to recognize that the factors that affect the academic areas may be different for marginalized subject areas in the schools. Our assumption is that if teachers teach to our test they will have a better program than the programs that preceded the assessment program. Our assumption is also that creating some accountability for student learning may be the only way to produce large-scale change in physical education at this time.

Opportunities for Teacher Development

The reform movement mobilized a very large group of state physical education professionals at all levels. An initiative set in motion by primarily college and university faculty was embraced by the professional community at all school levels. The policy initiatives created a professional energy in the state and opportunities for professional development at all levels.

With few exceptions some faculty from all teacher preparation institutions throughout the state are or have been involved in a phase of the program. The involvement of teacher education faculty has created a professional support for the program and has better assured the preparation of in-service teachers to work with the assessment program. With increased involvement of K-12 teachers the direct involvement of a large number of higher education faculty decreased. This necessitated running preservice opportunities for college and university faculty in the later years of the program.

An unanticipated benefit of the assessment program has been the positive effect of teacher participation in the development of the materials and training programs. A very large number of teachers were involved in developing and piloting materials at all levels. Teachers continue to be involved in making decisions about competency issues, developing the materials, or helping educate and train other teachers participated in a process that in and of itself was a tremendous growth producing experience.

Because announcements of teacher development opportunities were supported by the SDE and state policy, in-service and training programs were well attended by public school teachers. With few exceptions all of the high schools in the state attended at least one PEI and many schools were represented at the institutes for several years. All of the schools in the database reported in this monograph attended the data collection training, although not every teacher at a school did.

In this first set of data, significant positive relationships were established between teacher participation in both the institutes and training sessions and teacher scores (Mitchell et al., 2003). It is difficult to determine if the training programs facilitated more effective teaching or if it was the more effective teachers who took advantage of the teacher development opportunities. We can only speculate that both factors were related to effectiveness.

The Importance of Accountability

Questions concerning the necessity of accountability and more importantly accountability at the state level envelop this project. The characteristics of the effective programs identified in chapter 3 are consistent with the literature we have on effective programs and schools. What is perhaps new is that the state law and impending accountability facilitated an increased department cohesiveness necessary for effective programs in these high performing schools.

What differentiated teachers at high performing and low performing schools most in this study was that high performing schools did not wait to make sure the assessment program was going to be a reality before they began making change. Teachers at low performing schools took the popular attitude toward the many education initiatives that are handed down – they waited for it to go away. Not until it was clear that the physical education initiative was not going away did they begin to think about the changes they had to make. A reasonable conclusion is that these schools probably would not have made any change without the accountability created by the assessment program. What is more, the extent of change would not have been as great without state level accountability (see chapter 3).

The level of accountability applied to this first round of schools was not public reporting. Results of the assessment were sent in a report to the SDE and to school administrators. The SDE received both combined data from all of the schools as well as reports for each school by school and teacher. The only consequences for not doing well were those created by local school administrators. We have no data on this aspect of the program at this point but suspect that school administrators wanted their schools to do well. At the very least, the competitive climate that surrounds schools led many administrators to talk to their physical education teachers about the quality of the school physical education program.

Accountability can make change and certainly accountability is needed in physical education. What is at issue is the level of accountability needed to create accountability for student learning without changing what you want students to learn. It may be that providing teachers with a clearer idea of what students should be learning in their physical education programs and administrators with information on what students are learning in physical education programs may be sufficient to produce change in our field.

Accountability issues also surrounded the data collection. If teachers were able to submit data without accountability for its accuracy and compliance with protocols, the data submitted would most likely not be useful data. We have support from both Williams and Rink (chapter 5) and Stewart and Mitchell (chapter 4) that indicates that some teachers would either non- intentionally or intentionally hand in inaccurate data unless they were held responsible for its accuracy. Clearly some teachers did not expect anyone to look at the data they submitted; one even said so on camera when filming students.

The procedures and policies established to govern the data collection process and the monitoring committee entrusted to determine the accuracy of the data, played major roles in assuring that the data were accurate. We do have some questions about the accuracy of the data submitted by a few teachers in PI-2 (cognitive fitness) and PI-3 (outside activity) and will continue to develop procedures to ensure its accuracy. Practical procedures that ensure the accuracy of the data are essential for accountability. Accountability is essential to make change in a large number of schools.

