

Chapter 1: Developing and Implementing a State Assessment Program

Judith Rink and Lori Williams

University of South Carolina – Spartanburg

School reform efforts come and go with cyclic frequency. Many educators assume that if they wait long enough they will be untouched by their presence. The short lifespan of reform is apparent whether talking about comprehensive school reform efforts that are designed to make wide scale change across schools or less ambitious efforts limited to districts or individual schools.

The current school reform movement has had uncharacteristic staying power. Its onset can be traced to the publication, *A Nation At Risk*, in 1983 (National Commission on Excellence in Education), which painted a rather dismal picture of the education system in the United States. The subsequent legislation, Goals 2000: Educate America Act (U.S. Congress, 1994), outlined a comprehensive effort to fix America's schools. The uniqueness of the Educate America Act was its comprehensiveness. The Educate America Act was one of the first national efforts at comprehensive school reform to take a systemic approach to producing change in schools. Built into the legislation was an attempt to address multiple initiatives including parent involvement in schools, technology, student opportunity to learn, and teacher development. Additional initiatives included the development of curricular content standards and the assessment of those standards. A systemic approach acknowledges the complexity of school change and attempts to work with the entire system to institutionalize change (Fullan, 1991, 2001). The current standards, assessment, and accountability initiatives in many states have roots in this initial national legislation and in efforts to describe effective schools.

Inherent in the current movement is an assumption that good schools are effective schools (Cuban, 1998). Effectiveness is usually defined narrowly in terms of academic achievement. Other purposes of schools related to social justice, developing student social skills, student health, or the personal development of students has not been considered criteria important enough from a policy perspective to warrant assessment.

The narrow interpretation of school effectiveness has had an impact on physical education programs. At the national level, the importance of health and physical education has received support (U.S. Department of Health and Human Services [DHHS], 1996). At the state education level, the concern for student health or the development of a physically active lifestyle for students and adults has been treated mostly as a distraction from an academic agenda.

The standards, assessment, and accountability movement began with national efforts to describe what every student should know and be able to do. *Moving into the Future: National Standards for Physical Education*, like the materials

from other content areas, was developed by the content area national organization (National Association for Sport and Physical Education [NAPSE], 1995). The national standards were meant to guide rather than be a blue print for the establishment of state and local standards. Like many states, South Carolina used the national standards in most content areas, including physical education, as a beginning point to establish state standards.

Inherent in the standards, assessment, and accountability movement is the notion that schools and teachers should be responsible for outcomes and not process. Many states abandoned program requirements and the micro managing of the process of schooling for site-based management and an emphasis on outcomes (Holloway, 2000; Schmoker, 1996; Wholstetter, 1995). An extensive system of testing and accountability for outcomes, and in some cases public accountability (high stakes assessment) for those outcomes, took the place of extensive regulations describing time requirements for different subject areas.

South Carolina embraced the standards, assessment, and accountability movement. The state legislature mandated testing in all *core* subjects and the development of a report card for schools that describes student achievement for a particular school in those subjects. Physical education was left without a place at the table and no regulations mandating that schools even have a physical education program at the elementary and middle school levels. The exclusion of subject areas not considered *core* is typical of high stakes accountability systems (Linn, 2000). Increasing pressure for academic performance made physical education not only a marginalized subject area within the school curriculum but also an endangered subject area in the school curriculum.

It is not uncommon for physical education to be left out of school reform efforts. Ward and Doutis (1999) in their review of reform efforts and the role of physical education in those efforts concluded that by and large physical education has not been included in any of the recent and significant reform movements in education. They attribute this exclusion to the idea that physical education is not a core subject in the schools and physical educators have not made a case for being included. The discussion of physical education's participation in reform movements is extended by Rink and Mitchell (2002) who identify a lack of preparation for participation in the process and a reluctance of physical educators to participate in the political process that develops policy as additional factors related to the exclusion of physical education in more large scale reform efforts.

As a profession physical education has simply not been prepared to be included in most reform movements (Rink & Mitchell, 2002). In the standards, assessment, and accountability reform movement, a subject area has to have clear and measurable expectations for student achievement. Physical education as a field has been very reluctant to describe outcomes or goals with a useful level of specificity. As a field we have not recognized student achievement as an indicator of a "good" program (Ennis et al., 1997; Locke, 1992). Student achievement is the objective of the standards, assessment, and accountability reform movement.

Physical education has been included in the standards, assessment, and accountability movement to the extent that the national professional organization (NASPE) was asked to develop national content standards. Many states have also established state standards for physical education. If we assume that physical education has measurable outcomes or that states can define measurable outcomes from the standards, then what physical education lacks is a way to measure those

outcomes comprehensively and policy that creates some kind of accountability for the achievement of those outcomes.

The reform literature in our field and efforts to create change in physical education have primarily targeted the school or district level. Our literature focuses primarily on documenting the working conditions of physical educators and calls for curriculum change with little attention to how to produce that change (Rink & Mitchell, 2002; Ward & Doutis, 1999). Efforts to make change in schools have primarily been associated with teacher development orientations rather than changing policy (Faucette, 1984; O'Sullivan, Tannehill, Knop, Pope, & Henninger, 1999; Ward & Doutis, 1999). We have not addressed changing policy as a mechanism to produce change in physical education (Rink, 2001), nor have we been able to solicit support for large-scale change in school physical education programs.

The need for change in physical education school programs has been well documented (Rink, 1992; Siedentop, 1987; Stroot, 1994). How to produce that change is less clear. This monograph reports the efforts of physical educators in one state to change school physical education programs through what Fullan (2001) would call whole-school reform. Professionals at all levels in the state mobilized to become part of the standards, assessment, and accountability reform movement dominating the conduct of education in the state. The process involved developing materials, establishing state policy, and extensive teacher development.

Creating change in South Carolina involved building consensus, for a vision of what physical education programs should achieve, and obtaining the support of key policy makers. Policy makers included administrators, legislators, and the State Department of Education (SDE). It is a process that began with legislation in 1994 (South Carolina General Assembly, 1993) and has taken us to the first statewide data collection of high school performance in 2000.

