

Measuring Implementation of School Programs and Policies to Promote Healthy Eating and Physical Activity among Youth

Sarah A. McGraw, Ph.D.,*¹ Deborah Sellers, Ph.D.,† Elaine Stone, Ph.D., M.P.H.,‡ Ken A. Resnicow, Ph.D.,§ Sarah Kuester, M.S., R.D.,¶ Fred Fridinger, Dr.P.H., CHES,¶ and Howell Wechsler, Ed.D., MPH||

*New England Research Institutes, Watertown, Massachusetts 02472; †Educational Development Center, Newton, Massachusetts 02458-1060; ‡Division of Epidemiology and Clinical Applications, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892; §Department of Behavioral Sciences and Health Education, Rollins School of Public Health of Emory University, Atlanta, Georgia 30322; and ¶Division of Nutrition and Physical Activity and ||Division of Adolescent and School Health, Center for Disease Control and Prevention, Atlanta, Georgia 30341

Background. The measurement of program implementation and policy adoption is an essential evaluation component of any health intervention program. Data on program implementation are used to monitor program progress; identify elements of a program to be strengthened or eliminated; provide accountability; and help explain program effects.

Method. This paper reviews approaches to measuring the implementation of school-based programs and policy to promote physical activity and healthful eating among youth. Areas examined include classroom instruction, food service, physical activity classes, and school policies. Operational definitions of implementation and methods of collecting data are described and compared.

Results. Most implementation measures are focused on two dimensions: quantity (dose or completeness) and quality (fidelity). Data collection methods include the use of teacher self-report recorded through checklists, questionnaires, and interviews. Classroom observations by a trained observer are also used. Studies of policy development have used archival records and semi-structured interviews.

Conclusions. Considerable variability exists across studies in how program of implementation is defined and measured. This is in part due to the need to tie measures closely to the content and format of the intervention. More work is needed to assess and compare the reliability and validity of various approaches to measuring implementation.

© 2000 American Health Foundation and Academic Press

Key Words: implementation evaluation; policy adoption; process evaluation; school; physical activity; diet.

INTRODUCTION

Recent studies suggest that insufficient physical activity and unhealthy dietary patterns during youth contribute to increased mortality and morbidity in adulthood. School-based programs that are carefully designed and implemented can play an important role in promoting lifelong physical activity and healthy eating among young people [1,2].

Measurement of program implementation and policy adoption is an essential component of any public health or educational intervention and is critical to ensuring the success of these programs such as those conducted in schools [3–6]. Data about the implementation of any program can be used to improve the program by identifying elements that need to be strengthened or eliminated while it is ongoing or before it is to be implemented more widely [6–9]. Implementation data also may be used to develop benchmarks to monitor progress toward specified goals [4,6,9,10], to provide accountability to funding sources or other stakeholders [4,5,11–14], and to explain program effects [3,4,7,15,16]. Program implementation data from rigorously designed interventions that are carefully analyzed can demonstrate the impact of specific intervention strategies, explain study outcomes [7,15,17–23], and infer the success or failure of an intervention strategy [3,4,7,15,24,25]. Data on program implementation should be part of a process evaluation used to help determine how program elements impact the mediating variables hypothesized in theoretical models of interventions [26].

This paper will review approaches used to measure the implementation of school-based programs. These approaches are applied to classroom curricula, food service, physical activity classes, and school-wide policies. Although this paper will focus on school-based programs, measures of program implementation can be

¹ To whom reprint requests should be addressed.

applied to youth programs conducted outside of school settings [27].

The measurement of program implementation of school-based health interventions varies considerably across studies. Some variability is due to a lack of consistency and clarity in definition of terms—a symptom of a still-developing field. Some is inherent to the task of measuring program implementation because measures must be tied closely to the intervention being evaluated, and interventions vary widely.

In designing implementation studies of school-based physical activity and nutrition programs, researchers focus largely on two key constructs: quantity, or dose—how much of the intervention was delivered; and quality, or fidelity—to what extent the intervention was delivered as intended. These constructs are closely related and are sometimes used interchangeably.

To document program implementation, investigators have used three types of data collection: examining archival records, observing activities using trained observers, and self-reporting—typically by teachers or other staff through interviews with trained interviewers, self-administered questionnaires, logs, or focus groups. Each approach has different strengths and weaknesses [6,11].

Although examining archival records can be relatively inexpensive, investigators typically find this approach to be the least useful because existing records often do not provide the specific program implementation information needed. Using trained observers is generally considered the most objective data collection approach, but this method is more expensive than self-report logs or interviews, and the costs involved preclude collecting observational data on every teaching session or school activity [15]. Furthermore, the presence of observers in the classroom is intrusive; teachers may alter their teaching behavior to meet what they believe to be the expectations of the observers. Finally, investigators use checklists or logs that are maintained by school staff. This approach to data collection is relatively economical but may have shortcomings in the quality of the data and places the burden of completing records on the school staff [15].

Classroom Curricula

Measures of implementation of classroom curricula are reviewed in this section. A more detailed description of implementation measurements used in studies of classroom health education programs is provided in Table 1.

Quantity. The measurement of quantity, which is also called dose or completeness, has been defined in the literature in at least three ways:

- The proportion of the curriculum covered, which has been defined as the number of activities or specific lessons completed of the total required by the curriculum [15,17,19,28–35]. This is the most commonly used measure of quantity, but sometimes it also has been used as a measure of quality.

- The absolute amount of instruction delivered. This value is generally measured as the number of hours or minutes of classroom instruction or as a count of student activity sheets or workbook pages completed by the class [7,15,18,22,30,36].

- A subjective assessment of how much of the curriculum was covered [15].

Classroom observations to document quantity [15,23,35–38] have been used, but measures of quantity are most commonly obtained through a teacher's self-report [15,23,29,30,33,34]. Typically, teachers use a checklist or log to indicate the amount of time spent delivering the curriculum, the activities completed during each session, and the materials used [22,29]. Ideally, teachers should complete the checklists or logs soon after each session, because the accuracy of recollection decreases as the time between the event and the reporting of the event increases [39].

Quality. Quality, which is also called fidelity, has been defined in the literature in at least three ways:

- The proportion of individual curricular activities that were conducted as intended or as described in a curriculum guideline [7,29,34,36,37,40].

