QUEST, 2002, 54, 292-307

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 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.

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 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

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 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

(Cummings, 1998; Kahne et al., 2001) and (b) weaker evidence from one or more data sources. To acknowledge the "gold standard" of rigor, studies published in theory-research journals are identified by an asterisk (*).

Selected excerpts from the studies are also included to convey a sense of the authors' language in describing impact or process.

What was the impact of RM on program participants' improvement in in-program RM goals?

1. Strong evidence supported the following improvements:
   - Self-control improvement: 9 studies (DeBusk & Hellison, 1989*; Galvan, 2000; Georgiadis, 1990; Hellison & Georgiadis, 1992; Herbel & Parker, 1997; Kahne et al., 2001*; Kallusky, 2000; Lifka, 1990; Williamson & Georgiadis, 1992).

     During the first week of the program, it was common to witness students kicking balls away from each other . . . in a complete turnaround, by the fourth week of the program, students were actually handing balls that were kicked in their direction back to one another. (Galvan, 2000, p. 53)

   - Effort improvement: 6 studies (Compagnone, 1995; Hellison, 1978; Galvan, 2000; Herbel & Parker, 1997; Lifka, 1990; Wright, 2001).

     Before the . . . program was initiated, three . . . [of the four at-risk students in class] were off task for more than a third of the time, and one boy was off task for more than half of the recorded period. After the program, all four boys were on task from approximately 75 percent of the time. (Compagnone, 1995)

   - Helping others' improvement: 5 studies (Galvan, 2000; Hellison & Georgiadis, 1992; Walsh, 1999; Williamson & Georgiadis, 1992; Wright, 2001).

     Toward the end of the program, students called their own time outs and the caring aspect was shown when they gave each other ideas for different plays and positive reinforcement. (Williamson & Georgiadis, 1992)

   - Learned the principles of RM: 4 studies (Cutforth, 1997*; DeBusk & Hellison, 1989*; Hellison, 1978; Lifka, 1990)
   - Self-worth improvement: 4 studies (Hellison, 1978; Hellison & Georgiadis, 1992; Lifka, 1990; Puckett, 2000)
   - Self-direction improvement (e.g., working independently, setting and working on goals): 3 studies (Lifka, 1990; Williamson & Georgiadis, 1992; Wright, 2001)
   - Physical skill or fitness development: 3 studies (Galvan, 2000; Hellison & Georgiadis, 1992; Herbel & Parker, 1997)
   - Teamwork/cooperation improvement: 4 studies (Hellison & Georgiadis, 1992; Herbel & Parker, 1997; Puckett, 2000; Wright, 2001)
   - Communication skills improvement: 3 studies (Cutforth, 1997*; Kallusky, 2000; Lifka, 1990)

- Interpersonal relations improvement: 3 studies (Cutforth, 1997; Williamson & Georgiadis, 1992; Wright, 2001)
- Sense of responsibility improvement: 2 studies (Compagnone, 1995; Kallusky, 2000)
- Learned physical activity concepts: 1 study (Hellison, 1978)
- Sportsmanship improvement: 1 study (Herbel & Parker, 1997)

2. The following studies reported improvements that were supported by weaker evidence:
   - Self-control improvement: 4 studies (Cummings, 2000; Mulaudzi, 1995; Schilling, 2001*)
   - Self-direction improvement: 6 studies (Cummings, 2000; Galvan, 2000; Georgiadis, 1990; Herbel & Parker, 1997; Mulaudzi, 1995)
   - Helping others improvement: 2 studies (Mulaudzi, 1995; Schilling, 2001*)
   - Teamwork/cooperation improvement: 2 studies (Georgiadis, 1990; Mulaudzi, 1995)
   - Effort improvement: 2 studies (Mulaudzi, 1995; Schilling, 2001*)
   - Sense of responsibility improvement: 1 study (Mulaudzi, 1995)

What was the impact of RM on transferring RM goals to program participants' lives outside the program?