This monograph is a report of that journey and the initial results of a statewide assessment program. This first chapter begins with a brief historical overview of the process we used to initiate change and a description of the assessment and accountability program that is the center of the program. This chapter includes a description of what we considered to be the literature and research that informed the project and concludes with a brief description of each of the monograph chapters.

A project of this magnitude has involved many professionals at many levels of employment. These people are identified on Table 1. The use of the term "we" is used throughout this monograph to give credit to their participation and unwavering support.

Historical Overview

The public school programs in South Carolina have, as a whole, never ranked at the top of anyone's list for quality. In spite of the many well-intentioned efforts to reform schools, students in South Carolina are usually ranked at or near the bottom in achievement on national tests and indicators such as SAT scores. As a result of poor school performance, South Carolina policy makers sought change and embraced the standards, assessment, and accountability reform movement to produce that change. Legislation was established that holds schools accountable for student academic achievement through extensive testing and public reporting of school performance.

Table 1 Participants in the Development of the South Carolina Assessment Program

Jane Abbott	Barry Frishberg	Anne Long	Linda Wannamaker
Marie Andrus	Pat Frye	Tracey Long	Jackie Weedon
Scott Arrington	Steve Furney	Rachel Mahaffey	Eugene Wang
Jennings Austin	Ritchie Gabbei	Kathy Manos	Mary Werner
Suzan Ayers	Cindy Gallman	Gwen Massey	Peter Werner
Gina Barton	John Gentry	Karen McKinney	Jimmy White
Brenda Beaty	Pat Gold	Murray Mitchell	Cindy Wilkerson
Debbie Bernhagen	Mary Goodwin	Ronald Moore	Jacob Wilkerson
Marcia Berry	Jennifer Gorecki	Lynda Nilges	Lori Williams
Jean Blayds	Kathy Graham	Jenny Norris	Gary Wilson
Penny Bostain	Sherry Haithcock	Tracy Orvin	Kim Wilson
Eve Branyon	Jenny Hallman	Sue Owens	Chris Wirszyla
Walt Bray	Bob Hampton	Ezell Parker	Ann Witwer
Cathy Brooks	John Harper	Katherine	Toni Wood
Earl Brown	Laverne Harris	Pebworth	Judy Wyatt
Bonnie-jean	Joe Hauf	Kathy Peebles	Dan Young
Buckett	Brandie Hewett	Leslie Pizzuti	Freddie Young
Mark Buffamoyer	Pat Hewitt	Rich Rairigh	
Ellen Campbell	Cindy Hickman	Elaine Randles	
Adelaide Carpenter	Gina Hilts	Stephanie	<i>Special Thanks</i>
Don Carr	Tammy Hodges	Richardson	Barbara Ainsworth
Karen Carter	Dick Hohn	Megan Righter	Susan Agrusso
Marian Carwile	Athena Hortis	Dana Riley	Charlene Burgeson
Darla Castelli	Kathy Ingram	Judy Rink	Bill Chaiken
Judy Causey	John Jacumin	Debbie Rogers	Joanne Fraser
Gina Chapman	Kay Jackson	Karen Roof	Warren Giese
Steve Cooper	Lindsay Jameson	Phil Savitz	Sandra Lindsay
Dianne Cotney	Leonard Johnson	Miriam Sheldon	Cindy Saylor
Gylton Da Matta	Theresa Johnson	J.D. Simpson	Nicki Setzler
Regina Flynn	John Kading	Kate Stanne	Harry Stille
Damon	Mildred Kennedy	Karen Stevens	Inez Tennenbaum
Barbara Davis	Cay Kessler	Susan Stewart	Howell Wechsler
Lisa Davis	Cindy Kessler	Skip Strainer	
Marie Dawkins	Kym Kirby	Deborah Stroman	
Panayiotis Doutis	Kelly Kowalchick	Kathy Sullivan	
Ruth Earls	Maxie Krause	Sharon Supplee	
John Farrelly	Tom Langley	Mickey Taylor	
Rosa Ferguson	Sandra Lee	Josey Templeton	
Dave Fleming	Harry Lehwald	Lou Thomson	
Lori Florence	Renee Lemmon	Lula Thompson	
Lyman Foster	Andrew Lewis	Mary Lou Veal	
Karen French	Peggy Lineberger	Jennifer Walton	

The potential of the standards, assessment, and accountability movement to either positively or negatively impact physical education programs in the state was recognized early. The emphasis on accountability for academics only would threaten the status of any school program not included as part of the assessment and accountability program. We chose to make sure that we were a part of the movement rather than be negatively affected by it. The process we used to do this has spanned almost 9 years and is described in Table 2.

**Table 2 South Carolina Physical Education Reform Effort
Time Line of Events**

1994–1995	High school legislation established
1995	High school SDE committee establishes criteria (performance indicators) State Department of Education contracts with SCAHPERD to do teacher development for the new legislation
1995	First High School Physical Education Institutes (PEI's) 12 colleges and 24 faculty design and conduct the inservice 75 schools completed all year long in-service the first year 185 schools out of 217 have been represented (1995 -2001)
1996–1998	Formative assessment materials developed for the high school Fleming and Wirszyla studies conducted
1996–2000	State curriculum framework designed to be consistent with high school material
1998–1999	Assessment materials developed and piloted by state college and university faculty
1998–2000	Elementary and middle school writing teams determine indicators and develop assessment materials (primarily public school teachers)
1999	State Education Oversight Committee places physical education on the report card
1999	State Department of Education contracts with SCAHPERD to do state assessment. SCPEAP is established as a standing committee of SCAHPERD.
2000–2001	Cycle 1 high schools are trained spring 2000 to collect data in the fall and submit it spring 2001
January 2002	Schools receive reports of first data collection
Spring 2003	Cycle 2 high schools submit data

Developing a Shared Vision

In 1994 a coalition of state groups interested in the health and physical activity of children and youth sought legislation that would clearly articulate the nature of the one year required high school physical education program. The legislation was designed to move programs away from team sport dominated curriculums and throw-out-the-ball programs, toward instructional programs focusing on health related fitness. The legislation did not ask for funding and passed easily. The State Department of Education (SDE) appointed a committee of professionals with representatives from higher education, high school physical education, public school administration, and the SDE to draft the specific policy. The deliberations of this committee resulted in describing what students should know and be able to do as a result of the 1-year high school physical education requirement.