- The proportion of session plans and materials that were not modified (no activities were omitted or altered [19].

- A score determined by deducting points from a base score for each deviation from the curriculum judged to be detrimental and by adding points to the base score for each deviation judged to enhance the curriculum [41].

Data that are used to assess quality are sometimes obtained through logs in which teachers rate how closely the curriculum guide was followed during each session [29,30,34,35], but more commonly these data are obtained through observation. Observation forms can be designed with sufficient detail to allow researchers to group activities by key teaching tasks, with groupings often determined by the theoretical framework supporting the intervention. For example, in the *Gimme 5* study, tasks were grouped according to teachers' didactic behaviors, modeling, and use of social reinforcements [8]. This information can then be used to identify which tasks a teacher did or did not carry out and which teaching tasks were associated with the program's outcome.

Composite or multidimensional measures of implementation. Composite scores that combine measures

TABLE 1
Summary of Implementation Measures of Classroom Instruction by Study

Program	Design	Characteristics assessed	Data collection methods	Validity/reliability
• CATCH (Child and Adolescent Cardiovascular Health Project) [19,29,36,39,52]	<ul style="list-style-type: none"> • 96 schools randomized, • 56 intervention schools • 4 states (MN, CA, TX, LA) • Grades 3–5 • Cardiovascular risk behaviors • Policy, physical education, food service, families, classroom instruction 	<ul style="list-style-type: none"> • Quantity Percentage of activities completed 	<ul style="list-style-type: none"> • Teacher's checklists 	<ul style="list-style-type: none"> • Quality correlated with quantity [19]
		<ul style="list-style-type: none"> • Quality Percentage of activities modified 	<ul style="list-style-type: none"> • Observations 	<ul style="list-style-type: none"> • Interrater reliability (100% interrater agreement in third grade) [29]
• GIMME 5 [8,15,78]	<ul style="list-style-type: none"> • 16 schools randomized • 8 intervention schools • Georgia only • Grades 3–5 • Increase fruit and vegetable consumption • Classroom instruction 	<ul style="list-style-type: none"> • Quantity Percentage of activities completed 	<ul style="list-style-type: none"> • Teacher's self-report logs • Teacher's interview 	<ul style="list-style-type: none"> • Self-report measure correlated with two measures obtained from teacher interviews ($r = 0.51$; $P = 0.002$ and 0.66 $P < 0.001$) [15]
		<ul style="list-style-type: none"> • Quality Tasks grouped into 11 categories by function (e.g., procedural, didactic) Percentage of procedures completed 	<ul style="list-style-type: none"> • Observations 	<ul style="list-style-type: none"> • Interrater reliability—Fidelity Index Score: ($r = 0.93$); individual items ($\kappa = 0.68$) • Correlations with measures of quantity ($r = 0.12$ to 0.33) [15]
• SHEE (School Health Education Evaluation) [17,18]	<ul style="list-style-type: none"> • Naturalistic and experimental study • 1071 classrooms; 20 states • Grades K–11 	<ul style="list-style-type: none"> • Quantity Number of hours of instruction Percentage of curriculum completed 	<ul style="list-style-type: none"> • Teacher questionnaires 	<ul style="list-style-type: none"> • N/A
		<ul style="list-style-type: none"> • Quality Average rating of degree to which components were taught without modification 	<ul style="list-style-type: none"> • Teacher questionnaires 	<ul style="list-style-type: none"> • N/A
• SHCP (School Health Curriculum Project) [7]	<ul style="list-style-type: none"> • 10 schools randomized to condition; 5 intervention schools • Illinois only • Grade 5 • General health • Classroom curriculum 	<ul style="list-style-type: none"> • Quantity Amount of classtime devoted to health instruction 	<ul style="list-style-type: none"> • Questionnaire • (Data sheet on health instruction) 	<ul style="list-style-type: none"> • N/A
		<ul style="list-style-type: none"> • Quality Was activity implemented as planned, modified, not implemented 	<ul style="list-style-type: none"> • Teachers logs • (Teacher activities analysis reports) 	<ul style="list-style-type: none"> • Observations on 10 randomly selected days (2 per classroom); Observer and teacher reports agreement: 25–100% (mean 81.5%)
		<ul style="list-style-type: none"> Whether curriculum materials were used 	<ul style="list-style-type: none"> • Teacher logs (materials analysis report) 	<ul style="list-style-type: none"> • N/A

TABLE 1—Continued

Program	Design	Characteristics assessed	Data collection methods	Validity/reliability
• THTM (Teenage Health Teaching Modules) [30,33]	<ul style="list-style-type: none"> • Experimental and naturalistic study • Training substudy (85 of 111 experimental school teachers randomized to receive training) • Multiple health behaviors • Classroom curriculum 	<ul style="list-style-type: none"> • Quantity Number of minutes spent on required modules • Quality Lack of modification and high percentage of required activities taught 	<ul style="list-style-type: none"> • Implementation logs • Implementation logs • Level of use telephone interview 	<ul style="list-style-type: none"> • N/A • N/A
• Curricula diffusion (Project SMART, Growing Healthy, THTM) [33]	<ul style="list-style-type: none"> • 21 No. Carolina school districts randomized to condition • Offered choice of implementing 1 of 3 curricula • Grades K–9 • Multiple health behaviors • Classroom curriculum 	<ul style="list-style-type: none"> • Quantity Percentage of activities taught 	<ul style="list-style-type: none"> • Teacher checklists 	<ul style="list-style-type: none"> • N/A
• Changing the Course [41]	<ul style="list-style-type: none"> • Pilot study, 6 schools receive intervention, 26 teachers, 2 states • Grades 1–6 • Food choices • Classroom curriculum 	<ul style="list-style-type: none"> • Quality Augmentation score: Starting at 0; 1 pt added for lesson taught to which teacher added relevant content; 0.5 pt added for each lesson in which teacher altered instruction deemed enhancing Fidelity score: Starting with base score of 15 for lower and 16 upper elementary; 1 pt subtracted for each lesson not taught, 0.5 for each lesson taught but content omitted, 0.25 subtracted for each lesson taught with modifications deemed detrimental 	<ul style="list-style-type: none"> • Teacher logs 	<ul style="list-style-type: none"> • N/A
• Know Your Body, NY [22]	<ul style="list-style-type: none"> • 5 schools assigned to condition by district; 3 intervention • Grades 1–4 at baseline • CVD risk behaviors • School programs and classroom instruction 	<ul style="list-style-type: none"> • Degree of implementation Ratings intended as composite measure of quantity and quality Number of modules covered; number of activities covered; number of minutes spent teaching per week 	<ul style="list-style-type: none"> • Teacher questionnaire • Head teacher ratings • Project coordinator rating 	<ul style="list-style-type: none"> • Compared rating to mean number of pages completed in four student activity books for last year of study; Spearman rank correlations: project coordinator = 0.68; head teacher = 0.49; teacher self-report = 0.56
• Know Your Body, DC [23]	<ul style="list-style-type: none"> • 13 intervention schools, 82 teachers • Grades 4–8 • CVD risk behaviors • Classroom instruction 	<ul style="list-style-type: none"> • Quantity Activities taught • Quality Measures to assess teacher's performance and status as role model • Teacher effectiveness Composite score of quality and quantity 	<ul style="list-style-type: none"> • Teachers activity report • Count of workbook pages • Classroom observations • 20-item checklist 	<ul style="list-style-type: none"> • N/A • Checklist reviewed by expert panel of three to assess content validity • Agreement between observation scores and project staff opinions of overall effectiveness (gamma 0.80)