1. Strong evidence supported the following outcomes:
   - Self-control improvement in the classroom: 6 studies (Cutforth, 1997*; DeBusk & Hellison, 1989*; Galvan, 2000; Kallusky, 1991; Martinek et al., 1999; Mulaudzi, 1995)

     > Both teachers noted that the boys had a better understanding of their feelings and problems and were more willing to talk about them, had less of a chip on their shoulder. . . and had more trust in authority figures. (DeBusk & Hellison, 1989)

   - Effort improvement in the classroom: 2 studies (Martinek et al., 1999; Martinek et al., 2001*)
   - Self-esteem improvement in the classroom: 2 studies (Kallusky, 1991; Martinek et al., 1999)
   - Potential transfer of the value of violence prevention: 2 studies (Eddy, 1998; Wright, 1998)
   - Reduced dropout rate in high school: 1 study (Cummings, 1998)

     > When you compare a [high school] zero percent dropout rate [in former RM participants] to a 34.1 percent dropout rate [in the control group], it appears that there is a strong link between participation in [the RM program] and staying in school. (Cummings, 1998, p. 30)

   - Learning to be a positive force in the community: 2 studies (Kahne et al., 2001*; Walsh, 1999)
   - Reduction of classroom teacher reprimands and disciplinary office referrals: 1 study (Martinek et al., 1999)
   - Making more reflective and better choices in the classroom: 1 study (Martinek et al., 1999)
   - Maturity development in classroom: 1 study (Kallusky, 1991)

- Learned to solve problems outside the program: 1 study (Walsh, 1999)

2. Weaker evidence reflected some impact on transfer in these areas:
   - Increase in grade point average: 1 study (Martinek et al., 1999)
   - Less stubborn at home, more willing to share and help out: 1 study (Mulaudzi, 1995)
   - Less abusive language, willing to share more in classroom: 1 study (Lifka, 1990)
3. However, weaker evidence on transfer outside the program showed that some programs did not promote transfer:
   - At home or school (Schilling, 2001*)
   - In setting goals in the classroom (Martinek et al., 2001*)
   - On high school absenteeism or grades (Cummings, 1998)

What was the impact of RM-based cross-age teaching programs on cross-age teachers?

RM-based cross-age teaching programs provide advanced responsibility opportunities for those who have participated successfully in an RM youth program. Two studies (Cutforth, 2000; Cutforth & Puckett, 1999*) found the following kinds of impact on the cross-age teachers:

- Problem-solving (Cutforth, 2000; Cutforth & Puckett, 1999*).

  > Thus, in addition to providing a service for the children in the sports camp, the apprentice teachers were developing and learning analytical skills, moral acuity, and social sensitivity through their critical assessment and collective response to authentic problems. (Cutforth & Puckett, 1999*, p. 166)

- Helping, concern for others (Cutforth, 2000; Cutforth & Puckett, 1999*)
- Self-confidence development, interpersonal skill development, and enhanced enthusiasm for learning (Cutforth & Puckett, 1999*)
- Autonomy development, self-reflection development, and ability to accept constructive criticism (Cutforth, 2000)

What processes were experienced by program participants?

Although processes were included in program descriptions in many of the studies—for example, group meetings and keeping journals—only the following studies provided evidence to support processes experienced by program participants:

1. Strong evidence was provided for
   - Having fun/enjoyment: 5 studies (Kahne et al., 2001*; Schilling, 2001*; Puckett, 2000; Walsh, 1999; Williamson & Georgiadis, 1992)

     > . . . over half the participants reported that having fun led to their program commitment. (Schilling, 2001*, p. 362)

   - Interaction with a caring adult: 5 studies (Cutforth & Puckett, 1999*; Kahne et al., 2001*; Kallusky, 2000; Lifka, 1990; Walsh, 1999)
   - Sense of belonging: 4 studies (Galvan, 2000; Kahne et al., 2001*; Schilling, 2001*; Walsh, 1999)

Every student said they wanted to spend more time in the extended day program. (Walsh, 1999)

- Felt safe: 1 study (Wright, 2001).

2. Weaker evidence was also found for processes in these studies:
   - Interaction with a caring adult: 1 study (Cummings, 2000)
   - Sense of belonging: 1 study (Cummings, 2000)
   - Positive climate: 1 study (Walsh, 1999)

Discussion

The question "Is it working?" was raised earlier in reference to RM programs for underserved youth and seems relevant as the focus for a closing discussion. Yet perhaps a different question is more appropriate: Is this review of literature worth doing? We attempt to explore both questions here.

