

Youth Development and Physical Activity

Linking Universities and Communities

Don Hellison, PhD

University of Illinois at Chicago

Nick Cutforth, PhD

University of Denver

James Kallusky, EdD

California State University, Los Angeles

Tom Martinek, EdD

University of North Carolina at Greensboro

Melissa Parker, PhD

Jim Stiehl, PhD

University of Northern Colorado



Human Kinetics

Program Evaluation

Tom Martinek

Evaluating the processes and outcomes of programs for underserved youth is, like teaching and coaching, a messy business that does not always follow a step-by-step blueprint. Often the line between instruction and evaluation becomes blurred because both activities require observing, asking questions, and trying and retrying things. Once programs begin, however, people need to evaluate them to pave the way for change and improvement. Evaluation involves collecting information about program activities, characteristics, and outcomes. For this information to be useful, some kind of plan must be devised that fits the goals of your program. The purpose of this chapter is to explore ways that youth development programs such as those described in this book can be evaluated. The key is to identify processes and outcomes that will be useful to those directly involved in the programs—teachers, youth workers, university faculty, and university students. Both traditional and nontraditional methods will be described, along with various data sources (e.g., questionnaires, interviews, observations, journals). In addition, the chapter provides specific ideas that will enable you to evaluate your own programs.

Why Do We Evaluate?

Probably one of the most important reasons we evaluate programs is to improve them. The programs described in the previous chapters undergo continual change because we see that some ideas are not working or that the program's goals are not being met. Very often evaluation occurs on the spot because something needs to be fixed quickly. When things are not working because of poor matching of kids, unclear directions, coaches who are too competitive, or inappropriate learning tasks (too difficult or easy), immediate attention is required. Because we

believe that evaluation is ongoing in all we do in the clubs, being able to address on-the-spot problems requires careful attention or what Kounin (1970) called “withitness.”

Other changes are linked to long-term observation, where patterns (some good and some not so good) begin to emerge. This type of evaluation tends to be more formal and requires careful and systematic analysis of the program’s operation. For example, in one of the Coaching Clubs in Greensboro, office referrals were tracked throughout an entire school year. The data were organized around each nine-week grading period. Changes across the periods indicated how certain program goals were being met, especially those that focused on self-control and respect in the school setting (Martinek, McLaughlin, & Schilling, 1999). In general, evaluation results have been used in programs to

- assess need,
- document how well a program is working,
- determine how participants view the program,
- show the usefulness of a particular approach,
- examine how certain parts of the program (e.g., mentoring, apprentice teaching) are working, or
- provide information for modifying program goals.

Assumptions About Evaluation

Evaluating programs that use the responsibility model is difficult. Often we have struggled with some ideas that seem to work and also with those that have failed miserably. Getting it right takes time—sometimes years! Whether evaluation is done in a traditional or nontraditional way, the meaningfulness of the data relies on basic assumptions. Knowing these assumptions is critical in developing any kind of evaluation scheme, because they provide the guideposts for interpretation and application. Here are five assumptions that have guided all of the authors’ evaluation efforts over the past several years:

1. There is commitment to making a difference.
2. There is no one way of doing evaluation.
3. Evaluation does not have to be done by outsiders.
4. Findings are useful to stakeholders (kids, youth leaders, coaches, teachers, etc.).
5. The program is running smoothly.

Commitment to Making a Difference

Evaluating a program should result in making a difference in kids' lives. This type of commitment needs to be at the forefront of any evaluation plan. Without this commitment, the evaluator will assuredly short-circuit the process of tying the findings to program improvement. Motives vary for doing program evaluation. Some individuals evaluate because there is a need to be accountable—no argument from us on this one. Programs based on the responsibility model are often met with skepticism. Therefore, some program directors may feel they have to provide evidence that something is happening; they have to prove that kids are becoming more responsible! Others feel that funding agencies will want “hard data” to fund programs or because they want data to publish. Unfortunately, such motives often leave kids out of the picture. If kids are going to be the benefactors of responsibility-based programs, they must come first. This assumption encourages the staff to remain connected to the youngsters throughout the evaluation process.