of quantity and quality are sometimes used to measure implementation of classroom instruction [15,22,23,30,35]. In the *Know Your Body* study, Taggart and colleagues [23] created a composite measure of a teacher's "effectiveness" by combining a quality score (based on observers' ratings of a teacher's performance in the classroom) with a quantity score (the amount of curriculum taught based on teacher's self-report and observers' ratings). Connell and colleagues [17,18] created a composite measure of "extent of implementation" for the *School Health Education Evaluation (SHEE)* project by combining hours of instruction delivered, the percentage of required activities taught, and the extent to which components were taught without modifying plans or materials.

Strengths and weaknesses of classroom curriculum implementation measures. Only a few investigators have attempted to validate measures of implementation of classroom curriculum in the area of health education. In *Know Your Body*, Resnicow and colleagues [22] compared a relatively objective measure of quantity, the number of pages completed in student workbooks, with three less objective measures collected at the end of the school year. These measures were the teacher's self-report of the number of activities and the number of minutes spent teaching per week, the head teacher's rating of each classroom teacher reflecting both quality and quantity of implementation, and the project coordinator's assessment of quality and quantity. The Spearman rank order correlations indicated that the project coordinator's ratings had the highest correlation (0.68) with the number of pages completed in the activity books. The correlations for the teachers' and head teachers' ratings were only 0.56 and 0.49, suggesting that the project coordinator offered a more objective assessment of implementation than did the teachers' self-report and the head teachers' assessments.

Basch and colleagues [7] estimated the validity of teachers' reports of quality. Teachers were asked to report on which activities were planned, modified, or omitted. The teacher's self-report was validated by observing implementation in a randomly selected sample of classrooms. Average agreement between the self-reports and the observers was 81.5% (range: 25–100%).

In the *Gimme 5* study, Resnicow and colleagues [15] compared three measures of implementation: a measure of quantity based on the percentage of activities covered as recorded in the teacher's self-report questionnaires, a measure of quantity based on postimplementation interviews with teachers, and a third measure obtained through classroom observations by trained observers. The correlations between self-report questionnaires and two methods of coding interview data were 0.51 and 0.66, respectively. These were substantially higher than the correlations between each of

these measures and those obtained by trained observers (range: 0.12–0.33). The self-reported measures of quantity also were not associated with posttest measures of the study outcomes (student knowledge and fruit and vegetable consumption), whereas the measures obtained through teacher interviews and classroom observation were correlated with these study outcomes. The differences in correlations may be due, in part, to differences in methods of data collection (self-report versus observations). Resnicow and colleagues [15] concluded that postimplementation interviews and classroom observation yield the most valid data for measuring implementation.

If workbooks are used in the program, Resnicow and associates [22] found that counting the number of completed workbook pages may be a more objective measure of quantity than teacher self-report and may be the preferred method for measuring quantity. Workbooks, however, have their own sources of error—lost workbooks, for example, result in missing observations.

Studies reported by Resnicow et al., [15], Davis-Hearn and Baranowski, [8], and Basch et al. [7] suggest that classroom observations may be a better method than teacher self-reports for measuring quality, but these observations are intrusive and costly and cannot be conducted routinely for all teachers. Classroom observations place fewer reporting burdens on teachers, but teachers may find having an observer in the classroom inconvenient and inhibiting. The validity of observational data can also be challenged because the participants may modify their behavior due to the presence of an observer, particularly if plans for the observation are announced in advance [15,39].

Matheson and Achterberg [42] described an approach to collecting very detailed data on how a computer-assisted instruction program for nutrition education was implemented in a classroom. Their primary interest was to assess changes in cognitive structures as the students learned. They drew upon qualitative methods including semi-structured interviews analyzed employing a grounded theory approach and detailed observations of how individual students used the program and interacted with each other while using the program. They acknowledged that qualitative approaches can be costly and time consuming, requiring extensive observer time in the classroom and coding time for analyses, but they argued that the richness of the data justified the effort.

Checklists or logs maintained by school staff can be relatively economical compared to employing interviewers or trained observers to collect data. However, checklists, logs, or questionnaires place an additional response burden on the school's staff who are asked to complete the forms. The reliability and validity of self-reported data collected through checklists, logs, interviews and questionnaires can be challenged because of

recollection (school personnel might not have a precise memory of what was taught) and social desirability bias (teachers might report having taught more than they actually taught to make themselves appear to be more competent or cooperative). Some teachers may report teaching a topic even if they only made a brief mention of it without covering the detail suggested in the curriculum guide. This type of overreporting is more likely to occur when teachers are asked to report on a topic they value as one of lesser importance or priority. To counterbalance the potential weaknesses of methods relying on self-report, some studies combine independent observations with checklists, questionnaires, or interviews [15,22].

Currently, the science of documenting classroom instruction is relatively undeveloped. Those investigators who have attempted to assess the reliability and validity of approaches to measuring implementation have often used inconsistent definitions of quality and quantity and different methods of data collection. The lack of consistency makes an unqualified recommendation of any one approach difficult and suggests the need for more systematic study of the best methods and measures of implementation.