The committee designing the student performance indicators (initially called performance criteria) wanted the performance indicators to be measurable and achievable for all students participating in good programs. The materials were designed to prescribe outcomes but not curriculum. The performance indicators were intended to be culminating, few in number, critical for a comprehensive program, and measurable. Four performance indicators were identified below and are described in Appendix A.

- PI – 1 Demonstrate competency in two movement forms
- PI – 2 Design and develop a personal fitness program to reach a desired level of health related fitness
- PI – 3 Participate regularly in physical activity outside the physical education class
- PI – 4 Meet the health related fitness standard for their age and gender as described by Fitnessgram

The performance indicators were designed to give students skills to lead a physically active lifestyle. We asked ourselves the questions, “If teachers taught to these outcomes would it be a better program than what most of our schools were doing at the time?” and “Would the program contribute to developing a physically active lifestyle?” We did not want to focus on only fitness or student cognitive knowledge of fitness. We wanted programs to help students make the transition to being active outside of physical education. We considered competence in motor skills to be a major factor contributing to the development of a physically active lifestyle and we didn’t want to exclude this outcome because it was difficult to measure.

Teacher Development

At the time the high school physical education legislation was to go into effect, the SDE did not have a physical education consultant and asked the South Carolina Alliance for Health, Physical Education, Recreation, and Dance (SCAHPERD) to conduct teacher development for the new high school legislation. This was the beginning of a relationship between SCAHPERD and the SDE that would prove to be a major facilitator in the reform effort. The SDE contracted SCAHPERD to provide the teacher training for the new legislation.

Most of the college and university leadership of SCAHPERD had spent entire careers doing *one-shot* in-service programs with teachers that had little or no impact on what happened to students in physical education. We were not willing to participate in another professional development experience that had no impact on the quality of programs being offered students. We decided to set up the teacher development program as long-term with a few schools rather than short term with many.

The strategy for long-term professional development involved what we identified as the Physical Education Institute (PEI). This strategy included full-day sessions, delivered across an academic year, that were intended to build upon each other—in contrast to the usual professional development model of independent, occasional sessions.

The first Physical Education Institute (PEI) was designed as a program that would meet in five sessions scheduled throughout the year and was focused on helping teachers establish school and department policy and curriculum to meet the state standards (which were mostly the national ones) and the performance indicators. Teachers were required to sign up for all five sessions. At the time we expected ten schools to come. Seventy-five of the 203 high schools in the state were represented at the first institute. This was our first introduction to the power of policy and invitations to schools sent with SDE letterhead. One of the sessions of the PEI was the state SCAHPERD convention that served to introduce many practicing high school teachers to their professional organization. Schools received substitute pay if they attended all five sessions and completed the required assignments. This was the first and only year monetary incentives would be provided to teachers.

The Physical Education Institutes (PEIs) were designed to help programs do what they needed to be effective in getting students to meet the standards and the performance indicators. Teachers were pushed to make changes that would align their programs with the standards. Public reports of changes teachers made from one meeting to the next became a strategy used to encourage action to make change. We didn't want to dictate curriculum and instruction but we did want to equip teachers with strategies they could use to align their program with the standards. We wanted to move schools toward longer units, a broader curriculum, and a selective "choice" curriculum. We also wanted schools to teach for a physically active lifestyle. To facilitate change to allow for student choice, teachers were provided information on how to change the state scheduling software and how to work with guidance counselors and administrators. The committee reviewed the state textbooks for physical education and helped teachers to order them.

The first PEI was conducted by a coalition of college and university faculty from the state. Twelve colleges and universities and 24 faculty called together by SCAHPERD met on a regular basis to plan each session. Some of the teacher development sessions were held centrally in Columbia and some regionally. This was the only year of the program that university faculty dominated the conduct of the program. As public school teachers took ownership of the ideas they were quickly moved into leadership positions and assumed roles as presenters. University and college faculty make up only a small part of the work force of the project at this point.

PEIs have been conducted every year since 1995. With each year more schools were represented at the PEIs. To accommodate newcomers, several sections of

PEIs were conducted to meet the diverse needs of the first, second, and third year participants. Over the years all but a few high schools have been represented at the institutes. As schools have made changes, the nature of the institutes has changed from establishing policy and curriculum to teaching content effectively and assessing students both formatively and summatively. As schools began to take ownership of the ideas we set up site schools (where some parts of a good program were apparent) and demonstration schools (where most desirable parts of a good program were evident) so that teachers could see programs in action.

Across the years, we have tried various strategies to facilitate change. For example, one year we established an “essential friend” program and paired university faculty with a school that requested help. We found this to be an excellent approach but far too time consuming for college faculty with full time jobs. We designed a session called “Time Out For Planning” and encouraged entire departments to attend and to bring their administrators. Again, this was a wonderful opportunity for departments that brought their administrator and had most teachers in the department in attendance.

With each year of the teacher development program, group diversity in understanding the standards and the assessment program became more and more problematic. We began offering a variety of sessions from which teachers would choose what was appropriate for them. In later years, it was more difficult to get teachers released to come to in-service workshops or to participate long term in conducting the workshops. A school report card indicator entitled, *teacher attendance*, reported to parents the number of days teachers were actually in front of their classes without distinguishing the reasons for not being in front of their classes. Because of this public record administrators were hesitant to release teachers for any purpose.

The Need for Assessment Materials

The first set of assessment materials for the indicators was developed during the first year of the project. There were no funds for this effort so we entered into a partnership with a school district to develop the materials. The district was willing to give teachers time for the project during the school day and we invited a few college and university faculty to work with these teachers. The first set of assessment materials were not designed as summative assessment, but as formative assessment to be used to assess students over a unit of instruction (Appendix B). This first set of assessment materials served the purpose of defining competency for teachers and students.

It was important to us to provide assessment materials for a broad range of physical activities and therefore materials for 21 different activities were developed for performance indicator one. We wanted to increase and not limit the number of movement forms schools offered students as choices. The assessment rubrics took 2 years to develop and are good formative assessment tools that include activity specific affective and cognitive indicators.