There Is No One Way to Evaluate

Many people feel that they have to select a single best design when evaluating a program. Perhaps you were taught in your measurement class that there is one best design. We guess that the reasoning behind this is that for every problem there is a single best answer. However, we must account for many complex factors when examining programs for underserved youth. Creativity enables us to attend to the multiple roles, values, and situations inherent in youth work. Consequently, there is no best way to design or conduct a program evaluation. Michael Patton (1987) argued that maintaining “rigor” in design will lead you away from evaluating what is useful to the staff and the kids in your programs. A scientific or technical approach leaves little room for looking at the genuineness of any program because it attempts to mold and define the program to fit a model of how things should be evaluated. Therefore, modifying and matching evaluations to the uniqueness of the program enable one to be responsive to the needs and interests of those who play and work in our programs. Rather than being faithful to some model, we must be faithful to the characteristics of the program (Patton, 1987). We need to adopt a “real world” view so that we are able to recognize and deal with a multitude of choices and decisions in the evaluation process.

Evaluation Is an “Inside Job”

The authors of this book not only plan programs, but actually work with kids who attend them. This places us in a good position to know what, who, and when to evaluate. It also puts us closer to information

derived from evaluation. Outsiders who are not intimately connected with the program are often unable to effectively translate their findings to the situation. That is, outside evaluators are often unfamiliar with the social and economic constraints of the program's setting, the varied dispositions of the kids in the program, and personal values of the individuals running the program. Some of our colleagues in the sciences (e.g., physiologists, psychologists, biomechanists) would argue about potential biasing because they believe we are too close to our own data! We believe, however, that having biases is not necessarily bad. After all, why shouldn't we be biased, especially when it comes to working with underserved youngsters? Why shouldn't we deliberately look for things that we have struggled with in our programs? We know the types of information we already collect, such as journal entries, informal and formal observations, and other anecdotal records from individual and group discussion. If we truly want the evaluation strategy and the findings to relate to what we do with our kids, then we must be involved with the development, delivery, and application of the evaluation program.

Data Are Useful to All

Evaluation data are important and must be useful both to those who run the programs and to youngsters. Program staff should be able to make sense of the data so they can adjust the program. Likewise, youngsters and others (parents, teachers, mentors, volunteers) should benefit from the products of evaluation. Many types of data are available to the program evaluator. Deciding what to use will depend on the goals of the program and the feasibility of obtaining the information from participants and other data sources. No matter what kind of data are available, it would be foolhardy to gather information that was not useful or interpretable to those who run the programs. This requires that the results of any evaluation effort be communicated in a way that all parties can understand.

The Program Is Running Smoothly

Until the program is running relatively smoothly, traditional evaluation methods are not very helpful. That is, there is little to be gained in doing pre and post measures until the program leaders clearly know what they are evaluating! This is not to say that informal, ongoing evaluation isn't a must—it is vital to providing positive day-to-day experiences for kids. Unfortunately, there is often a rush to put a traditional evaluation scheme in place at the beginning of a program. A design is created, measures are selected, a timetable is established for

data collection, and analysis procedures are predetermined. However, this type of traditional planning is not very useful in the long run. A main reason for this is that it takes time to get the program right. There are many “bumps in the road” that have to be dealt with before one knows what to evaluate. One of these bumps is getting kids to understand the purposes of the clubs. Program leaders continually use various trial-and-error methods and some kids don’t attend regularly. Another bump is not using effective teaching strategies, or selecting the wrong strategy to use. Kids bring varied and unexpected dispositions into the gym, requiring on-the-spot changes. In short, those who are just starting programs need to approach evaluation with a clear understanding of what they are doing.

Evaluating Youth Programs

The remainder of this chapter describes evaluation approaches through examples of programs that have used the responsibility model with underserved youth. The examples show various evaluation designs and tools for collecting information about each program. The examples are to serve as guides for developing your own ideas for evaluation.

There is always a tendency to compare programs to a set of standards or criteria such as those presented in chapter 3. This is a legitimate way of describing the quality of programs. However, another way is to look at specific outcomes or what kids gain from participating in a responsibility-based program. This chapter attempts to do both without adhering to any set of criteria for either process or outcome measures. Although programs described in the text use the responsibility model, they differ in the ways of using the model and determining its effectiveness. Each example includes a description of

- what is being evaluated and why,
- the data collection process and analysis,
- the results, and
- the lessons learned from the evaluation process.

