

# Responsibility-Based Youth Programs Evaluation: Investigating the Investigations

Don Hellison and David Walsh

The personal-social responsibility model (RM), a physical activity program model originally intended for underserved, has been used in physical activity and other programs for a wide range of children and youth in a number of states and several other countries. While RM's presence in practice is generally acknowledged, some in the academic community have criticized its weak empirical base. We address this criticism by reviewing 26 studies that have investigated the impact of RM on underserved and/or at-risk youth since its inception. Because field research encompasses a wide array of approaches, issues related to research design and methodology are explored in an effort to provide a rational basis for this work. Implications for theory, practice, and public policy are drawn from this review.

Sometimes you have to build something to see if it will work . . . and then you have to study the hell out of it . . . this kind of approach does not represent a weak alternative to conducting controlled experiments but a different option altogether. (Schoenfeld, 1999, p. 12)

Back in 1970, in a gloomy high school physical education gym in a low income neighborhood in Portland, Oregon, an attempt at alternative youth program exploration began, based on one person's convictions and steered by some rudimentary self-reflection. Within three years, replete with detours and dead ends, an early version of the personal-social responsibility program model (RM) emerged (Hellison, 1978). Now, some 32 years later (and counting), RM development continues (e.g., Hellison, 1985, 1995; Martinek, Schilling, & Johnson, 2001). The following model shows the latest version:

Don Hellison is with the Jane Addams College of Social Work and the College of Education at the University of Illinois at Chicago. E-mail: <hellison@uic.edu>. David Walsh is with the College of Education at the University of Illinois at Chicago.

## The Person

### Convictions

- Teaching life skills and subject matter rather than
- Lessons learned in the aspects of the program
- Instructional strategies from the program leader
- For any of these convictions recognize and respect the decision making of each

### Goals

1. Respect for the rights of
  - Self-control of tempo
  - Respect everyone's
  - Involvement in peace
2. Effort and teamwork
  - Self-motivation to effort
  - Cooperation & coaching
3. Self-direction
  - Independent work
  - Goal-setting progress
4. Helping and leadership
  - Sensitivity and respect
  - Contribution to the
5. Outside the gym
  - Trying these ideas on
  - Being a role model

### Strategies

- Lesson format: awareness
- Instructional strategy categories, individual decisions

At first recognized as a program for adolescents, RM soon spread to elementary school children in classrooms (Hellison, 1995). The development of RM (Hellison, 1995) (e.g., Oslin, Collier, & Mitchell, 1997) was formed in 1997 to extend the program to community youth organizations (e.g., Oslin, Collier, & Mitchell, 2000). RM also began to appear in other parts of the world such as New Zealand

## The Personal-Social Responsibility Model

### Convictions

- Teaching life skills and values must be integrated with the physical activity subject matter rather than taught separately.
- Lessons learned in the gym must be taught so that they can transfer to other aspects of the program participants' lives.
- Instructional strategies must be based on a gradual shift of responsibility from the program leader to program participants.
- For any of these convictions to be successful, the program leader must recognize and respect the individuality, strengths, opinions, and capacity for decision making of each program participant.

### Goals

1. Respect for the rights and feelings of others
  - Self-control of temper and mouth
  - Respect everyone's right to be included
  - Involvement in peaceful and democratic conflict resolution
2. Effort and teamwork
  - Self-motivation to explore self-effort, try new tasks, persist in tasks
  - Cooperation & coachability (when working with peer leaders)
3. Self-direction
  - Independent work
  - Goal-setting progression with courage to resist peer pressure
4. Helping and leadership
  - Sensitivity and responsiveness to others' needs and interests
  - Contribution to the well-being of both individuals and the group
5. Outside the gym
  - Trying these ideas outside the physical activity program
  - Being a role model

### Strategies

- Lesson format: awareness talk, lesson, group meeting, reflection time
- Instructional strategy categories: awareness strategies, direct instruction, strategies, individual decision-making strategies, group decision-making strategies

At first recognized as a program model for so-called at-risk and underserved adolescents, RM soon spread to more affluent communities and to programs for elementary school children as well as older kids, even finding its way into a few classrooms (Hellison, 1995). A few adventurous coaches also found useful applications of RM (Hellison, 1995) as did a few physical education teacher educators (e.g., Oslin, Collier, & Mitchell, 2001). A national partnership of six universities was formed in 1997 to extend the implementation of RM to more schools and community youth organizations in underserved neighborhoods (Hellison et al., 2000). RM also began to appear in youth programs of other countries outside North America such as New Zealand, England, and Spain.