During the third year of the project it became clear that the state was moving toward high stakes assessment. The state legislature appointed an Education Oversight Committee (EOC), of mostly non-educators, whose task was to develop a state report card for schools (South Carolina State Department of Education, 1998). The SDE was working feverishly to develop assessment materials for all of the “core” content areas. We lobbied the SDE for support to begin to develop summative

assessment materials for the high school physical education program and received minimal funds.

Again a call went out to college and university faculty, this time to help develop and pilot the high school assessment material. We brought in a consultant (Mary Lou Veal, University of North Carolina at Greensboro) in physical education assessment to help us define the task and suggest an approach. Ten faculty throughout the state worked for a year to develop assessment materials for all four indicators. We had to change our mindset from formative assessments used by teachers during instruction to developing formal assessment tasks with measurable criteria and standardized protocols. The assessments would not be used over a period of time but would need to be a one time practical measurement of each indicator.

In the spring of the second year of development, college and university faculty piloted the assessment material in ten large high schools throughout the state. In some cases a large pilot data collection team spent several days at a school. Sample protocols and scoring rubrics for PI-1 are provided in Appendix C. When the high school assessment materials were finished, we went back to the SDE to ask permission to develop performance indicators and assessment materials for the elementary and middle school programs. The programs at these levels were developed very differently than the high school materials and will not be reported in this monograph.

Obtaining Accountability

The state legislation mandating the outcomes of the high school physical education program was a step toward accountability but it was not sufficient to ensure change. Early research on the project indicated that the state mandate was only partially effective in creating a climate for change (Fleming, 1998; Wirszyla, 2002). We needed to establish an assessment program that created some accountability for having a good program. To do this we would need policy that created accountability, summative assessment materials, and a strategy for conducting a large-scale state assessment.

We started a campaign, with health education, to be included on the state report card. Our efforts were hindered by the fact that the state accountability legislation specifically mandated that scores for core academic subjects would be on the report card and provided the funding for only those subjects. Other content areas were not permitted to use the extensive funding provided for the core subjects.

It took 2 years of letters, e-mail, and fax campaigns to get a meeting with the Education Oversight Committee (EOC), the organization responsible for selecting school report card content. Our meeting with the committee focused on our contribution to academics, the health consequences of a lack of physical activity, the special needs for accountability of marginalized content areas in the schools, and the responsibility of the committee for unintended consequences of only assessing core subjects. We reinforced the idea that we were ready. The committee voted immediately to give physical education one line on the report card in a section providing the public with a list of indicators of a good school. We requested to report the percentage of students who met the state standards for that school, the state average, and whether the school increased or decreased from previous performance on a once every three years data collection cycle. Data collection was to

begin with the 2000–2001 school year for the high school. Later the EOC would make the decision to convert all scores on the report card to a five level Likert scale.

SCAHPERD would be the contracting agency to develop and conduct the assessment program in physical education. The South Carolina Physical Education Assessment Program (SCPEAP) was established as a permanent sub committee of SCAHPERD to conduct the assessment program. The first meeting of an appointed policy board of 21 people was held in January 1999.

How the Assessment Program Works

The South Carolina Physical Education Assessment Program (SCPEAP) is unique in that student performance is used to do program evaluation. A sample of classes from all teachers at a school is assessed on each of the indicators to determine school effectiveness. Program assessment does not provide information on every student but rather makes the assumption that if the program is good, individual students will achieve. The program was designed to try and avoid many of the potential problems of high stakes assessment.

The assessment program is governed by a policy board (21 members) of administrators, K-12 teachers, university faculty, SCAHPERD representatives, and SDE representatives. The first task of the policy board was to establish policy manuals that would govern the collection, analysis, and reporting of data. SCPEAP maintains an office with a part-time administrator and secretary who are available to answer any questions teachers have during the data collection process.

Schools Are Assessed Every 3 Years

School districts were divided into three groups by enrollment and then randomly assigned to one of three cycles. Each cycle is assessed every 3 years. The data reported in this monograph are from Cycle 1 high schools. We felt that we could do a better assessment of fewer schools if we didn't collect data on each school every year. Two years would give schools time to improve their programs and would not consume an inordinate amount of instructional time.

Teachers Are Trained to Collect Data

Physical education teachers as a whole have not been accustomed to assessing students and are not accustomed to using standardized tests in a standardized way. If judgments are to be made about how well a school meets the state standards then the assessment material had to be standardized and teachers had to follow the protocols. Two-day data collection training sessions were designed to help teachers collect data, score student performance, and submit data to SCPEAP. Training sessions were held in the early summer after school was out. Up to two teachers from each school in the cycle (1/3 the schools) were invited to the training session. With some prodding and reminders, all schools were represented at the data collection training in the summer of 2000.

The training sessions were designed as hands-on experiences and teachers were asked to bring video cameras so they could actually practice setting up the camera, video recording and using the protocols. Many cameras arrived in unopened boxes. Teachers were given assessment manuals with explicit instructions

on administering and scoring each assessment as well as a user friendly training CD-ROM that had video clips of each of the performance levels for many of the activities.

Schools Submit a Sampling Plan

In the fall of the year a school is going to be assessed, they submit a sampling plan describing their teaching schedules and designating which classes for a teacher will be assessed. All teachers must submit data on two classes for each performance indicator, providing they teach at least two classes. Administrators must sign the sampling plan. A sampling plan committee of SCPEAP reviews the plans and approves them. The sampling plan then becomes an assessment contract defining the data that must be submitted by a school in the spring.

Teachers Collect the Data Any Time During the School Year

The administration of and preparation for the Palmetto Achievement Challenge Test (PACT), which is the state test of basic academic skills in South Carolina, consumes an inordinate amount of instructional time. Students and faculty are so nervous when the test approaches little gets done academically. We wanted testing in physical education to be part of the instructional process and we wanted the teacher and program to decide which content would be offered and when it would be offered and assessed. We also wanted teachers to collect and analyze the data so that it could be a learning experience for both and so that it could provide immediate feedback to the teacher and the student. What is on the test and how well students did should not be a secret disclosed only in the distant future. Teachers score student performance using scoring rubrics, test keys, and scoring sheets provided.