The examples are presented in progression starting with traditional approaches and then ending with nontraditional, creative forms of evaluation. In each example, one or multiple approaches are emphasized to provide you with detailed information for use in your own programs. An overview of the five examples is presented in table 12.1.

Overview of Evaluation Programs

Description	Evaluator(s)	Topic	Methods	Findings	Application
TRADITIONAL MODELS					
Chicago Coaching Club	T. Cummings (1997)	Club impact on future school performance.	Participant and control group comparisons.	Dropout rate decreased, no differences were found for grade retention and absenteeism.	Strong link between being in the club and staying in school.
Chicago Martial Arts Club	P. Wright (1998)	Club impact on conflict resolution skills.	High- and low-attendance group comparisons.	Conflict resolution skills improved.	Program promoted positive values and leadership.
LESS TRADITIONAL MODELS					
Chicago apprentice	N. Cutforth & K. Puckett (1999)	Impact of apprentice teacher program on leadership and commitment.	Descriptive study of veteran club members who participated in a summer apprentice teacher program.	Some kids were effective teachers and showed strong commitment to club values.	Apprentice teaching appears to foster leadership and commitment to club values.
Greensboro veteran club members	T. Schilling (1999)	Describe why kids "stay with" the club over time.	Descriptive study of how kids view commitment and what factors influence levels of commitment.	Program structure and environment, relationships, and personal characteristics influenced commitment.	Including fun activities and clear goals, giving students a voice in the program, and developing relationships.
CREATIVE MODEL					
Greensboro sport	I. Martinek, T. Schilling, & D. Johnson (1999)	Self-designed case study of club members and their ability to transfer.	Matrix analysis of data obtained from participant portfolios.	Some goals were met both in gym and classroom; transfer of goals was indicated.	Transfer of club goals needs to be focus of club and mentor programs.

Example One: Comparing School Outcomes Using Club Members and Control Groups

Teresa Cummings's (1997) study of "graduates" from a basketball Coaching Club at a Chicago elementary school (see chapter 7) is an example of how a traditional program evaluation looks. The study focused on the impact that the club experience had on the graduates' attendance, grades, and dropout rate. In essence, Cummings wanted to know whether the club experience was a significant factor in future school performance.

Using an experimental-control group design, Cummings compared a control or "nonparticipant" group with former club members. Both groups were made up of kids with similar backgrounds. She analyzed data by graphing group differences (percentage dropping out) and using statistical techniques for looking at absenteeism, grades, and retention. Past evaluation of clubs had been qualitative and provided various images of how the program was working for kids and staff members (Debusk & Hellison, 1987; Mulaudzi, 1995). However, quantitative evaluation of dropout rate, retention, grades, and absenteeism had not been done during the program's six-year tenure.

Cummings found a notable difference between the two groups, with none of the club members dropping out of school as compared with the nonparticipant group, which had a 34 percent dropout rate. She also found no significant difference between the two groups for grade retention and absenteeism. In other words, club members and nonparticipants repeated a grade and were absent about the same amount of time.

Cummings's evaluation suggested a strong link between being a basketball club member and staying in school. Unfortunately, there was no way of determining whether being in the club was a reason for staying in school. Knowing this would help determine whether there was an actual cause-and-effect relationship between club participation and dropout rates. Qualitative research via interviews would help in determining this. In presenting this to program staff, Cummings stressed the importance of doing some sort of qualitative follow-up (i.e., interviews) to identify the reasons for staying in school. Although keeping at-risk youngsters in school is an important goal for many programs, it was not a major goal in Hellison's early inception of the responsibility model. Rather, he wanted to focus on getting kids to feel good about themselves, clarifying their strengths and identity, and giving them purpose in life.