# Physical Education Programs: Investigations

## Physical

Physical activity program  
based in physical activity  
youth in a number of  
years in practice is gener-  
ally criticized its weak  
evidence 26 studies that have  
at-risk youth since its  
wide array of approaches,  
explored in an effort to  
test theory, practice, and

... if it will work . . .  
this kind of approach  
using controlled experi-  
ments (Held, 1999, p. 12)

Physical education gym in a low  
cost alternative youth pro-  
grams and steered by some  
with detours and dead ends,  
program model (RM) emerged  
(e.g., Hellison, 1995), RM development con-  
gress, & Johnson, 2001). The

Physical Work and the College of  
Education@uic.edu>. David Walsh  
at Chicago.

RM has been recognized by scholars as an exemplary curriculum model (Bain, 1988; Steinhardt, 1992), an influential humanistic and social development model for physical education (Siedentop, 1990), a way to use sport and exercise to promote life skills (Hodge & Danish, 1999), an alternative approach for special populations (Lavay, French, & Henderson, 1997; Siedentop, Mand, & Taggart, 1986; Winnick, 1990) and for discipline problems in public school physical education (Pangrazi, 2001; Rink, 1993). Kirk (1992, p. 4) observed that "[RM offers] genuinely alternative forms of social organization in physical education classes in an attempt to constructively redress the social conditions that places some young people's well-being at risk."

### But Does it Work?

While the development of this model has appeared to answer the question "What's worth doing?" for a number of teachers and youth workers, the companion question "Is it working?" (Hellison & Templin, 1991) has raised some concerns in the academic community. For example, scholars such as Shields and Bredemeier (1995) and Newton and her associates (Newton, Sandberg, & Watson, 2001) have lamented the lack of evidence to support RM's claims. Practitioners, on the other hand, appear to be more concerned with whether the model makes more sense than what they are currently doing and whether they can implement it (e.g., Zavacky, 1997).

To better address the "Is it working" question, this essay draws on 26 empirical studies of the impact of RM on underserved and so-called at-risk youth that have been conducted since its inception. The chronology of these studies matters, because RM as well as the research questions, methodologies, and designs for this work have gradually evolved. However, chronology was a relatively minor influence, since the majority of studies were published fairly recently as shown below.

- 1970s: 1 study.
- 1980s: 2 studies.
- 1990-1995: 6 studies.
- 1996-1999: 11 studies.
- 2000-2001: 6 studies.

Slavin (1987) questioned the wisdom of including all studies as we have done here when he utilized a best-evidence approach in conducting meta-analyses by screening for appropriateness of research designs and methods. To at least minimally acknowledge the best-evidence issue, the 26 studies were categorized by review process and publication status.

- 6 published articles in theory or research-based peer reviewed journals (Cutforth, 1997; Cutforth & Puckett, 1999; DeBusk & Hellison, 1989; Kahne et al., 2001; Martinek et al., 2001; Schilling, 2001)
- 7 published reports in practice-based peer reviewed journals (Compagnone, 1995; Cutforth, 2000; Georgiadis, 1990; Hellison & Georgiadis, 1992; Lifka, 1990; Martinek, McLaughlin, & Schilling, 1999; Williamson & Georgiadis, 1992)
- 3 published peer-reviewed books or chapters in books, not primarily textbooks (Hellison, 1978; Kallusky, 2000; Wright, 2001)

- 8 unpublished theses (Cummings, 1995; Puckett, 2000)
- 1 unpublished peer-reviewed article (Hellison & Parker, 1997)
- 1 unpublished manuscript

Only the 6 studies that met the "gold standard" for reviewed professional decisions). In fact, ten of merit has not been included for a variety of reasons that may in fact restrict implementation. For example, Cummings' (1995) study of school grade point average program participants at a Robert Donmoyer (1999) expressed his frustration about the role of the gatekeeper in what research is and what is not.

Schon (1995) put it on the high ground [i.e., unimportant problems in a swamp of important problems] in any way he can describe the data sources (including each study's credibility).

R

The literature review research, but its utility as an alternative, meta-analytic studies amenable to statistical analysis (1987).

Some literature review is particularly important story. One of the most important search methods but show that RM, has been criticized.