Relationships Among Indicators and Factors Related to Teacher and School Performance

Relationships among the performance indicators and between the total school physical education score and several important school performance scores, school characteristics, and work conditions were investigated for this database. Because SCPEAP is designed to assess a program and not individual students, it was determined that the class would serve as the descriptive unit. Some variables examined could only be examined at the school level (SES), while others could be examined at the level of the teacher. Our conclusions are about class, teacher, and school performance, not performance by individual students.

Relationships Among the Performance Indicators

In this data base both school motor skill performance and school fitness scores were related to each of the other indicators. Motor skill and fitness were also related at the teacher level. It is difficult to determine whether differences are a *teacher effect*, although variability within a school would certainly support a role for the teacher. It is also difficult to determine program effects from this data set. Variability among programs may in fact be due primarily to other factors. What we can conclude is that classes of fit students were more likely to be competent movers, participate in physical activity outside of physical education, and have the ability to design their own personal fitness program.

Whether fit students are more skilled or more skilled students are fit is intriguing and an important issue for our field. While most physical educators have supported a strong relationship between skill in motor skills and a physically active lifestyle, many health related professionals have talked more about engaging students in physical activity rather than developing skill in motor skills in physical education classes. The role of motor skill performance in both developing and maintaining fitness and in developing a physically active lifestyle may be one of the most significant questions needing investigation in our field. The results of this monograph would indicate there is a strong relationship between the two.

The Physical Education Score and Factors Related to Performance

The studies reported in this monograph sought to establish relationships between school characteristics, work place conditions, teacher characteristics, and physical education performance. The results identifying the relationships between school characteristics, class size, teacher training, and teacher gender are probably the most important.

School characteristics. A continuing debate in academic programs often played out in the popular media concerns the “fairness” of holding schools accountable for student academic performance when there is a very high relationship between school performance and the socio-economic status (SES) of the school. For the data base in this study a high significant relationship between the poverty index of a school and the overall school physical education score was not identified. However, the poverty index did have a low positive relationship to both the cognitive performance indicator and the fitness indicator. A positive and significant relationship between the overall physical education school score and three different academic performance scores for a school were identified (overall school score, 10th grade exit exam and 12th grade exit exam) (see chapter 2).

The results identified in this monograph can be interpreted in several ways, but mostly it means that teachers and schools are more in control of how students do in physical education than they might be in the core academic areas. In the core academic areas high relationships exist between SES school variables and academic performance on state tests. This is particularly true in South Carolina. In physical education these relationships are considerably weaker. The wide range and great variability in teacher scores within the same school and school scores within similar economic classifications support this idea. Several low SES schools were high performing schools.

Class size. There was no difference between the academic class size mean and the physical education class size mean for this database. There was a relationship between class size and school physical education score but it was a low one. Very few of the schools in this study had large class sizes, which makes the overall role of class size difficult to determine. Most of the identified relationship is probably due to the high negative relationship between the motor skill performance indicator and class size (see chapter 2). Instruction in motor skills was most affected by class size, which should be expected.

Teacher training. A positive relationship between class performance and attendance at both physical education institutes and data collection training programs was identified. Again however, we will not be able to determine if the more effective teachers went to these sessions or if going to these sessions made the teachers more effective. It is most likely a combination of both factors. Those teachers who attended PEIs and data collection training sessions had classes that scored better.

Teacher gender. The male/female teacher issue continues to be problematic for our field. Having evidence that suggests females are more effective teachers affirms what many have suspected (see chapter 2). Coaching responsibilities in many situations do consume more time for most men than they do for most women. On the other hand the data on high performing schools indicates that both men and woman at these schools coached and all were effective teachers and took pride in their teaching. This was a change from Wirszyla (2002) who identified in the first years of the project that many men in the schools studied were not concerned with doing a good job in their teaching. Although the evidence is not conclusive at this point it would seem as though some kind of accountability may motivate men to do a better job with their teaching. Continuing to blame the problem on coaching responsibilities may be too simplistic of an answer. The profession would be served by continued investigation of the gender issue.

The Appropriateness of the Measures, Materials, and Processes

Physical education programs should be addressing the national standards, which were adapted in South Carolina. The power of the standards is their goal setting ability. They define what a student needs to know and be able to do in order to lead a physically active lifestyle. If teachers were to teach effectively to the standards most of us would agree that more students would probably lead a physically active lifestyle. The South Carolina high school performance indicators do not begin to measure the standards as they are written in the national document *Moving into the Future: Content Standards for Physical Education* (NASPE, 1995) or the state document (South Carolina State Department of Education, 2000). The South Carolina indicators are more narrow in their scope, and from the perspective of the intent of the standards themselves are minimal expectations.