Teachers Submit Data in the Spring

School data are due to the SCPEAP office at the close of the school year. Schools must submit a computer generated role sheet, exception sheet for any students not assessed on the role sheet, a score sheet for each class, and records of original student performance (videotape of movement forms assessed, the scored written test, outside participation confirmation sheet and contract, Fitnessgram scores, and a video tape of the curl ups). Administrators must sign off on the submitted data.

A Monitoring Committee Determines the Compliance and Accuracy of the Data

For 2 weeks during the summer, a monitoring committee of peers meets to determine the degree to which teachers have followed the testing protocols and to determine the accuracy of the data submitted. Approximately 12–16 teachers and college and university faculty are trained in the use of the assessment materials and the policies for evaluating school materials. They compare the classes submitted by a school with the sampling plan submitted to determine if teachers submitted the required classes. Data for each performance indicator for each class are

examined. The monitoring committee scores a 25% sample (no less than 7) of students. They compare their scores with the teacher's scores. If a teacher's scores are in agreement (80% or more) with the monitoring committee's score, the teacher's total scores for a class are accepted as being accurate. If there is less than 80% agreement between the teacher's scores and the scores given by the monitoring committee, an additional sample of 25% is drawn. If there is still not agreement a second team from the monitoring committee scores the sample. If total agreement for both teams combined is not at or above 80% the data set is considered inaccurate and unaccepted. Unaccepted data is scored as a "0" for that class.

The monitoring committee also judges the extent to which teachers adhere to the protocols of the assessment materials and identifies any protocol violations. For this first data set, data were accepted as following protocols if the compliance problem did not affect student scores and if the monitoring committee could find a way to assess student performance. Data with minor protocol violations (e.g., forms not submitted correctly or student names not announced on video) were designated accepted/noncompliant. Data that could not be analyzed or data for which student scores were compromised is scored a "0."

Data Are Analyzed and Submitted to the SDE

When the monitoring committee has finished its work, each class of submitted data is determined to be accepted/compliant, accepted/noncompliant, not accepted/noncompliant or not submitted. Final scores for each class are entered into a database by school, teacher, performance indicator, and in the case of Performance Indicator One by activity. A state report of total scores and scores by performance indicator is created. Individual school reports describe total school performance, school performance by indicator, and school performance by teacher. School reports are sent to building principals and the district superintendent of each district. The total database is delivered to the SDE.

The Selection of Indicators and Design of Assessment Materials

Apart from the policy issues that surround the idea of state level assessment, there are important practical issues related to developing assessment materials. Unlike other content areas, physical education does not have a great deal of standardized assessment materials that would appropriately assess the standards, nor do we have the option to hire commercial companies with the potential to develop them. Assessment with accountability has the power to change programs in the direction of the materials you design. Hence, what you measure, how you measure it, and how you interpret what you measure are critical.

The state of South Carolina adapted the seven national standards (NASPE, 1995) with some adaptations in the grade level material. These were published in the document, *South Carolina Physical Education Content Standards* (2000). Only the first four standards are reflected in the South Carolina assessment material. The decision not to formally assess the affective standards (5-7) was based on philosophical, political, and practical perspectives. Although all the professionals in the state are most supportive of the affective standards, the affective standards

were not considered unique contributions of physical education. All school programs share them but no other content area in the state is assessing affective standards. In fact, physical education is the only content area that has included affective standards in their national standards. This may be due to the idea that it is not only difficult to assess individual students in the affective domain, it may be inappropriate to do so.

Selection of Performance Indicators and Assessment Tools

Assessment programs must select performance indicators that are important and *representative* of achievement in standards. Selecting performance indicators that most reflect standards is a difficult undertaking. Physical education as a field has a long history of using written tests and fitness tests as indicators of achievement in physical education when more formal measures are required. The decision to assess outcomes more comprehensively and in particular to include both psychomotor indicators and participation outside of the physical education class involved a leap of faith. We believed that we could do more than what has been done in the past in large-scale assessment projects. The decision also represented a commitment to the role of motor skill competence in developing a physically active lifestyle and the importance of helping students make the transition to a physically active lifestyle.

Setting expectations for performance was a difficult task. A major assumption of the program is that if expectations are reasonable and achievable then there is no reason why programs should not be held accountable for achieving them. Holding students accountable for achievement is an established practice in education. South Carolina has only a 1-year high school physical education requirement. Performance indicators were developed that represented what we felt were reasonable expectations for a program to accomplish within a year.

State level assessment requires you to put a value on a level of performance. Ultimately someone has to decide what does or does not meet a standard. Setting goals and standards is slippery business. Discussions about setting levels of competency for the indicators revolved around whether materials should be criterion or norm based. The decision was to make the assessment criterion based and to write the performance indicators so that they described that criterion. The performance indicators were selected because each was felt to be a contributor to developing a physically active lifestyle.

Performance Indicator 1—Movement Competency. We defined competency in an activity as the ability to independently and safely participate in the activity with enough skill to make it an enjoyable experience and to perform the activity with continuity. In some activities it would be difficult to get students with little prior experience to become competent. For some it would not be difficult. Scoring rubrics for assessing competence in an activity in an authentic setting were developed.

Currently the measures for competence in an activity do not include what might be considered advanced skills, such as offensive skills like the volleyball spike or tennis smash, which may be a concern for some professionals. The concept of educating students to be participants became the criterion through which we filtered our expectations for student performance. We wanted students to be

participants and we tried to stay true to that in the levels of achievement we designated as *competence* in the assessment materials.

Very specific protocols to standardize testing were developed and tried with small groups of participants, revised and tried and revised again. Content validity was established with a committee of qualified professionals (e.g., content specialists, high school coaches and teachers, university teacher educators) and an extensive process of continuous revision and repeated administrations of the assessment materials. Piloting materials with students on a regular basis was a critical part of the process. The initial materials were often inappropriate. The assessment task, the protocols, the criteria for good performance and the levels of performance all had to be refined extensively and continuously. No materials were approved for use unless they could be used by trained observers with an inter-observer agreement of over 80 percent. Appendix D lists the reliability levels for the observation tools. Reported inter-observer agreement scores were obtained with graduate assistants at the university.