In the spirit of the essay is to evaluate the effectiveness of categorizing an

### Program Evaluation

The purpose of all programs were conducted

- 8 unpublished theses and dissertations reviewed by faculty committees (Cummings, 1998; Eddy, 1998; Galvan, 2000; Kallusky, 1991; Mulaudzi, 1995; Puckett, 2000; Walsh, 1999; Wright, 1998)
- 1 unpublished peer-reviewed paper presented at a national convention (Herbel & Parker, 1997)
- 1 unpublished manuscript (Cummings, 2000)

Only the 6 studies in theory or research-based peer reviewed journals meet the "gold standard" for rigor (although the 3 books/chapters and 6 articles in peer-reviewed professional journals are sometimes counted in university promotion decisions). In fact, ten are unpublished, suggesting that an even less rigorous test of merit has not been met. These so-called less rigorous studies have been included for a variety of reasons, the most important being that what passes for rigor may in fact restrict important evidence and alternative research designs. For example, Cummings' (1998) longitudinal quasi-experimental comparison of high school grade point average, absenteeism, and dropout rate between former RM program participants and non-RM classmate cohorts would have been excluded. Robert Donmoyer (1996), former editor of the *Educational Researcher*, expressed his frustration about the review process: "[The problem is] to figure out how to play the gatekeeper role at a time when there is little consensus in the field about what research is and what scholarly discourse should look like" (p. 19).

Schon (1995) put the dilemma more bluntly: "Shall [the researcher] remain on the high ground [i.e., the research university] where he can solve relatively unimportant problems according to his standard of rigor, or shall he descend to the swamp of important problems [e.g., youth programs] where he cannot be rigorous in any way he can describe" (p. 28). Information on research design and strength of data sources (included below) permits the reader to make further judgments of each study's credibility.

### Revisiting the Literature Review

The literature review is a staple of graduate student projects and published research, but its utility depends on painstaking analysis and synthesis. A popular alternative, meta-analysis (Cooper & Hedges, 1984; Glass, 1976), is limited to studies amenable to statistical manipulation and is not without controversy (Slavin, 1987).

Some literature reviews, because of the way they are constructed, tell an important story. One example is Bryra's (2000) review of research investigating Mosston's spectrum of teaching styles, which not only critiqued the studies' research methods but showed chronological development over a 30-year period. This review is particularly relevant to our work here, because Mosston's model, like RM, has been criticized for lack of evidence (e.g., Metzler, 1983).

In the spirit of telling what we hope is an important story, the purpose of this essay is to evaluate the impact of RM on underserved and at-risk youth by systematically categorizing and synthesizing all available studies.

### Program Evaluation Focus

The purpose of all 26 studies was program evaluation. Twenty of these programs were conducted before or after school, at lunchtime, or in summer, while 6

an exemplary curriculum model  
 anistic and social development  
 way to use sport and exercise to  
 alternative approach for special  
 ; Siedentop, Mand, & Taggart,  
 s in public school physical edu-  
 p. 4) observed that "[RM offers]  
 in physical education classes in  
 nditions that places some young

k?

appeared to answer the question  
 and youth workers, the compan-  
 , 1991) has raised some concerns  
 s such as Shields and Bredemeier  
 Sandberg, & Watson, 2001) have  
 claims. Practitioners, on the other  
 the model makes more sense than  
 can implement it (e.g., Zavacky,

tion, this essay draws on 26 em-  
 ed and so-called at-risk youth that  
 onology of these studies matters,  
 ethodologies, and designs for this  
 ogy was a relatively minor influ-  
 ed fairly recently as shown below.

including all studies as we have  
 oach in conducting meta-analyses  
 igns and methods. To at least mini-  
 e 26 studies were categorized by

ch-based peer reviewed journals  
 ; DeBusk & Hellison, 1989; Kahne  
 ing, 2001)

r reviewed journals (Compagnone,  
 Hellison & Georgiadis, 1992; Lifka,  
 g, 1999; Williamson & Georgiadis,

pters in books, not primarily text-  
 Wright, 2001)

took place during in-school PE. Since the 26 studies include social program evaluation, case studies, and comparative analyses, a brief description of these approaches and related issues follow.

Greene (2000) divided social program evaluation research into four contemporary approaches: postpositivist, pragmatic, interpretist, and critical. Although policy and decision makers and funders generally favor the postpositivist approach with its experimental or quasi-experimental designs and quantitative outcomes (e.g., data on teen pregnancy reduction), most of the studies under review fall into either the pragmatic or interpretist categories. This conflict is common, because, for many program evaluators, the postpositivist approach "simply does not transfer well to real world contexts" (Greene, 2000, p. 983; see also McLaughlin, 2000; Schon, 1995). Indeed, these alternative approaches arose in response to "the failure of classic experimental science to provide trustworthy, timely, and useful information for program decision-making . . . [for example] the Head Start evaluation" (Greene, 2000, p. 983).