Physical Activity Programs

Providing students with a substantial percentage of their recommended amount of physical activity is one of the major goals of school-based physical education programs [1]. Measures of both the quantity and the quality of implementation are used to document the implementation of these programs. Approaches to documenting implementation of physical education programs in schools, as well as physical-activity-related cocurricular activities such as recess, are described in this section.

Quantity. Investigators examine the amount of physical activity that students engage in or are given the opportunity to engage in at school in at least two ways:

- absolute number of minutes of opportunities for physical activity that are made available to students during the school day [43,44];
- absolute number of minutes of actual physical activity and moderate to vigorous physical activity that students participate in [43,45].

Teachers' logs have been used to record the minutes of opportunity students have to engage in physical activity during physical education classes and recess breaks. The PARC (Physical Activity Record of Classes) form was designed to record minutes of physical activity opportunity by time period [45]. Information about the context of the activity was recorded by noting whether activity was "structured" time for physical education

and, if so, whether the physical education was taught by a classroom teacher or physical education specialist.

Data on the actual amount of physical activity engaged in by students have been collected through observation and student self-report. The SOFIT (System for Observing Fitness Instruction Time) instrument has been used to track quantity of physical activity by observation in several research studies [43–47]. Trained observers track randomly selected students during a class period. At timed intervals, the observers record three features: the level of intensity of activity (e.g., very active) and body position (e.g., lying, standing) during the observation interval, the lesson context during the interval (e.g., general information, knowledge, or motor activity instruction), and the teacher's behavior (e.g., promoting fitness, demonstrating, observing). At the conclusion of the session, observers record the lesson length. The SOFIT data can be used to estimate the mean number of minutes of physical activity, the mean number of minutes of moderate to vigorous physical activity, and the proportion of class time devoted to different types of activities.

Asking students to record their level of physical activity on an ongoing basis is a low-cost approach. For example, in the *DUCK Walking Project (Discover and Understand Carolina Kids)*, elementary school classrooms used a simple log to record the date and the number of miles walked by the class (Jennifer J. Stellar, *DUCK Walking*, unpublished manuscript). However, this approach places the response burden on the students asked to complete the forms and the teachers asked to monitor their completion.

Quality. Investigators define the measure of quality used in physical education or physical activity programs as the degree to which physical education lessons are implemented as intended. This definition may include prespecified objectives that the teacher must meet such as allowing a cool-down period at the end of a session, encouraging students to be physically active during the session, demonstrating enthusiasm, or having adequate equipment available. In one study, McKenzie and colleagues [46] reported items related to quality as distinct characteristics rather than combining the items into a quality scale. They employed trained observers to use the Lesson Observation Checklist [45,46] to rate the class on such items as teacher involvement, clarity of instruction, extent to which safety guidelines were followed, appropriateness of group size, and the extent to which the teacher rewarded students for moderate to vigorous activity.

Strengths and weaknesses of measures of physical activity instruction. McKenzie and associates validated the SOFIT instrument by comparing heart rate data for children ages 4–9 for the instrument's five levels of physical activity based on the intensity of the

activity and body position (lying, sitting, standing, walking, very active). Because heart rates increased at each level of activity, the investigators suggested that the instrument was a valid indicator of level of physical activity. The SOFIT items that pertained to the teacher's behavior and lesson context discriminated between those classes with fitness as the primary focus of the lesson and those without such instruction [46]. Teachers in the fitness class promoted more physical activity among their students and the students in the classes were more physically active than those in the nonfitness classes. The items also discriminated between those classes taught by physical education specialists and those taught by teachers trained in other subjects. When investigator's calculated interobserver reliability for the SOFIT they found the ratings to be generally high [45,46]. In one study, agreement exceeded 88% (range: 88.3–91.8%) for each of the three observation categories: level of student activity, lesson context, and teacher behavior [46].

The validity and reliability of log sheets used in the DUCK walking program to record miles have not been ascertained. It is unknown whether young students can reliably track and record the miles they walk without the close supervision of a classroom teacher or some other adult. The validity and reliability of the PARC and Lesson Observation Checklist have also not been reported. The data obtained from the PARC form and the DUCK form are subject to limitations of self-reported data, including faulty recall of the number of minutes of physical activity and overreporting the number of minutes for social desirability purposes. The limitations of the Lesson Observation Checklist include the potential cost of using trained observers. Observers, however, can complete the checklist while they are completing the SOFIT instrument, which obviates the need for training additional observers and scheduling additional observations.

Food Service Programs

Measures of both the quantity and the quality of implementation of food service programs in schools are reviewed in this section. The major focus in this area has been to assess the implementation of menu and food preparation changes in school cafeterias and exposure to these changes. In addition, measures of implementation of nutrition education interventions sponsored or facilitated by the school food service program are described.

Quantity. The measures of quantity employed to assess implementation of food service programs differ slightly from those used to measure implementation of classroom or physical education; the measures assess exposure to changes in menu and food preparation techniques by documenting purchasing or consumption of school meals. This is important, because unlike the

classroom where students have less choice about their exposure to the curriculum, students have more choice about their exposure to the school lunch. They can opt to avoid the altered school meals by bringing food from home. These measures of quantity for food service programs include

- student participation in purchasing school meals or the actual percentage of students attending school who purchase the meal [43–52];
- student consumption of school meals [50–52].

In the *Go For Health* study [51,53], investigators used a randomly selected group of children to observe the number of children who ate lunches prepared by the school and the number of children who ate lunches prepared at home. In a more refined approach, investigators measured plate waste, which can be either observed or weighed [54,55].

Quality. In the literature, investigators have defined quality of school food preparation as

- the extent to which staff carry out recommended food purchasing and preparation practices [48,49,53,56,57];
- the actual nutritional quality of school meals that is assessed either by chemical analysis or menu analysis [48,50–52].

Investigators have identified different practices of food purchasing and preparation that can be carried out to reduce the fat, saturated fat, cholesterol, sodium, and sugar content of school meals and to increase fruits, vegetables, grains, and calcium-rich foods [48,49,53,56,57]. Researchers have used interviews and observation to identify how many and which of these practices are implemented and how often they are carried out [49,58,59].