Performance Indicator 2—Cognitive Fitness. The inclusion of the cognitive fitness indicator reflected a commitment to the importance of health related fitness cognitive knowledge. We wanted students to be able to design their own health related fitness program based on Fitnessgram case study results from scores they had to interpret. We were reluctant to use a short answer exam. We needed a constructed response design that we could use to develop equivalent forms. We knew from earlier research that teachers were supportive of this indicator in that they thought it was appropriate and didn't feel that students would have great difficulty with it (Fleming, 1998). Repeated pilot administrations of the written test were conducted to refine it. An agreed upon competence level of 70% was set and a standardized scoring system was designed and shared with teachers.

Performance Indicator 3—Outside Activity. Unlike PI-2 (cognitive fitness), teachers did not support requiring students to be physically active outside of physical education. They were reluctant to accept responsibility for what students did outside of physical education. Lack of teacher support was discussed by the Policy Board of SCPEAP before the first data collection. The decision was made to maintain the indicator the way that it was because of a shared perception that this indicator was a critical component in helping students make the transition to a lifetime of physical activity. Students contract with the teacher to participate in physical activity outside of physical education. They include in their contracts the identification of adult contacts to verify their participation. Teachers must confirm student participation with the adult contact.

Many professionals see the overriding goal of physical education as the development of a physically active lifestyle. Teachers did not see helping students to make the connection between what is done in physical education class with what they do outside of physical education. This was considered problematic in our discussions and would need to be addressed in teacher development. 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. The expectation for this indicator was that students would participate regularly (at least three times a week) in moderate to vigorous physical activity outside of physical education class for a period of nine weeks. Programs were encouraged to set their own policies on exactly what activities they would accept or not accept for this indicator.

Performance Indicator 4–Fitness. The fourth performance indicator required students to meet the health related fitness standard for their gender and age as defined by Fitnessgram. Lengthy discussions ensued about the appropriateness of this indicator. In the pilot testing it was clear that in some schools very large percentages of students were meeting the standard and in other schools very few were. We were very aware that even as early as 1995 when the indicators were being designed, there was a great deal of disagreement in the field about holding all students to a single expectation in fitness (1995 national standard 4).

The fitness standard was maintained based on the assumption that the standard was achievable by most students. Teachers were encouraged to design personal standards for students for whom the published standard was not appropriate. Individual student scores were not being reported. The extent to which all individuals are forced to meet a single fitness standard may not be appropriate. The extent to which a program develops fitness, however, was felt to be a valid measure of an effective physical education program.

The Literature that Informed Us

Our work has been informed by both the literature on standards, assessment, and accountability as well as the literature on school change and effective schools. Not only have we been informed by that work but we have tried to use it in developing SCPEAP. We did not initiate the standards, assessment, and accountability movement in our state but we realized early that we could be a victim of it. The issue for us was not so much “Should assessment with accountability be used as a reform strategy in the schools of South Carolina,” but rather, “Should physical education be a part of the high stakes reform effort in South Carolina?”

The Standards, Assessment, and Accountability Movement

In spite of its many critics, the standards, assessment, and accountability reform movement has been shown to be an effective way to produce positive change in school programs (Fullan, 2001; Haertel, 1999). Accountability may even be necessary to initiate change and we had no reason to believe that this approach would not also have a positive effect on physical education programs. High stakes accountability is a form of external pressure. Fullan (1991, 2001) suggested that the presence of external pressure is often necessary for reform to be initiated. External pressure, from policy makers, legislators, and parents, has been identified as a key factor associated with the initiation phase of reform. Fullan (2001) also suggested that a balance between internal pressures (e.g., high administration and teacher expectations of student performance) and external pressures are more likely to result in second order change in schools (alterations in the fundamental structure).

The standards, assessment, and accountability movement works positively in that it can develop a shared vision of what programs ought to be doing by focusing teaching on achieving defined outcomes and creating some kind of accountability for achieving those outcomes (DeStefano & Prestine, 1999; Fullan, 1991). Inherent in this movement is the potential to create high expectations, clarify the purpose of programs, and focus teachers on the learning experiences of all students (Fullan, 1991). South Carolina high schools did not have a shared expectation and

purpose, nor could they be characterized by a focus on learning experiences for all students.

A position paper by the American Educational Research Association (AERA, 2000) reiterated the potential value of high stakes assessment. "It is hoped that setting high standards of achievement will inspire greater effort on the part of students, teacher, and educational administrators. Reporting results may also be beneficial in directing public attention to gross achievement disparities" (AERA, 2000, pp. 1-2). The position paper also issues cautions, such as, (a) the problem of high stakes decisions based on a single test, (b) providing adequate opportunities to learn, (c) using test scores for purposes for which they were not intended, and (d) the importance of providing guidelines for special education students, and establishing policy regarding the use of high stakes assessment.

Many opposed to the standards-based accountability movement believe the notion of standards is contrary to the belief that the greatest importance in education is creating flexible, lifelong learners, who can adjust to a changing world (Sheldon & Biddle, 1998). High stakes assessment can reduce curriculums to "teaching to the test" (Darling-Hammond, 1997; Marzano, 2000; McNeil, 2000). It is not only likely, it is probable that teachers will do what they need to do to have their students do well on the test (Linn, 2000). Opponents fear that there will be a lack of meaningful connections between content and application for the student when curricula are narrowed.

In an attempt to deal with some of the problems of state mandated accountability and assessment, a committee was commissioned by several national education associations to recommend how state assessment and accountability could be used to improve instruction as well as provide states with information they needed for accountability (Commission on Instructionally Supportive Assessment, 2001). Among the committee recommendations is a suggestion to very carefully select what is assessed and to provide teachers materials to assess what is not assessed by the state, both within and across different sets of standards.

We tried to make sure that if teachers did teach to the test that it would be an improvement in the outcomes of most programs. We tried to make sure that the assessment materials themselves were authentic in nature, so that what was being assessed was what we wanted students to be able to do in real life.