The difficult part of having standards and assessing them is that even though you are measuring indicators you want teachers to teach to the standards. Indicators should be samples representative of a standard. In order to help teachers teach to the standards, teacher development programs, particularly the PEI's were designed to create an understanding of the intent and implications of the standards for curriculum and to a lesser extent instruction. This was easier to do before the onset of the assessment program. With the introduction of the assessment program many teachers at teacher development sessions became more concerned with receiving information on how to teach to the indicators rather than teaching to the standards. This is in spite of our efforts to tie the indicators to the standards and the standards to the notion of a physically active lifestyle.

The teachers' focus on the indicators and not the broader standards may be a phase in program development and teacher understanding of the reform. It may also be a permanent and unavoidable result of more accountability for the indicators. It is unclear whether we can change this focus, but we are encouraged by the idea that there were programs able to integrate the indicators into both curriculum and instruction. Teachers in high performing programs were able to focus on broader perspectives in the design of their programs (see chapter 3). Teacher difficulty in seeing or being able to use big ideas certainly is not unique to physical education teachers.

Performance Indicator 1 (Movement Competence)

The current data on student performance in PI-1 is cause to reflect on the decisions made in respect to competency. If the notion of competency is truly criterion based then it should not change with the scores students receive. However, if large numbers of students either are considered competent or non-competent in an activity should you change the criterion for that activity? For example, if over 90% of the students were considered competent in ballroom dancing is the level of competency set for ballroom dancing too low or do you celebrate the idea that most students are competent in this activity? If only 52% of the students could achieve competency in golf is the level of achievement necessary to achieve competence for golf too high?

Physical educators in South Carolina established levels of competency based on a notion of the level of skill necessary to be a participant in an activity. At some point levels of competency attached to being an active participant need to be based on more trustworthy information. What our profession most needs is a study of the skill level of the participant. We need to know what level of skill is related to continuous participation in an activity. We need to determine the level of skill people have who participate in an activity regularly.

Performance Indicator 2 *(Cognitive Fitness)*

The cognitive performance indicator asked students to design a personal fitness program and had the highest level of student competency (57%) of all of the indicators (see chapter 4). Student competency was based on achieving a score representative of at least 70% on the written test. The test did identify problems students had in applying concepts as well as misconceptions students had of some critical concepts in the development of fitness. To this extent the measure of fitness knowledge was effective and appropriate.

The study of the cognitive written test did uncover problems in the knowledge base of students (see chapter 4). Students had difficulty connecting concepts and could not apply concepts they did know as well as they should have been able to. There is also some concern that in spite of the idea that most students were able to use fitness concepts to develop a personal fitness program, the programs they designed were more associated with narrow perspectives of conditioning and exercise rather than broader notions of being physically active. Corbin, Pangrazi, and Welk (1994) distinguished two models for working with fitness. The first is what they call the Exercise Prescription Model (EPM), which focuses on higher intensity and shorter duration activities. In contrast to the EPM, the Lifetime Physical Activity Model (LPAM) focuses on activities of longer duration and lower intensity more characteristic of lifestyle activities. In the article the authors suggest that the EPM is more appropriate for adults and the LPAM more appropriate for children. Clearly the current trend is more toward lifestyle issues rather than conditioning to maintain health. Few students in this first data base included activities other than conditioning exercises in their responses (e.g., bicycling, swimming, yoga). Instead of focusing on how to develop fitness components through conditioning programs, we wonder if we should be helping teachers to focus students on how to maintain a physically active lifestyle, the LPAM model? A more reasonable solution might be to focus on both as separate and important components of an education for a physically active lifestyle.

In spite of our efforts to design a written test for health related fitness that required constructed response answers that could not be easily memorized, it may still be too easy for teachers to teach to a narrow perspective of the content. The test questions will require revisiting after each school has been assessed one time.

For many teachers, teaching cognitive knowledge to this level of understanding was a new experience. Teachers were supportive of this indicator and teachers who taught the fitness enjoyed teaching the fitness in spite of some resource problems in schools. Issues related to student performance in this indicator are most likely issues related to how well the material was understood and taught by teachers (see chapter 4). Teacher knowledge of fitness is an area that needs investigation.