A systematic research design to guide all RM studies would have reduced the patchwork of methodologies and tightened the research focus while at the same time restricting the exploration of new designs and methodologies. Instead, RM's thirty year chronology has spun off a number of ideas and approaches. Program evaluation was an ongoing process, especially through the latter part of this period, and was influenced by research designs and methods available at the time of the evaluation and by independent decisions made by the researchers. The recent emergence of many new research designs and methodologies in the fields of education and youth development, accompanied by controversy and in some cases acrimony among scholars, has widened the playing field but elevated the ambiguity in the selection and application research designs and methodologies.

The preponderance of investigations in our analysis were case studies utilizing quantitative and qualitative data sources. The few noncase studies included comparisons, but only one had a separate control group. Case studies encourage the "triangulation" of data sources and an investigation of processes as well as outcomes. In the reviewed studies, utilizing mixed data sources permitted drawing inferences from, for example, field notes, interviews, and journal entries as well as number of disciplinary referrals, voluntary attendance, and pre-post psychometric questionnaires. Triangulation of these data enabled researchers to identify changes in participant attitudes, feelings, intentions, and behaviors. Despite these advantages, the interest in post-positivist program evaluations among many funders and policy makers suggests that future RM studies should include a more equitable balance of research designs.

The Process Issue

One advantage of qualitative program evaluation research, such as the case study, is the potential for linking processes to outcomes, yet none of the studies directly addressed this question. Our analysis attempted to separate process and impact, but, as we pointed out, context is difficult to separate from impact in case studies. McLaughlin's (2000) argument that process is often product in good youth organizations emphasizes the difficulty of making this distinction. Processes are ongoing experiences that program participants engage in as integral parts of a program. These experiences—e.g., a positive climate, a sense of belonging and safety,

a caring adult—are processes, but they may also contain what program leaders want kids to take away with them. Hal Adams (personal communication, 1999), a veteran of inner city work and university-community collaboration, insisted that these process-oriented experiences are the very heart of RM youth programs; they are the point! Even seemingly concrete outcomes such as self-control improvement and learning to help others are as much process as product, because qualitative evaluation of impact included field notes and journal records of the ups and downs of these experiences in a variety of situations over time, thereby blurring the distinction.

Despite the progress of program theory evaluation (Rogers, Hasci, Petrosino, & Huebner, 2000), measuring process as a causal factor in outcomes in fraught with problems. For example, Schilling (2001) reported her frustration in attempting to measure the linkage between program leader qualities and participants' commitment to the program.

Support for RM theory

As with other program models, RM requires faithful implementation by the program leader. Otherwise, program evaluations may reflect poor implementation rather than a faulty program model. Only one study (Puckett, 2000) reported serious difficulties in implementation (and her data did contain some evidence of positive impact despite implementation problems).

None of the studies contained sufficient controls to permit generalizations to populations. The data in these studies can, however, contribute to program model theory (Hellison, 1995; Hellison et al., 2000), something Yin (1984) calls analytical generalization.

RM theory describes RM goals as "a loose teaching-learning progression" (Hellison, 1995, p. 12). This loose progression is reflected in the findings, with beginning responsibilities of self-control and effort—including similar concepts—receiving more support than advanced responsibilities of self-direction and helping others. However, RM theory also posits that its impact on participants depends in part on program longevity. Surprisingly, comparisons of studies according to program longevity did not reveal many differences in participants' in-program differences. However, longer programs did report substantially more transfer outside the program.

Transfer outside the gym was not formally added to RM until the early 1990s (Hellison, 1993), although it was hinted at in earlier RM literature and particularly in a brief chapter titled "going beyond the levels" (Hellison, 1985). Despite being a late addition and its position as the last responsibility in the loose progression, transfer received somewhat more support across all studies than its recent emergence and position warrants, although three of the studies that focused on transfer reported that none had taken place. These findings are confounded by the work of Martinek and his associates (1999, 2001), who added a mentoring component to RM specifically to address transfer.

RM theory states that the number of participants in the program (i.e., class size) influences the potential for full implementation of the model. A few programs exceeded the theoretically ideal class size (15), yet greater in-program improvements were to some extent overrepresented in studies of the larger groups.

However, none of these larger group studies reported much evidence of transfer, a finding supported by Mrugala's (2002) survey of in-school PE teachers (with larger classes) who use RM but tend to exclude transfer in their implementation.