To assess the nutrient composition of food served in schools, investigators may send randomly selected samples of food items to a laboratory for chemical analyses [60]. In menu analysis, detailed information on menus, recipes, and food preparation techniques is entered into a dietary analysis program that computes nutritional content based upon the food composition data in its nutrient database. Many computerized programs are available for completing these analyses [61,62]. These programs vary in their cost, in the number of foods and nutrients in their database, in their use of non-USDA/vendor-specific foods, in how missing data are handled, and in their ease of use. In addition, these programs may also vary in their estimation of nutrient content or dietary components [63]. Investigators use a variety of methods to collect data on school-meal menus and recipes including collecting usual or planned menus, interviewing food service staff, and observing food services on-site. Some investigators also use combinations of these methods [49,50,64].

Nutrition Education

The school cafeteria is an ideal location for nutrition education [1,49,52]. In the *Child and Adolescent Cardiovascular Health (CATCH)* study, for example, investigators measured quantity and quality of an educational program in a cafeteria setting using activity checklists [36,49].

Quantity. The quantity of educational programs delivered through the cafeteria has been defined as the number of promotional activities delivered by school staff in the cafeteria. In the *CATCH* study, school staff reported the number of promotional activities each month on a checklist form completed by school staff [36,49]. By documenting the nature and timing of cafeteria activities investigators can analyze their impact or explore the link between the number or type of promotional activities and the degree of school meal participation.

Quality. Investigators should also consider measures of quality when documenting the implementation of nutrition education activities conducted by food service personnel. As in classroom instruction and physical education, these measures are defined as the extent to which the activities were conducted as intended.

Strengths and weaknesses of food service program measures. The validity of different measures of food choice and consumption by students and the food purchasing and preparation practices of food services staff have not been reported in the literature. Thus, how many selection and preparation practices or what specific practices are needed to achieve nutritional quality is not clear.

Although direct chemical analysis of food items served to students is more expensive than a computer analysis [60], investigators agree that it is a more accurate measurement for some food components, particularly micronutrients or trace compounds. Computerized analysis requires the selection of a suitable dietary analysis program as well as careful collection of representative menu and recipe data. Ebzery and colleagues [64] have summarized the advantages and disadvantages of various methods for collecting the information required to complete computerized dietary analysis of menus and recipes. In addition, other reviews outline the advantages and disadvantages of particular programs [61] and provide guidelines for selecting the program appropriate for a particular purpose [62].

Policy Adoption and Implementation

Putting explicit policies in place is vital to implementing an effective school health program and ensuring its long-term sustainability [1,2,65]. The Center for Disease Control and Prevention outlined policies to sustain school programs to promote physical activity and

healthy eating [1,2]. These guidelines address curricular requirements and staff qualification, collaboration among different school programs, and a supportive environment that promotes physical activity and healthy eating such as having healthy snacks in vending machines and a variety of competitive and noncompetitive intramural sports and recreation activities. Approaches used to document policy adoption and implementation are covered in this section.

The constructs of quantity and quality used to measure implementation of classroom instruction, physical activity, and food service programs are generally not applicable to recording policy adoption and implementation. Instead, this area of investigation focuses on measuring the adoption of formal policies and the steps that should occur to increase the chances of getting a new policy adopted and implemented.

To document existing policies, investigators have used data sources similar to those employed in most implementation measures, for example, structured interviews [58,66] and self-administered questionnaires or a combination of both [58,67,68]. In the *School Health Policies and Programs Study*, Kolbe and associates [58] used a combination of self-administered questionnaires on policies at the state and district levels and in-person structured interviews at the school and classroom levels.

Investigators have documented the development and adoption of policy to the extent that school leadership gives its endorsement and provides strong support for proposed policies, the school staff is involved in identifying, writing, or adopting specific policies, and the school staff is involved in planning the program to implement the policies [69,70]. The steps toward adoption of a policy might be documented through records generated during the adoption process. Such records, which often exist as part of school or other archives [25,71], include

- letters of commitment by the school administration to indicate interest in a policy;
- minutes of meetings held to discuss the policy that record the presence of the individuals and institutions actively involved in the discussions;
- annual reports with a description of policies (existing and newly adopted);
- mission statements;
- program goals and objectives.

Milio [71] notes that policy adoption is greatly facilitated when there is greater agreement between policy and program planners about the nature of the problem to be addressed by the policy, the importance of the problem, potential solutions to the problem, and who is responsible for making decisions. Information for documenting the level of agreement on these issues might

be available in existing school archives or records. Additional interviews with key participants might be required to obtain more detailed information. Semi-structured interviews with open-ended questions can be used to allow the respondent to describe his or her perceptions of the conditions related to policy adoption in his or her own words.

Strengths and weaknesses of approaches to documenting policy. Archival records such as documents containing a description of policies, meeting minutes, and written statements of support can be a potentially low-cost source of information if documents are reliably maintained and easy to retrieve. The information on the records must be abstracted and organized through some coding system. The location and abstraction of data from archival records can be time-consuming if records are incomplete or disorganized. The completeness or accuracy of archival documents is not always certain and can be difficult to verify.

The validity and the reliability of the surveys used to document policies have not been demonstrated in the literature. Thus, the items on the survey are probably subject to all the potential limitations of self-reported data, which include the tendency to seek social approval by overreporting the existence of formal or informal policies and recall bias.

Semi-structured interviews can be long and require careful training of the interviewer to ensure that he or she understands the intent of each question and correctly uses probes to help the respondent amplify comments. Semi-structured interviews also require that the interviewer take thorough notes. Preferably each interview should be tape-recorded and transcribed. Once the interviews are completed, coders must take the responses to each item and synthesize them through a coding process, a process similar to that used for coding policy documents. Coders must be carefully trained to use the same criteria or definitions to categorize responses.

The advantage of a semi-structured interview is that detailed responses are obtained in the respondent's own words. This qualitative approach has the potential for deeper insight into the perspective of key individuals and more detailed information about their understanding of, and reactions to, policy adoption [72]. Some researchers suggest that qualitative data collection is the most appropriate approach to documenting policy development and implementation [71, 73].