Performance Indicator 3 (Outside Activity)

Many teachers were not supportive of the third performance indicator requiring students to be physically active outside of physical education. Clearly they did not accept responsibility for what students did outside of physical education and they were not equipped with teaching strategies to attain high levels of student compliance for this indicator. The idea that teachers do not see helping students to make the connection between what students do in physical education with what they do outside of physical education is problematic. Unless teachers make a concerted effort to help students make the transition to what they do outside of class than it is unlikely that most students will make the transition.

A lot of the teachers' lack of support for PI-3 can be attributed to teachers not feeling in control of what students do outside of class. This was a persistent feeling among teachers in both formal interviews and informal discussions (see chapter 3). Teachers do not want to be held accountable for what students do outside of their class and they do not see that what they do as teachers inside of class can have an effect on what students do outside of class.

A concern often expressed in training programs was the difficulty of students in rural areas or the inner city to access opportunities for physical activity. There was little relationship between school SES and student performance in PI-3 which means that community resources most likely did not have a significant effect on competency for this indicator (see chapter 2). Teachers in low economic areas found ways to create opportunities for students to be participants outside of the physical education class.

Student achievement in PI-3 (outside activity) appears to be largely attributable to the teacher. These data indicate that some teachers were effective at getting very high levels of student compliance with PI-3. Documentation of how teachers who have been successful getting high levels of compliance in this indicator will be a useful contribution to the physical education and public health policy literature.

Performance Indicator 4 (Fitness)

Students were considered competent in fitness only when they reached the healthy fitness zone in all five components. The state school average was 28%, which is not so different from other large data bases in the fitness area. The state average includes classes of unacceptable data for which a zero was scored for a class. Some of the low scores may be attributable to data that was not accepted, because the teacher failed to administer the curl-ups correctly or failed to follow other protocols, which may mean that the state average is actually even higher. Nevertheless the scores were not encouraging and the issues involved in requiring students to be fit in each component of health related fitness will need to be revisited.

There was a great deal of variability in class scores for fitness. Some schools were able to get a large percentage of their students in the healthy fitness zone but most did very poorly. One of the key inhibitors identified in this data collection was the relationship of fitness with SES. There was a negative relationship between the scores on the Fitnessgram and the poverty index (SES). Students in low SES schools are less fit. This relationship is supported by the literature and is attributed

to differences in diet and physical activity levels of low SES students, particularly African American females (U.S. Department of Health and Human Services, 1996).

Policy makers in South Carolina are not ready to interpret the low school fitness scores as evidence of an unreasonable expectation. Two key strategies for affecting the fitness level of students would be to focus on early intervention in the K-12 program and strategies for encouraging students to be physically active and to work on fitness outside of class time with accountability for personalized goals in class.

At this time many physical educators do not think it is reasonable to hold a teacher accountable for high student fitness levels. They also do not see their programs extending beyond the physical education class. Unless we can help teachers to adapt more effective strategies for helping students to be participants outside the physical education class we will probably not be effective in changing fitness levels of most students with limited program time.

Appropriateness of Setting Benchmark Scores

Academic performance of a school is converted into a five level Likert scale for the school report card. Benchmark scores were not established for this data set nor were they converted to a Likert scale score. School scores were reported in terms of the percentage of students competent for each indicator and were weighted for a final score. Studies of high stakes assessment in academic testing would seem to show that school scores are likely to initially go up with the first administrations of the assessment program and then level out (Linn, 2000). There is every reason to believe that the same will not be true for physical education. Teachers will begin to teach to the test and will probably become better at achieving student competency in all of the indicators. They will learn what it is that is important to teach and with practice get better at teaching it. The present scores suggest that there is room for improvement in the scores without the test "topping out." This means that the measures and performance levels set for competency can accommodate improvement. Low scores on indicators such as fitness scores (28%) will also have to be revisited to determine their appropriateness.

The Design of the Assessment Program

The assessment program was designed as a program evaluation. Unlike many program evaluations it uses student performance scores to determine program effectiveness. Unlike many state level assessment programs it samples teachers and classes and assesses a school only every third year. There were many reasons for sampling classes rather than testing all of the students. Among the primary reasons were the cost of handling data from more classes and the time the teacher would have to spend in formal assessment. As with any sampling program there are many issues regarding the representativeness of the sample.