Example Two: Comparing Aggression Levels Using Subgroups Within a Martial Arts Club Program

Paul Wright's study (1998) of a Chicago middle school Martial Arts Club is another example of a traditional two-group design. Paul evaluated two groups of students based on their attendance records at the club; no actual control group was used. He focused on the number of aggressive acts by students and their responses to confrontation. The students were from an elementary school located in the northwest part of Chicago; all were participants in the Martial Arts Club described in chapter 7. The teachers were becoming concerned over the increased amount of fighting among students at their school. In addition, physical and verbal confrontations were becoming more extreme, sometimes involving the use of weapons. Some teachers had been assaulted by students.

Wright was trying to see whether the frequency, types, and causes of conflict and the use of conflict resolution skills were related to attendance in the Martial Arts Club for the previous two years. Unlike Cummings's use of independent groups, club members were divided into three groups based on their attendance records during the past two years. Wright felt that students with low attendance rates could serve as a control group. He was especially interested in how the club members were able to stay away from conflict in their schools and neighborhoods.

He found that a variety of circumstances sparked conflicts. Examples of situations leading to conflict were basketball games, classroom disputes, stolen property, gang-related problems, and sticking up for someone. The highest number of incidents were the result of actions during basketball games (e.g., fouls, ball hogging). Outside the school, incidents such as being jumped or attacked by rivals and fighting with friends and relatives were also frequent, indicating that the participants' natural environment was violent and unsafe.

Wright also created a conflict resolution scale that categorized the types of conflict resolution skills. He found that participants who had the lowest attendance fell between violent and aggressive on the conflict resolution scale. Students in the middle attendance group were neutral on the conflict resolution scale. The group with the highest attendance were more assertive and peaceful on the scale, in terms of conflict resolution, than the other two groups.

One concern with using this scale related to categorizing responses that contained two or more types of conflict resolution. For example, one of the youngsters and a friend were assaulted one day. The youngster first tried to walk away but eventually physically defended himself after being attacked. He then left and called on some of his friends to help him retaliate. Although he started out as being assertive, the scenario ended

with revenge. Because each student was only assigned to one conflict resolution category, Wright decided to maintain consistency by assigning the students according to the most extreme behavior. This may not have allowed us to see the whole picture. Therefore, it may be useful to include another category referring to the youngster's initial response.

Although club attendance and conflict resolution appear to be related, causality could not be inferred. This study used a descriptive design rather than an experimental one. It is possible that students who already have positive values and conflict resolution skills are drawn to the club and stay with it over time. Similarly, some students may not commit to the club and stop attending because the club values and their own values do not match. Thus, rather than serving as the impetus for changing values, the club may serve as a supportive and reinforcing environment in which youngsters can comfortably accept and act on their own sense of right and wrong. This may be different from what they encounter in their natural environment.

Example Three: Examining the Social and Psychological Outcomes of an Apprentice Teacher Using Journals, Interviews, and Questionnaires

Nick Cutforth and Karen Puckett (1999) evaluated a specialized program called the Apprentice Teacher Program. The program was designed to give urban youngsters the opportunity to teach basketball to young children attending a summer basketball camp. Unlike Cummings's and Wright's programs, this one focused on one group of 11 kids. Both quantitative and qualitative methods were used to extract information about experiences in this sport camp for younger kids. This program was a nice example of how quantitative and qualitative data could be combined in the evaluation process.

All of the students were in or had been in the before-school Coaching Club previously described in Cummings's evaluation program. Their responsibilities were to teach basketball and the responsibility goals to approximately 40 eight- and nine-year-olds from a nearby housing project for an hour a day, four days a week, over a five-week period.

The evaluators used informal observations, entries from a personal journal kept by the director, and interview responses from the director and apprentice teachers before and after the program as the main data sources for the evaluation. In addition, the director took attendance of the apprentice teachers and graded (i.e., A, B, C, D, F) their effectiveness each day.

An important aspect of this evaluation was that the evaluators knew the teachers and had worked with them in previous years in the

Coaching Club. This rapport help considerably in getting the apprentice teachers to respond with candor during the interview and thereby increasing the validity of the data.

The evaluators concluded from their data that the "core" program staff consisted of eight reliable teacher apprentices, the program director, and the graduate assistants. The journal entries and interview responses of the director indicated that the greatest challenge was to make the apprentice teachers aware of their leadership role, particularly the need to take charge of the sessions. Some teachers were organized and seemed to get down to business teaching skills and motivating the children. However, some teachers needed to be checked consistently throughout the camp to be sure they were prepared for the morning activities. Others simply needed confirmation that they were on track. The evaluation data showed that training in managing student behavior and teaching basketball skills was needed. It was also concluded that ways to dilute the competitive edge of some of the teachers needed to be explored.