RECOMMENDATIONS

The discipline of documenting program implementation and policy adoption is in an early stage of development. Valid and reliable instruments are needed to assess key components of multifaceted, school-based

physical activity and nutrition interventions. Also, researchers need to describe their assessment techniques in enough detail, including study instruments, to allow a reader to determine how implementation characteristics were defined and how data were obtained. Few studies to date have described such key properties as the reliability and validity of instruments used; even fewer report on the quality of the process data by noting, for example, response or return rates for self-reported or observational data.

Greater conceptual clarity is needed in defining key implementation constructs. Measures of quantity and quality, in particular, are often confused or blended. This confusion arises, in part, because these two concepts are closely related; a program cannot be delivered with high quality if the quantity of the program activities delivered is less than that intended.

Composite indices to assess program implementation might be considered. One approach might be to combine measures of the quality of each unit of the program delivered (e.g., the quality of each classroom session) and the number of participants to whom the intervention was actually delivered (e.g., those exposed to the program), adjusting for the number of participants to whom the intervention should or could have been delivered. However, before measures of quantity and quality can be combined, considerably more thought must be given to a number of methodological concerns. Investigators need to be clear about their definitions of quality and quantity, articulating how these concepts are related to each other. They need to determine which measures of quality and quantity are more important for each component of a program. Finally, the relative importance or weights given to items to be combined and how they are to be combined (e.g., by adding or multiplying) must be determined.

Researchers should further explore techniques for enhancing the validity and reliability of methods of documenting program implementation. Postimplementation interviews with school staff are one of the most appealing approaches because they are relatively less costly and burdensome than observation or regular self-administered checklists completed during the program [15]. It is important, however, to also employ some more objective measures, such as independent observations, when documenting implementation. Future researchers should investigate the levels of agreement between self-report and observation and techniques for improving self-reports [15]. More detailed qualitative studies of what actually happens in the classroom when a program is implemented will help to develop more valid and reliable methods for measuring implementation and classroom context [42].

Documenting program implementation and policy adoption is an essential endeavor for both researchers and practitioners alike. It is especially important when

program delivery requires a complex set of procedures or behaviors [26]. Information about the quality and quantity of the activities conducted is critical to describe and understand the nature of the program delivered, interpret program results, and make modifications to improve future programs. When resources to support program implementation are limited, less costly methods to obtain data on implementation may be used, but these methods may result in less precise measures of implementation. These limitations may be acceptable when implementation is documented for administrative purposes or for uses outside a large research study. More precise and perhaps more costly measures of program implementation are feasible within the context of a large research study with staff to support the collection and analyses of the implementation data.

Selecting Approaches

In light of the inadequate development and testing of measurement techniques in the field, it is difficult to recommend specific approaches for assessing implementation. However, the following is a list of some questions to be answered before selecting specific approaches for assessing program implementation.

What is to be measured? Data collection methods must vary according to the efforts to be documented. Archival records or interviews are appropriate for documenting policy adoption, whereas observation methods are more appropriate for documenting implementation of instructional efforts.

How much detail is needed? When less precision is required and resources are limited, some degree of reliability or validity may be sacrificed to save costs or reduce the burden on staff. The disadvantage of such a decision must be considered—less precise data will give a less accurate picture of implementation [4, 5]. For example, the data required for reporting or for administrative purposes may not need to be as precise as data required for research purposes. Accordingly, a school district that wants to monitor program progress for reporting to a funding source may need less precise data than a researcher who is assessing the effectiveness of a new intervention.

How often should data be collected? Process data are collected with varying frequency ranging from every session [29] to once at the end of the program [15, 23, 32]. Frequency is driven by four factors: how often data are needed for reporting or evaluation purposes, the need to minimize respondent burden, the requirement to control costs, and the specificity and reliability of the data required. Collecting data at multiple time points can enhance validity by offering a more complete picture of the implementation process as it occurs over time [74].

Data to estimate dose–response relationships or test causal relationships may require more frequent collection than is necessary for administrative purposes such as demonstrating that program goals and objectives have been met.

Can high-quality observational data be obtained and when should it be obtained? Systematic observation of program implementation is an important method of obtaining data on what is actually occurring in the school or classroom. Some evidence indicates that classroom observations yield more valid and reliable information than does self-reported information collected through teacher checklists [15]. However, high-quality observational data can only be obtained if a number of conditions are met [75]:

- The behaviors to be observed must be precisely defined.
- Observers must be trained to follow the observational protocol and minimize subject reactivity such as changes in behavior of those being observed that are due to the presence of the observers.
- The observation instruments must have adequate reliability and validity [5, 76, 77].
- A system for selecting sessions or units to be observed must be established to minimize bias. This system could include making unannounced observations or a random sampling of teaching sessions and teachers [29] to assess sampling error and control selection biases.
- The program has sufficient resources to pay for trained, independent observers, which is a considerably more costly approach to data collection than is collecting self-reported information.
- A system must be established to periodically assess, and correct as necessary, interobserver reliability in the field and observer drift.

SUMMARY

Measuring program implementation and policy adoption is a critical component of evaluating school programs to promote physical activity and nutrition. Data on program implementation are useful for monitoring program progress, refining program elements, and explaining program effects. Most comprehensive school health programs are multifaceted with components to address classroom curricula, physical activity, food service, and policy. To capture information about each of these diverse elements, researchers will have to use multiple data collection approaches.

Most efforts to measure program implementation have focused on the quantity and quality of implementation. For implementation of classroom curricula, investigators measure the hours of instruction, the percentage of activities covered, and ratings of the extent

to which activities were completed. Implementation measures for physical education programs include counts of the opportunities for physical activity and of the actual minutes of physical activity in which students engage at school. Measures of implementation for food service programs include changes in food purchasing and preparation practices, chemical analyses of food served, and documentation of the number and types of activities to promote healthy foods served through the school meals. Data to measure program implementation can be obtained through a variety of approaches including archival records, self-reports recorded on logs or questionnaires, structured and open-ended interviews, and observations conducted by trained observers. Measures of policy adoption and change can include documenting the nature of written policies and key decisions made at meetings or other events leading up to policy decisions.

The assessment of program implementation and policy adoption in school settings is still underdeveloped. More attention should be paid to clearly and consistently defining constructs used to measure implementation, thoroughly describing assessment methods and tools used, and assessing the reliability and validity of assessment tools.