Critics of high stakes assessment are concerned that the assessment of students has little value other than for policy decisions. In most assessment programs teachers do not know what is on the test and they do not receive the scores of their students until well after the completion of the school year. We wanted teachers to know what was on the test and we wanted teachers to assess student performance so that they would get immediate feedback on student performance.

High stakes assessment in core subjects in the state of South Carolina had the potential to reduce or even eliminate physical education programs in our state, as efforts and resources were all being reassigned to those curriculum areas that "count." Program time was being reduced, teachers were being shifted to work in academic areas, and students were being pulled out of physical education classes for remedial help in academic areas. Some students were never in attendance in the physical education classes in which they were enrolled. Establishing an assessment program in physical education with some accountability for performance was a way to avoid further program erosion.

Systemic Change

We have been most conscious of the need for comprehensive (Odden & Marsh, 1988), systemic change (Smith & O'Day, 1991), more recently characterized as school-wide change (Fullan, 2001). Systemic change recognizes that if change is going to occur then you must work with all parts of the education system. Systemic change depends on top-down policy to initiate change and collaborative bottom up processes to implement and maintain change (Fullan, 2001; Odden & Marsh, 1988; Smith & Oday, 1991). When we set about to change physical education we realized early that we would have to attend to issues of systemic change. Systemic change in action is a little bit like being the boy who tried to hold water back in the dyke by putting his finger in one of the holes. To produce the kind of change we were looking for we would have to address many concerns simultaneously. The process we used to create change is described below.

Create policy. Top-down policy has been the one area that physical educators have avoided as a mechanism for change. Several authors would attest to the idea that both top-down and bottom up efforts may be essential for change (Firestone, Mayrowetz & Fairman, 1998; Locke, 1992).

Physical education teachers in our state go largely unsupervised. In all but a few cases, they operate as loners in a school and district. Normally, district administrators or district supervisors of physical education hold subject areas within the school accountable for good programs. In South Carolina schools this level of supervision and accountability is largely non-existent. We have few district coordinators who are specialists in physical education. Our response was to seek to create top-down policy at the state level that would generate some kind of accountability for schools and districts to hold teachers and programs accountable for change. We also needed to seek funding for that policy.

Because physical education is a marginalized subject, we could not wait for policy to be established for us by someone else at the state level. Nor could we wait to be invited to participate. It was never the intention of the SDE to eliminate the "special" or non-core subjects in the development of assessment materials or accountability measures. In fact the leadership at the state level was philosophically very supportive of assessment in all school subjects. Physical education and other non-core subject areas were more of a "to do later" on their list. Without proactive efforts on our part it is likely that the standards and accountability reform effort would be over before physical education ever got "our turn."

To get support for policy, coalitions with the many health and physical activity groups in the state had to be established. South Carolina has very active and strong coalitions for physical activity. These coalitions proved vital to our success with the legislature (see Rink & Mitchell, 2002) and vital to our efforts to procure funding from outside the legislature and the SDE.

Create a large base of support. We needed the support of many professionals, not only to facilitate them buying in to what we were doing (Hall & Hord, 2001), but also because carrying out a program of this magnitude required a lot of people committed to work with few monetary rewards. It is ultimately the teacher who is responsible for implementing any program (Abelmann & Kenyon, 1996; Odden & Anderson, 1986). Program efficacy was dependent on a large enough number of teachers willing to support the program and not subvert it (Fullan, 2001; Illinois State Board of Education, 2002).

Public school as well as college and university involvement is essential and both play different roles in achieving successful reform (Goodlad, 1994; Illinois State Board of Education, 2002). We worked quickly to move the program from a college and university dominated effort to a shared responsibility of all the professionals in the state. K-12 teachers were involved in developing the performance indicators and assessment materials and in setting up the teacher development and teacher training programs. The involvement of the college and university faculty served to initiate the effort and to prepare the pre-service teachers to work with the standards and assessment material.

At one point in the project almost 200 state professionals were working on some aspect of the program. A defining moment in the program was the point at which teacher development and training programs were conducted almost entirely and most competently by practicing public school teachers. Building this base of support and helping teachers to develop the confidence to be major contributors took time (years) (Gredler, 1996; McLaughlin, 1976) and specific strategies for inclusion that were more effective with some groups than others.

Administrative support. The literature on school change is very clear about the role of school administrators and the importance of getting their support for change initiatives (Coffey & Lashway, 2002; Fullan, 2001). Prior to the first data collection the program director and administrator conducted regional information sessions throughout the state just for administrators. The purpose of these sessions was to communicate the standards, share with them how physical education had changed since their personal experiences in physical education class, and give them an understanding of how the assessment program worked. We wanted administrators to realize the program would help them do their jobs without putting an undue burden on them. The real purpose was to prevent administrators from derailing the project before it even began.

In most reform efforts the involvement of administrators is essential. In South Carolina school administrators were already finding it difficult to handle all of the expectations coming down from the state house with few resources. We made a deliberate decision not to directly seek administrator support at the state level, but rather assume that they would not play non-supportive roles.

We were very aware that if the assessment program were to work we would have to make it administrator friendly. We would need to limit the direct involvement of the administrator in assessing physical education and shift most of that responsibility to the teacher. SCPEAP would do all of the work usually “dumped” on schools to conduct a teacher development and assessment program. SCPEAP would need to develop and conduct in-service programs to help teachers align their programs with the standards. We would also have to train teachers to collect and submit data, and communicate directly with teachers to help those programs that needed help with the assessment process. Most of all we would need to try and send teachers to administrators with solutions to problems (not just problems) they might be having in implementing the program.

Create reasonable expectations. The individual teacher is the most important player in a reform effort (Hall & Hord, 2001). Program efficacy was dependent upon the teacher understanding and buying into the intent of both the outcomes as well as the need for assessment. We wanted to create *expectations* that were reasonable for school programs and goals that teachers could buy into. Projects like the national curriculum in England were initially easily derailed due to

unreasonable expectations (Calderhead, 2001). We spent a lot of time at the elementary and middle school levels helping teachers involved in the program work through the decision making process to set those expectations. The high school indicators were legislated.

We would need to have an assessment program that teachers could perceive as fair and reasonable. We wanted to influence outcomes of programs but at the same time give teachers the flexibility in how they achieved those outcomes. We also wanted the assessment program to be reasonable so we chose to limit the number of outcomes assessed and the amount of time in any three-year period devoted to assessment.