The evaluation scheme of Cutforth and Puckett provided a nice baseline of information for the Apprentice Teacher Program. Overall, the evaluation showed that the program increased the apprentice teachers' awareness that they could contribute to other people's lives. They also knew that they had done something good that summer and began to see an alternative future for themselves. This appeared to foster an eagerness to stay with the program and be a part of it the next year.

Example Four: Using Interview and Q-Sort Techniques to Look at Commitment to Program Goals

Tammy Schilling's (1999) evaluation program looked at "veteran" club members of University of North Carolina at Greensboro's Project Effort (see chapter 10). Because the program originated with participants in an elementary school in a poor area of Greensboro, the program staff were interested in seeing why club members stay with the program into their high school years. Therefore, the purpose of Schilling's evaluation program was to determine program involvement and commitment over the length of the program's history (i.e., five years) through the participants' perspective.

Schilling chose a case study method that allowed for an in-depth look into the experiences of individual participants. Data collection included individual and focus interviews and a card-sorting task. The focus group interview involved asking questions to all seven of the participants as a group. Schilling used this method because she wanted to ask

them about barriers (e.g., gang involvement, school activities) to program commitment and to compare their own and the staff's commitment levels. She felt that the participants would be more likely to answer the questions honestly and to elaborate on their feelings in the supportive presence of the other kids, rather than by themselves.

For the card sort, Schilling took main themes (e.g., "learning new activities" and "having a mentor") or ideas from the participants' responses in an earlier interview and asked them to rate the relevance and importance of each theme with respect to their personal involvement. The card sort allowed Schilling to look across the program participants and note the aspects that were consistently most and least important to the participants' involvement.

As with the other studies, it was essential for Schilling to use multiple data sources to find out about the participants' program commitment, particularly because the program participants were youngsters. Therefore, she interviewed the parents/caregivers to get their perceptions of their child's program involvement and commitment. Two of the program leaders were also interviewed. The leaders developed commitment profiles for each participant that included things such as commitment ratings of factors most important to the participant, change over years of involvement, parts of the program that they like/dislike, and examples of the participant transferring the goals of Project Effort outside the gym.

The analysis consisted of two phases. The first phase included the development of historical profiles and the card-sorting task. This information provided a backdrop for subsequent information regarding the participants' commitment.

The second phase consisted of a content analysis that included the identification of main data themes (i.e., meaningful quotes) from all the participants' data regarding commitment. Schilling also examined which themes were validated by the parent/caregiver, leader, and/or focus group data. Finally, a cross-case analysis was conducted, which allowed Schilling to look across the cases of all five participants to note similarities or interesting trends.

The results suggested some considerations for program development. The participants noted that important factors related to their program involvement were opportunities to try out new activities and participate in specific activities (e.g., basketball, swimming), having fun, having a voice in how the program runs, staying out of trouble, and having program goals. The participants also perceived that their commitment across a multiyear period was influenced by program structure (e.g., type of activity), program environment (e.g., having fun), relationships (e.g., with staff and other participants), and personal characteristics (e.g.,

lack of alternative options for participation). The results reinforced what research has consistently shown—make it fun and they will come! However, it was clear that participating in fun activities was just a part of the big picture. Developing close relationships with other participants and the staff members and being given responsibility and leadership opportunities were also central components of the participants' commitment across time.

Example Five: Evaluating Program Outcomes Using Matrix Development From Journal Entries, Interview Responses, and Informal Observations

In this evaluation program, a more "creative" approach was used by me along with Tammy Schilling and Dennis Johnson to find out whether the goals of a program were being met. By creative, I mean that a self-designed strategy was used, one that strayed from traditional methodologies. The strategy focused on the development of a matrix that helped the evaluators look at how each club member was meeting goals. A main issue with the staff was to find ways of organizing and interpreting an ample amount of information that was collected during the program. Again, Project Effort was the program being evaluated. The evaluators were particularly interested in seeing how well the values learned in the sport club were being applied to the classroom and elsewhere.