REFERENCES

- Centers for Disease Control and Prevention. Guidelines for school and community programs to promote lifelong physical activity among young people. *MMWR* 1997;46(No. RR-6):1-36.
- Centers for Disease Control and Prevention. Guidelines for school health programs to promote lifelong healthy eating. *MMWR* 1996;45(No. RR-9):1-41.
- Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med* 1986;15:451-74.
- Rossi PH, Freeman HE. *Evaluation: a systemic approach*. Vol. 5. Newbury Park (CA): Sage, 1993.
- Windsor RA, et al. *Evaluation of health promotion and Windsor programs*. Mountain View (CA): Mayfield, 1984.
- Israel BA, et al. Evaluation of health education programs: current assessment and future directions. *Health Educ Q* 1995;22(3):364-89.
- Basch CE, et al. Avoiding type III errors in health education program evaluations: a case study. *Health Educ Q* 1985; 12(4):315-31.
- Davis M., et al. Gimme 5 fruit and vegetables for fun and health: process evaluation. *Health Educ Behav* 2000; 27(2):167-176.
- Stone EJ, et al. Operational design and quality control in the CATCH Multicenter Trial. *Prev Med* 1996;25:384-99.
- Williams W, Implementation analysis and assessment. In: Sechrest L, et al. editors. *Evaluation studies: review annual*. 1979, Sage Publications: Beverly Hills (CA); 1979:267-92.
- Scheirer MA, Designing and using process evaluation. In: Wholey JS, Hatry HP, and Newcomer KE, editors. *Handbook of practical program evaluation*. San Francisco (CA): Jossey-Bass, 1994: 40-68.
- McGraw SA, et al. Methods in program evaluation: the process evaluation system of the Pawtucket Heart Health Program. *Eval Rev*, 1989;13(5):459-83.
- Dehar M-A, Casswell S, Duignan P. Formative and process evaluation of health promotion and disease prevention programs. *Eval Rev*, 1993;17(2):204-20.
- McGraw SA, et al. Design of process evaluation within the child and adolescent trial for cardiovascular health (CATCH). *Health Educ Q Suppl*, 1994;S2:S5-26.
- Resnicow K, et al. How to best measure implementation of health curricula: a comparison of three measures. *Health Educ Res*, 1998;13:239-50.
- Murray DM, et al. Results from a statewide approach to adolescent tobacco use prevention. *Prev Med*, 1992. 21:449-72.
- Connell DB, Turner RR. The impact of instructional experience and the effects of cumulative instruction. *J Sch Health* 1985; 55(8):324-31.
- Connell DB, Turner RR, Mason EF. Summary of findings of the school health education evaluation: health promotion effectiveness, implementation, and costs. *J Sch Health* 1985;55(8): 316-21.
- McGraw SA, et al. Using process data to explain outcomes: an illustration from the child and adolescent trial for cardiovascular health (CATCH). *Eval Rev* 1996;20(3):291-311.
- Nader PR, et al. The effect of adult participation in a school-based family intervention to improve children's diet and physical activity: the Child and Adolescent Trial for Cardiovascular Health. *Prev Med* 1996;25(4):455-64.
- Pentz MA, et al. Effects of program implementation on adolescent drug use behavior: the Midwestern Prevention Project (MPP). *Eval Rev* 1990;14(3):264-89.
- Resnicow K, et al. A three-year evaluation of the Know Your Body program in inner-city schoolchildren. *Health Educ Q* 1992;19(4):463-80.
- Taggart VS, et al. A process evaluation of the District of Columbia "Know Your Body" project. *J Sch Health* 1990;60(2):60-6.
- Resnicow K. Evaluation and research issues for the Committee on Comprehensive School Health Promotion. In: Allensworth D, et al., editors. *Schools & health: our Nation's investment: Report from the National Academy of Science, Institute of Medicine*. Washington (DC): National Academy Press; 1997, 6/1-6/18.
- Steckler A, Goodman RM. How to institutionalize health promotion programs. *Am J Health Promot* 1989;3(4):34-44.
- Baranowski T, et al. Theory as mediating variables: why aren't community interventions working as desired. *Ann Epidemiol* 1997;S7:S89-95.
- Cullen KW, Bartholomew LK, Parcel GS. Girl scouting: an effective channel for nutrition education. *Journal of Nutrition Education* 1997;29(2):86-91.
- Rohrbach LA, Graham JW, Hansen WB. Diffusion of a school-based substance abuse prevention program: predictors of program implementation. *Prev Med* 1993;22:237-60.
- Edmundson EW, et al. CATCH: classroom process evaluation in multicenter trial. *Health Educ Q Suppl* 1994;S2:S27-50.
- Ross JG, et al. Teenage health teaching modules: impact of teacher training on implementation and student outcomes. *J Sch Health* 1991;61(1):31-4.
- Gunn WJ, Iverson DC, Katz M. Design of the school health education evaluation. *J Sch Health* 1985;55(8):301-4.
- Parcel GS, et al. School promotion of healthful diet and physical activity: impact on learning outcomes and self-reported behavior. *Health Educ Q* 1989;16(2):181-99.
- Smith DW, et al. Teachers' use of health curricula: implementation of growing healthy, project SMART, and the teenage health teaching modules. *J Sch Health* 1993;63(8):349-54.