Each teacher in this data collection was required to collect data on two different movement forms and each of the other performance indicators. The policy of SCPEAP requires data collection on a minimum sample of 25% of all the classes a teacher teaches. For many of the teachers represented in this data collection the sample was larger than 25% of the classes they teach. Our sense is that the teacher and school scores are representative of the school programs and the work of the

teacher. However, we are sensitive to the idea that we will have to determine that in a more formal manner.

A bigger issue relates to the “every three years” part of the program. Will schools revert back to what they were doing after they are assessed? Will they make an effort only for the year in which they are to collect data? The three-year period should give poor programs the time they need to improve what they are doing. Whether schools use the time to grow or just as a “breather” between assessments will depend on many factors. One of the most important factors is the extent to which teachers view the assessment program as a relatively permanent part of the education system. How administrators and the community see the physical education score on the school report card and the amount of help teachers receive to improve their programs will also be factors.

The most important factor affecting the degree of change teachers and administrators are willing to make as a result of the assessment program is a very practical one. For the past several years the state budget has been in an economic crisis. For the past two years the administration of SCPEAP has not known and schools have not known until way into the summer months whether or not the high school assessment program was going to be funded for the next year. This kind of uncertainty and last minute implementation has a very negative effect on school and teacher support. Getting new funds to support the assessment program in a state budget crisis has been most difficult. If administrators do not think that the assessment program will continue, school support for improving the physical education program is likely to be withdrawn.

A second factor that will affect the extent to which teachers and administrators are willing to make changes as a result of the assessment program is related to community reaction to the report card scores in physical education. The only real accountability part of this program is public reporting of the physical education program evaluation score to parents and the community. If an administrator or a community doesn't see a poor physical education score as a problem then there will be no incentive for a program to improve. Because the scores for this data base were not made public, we do not have any basis to judge whether any community cares enough about physical education to be concerned about a poor score. We do have some anecdotal indications that school administrators do not want a poor physical education score.

The concern school administrators and the community has for a low physical education score is likely to be most related to whether a school is a high performing or low performing school in the core academic areas reported on the state report card. High performing schools do not want to be considered inadequate in anything and are likely to make sure that they do not receive low scores for their school in physical education. On the other hand, low performing schools have low scores in many areas and administrators and the community are less likely to be concerned about a low score for physical education.

The final factor that will affect whether or not schools make change as a result of the assessment program is whether low performing schools are helped to get better by either the local district or by programs run by the assessment program. Plans are to target poor programs and offer them the help they need to improve. It will be critical to make low performing schools aware of what is inhibiting their efforts to develop a good program. It will also be necessary to give low performing schools some strategies for coping with these problems.

The Assessment Materials

One of the most important and generalizable findings of this study is that teachers can use a set of scoring rubrics accurately to observe student performance in authentic settings. This finding is strongest when teachers do not have to make difficult discriminations, when they are given good materials, and when they are held accountable for the accuracy of their observations (see chapter 5). This means that the assessment of motor skills, which has often been neglected in formal programs of evaluation, can be done accurately in authentic settings.

Assessing motor performance has always been problematic for physical education because there are no permanent products. The decision to video tape, have teachers score students from the tape, and confirm the accuracy of the teacher observations of student performance required the extensive development of standardized protocols, policy and procedures. Materials had to be created that would standardize the process of data collection and would make it possible to confirm teacher scores for individual students. Videotaping also requires the teacher to do a formal assessment of a one-time performance rather than to observe students over a period of time. Assessing student performance using videotape is the most time-consuming part of the process for the teacher and confirming the accuracy of those assessments is the most expensive part of the process for the assessment program. The ability to confirm the accuracy of teacher assessments is essential if good data are to be collected.

Unlike state assessment materials in other subject areas, the physical education assessment materials were developed on a very tight budget. At the time they were being developed there were few models or national materials to use to guide their development. Our materials were not developed by experts in assessment or in most cases professionals who have skills in writing for others. They were not developed with a great deal of funding and depended on volunteer help over a period of years. The materials are copied and not printed. In spite of the limitations of the process used to develop the materials, the results of this monograph would seem to clearly indicate that teachers could use them and that in general they were appropriate. The materials were piloted prior to their use and rely on content validity and observer agreement as measures of validity and reliability respectively. This is not to say that the materials for the program were not thoughtfully produced. They were, but they were produced primarily by professionals whose expertise did not necessarily lie in the development of assessment materials. On the other hand, we weren't quite sure where to look for this expertise in physical education.