We developed a participant-goal matrix (Demos, 1989; Van Tulder, Van der Vegt, & Vecnman, 1993) to organize and interpret student portfolios. The portfolios included informal notes taken by staff, journal entries, and end-of-year interview responses. The matrix allowed staff to see how each club member performed in relation to the values and goals of the club. The matrix also enabled us to see how well the club members transferred the values and goals of the responsibility model to the classroom setting.

All of the information was placed in a student folder that served as the portfolio. Using a specified procedure, a matrix was created from the contents of the portfolio. The matrix allowed the users to compare common and contrasting patterns. The procedure consisted of first having staff review all the data (journal cards, mentor journals, etc.) collected for each student. After each student's portfolio was reviewed by a staff member, a code was assigned to each program goal for the gym and classroom settings. For example, a student was coded for showing self-control and respect for others as these items related to the two settings. Then the student was coded for trying things out, and so on.

A student was coded with a plus symbol (+), an asterisk (*), or a minus symbol (-) for each goal category. A plus meant that the student demonstrated that goal most of the time. An asterisk meant that each

goal was demonstrated some of the time. A minus meant that each goal was demonstrated little of the time. All the staff evaluators reviewed and coded each student. Table 12.2 provides a sample of the matrix.

The data showed that the youngsters were able to acquire a degree of persistence at learning tasks in the gym and classroom. Their ability to set some goals and respect others in the gym setting also appeared to be augmented by the program.

The club members struggled, however, to transfer some of the club's values to the classroom. For example, goal setting was a persistent problem for the kids. Likewise, incidences of self-control, respect, and caring for others were not as high in the classroom as in the gym. The evaluators felt that this inability to transfer the values of the gym to the classroom also partially reflected a lack of ownership in the values of the club.

Principles Learned From the Evaluation Process

All of the leaders discussed in this chapter have applied certain principles for evaluating their programs. Some of these principles are common for all evaluators, and some are unique to the setting. In all cases, they reflect the essentials for conducting program evaluation: use of control groups, accessing various data sources, organizing and

Table 12.2

Program-Goal Matrix

Student	Personal responsibility				Social responsibility				Ownership
	Effort (trying things out)		Self-direction (set goals)		Self-control		Caring		Ability to transfer
	G	C	G	C	G	C	G	C	G to C
Rayshone	+	*	+	-	+	+	*	*	M
Lateesha	+	+	*	+	+	*	+	*	H
Leonard	*	-	-	-	+	-	-	-	L

+ Showed most of the time. * Showed some of the time. - Showed little of the time. C = classroom; G = gym.

analyzing data, and understanding limitations of applying results. We hope you can apply some of these principles to your program.

Principles Applied to Using Control or Comparison Groups

Some individuals feel that if you don't have a true control group, you cannot find answers to your questions. However, Wright showed that with a little creativity, you can compare different groups of kids within one program and discover possible relationships between the variables of interest (e.g., conflict resolution skills, referrals). By simply selecting kids with low attendance, he was able to establish a so-called control group.

Cummings's study presents a cautionary principle in the traditional use of a control or comparison group. It is important to recognize the difficulty of using a control group of underserved youngsters. This is especially problematic when the participants are volunteers; that is, they choose to come to a program. Trying to control for prior achievement, grades, socioeconomic status, or other factors does not account for the obvious motivational levels of the kids who elect to participate in the clubs or the teacher, principal, and perhaps parents who urge them to attend. The most common solution would be to randomly assign all the kids who wanted to be in the club to the program or to a "waiting list" group. This assumes a large number of kids want to be in the club, which may not always be the case. The waiting list or control group supposedly becomes equivalent to the treatment or program group.

Although this strategy seems sound from a pure research perspective, there are two problems with it. The first is an ethical concern: What happens to the kids on the waiting list? Are they to be ignored or put on hold just for the sake of research? The second concern relates to more practical issues regarding group comparisons. It is extremely difficult (if not impossible) to determine a "true control" group in this type of work. Although poverty, racism, and societal indifference are common denominators among underserved youth, each youngster brings into a program circumstances and daily experiences unique in severity and scope. They also bring in various levels of adaptability that allow them to navigate through a social system fraught with economic and geographic barriers. Thus, looking at how and why individual kids respond to the club becomes extremely important. This is not to say that group comparisons cannot highlight possible causes. However, making comparisons, such as those in Cummings's evaluation, must be done with caution and insight into the true meaning of any differences found or not found.