34. Basen-Engquist K, et al. The effect of two types of teacher training on implementation of smart choices: a tobacco prevention curriculum. *J Sch Health* 1994;64(8):334-9.
35. McCormick LK, Steckler AB, McLeroy KR. Diffusion of innovations in schools: a study of adoption and implementation of school-based tobacco prevention curricula. *Am J Health Promot* 1995; 9(3):210-9.
36. Perry CL, et al. The child and adolescent trial for cardiovascular health (CATCH): intervention, implementation, and feasibility for elementary schools in the United States. *Health Educ Behav* 1997;24(6):716-35.
37. Perry CL, Murray DM, Griffin G. Evaluating the statewide dissemination of smoking prevention curricula: factors in teacher compliance. *J Sch Health* 1990;60(10):501-4.
38. Botvin GJ, et al. Long-term follow-up results of a randomized drug abuse prevention trial in a white middle-class population. *JAMA* 1995;273(14):1106-12.
39. Lytle LA, et al. CATCH: challenges of conducting process evaluation in a multicenter trial. *Health Educ Q Suppl* 1994;S2:S129-42.
40. Botvin GJ, et al. Preventing adolescent drug abuse through a multimodal cognitive-behavioral approach: results of a 3-year study. *J Consult Clin Psychol* 1990;58(4):437-46.
41. Contento IR, et al. A formative evaluation of the American Cancer Society changing the course nutrition education curriculum. *J Sch Health* 1992;62(9):411-6.
42. Matheson D, Achterberg C. Description of a process evaluation model for nutrition education computer-assisted instruction programs. *J Nutr Educ* 1999;31(2):105-13.
43. McKenzie TL, et al. Effects of a curriculum and inservice program on the quantity and quality of elementary physical education. *Res Q Exerc Sport* 1993;64:178-87.
44. McKenzie TL, et al. School physical education: effect of the child and adolescent trial for cardiovascular health. *Prev Med* 1996;25(4):423-31.
45. McKenzie TL, et al. CATCH: physical activity process evaluation in a multicenter trial. *Health Educ Q Suppl* 1994:S73-90.
46. McKenzie TL, Sallis JF, Nader PR. SOFIT: system for observing fitness instruction time. *J Teach Phys Educ* 1991;11:195-205.
47. McKenzie TL. Observational measures of children's physical activity. *J Sch Health* 1991;61:224-7.
48. Snyder MP, Story M, Lytle L. Reducing fat and sodium in school lunch programs: the Lunchpower! intervention study. *J Am Diet Assoc* 1992;92:1087-91.
49. Raizman DJ, et al. CATCH: food service program process evaluation in a multicenter trial. *Health Educ Q Suppl* 1994:S51-71.
50. Osganian SK, et al. Changes in the nutrient content of school lunches: results from the CATCH Eat Smart food service intervention. *Prev Med* 1996;25(4):400-12.
51. Parcel GS, et al. School promotion of healthful diet and exercise behavior: an integration of organizational change and social learning theory interventions. *J Sch Health* 1987;57(4):150-6.
52. Luepker RV, et al. Outcomes of a field trial to improve children's dietary patterns and physical activity. *JAMA* 1996;275(10): 768-76.
53. Simons-Morton BG, et al. Promoting physical activity and a healthful diet among children: results of a school-based intervention study. *Am J Public Health* 1991;81:986-91.
54. Downey AM, et al. Implementation of Heart Smart: a cardiovascular school promotion program. *J Sch Health* 1987;57(3): 98-104.
55. Sandoval WM, Lockner DW, Adkins EW. Modified school lunch menus based on dietary guidelines: II. Acceptability as determined by plate waste. *Sch Food Serv Res Rev* 1986;10(1):31-4.
56. Nicklas TA, et al. Reducing total fat, saturated fatty acid, and sodium: the CATCH Eat Smart school nutrition program. *Sch Food Serv Res Rev* 1992;16:114-21.
57. Lytle L, Kelder SH, Snyder MP. A review of school food service research. *Sch Food Serv Res Rev* 1992;16(1):7-14.
58. Kolbe LJ, et al. The school health policies and program study (SHPPS): context, methods, general findings, and future efforts. *J Sch Health* 1995;65(8):339-43.
59. Burghardt JA, Devaney BL. School nutrition dietary assessment study. *Am J Clin Nutr* 1995;61(1(S)).
60. Frank GC, et al. Cardiovascular health promotion for school children. Part 2 observations on institutional food service change affecting student eating behavior. *Sch Food Serv Res Rev* 1989;13:137-45.
61. Lee RD, Nieman DC, Rainwater M. Comparison of eight micro-computer dietary analysis programs with the USDA Nutrient Data Base for Standard Reference. *J Am Diet Assoc* 1995;95: 858-67.
62. Hoover LW. Nutrient data bases: choosing wisely. *Clin Nutr* 1987;6:198-207.
63. Eck LH, et al. A comparison of four commonly used nutrient database programs. *J Am Diet Assoc* 1988;88(5):602-4.
64. Ebzery ML, et al. School meal data collection and documentation methods in a multisite study. *Sch Food Serv Res Rev* 1996; 20(2):69-77.
65. Schmid TL, Pratt M, Howze E. Policy as intervention: environmental and policy approaches to the prevention of cardiovascular disease. *Am J Public Health* 1995;85(9):1207-11.
66. Shannon B, et al. The status of school-based nutrition education at the state agency level. *J Sch Health* 1992;62(3):88-92.
67. Elder JP, et al. Tobacco use and measurement, prediction, and intervention in elementary schools in four states: the CATCH study. *Prev Med* 1996;25(4):186-94.
68. Elder JP, et al. Process evaluation of environmental factors and programs. *Health Educ Q Suppl* 1994:S107-28.
69. Breau MLM. Lessons from the food and nutrition policy for New Brunswick schools. In: Fourth Canadian Conference on Health Promotion. June, 1996. Montreal (Canada), 1996.
70. Parcel GS, Simons-Morton BG, Kolbe LJ. Health promotion: integrating organizational change and student learning strategies. *Health Educ Q* 1988;15(4):435-50.
71. Milio N. Making healthy public policy: Developing the science by learning the art: an ecological framework for policy studies. *Health Promot* 1988;2(3):263-74.
72. Steckler A, et al. Toward integrating qualitative and quantitative methods: an introduction. *Health Educ Q* 1992;19(1):1-8.
73. Patton MQ. *Qualitative evaluation methods*. Beverly Hills (CA): Sage, 1983.
74. Brink SG, Levenson-Gingiss P, Gottlieb NH. An evaluation of the effectiveness of a planned diffusion process: the Smoke-free Class of 2000 project in Texas. *Health Educ Res* 1991;6:353-62.
75. McKenzie TL, et al. BEACHES: an observational system for assessing children's eating and physical activity behaviors and associated events. *J Appl Behav Anal* 1991;24:141-51.
76. Carmines EG, Zeller RA. *Reliability and validity assessment*. Beverly Hills (CA): Sage, 1979.
77. Scheirer MA, Rezmovic EL. Measuring the degree of program implementation. *Eval Rev* 1983;7(5):599-633.
78. Baranowski T, Davis Hearn M, Resnicow K, Baranowski J, Doyle C, Smith M, Lin L, Wang DT. Gimme 5 fruit and vegetables for fun and health: Outcome evaluation. *Health Educ Behav* 2000;27(1):96-111.