. . . [I]n the field, evaluators rarely practice a "pure" form of their craft, either philosophically or methodologically. The complex, pluralistic demands of evaluation field contexts evoke instead multiple, diverse frames for guiding practice and invite dialog among them. (Greene, 2000, p. 988)

### Case Studies

Of the 26 program evaluations, 21 were case studies. The case study has "proven particularly useful for . . . evaluating programs" (Merriam, 1998, p. 41). According to Yin (1984), "A case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used" (p. 23).

Nine of these case studies were entirely based on qualitative data sources—for example, field notes, interviews, journals, focus groups—but ten included one or more quantitative data sources such as attendance and behavior tallies. Two case studies utilized about the same number of qualitative and quantitative data sources. This mix is not unusual. As both Yin (1984) and Greene (2000) pointed out, a case study is not defined by the nature of the data source.

The case study approach and the use of multiple data sources are particularly useful in research on RM-based youth programs, for the following reasons:

- Experimental designs pose a number of sometimes insurmountable problems in real life settings, especially when the impact of the program is difficult to separate from contextual influences (Kahne & McLaughlin, 1998), whereas case studies, according to Collins and Noblit, "reveal not statistic attributes but understanding of humans as they engage in action and interaction within the contexts of situations and settings. Thus . . . one can better understand how an intervention may affect behavior in a situation" (as cited in Merriam, 1998, pp. 41-42).

- Case studies explore processes as well as outcomes; in fact, as McLaughlin (2000, p. 24) stated, "process is product" in a quality youth organization." Such questions as how much of the program model has been implemented and what

processes influence (Merriam, 1998; Yin, 1984).

- Each program has varying characteristics. Length of program, number of sites, can be identified for

- Multiple RM goals such as reducing violence, can be identified for the program.

- Unintended

Use of the case study approach in thematic research. Case studies or a more locally focused range of population

However, Merriam's analysis can also be conducted to whole population beyond one case.

### Noncase Studies

This review of nonrandomized controlled studies that performed their own controls in their own settings (Merriam, 1998). These studies include primary referrals, Li

### Research Questions

Four program models were identified. The number of studies included (several studies included)

- What was the program RM?
- What was the impact on participants' lives?
- What was the role of teachers?
- What processes were identified?

### Findings

Data-based findings. These four categories were identified from multiple data sources.

processes influenced the results can also be answered by a case study (Merriam, 1998; Yin, 1984).

- Each program under investigation is to some extent a unique case with varying characteristics—e.g., program leader, participants' age and background, length of program, context. By treating them separately at first, these differences can be identified followed by relevant comparisons across cases.

- Multiple data sources are particularly useful for difficult-to-measure RM goals such as respect for the rights and feelings of others and transfer outside the program.

- Unintended outcomes need to be evaluated (Scriven, 1973).

Use of the case study raises the issue of generalizability, a hallmark of systematic research. Case studies can provide support for a previously developed theory or a more locally focused theory-in-action but are not generalizable to a wider range of populations and settings (Patton, 1990; Yin, 1984; see also Greene, 2000).

However, Merriam (1998) and Firestone (1993) argued that cross-case analysis can also be conducted by identifying themes across cases. Although not applicable to whole populations, cross-case analysis provides expanded evidence beyond one case.

### *Noncase Studies*

This review also included one quasi-experimental study characterized by nonrandomized control and experimental groups (Cummings, 1998) and four studies that performed comparative analyses within one group, including subjects as their own controls in three of them (Martinek et al., 1999, 2001; Walsh, 1999; Wright, 1998). These studies relied more on quantitative data sources—for example, disciplinary referrals, Likert-type scale questionnaires, and school transcript records.

## **Reviewing the RM Literature**

### *Research Questions*

Four program evaluation research questions were investigated in the 26 studies. The number of studies addressing each question are shown in parentheses (several studies included more than one research question).

- What was the impact of RM on program participants' improvement in in-program RM goals? (19)
- What was the impact of RM on transferring RM goals to program participants' lives outside the program? (11)
- What was the impact of RM-based cross-age teaching programs on cross-age teachers? (2)
- What processes were experienced by program participants? (12)

### *Findings*

Data-based findings were grouped according to the four research questions. These four categories were further divided into two parts: (a) strong evidence from multiple data sources or, in two studies, strong evidence from one data source