Principles Involved in Using Various Data Sources

Where can data come from for program evaluation? This critical question often leaves evaluators scratching their heads. There are a multitude of data sources, and describing each one would exceed the page limitation of this text. However, we can offer several principles about accessing certain data sources described in the previous studies.

One data source is school records. Of course, accessing these requires clearance from administrators, kids, and parents. However, these data can be quite useful in determining overall school performance. School records can augment much of the qualitative information acquired by other evaluators of the various clubs using the responsibility model. However, the accuracy of measures such as grades, dropout rate, and absenteeism depends on the diligence of those who keep student records (i.e., teachers, school and central office staff). Computer retrieval systems are helpful but only to the extent to which data are carefully entered into the data bank.

Another source for data is questionnaires. Some questionnaires can come from previous work or can be devised by the evaluator. You will often find yourself stuck in the evaluation process when you cannot find any literature, scales, or past examples regarding what you are trying to find out. At these times it is particularly important to use pilot work to get a feel for what is going on and even to create your own scale for measuring or categorizing responses. Wright used pilot work to create a conflict resolution scale that allowed him to use categories that made sense for his work and the club participants. He found that the scale had to be brief and worded so the kids could understand.

Journals are another big source of data. Participant-observers like program directors, staff (mentors, teachers), and kids (i.e., see studies by Cutforth & Puckett, 1999, and Martinek, Schilling, & Johnson, 1999) are encouraged to keep personal journals. Journal entries usually include such things as feelings about what happens in the club, how kids are doing, and impressions of staff performance. The entries are an important data source for evaluating many segments of the program. Journals can include quantitative assessment (i.e., grades of performance) as well as narrative data. The narrative entries help to embellish the quantitative data. As demonstrated by Martinek, Schilling, and Johnson (1999), portfolios can be developed to provide profiles on how each club member is doing. The portfolios can contain journal entries from mentors, teachers, and staff who work with the youngsters in different capacities.

A final data source used in these studies was interview responses. Cutforth and Puckett (1999) and Martinek, Schilling, and Johnson (1999)

applied several principles to maximize the information gained from interviewing kids and staff members:

- Practice your interviews. This helps you avoid pitfalls such as poor phrasing and prompting, inappropriate questions. Although many texts provide some guidance on how to conduct interviews, these guidelines don't always apply very well with underserved kids.
- Employ a "multiple meeting" approach. Use information from previous meetings to prepare for the next meeting. This provides coherence among the various sources of information. Multiple meetings also mean that you must fit into the schedule of those you are interviewing.
- Consider parents' perceptions of their child's involvement in after-school clubs. Parents provide a much richer picture of how the values are being transferred outside the club setting.
- A wonderful by-product of interviewing kids is the extended experience of working with the club members. There is much value gained in the time spent with kids during data collection; you learn more in the extra time spent in the car or playing basketball than you normally do in running the club.
- Kids and parents should be interviewed by individuals who know them. A trusting relationship between the interviewers and the interviewees is a must if responses are to be believable. In the programs described in this chapter, the interviewers had worked with the youngsters in the clubs, thus establishing a positive rapport with them.

Principles Applied to Organization and Analysis of Data

The organizing and analysis of data can take many forms. The decision on what method to use will depend on the type of data collected and how you want to use them. Charts, biographies/stories, case studies, tables, metaphors, and themes are but a few of the ways in which qualitative and quantitative can be organized (Patton, 1987). As you become more experienced in evaluating programs, you will develop a broad repertoire of methods. Many evaluators have used a combination of methods to analyze and present their findings. For example, Cummings (1997) found that simply graphing dropout rate for both groups was all that was needed. One could then simply "eyeball" the graphic differences. On the other hand, statistical analyses highlighted group and time differences for grade retention and absenteeism.