According to RM theory, small class size, when accompanied by program longevity, leads to "a sense of belonging." Three of the longer studies did include evidence suggesting that participants gained "a sense of belonging," thereby providing a bit of support for this relationship.

Cross-Case Analysis

Although not applicable to whole populations, cross-case analysis provides expanded evidence beyond one case. The mixed data source evidence across cases can be quantified according to the terminology used in each study to describe its findings. As shown below, a cursory quantitative analysis revealed a range of program participant experiences and improvements sprinkled across the 26 program evaluations, rather than a tight group of substantial findings:

- 14 supported self-control improvement.
- 12 supported effort improvement.
- 8 supported self-direction improvement.
- 7 supported helping others improvement.
- 6 supported transfer to more self-control in the classroom.
- 6 supported experiencing a caring adult.
- Although only two studies focused on the impact of cross-age teaching, both reported very positive results.
- 27 other findings associated with personal and social development were scattered across the 26 studies.
- Personal development included sense of responsibility, self-worth, self-esteem, self-confidence, choice-making, autonomy, self-reflection, maturity, problem solving, and less stubbornness.
- Social development included sense of responsibility, interpersonal relations, communication skills, teamwork, cooperation, sportsmanship, and being a positive force in the community.

Policy Implications

What are the policy implications of these findings? Joe Kahne (1996) argued that "in addition to assessing a policy's potential effects on test scores, graduation rates, or some other educational goal, one must assess whether achieving that goal will promote a given conception of the good society" (pp. 2-3). According to Kahne, current policies reflect utilitarian and rights-base theories, even though policymakers may not be aware of these theories. Less evident in policy work are democratic communitarian and humanistic theories, despite their relevance in defining the good life and the good society. It doesn't take a great deal of analysis of the data to notice the emphasis on democratic/communitarian and humanistic values in the RM studies, an emphasis clearly evident in RM theory and even pre-RM literature (Hellison, 1973).

The preference of policy makers for postpositivist research designs fits well with Kahne's analysis. Outcomes such as test scores and graduation rates can easily be reduced to numbers, whereas outcomes representative of the good life and

good society are more amenable to qualitative treatment than to some form of reductionism.

So . . . Is It Working?

The "is it working" question remains a work in progress due to methodological issues and gaps in the evidence, but these 26 studies, however limited, do enhance the theoretical and practical potential of RM as a program model for underserved and at-risk youth. Moreover, future research can focus on the methodological shortcomings and evidence gaps as well as to build on what has already been accomplished. Not surprisingly, it seems to us to be worth doing.

References

Bain, L. (1988). Curriculum for critical reflection in physical education. In R.S. Brandt (Ed.), *Content of the curriculum:1988 ASCD yearbook* (pp.133-147). Washington, DC: Association for Supervision and Curriculum Development.

Byra, M. (2000). A review of spectrum research: The contributions of two eras. *Quest*, **52**, 229-245.

Compagnone, N. (1995). Teaching responsibility to rural elementary youth: Going beyond the at-risk boundaries. *Journal of Physical Education, Recreation, &* Dance, **66**(8), 58-63.

Cooper, H., & Hedges, L.V. (1984). (Eds.). *The handbook of research synthesis.* Chicago: Rand-McNally.

Cummings, T. (1998). *Testing the effectiveness of Hellison's personal and social responsibility model: A dropout, repeated grade, and absentee rate comparison.* Unpublished master's thesis, California State University, Chico.

Cummings, T. (2000). *The influence of an inner city extended day program in dropout prevention.* Unpublished manuscript.

Cutforth, N. (1997). What's worth doing: Reflections on an after-school program in a Denver elementary school. *Quest*, **49**, 130-139.

Cutforth, N. (2000). Connecting school physical education to the community through service-learning. *Journal of Physical Education, Recreation, and Dance,* **71**(2), 39-45.

Cutforth N., & Puckett, K.M. (1999). An investigation into the organization, challenges and impact of an urban apprentice teacher program. *The Urban Review*, **31**(2), 153-172.

DeBusk, M., & Hellison, D. (1989). Implementing a physical education self-responsibility model for delinquncy-prone youth. *Journal of Teaching in Physical Education*, **8**(2), 104-112.