The program time given to physical education in South Carolina is very limited. We did not seek to set program goals beyond what could be reached by the present structure. Rather we sought to improve the quality of what presently exists with a long-term goal of asking for more program time. Teachers would not buy into expectations that they perceived to be hopelessly unachievable in their present conditions (Calderhead, 2001; Coffey & Lashway, 2002; Odden & Anderson, 1986).

Assessment as part of a good program. Assessment itself was not the goal, program improvement was. The assessment program was set up to be both a teacher development program as well as an assessment program. Assessment has the power to change teaching because it focuses teachers on what is important to teach and gives teachers feedback on the teaching process (Edmonds, 1979). As such, assessment is an invaluable part of the teaching-learning process. In working with teachers we found that few were actually assessing students on any outcomes related to the standards prior to the assessment program. Having teachers do the assessment as part of their classes and encouraging teachers to share expectations with students and use formative assessment throughout the instructional process had the potential to change the effectiveness of instruction (Wood, 1996).

Work at several levels with teachers. The ideal scenario for successful school reform is the balance between external pressure for change and internal support (Fullan, 2001; Odden & Anderson, 1986). We knew that it was unlikely teachers would find a great deal of initial support within their schools. We had to provide a lot of that support. Teachers needed immediate help in how to align their curriculums with the performance indicators. They needed long term sustained help in how to actually teach to the standards and how to teach specific content, and they needed a lot of help in doing assessment (Illinois State Board of Education, 2002).

We had to facilitate change by helping teachers work with the typical barriers to change such as scheduling problems, large classes, isolation, lack of department cohesion, lack of administrator support, and so forth (Bernauer & Cress, 1997; Ennis, 1992; MacKenzie, 1983). We had to share with teachers what we knew about effective programs (Coffey & Lashway, 2002; Edmonds, 1979; MacKenzie, 1983). Successful programs would have to 1) develop strong leadership in the department, 2) have a clear emphasis on learning, 3) monitor student progress regularly, and 4) have high expectations for teachers and students (Edmonds, 1979). Teachers needed training in how to collect and submit assessment data. We purposely separated teacher development sessions (PEI) from teacher training for data collection so that teacher development sessions could address the larger perspective of teaching to the standards and not just the collection of data on the performance indicators.

Evaluation of the project. From the very beginning of the project we have made an effort to collect data on what we are doing and how we are doing it. Even though state departments of education do not typically include substantive assessments of their programs, we felt it was important to know whether or not what we were doing was effective in making the kinds of changes we hoped to make. The project has maintained up-to-date and continuous records on all participants and their roles. We know who has attended what sessions from each school. All teacher development programs were evaluated by teachers at the end of each session which helped us to know what teachers felt they needed and what was important to them at different stages of their experience with the program (Abelmann & Kenyon, 1996; Fullan, 2001).

After the first year we began a series of research studies on the effects of the program, some of which are reported in this monograph. The research component of the program has only recently been given some funding from the Centers for Disease Control and Prevention. From the onset we set up a research agenda and have primarily used doctoral students at the University of South Carolina to carry out that agenda.

Be patient. We knew that change took time (Hall & Hord, 2001). We have gained a new appreciation for just what this idea means. It took us two years to bring in public school teachers to leadership positions. It took three years to develop assessment material. It took five years to get policy creating some accountability. It took another three years to fully implement the assessment program at the high school level only after the program was "ready to go." What has sustained us is that we had glimpses of change and the potential of what we were doing for creating change. In one sense we are at the very beginning of the change process. We have teachers and programs at all levels of Hall and Hord's Stages of Concern (2001) after seven years of work and yet we think we have seen change. This monograph documents some of it.

An Introduction to the Monograph Chapters

This monograph presents the results of four studies done on the first state data collection at the high school level. One third of the high schools in the state ($n = 62$) and 160 teachers submitted student assessment data on all four indicators. There are many questions we can ask regarding this first data set. The studies represented in this monograph are those we felt to be first steps in understanding the results and implications of the standards, assessment, and accountability program in our state.

In chapter 2, *Student Performance Data, School Attributes, and Relationships*, Mitchell, Castelli, and Strainer present the results of the state assessment, and the relationships among performance indicators. They also identify the school and teacher factors related to school performance. The entire database of 62 schools and 160 teachers was used for this study. School characteristics include class size (academic and physical education), school enrollment, poverty index, absolute report card grade for a school, and academic exit exam scores for the 10th and 12th grades. Teacher characteristics include gender, attendance at assessment data collection training, and attendance at PEI training.

In chapter 3, *A Comparison of High and Low Performing Secondary Physical Education Programs*, Castelli and Rink describe the differences between schools that did well in the state assessment and those that did not. Teacher surveys, teacher interviews, and school and SCPEAP documents provide data for both qualitative and quantitative analyses of differences. Four high performing and four low performing schools were selected for the study. Profiles of high and low performing schools are established.

In chapter 4, *Instructional Variables and Student Knowledge and Conceptions of Fitness*, Stewart and Mitchell describe high school students' knowledge and conceptions of fitness and the instructional variables related to teaching fitness. Primary data sources included a teacher self-report survey distributed to all teachers submitting data for Performance Indicator Two – Cognitive Fitness and the written tests taken by students and submitted by teachers. Sixty-one teachers completed the survey and a sample of one hundred eighty student written tests were analyzed from three different forms of the test.

In chapter 5, *Teacher Competency Using Observational Scoring Rubrics*, Williams and Rink determine the ability of teachers to follow testing protocols, and to score student psychomotor performance accurately using scoring rubrics developed for Performance Indicator One – Movement Competence. Relationships between teacher compliance with data collection protocols and teacher accuracy were established with the variables teacher gender, student performance, activity and teacher training. The study also analyzes the specific problems teachers had in following testing protocols.

In chapter 6, *Insights and Reflections on a State Assessment Program*, Rink and Stewart synthesize the South Carolina experience with state level assessment. Implications of the results of the four studies for practice are explored. The chapter concludes with a discussion on the generalizability of the South Carolina experience to other states.