In Cutforth's and Puckett's study (1999), attendance data, participant interviews, and the director's impression of how the apprentice teachers were fulfilling their teaching roles were interconnected to provide a true view of their commitment to the program. Schilling (1999) organized her interview and journal data around themes and categories that helped to explain effectiveness and commitment of veteran club members. Both presented their information through case descriptions. Each case provided an in-depth and unique characterization of each club member's sense of commitment to the responsibility model and the clubs.

Finally, information is often gathered throughout the course of the program. Martinek, Schilling, and Johnson (1999) organized the data into student folders or portfolios. At the end of the program, data from each portfolio were transformed into a participant-goal matrix. This matrix format requires that a simple but useful rubric be developed so that material can be evaluated. The rubric contains basic guidelines for assigning a value or code to the portfolio content. The participant-goal matrix allows users to look at how each club member did during the year. The matrix provides a general picture of how the program impacted the club members' values and their ability to transfer these values to the classroom setting. Pre and post measures do not have to be used with the matrix analysis. Rather, an overall assessment based on the content of each member's portfolio can be derived from this strategy.

Principles Applied to the Use of Information

The main premise of this chapter is that program evaluators should be concerned with how program leaders, kids, and policymakers will use the processes and findings. Three main principles guide this utility emphasis. First, utility must be in the eyes of the beholder. That is, utilization means different things to different people in different settings, and it is an issue to negotiate between evaluators and users. Knowing the needs of the program leader, the kids, and perhaps the community will clarify how the information can be used.

A second principle is that the users of the information should be identified. They are to be real, visible, specific, and caring human beings—not general or abstract audiences, organizations, or agencies. And finally, the evaluator must react, adapt to the unexpected, and interact with all those who will benefit from the findings. Design and redesign, focus and refocus, analysis and reanalysis, and interpretation and reinterpretation are all part of this process. The interplay of these processes will ensure a tighter bond between utility and user. I hope the ideas presented in this chapter will enhance your understanding of how evaluation can be meaningful and useful.

Library of Congress Cataloging-in-Publication Data

Youth development and physical activity : linking universities and communities / Don Hellison ... [et al].

p. cm.

Includes bibliographical references and index.

ISBN 0-7360-0160-3

1. Physical education of socially handicapped youth--United States. 2. Community and school--United States. 3. Student service--United States. I. Hellison, Donald R., 1938-

GV444 .Y68 2000

378.1'03--dc21

00-031926

ISBN 0-7360-0160-3

Copyright © 2000 by Don Hellison, Nick Cutforth, James Kallusky, Tom Martinek, Melissa Parker, Jim Stiehl

All rights reserved. Except for use in a review, the reproduction or utilization of this work in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including xerography, photocopying, and recording, and in any information storage and retrieval system, is forbidden without the written permission of the publisher.

Acquisitions Editor: Scott Wikgren; **Managing Editor:** Amy Flaig; **Assistant Editors:** Derek Campbell, Mark Zulauf; **Copyeditor:** Julie Anderson; **Proofreader:** Jim Burns; **Indexer:** Betty Frizzell; **Graphic Designer:** Robert Reuther; **Graphic Artist:** Dawn Sills; **Cover Designer:** Fred Starbird; **Art Manager:** Craig Newsom; **Illustrator:** Accurate Art; **Printer:** Versa Press

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Human Kinetics

Web site: <http://www.humankinetics.com>

United States: Human Kinetics, P.O. Box 5076, Champaign, IL 61825-5076
800-747-4457

e-mail: humank@hkusa.com

Canada: Human Kinetics, 475 Devonshire Road Unit 100, Windsor, ON N8Y 2L5
800-465-7301 (in Canada only)

e-mail: humank@hkcanada.com

Europe: Human Kinetics, P.O. Box JW14, Leeds LS16 6TR, United Kingdom
+44 (0)113-278 1708

e-mail: humank@hkeurope.com

Australia: Human Kinetics, 57A Price Avenue, Lower Mitcham, South Australia 5062
(08) 82771555

e-mail: liahka@senet.com.au

New Zealand: Human Kinetics, P.O. Box 105-231, Auckland Central
09-309-1890

e-mail: humank@hknewz.com