Donmoyer, R. (1996). Educational research in an era of paradigm proliferation: What's a journal editor to do? *Educational Researcher*, **25**(2), 1925.

Eddy, M.H. (1998). *The role of physical activity in educational violence prevention programs.* Doctoral dissertation, UMI dissertation services, Ann Arbor, MI.

Firestone, W. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher*, **22**(4), 16-23.

Galvan, C. (2000). *The impact of the responsibility model on underserved students in physical education classes: A university community collaboration.* Unpublished master's thesis, California State University, Los Angeles.

Georgiadis, N. (1990). Does basketball have to be all W's and L's: An alternative program at a residential boys' home. *Journal of Physical Education, Recreation, and Dance*, **61**(8), 42-43.

Glass, G.V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher*, **5**(1), 3-8.

Greene, J.C. (2000). Understanding social programs through evaluation. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of Qualitative Research* (2nd ed.; pp. 981-999). Thousand Oaks, CA: Sage.

Hellison, D. (1973). *Humanistic physical education.* Englewood Cliffs, NJ: Prentice-Hall.

Hellison, D. (1978). *Beyond balls and bats: Alienated (and other) youth in the gym.* Washington, D.C.: American Association of Health, Physical Education, & Recreation.

Hellison, D. (1985). *Goals and strategies for physical education.* Champaign, IL: Human Kinetics.

Hellison, D. (1993). Humanistic physial education in schools: Past, present, and future. In R. Lidor, D. Ben-Sira, & Z. Artzi (Eds.), *Physical activity in the lifecycle: Proceedings of the 1993 FIEP World Congress* (pp. 23-28). Israel: Wingaate Institute.

Hellison, D. (1995). *Teaching responsibility through physical activity.* Champaign, IL: Human Kinetics.

Hellison, D., Cutforth, N., Kallusky, J., Martinek, T., Parker, M., & Stiehl, J. (2000). *Youth development and physical activity: Linking universities and communities.* Champaign, IL: Human Kinetics.

Hellison, D., & Georgiadis, N. (1992). Basketball as a vehicle for teaching values. *Strategies*, **4**(1), 1-4.

Hellison, D., & Templin, T. (1991). *A reflective approach to teaching physical education.* Champaign, IL: Human Kinetics.

Herbel, K., & Parker, M. (1997). *Youth, basketball, and responsibility: A fairy tale ending?* Presentation at National Convention of the American Alliance of Health, Physical Education, Recreation, & Dance, St. Louis, MO.

Hodge, K., & Danish, S. (1999). Promoting life skills for adolescent males through sport. In A.M. Horne & M.S. Kiselica (Eds.), *Handbook of counseling boys and adolescent males: A practitioner's guide* (pp. 55-71). Thousand Oaks, CA: Sage.

Kahne, J. (1996). *Reframing educational policy: Democracy, community, and the individual.* New York: Teachers College Press.

Kahne, J., & McLaughlin, M. (1998). Framing issues for policy and practice through theory-driven evaluation. *New Designs for Youth Development*, **14**(1), 17-22.

Kahne, J., Nagaoka, J., Brown, A., O'Brien, J., Quinn, T., & Thiede, K. (2001). Assessing after-school programs as contexts for youth development. *Youth and Society,* **32**(4), 421-446.

Kallusky, J. (1991). *A qualitative evaluation of a physical education mentoring program for at-risk children.* Unpublished master's thesis, California State University, Chico.

Kallusky, J. (2000). In-school programs. In Hellison, D., Cutforth, N., Kallusky, J., Martinek, T., Parker, M., & Stiehl, J. *Youth development and physical activity: Linking universities and communities* (pp. 87-114). Champaign, IL: Human Kinetics.

Kirk, D. (1992). *Articulations and silences in socially critical research on physical education: Towards a new agenda.* Paper presented at the AARE Annual Conference, Geelong, Australia.

Lavay, B.W., French, R., & Henderson, H.L. (1997). *Positive behavior management strategies for physical educators.* Champaign, IL: Human Kinetics.

Lifka, R. (1990). Hiding beneath the stairwell: A dropout prevention program for Hispanic youth. *Journal of Physical Education, Recreation, and Dance*, **61**(8), 40-41.

Martinek, T., McLaughlin, D., & Schilling, T. (1999). Project Effort: Teaching responsibility in the gym. *Journal of Physical Education, Recreation, and Dance*, **70**(6), 59-65.