The Generalizability of the Program

There was no master plan to make change in physical education programs in South Carolina. Few decisions were made a-priori. We set out to make state level change in the high schools but did not set out to do a state assessment program at all levels. We created a state level assessment program because for us it seemed like the only avenue to make the kind of change we wanted in the education climate of the time. We created the program by taking advantage of the opportunities given to us and creating the opportunities where they did not exist. We grew to

understand and then to work within the political environment of the state to create policy and to get the support we needed for the program.

Given the scant resources and an ever-volatile political environment, we have been largely surprised by how well it has worked and the results that we have achieved thus far. In spite of the fact that it took seven years to send the first report to the schools, we know that this report of the first data collection is just a beginning report of the effects of a state level assessment. We have been encouraged by the results. We are also fully aware that the more important questions and consequences of a program of this sort lie ahead of us. We approach the future with a cautious optimism.

Each state is unique in terms of the political environment that governs education and the manner in which the business of education is conducted. South Carolina has embraced the standards, assessment, and accountability movement. It is likely that the new federal initiative *No Child Left Behind Act of 2001* (U.S. Congress, 2001) will push other states not now in this movement, into this movement. We are in an educational climate of accountability and evidence (Slavin, 2002). Educational programs that are not willing to both identify their outcomes and measure the extent to which they achieve those outcomes are at a real disadvantage.

To the extent that a state shares the characteristics of South Carolina, our experience will be directly generalizable. Many other states have competent supervisors and established supervision systems for physical education at the district level. District personnel in charge of physical education in South Carolina usually do not have their preparation in physical education and have multiple job responsibilities. Administrators at the school level have either been unwilling or unable to play strong supervisory roles for the physical education program. In this context, South Carolina was forced to work directly with the teacher from the state level. States with a strong supervisory system at the district level for physical education programs can, and probably should, consider other models.

More important for other states is what we have learned from doing this that will be of benefit to other states. We have learned that identifying expectations for student performance, developing assessment materials to measure the degree to which students meet those expectations, and developing some kind of accountability for achievement of those expectations has great potential to produce change in physical education programs. We have learned that at the early stages of the program the changes made have been overwhelmingly positive and very well received by teachers. We have been encouraged by the results.

Research on changing policy to achieve change in physical education is new. We sought to make change through changing state education policy. This was relatively unknown territory for physical educators. We have learned that you have to be willing to participate as a player in a complex political process governing education to achieve change. The process and the players in each state are likely to be different and the process seems far more an art than a science. The reality of today's educational climate is that physical educators need to know how policy is created and who the key players are for the policy being sought.

What we have most learned from this experience is patience. Each day we gain a greater appreciation for the amount of time it takes to produce change. It

was eight years between the enactment of the law describing what the high school program should be and the first set of high school data. The eight-year period was not a steady climb to success, but rather a roller coaster of successes and disappointments. We are just now beginning to see real change in some schools and our position in the state as an assessment program has not been financially secured.

We have viewed our mission as one of changing school culture in physical education from non-teaching and non-outcome oriented programs to instructional programs that are outcome oriented. The real change in schools is expected to come within the next five-year period when all schools will have been assessed twice. Our perspective on how long it takes to create change has changed.

In recent years, lack of physical activity has been identified as a major contributor to chronic disease and a long list of other medical problems. The financial and human cost to the country in terms of the percentage of the gross national product devoted to medical care has encouraged policy makers to search for ways to increase the physical activity levels of youth and adults. Physical education is not the only contributor to this goal but certainly should be considered and supported as a major contributor to this goal.

As a profession, we need to know how to help physical education programs be more effective in developing a physically active lifestyle. Our knowledge base in how to produce change is not extensive. What we have tried to do in this monograph is to introduce the idea that a state level assessment program for defined outcomes may be a way to improve high school programs. We have also suggested that because physical education has not had a shared vision of what good programs should be and because they have had no accountability for student learning, it may be the only way to initiate large-scale change in our field.

Cuban (1998) defines success of a reform movement in terms of five criteria: the reform's effectiveness, fidelity, popularity from a policy maker's perspective, adaptability, and longevity from a practitioner's perspective. In the future, each of these characteristics will need to be explored relative to this project if the research is to make a valuable contribution to the reform literature in physical education. This monograph is a beginning.