Martinek, T., Schilling, T., & Johnson, D. (2001). Transferring personal and social responsibility of underserved youth to the classroom. *The Urban Review*, **33**(1), 29-45.

McLaughlin, M.W. (2000). *Community counts.* Washington, D.C.: Public Education Network.

Merriam, S. (1998). *Qualitative research and case study applications in education.* San Fransisco, CA: Jossey-Bass.

Metzler, M. (1983). On styles. *Quest*, **35**, 145-154.

Mrugala, K. (2002). *An exploratory study of responsibility model practitioners.* Unpublished doctoral dissertation, University of Illinois at Chicago.

Mulaudzi, L. (1995). *A program evaluation of an implementation of a responsibility model for inner-city youth.* Unpublished master's project, University of Illinois at Chicago.

Newton, M., Sandberg, J., & Watson, D.L. (2001). Utilizing adventure education within the model of moral action. *Quest*, **53**, 483-494.

Oslin, J., Collier, C., & Mitchell, S. (2001). Living the currriculum. *Journal of Physical Education, Recreation, and Dance*, **72**(5), 47-51.

Pangrazi, R.P. (2001). *Dynamic physical education for elementary school children* (13th ed.). Needham Heights, MA: Allyn and Bacon.

Patton, M. (1990). *Qualitative evaluation methods* (2nd ed.). Thousand Oaks, CA: Sage.

Puckett, K. (2000). *Combining the responsibility model and adventure education in an inner city physical education program.* Unpublished master's project, University of Illinois at Chicago.

Rink, J. (1993). *Teaching physical education for learning* (2nd ed.). St. Louis, MO: Times Mirror/Mosby.

Rogers, P.J.. Hasci, T.A., Petrosino, A., & Huebner, T.A. (2000). (Eds). *Porgram theory in evaluation: Challenges and Opportunities.* San Fransisco: Jossey-Bass.

Schilling, T.A. (2001). An investigation of commitment among participants in an extended day physical activity program. *Research Quarterly for Exercise and Sport*, **72**(4), 355-365.

Schoenfeld (1999). Looking toward the 21st century: Challenges of educational theory and practice. *Educational Researcher*, **28**(7), 4-14.

Schon, D. (1995). The new scholarship requires a new epistemology. *Change,* **27**(4), 27-34.

Scriven, M. (1973). Goal-free evaluation. In E.R. House (Ed.), *School evaluation: The politics and process* (pp. 219-328). Berkeley, CA: McCutcheon.

Shields, D.L.L., & Bredemeier, B.J.L. (1995). *Character Development and Physical Activity.* Champaign, IL: Human Kinetics.

Siedentop, D. (1990). *Introduction to physical education, fitness, and sport.* Mountain View, CA: Mayfield.

Siedentop, D., Mand, C., & Taggart, A. (1986). *Physical education: Curriculum and instruction strategies for grades 5-12.* Palo Alto, CA: Mayfield.

Slavin, R.E. (1987). Ability grouping: A best-evidence synthesis. *Review of Educational Research*, **57**, 293-336.

Steinhardt, M. (1992). Physical education. In P.W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 961-1001). New York: Macmillan.

Walsh, D. (1999). *A comparative analysis of extended day programs for inner city youth.* Unpublished master's thesis, University of Illinois at Chicago.

Williamson, K.M., & Georgiadis, N. (1992). *Journal of Physical Education, Recreation, & Dance*, **63**(9), 14-18.

Winnick, J. (1990). *Adapted physical education and sport.* Champaign, IL: Human Kinetics.

Wright, P.M. (1998). *The impact of a responsibility-based martial arts program on violence prevention.* Unpublished master's thesis, University of Illinois at Chicago.

Wright, P.M. (2001). Violence prevention: What can coaches and sport educators do? In B.J. Lombardo, T.J. Caravella-Nadeau, H.S. Castagno, & V.H. Mancini (Eds.), *Sport in the twenty-first century: Alternatives for the new millennium* (pp. 189-202). Boston: Pearson.

Yin, R.K. (1984). *Case study research: Design and methods.* Beverly Hills, CA: Sage.

Zavacky, F. (1997). Motivating the l. *Teaching elementary physical education*, **8**, 30-31.