

Challenges in Evaluating Character Education Programs

Leslie Laud

Teachers College, Columbia University

Marvin W. Berkowitz

University of Missouri - St. Louis

Based on information gathered from eight independent character education program evaluations, this article offers suggestions on how to more effectively evaluate the following areas of character education programs: implementation, outcome definition, instrumentation, developmental considerations, and longitudinal concerns. Researchers must meet the challenge of adjusting their methods of evaluating these five areas to meet the unique needs of character education program evaluation if we are to gain better understandings of these programs' performance. This article concludes with an analysis of how one exceptional and exemplary program evaluation successfully met these five challenges.

Interest in character education has steadily increased over the past decade in the school systems, among policy makers and in academia. Considering the unstable history of character education (McClellan, 1999), the recent passage of supportive national legislation raises concerns that character education may become another legislated, yet short-lived, educational policy as in the past. In response, a stronger research base is being proposed as one way to ensure that it secure a more permanent place in our educational process. In a recent synopsis of the history of the movement, Leming (1993), voicing this concern, advised "The current revival of interest in character education, if it is to succeed, will have to successfully address the question of the assessment of program effectiveness" (p. 63).

This article uses Thomas Lickona's (1991) definition of character education as a broad developmental phenomenon designed to help children know the good, desire the good and do the good. Studies examined in this article were program evaluations gathered from organizations that have worked toward this goal and documented their efforts. Both qualitative and quantitative research approaches were used in these studies. This article focuses most heavily on a quantitative perspective, because this is the perspective that has most strongly driven funding and dominated the field. As more qualitative studies emerge and capture equal attention, more research exploring and supporting this perspective should be done. At this time, however, the most convincing and extensive evaluation data are quantitative.

To narrow the vast terrain of programs and innovations that we recognize as potentially developing "character," this article examines only those that focus on the direct development of most major traditional character traits, such as respect, responsibility, kindness, and honesty. These approaches are usually categorized as "character education."

However, the material herein may provide insights that could prove useful to the wider array of programs designed to foster any of the many facets of personality or behavior that constitute character (Berkowitz, 1997).

In order to explore the state of character education program evaluation, fifteen organizations were initially contacted and requested to send evaluation reports. The eight studied in this article were the only ones to respond. Descriptions of these programs and the specific approaches used in their evaluations appear in Table 1. The evaluation reports were analyzed, generating the issues raised in this article.

Before data purported to address program effectiveness can be meaningfully interpreted, the underlying issue of whether the evaluation methods themselves are successfully measuring what they aim to measure requires scrutiny. Although this issue applies when evaluating any field, character education evaluation presents exceptionally difficult and unique challenges to effective evaluation. This article proposes that methods currently used need to be better adapted to the many complexities inherent in assessing character education in the following five areas: implementation, outcome definition, instrumentation, developmental considerations, and longitudinal concerns. The challenges character education poses in each of these areas is explored, using the eight evaluation reports as examples. The Child Development Project evaluation will be addressed predominantly in a separate section at the end of this paper, serving as an example of an effective evaluation approach. Therefore, it will not be cited extensively in the discussion of the five challenges. Whether, or to what degree, efforts to develop character are effective cannot be determined until present evaluation methods are more effectively adjusted in these five areas.

Table 1
Methods Used for Program Evaluation

Program Name	Evaluation Methods							
	<i>S</i>	<i>I</i>	<i>O</i>	<i>DC</i>	<i>IS</i>	<i>AD</i>	<i>C</i>	<i>P</i>
Baltimore Schools ¹	+	+	+					
PREP ²	+	+	+	+				
Megaskills ³		+				+	+	
Jefferson Center ⁴	+	+		+				
Character Education Institute ⁵	+	+		+	+	+	+	+
Quest ⁶	+	+		+	+	+	+	+
Heartwood ⁷	+	+						+
Child Development Project ⁸	+	+	+	+	+	+	+	+

Note. *S* = Survey; *I* = Interview; *O* = Observation; *DC* = Descriptive comparisons of data; *IS* = Inferential statistics; *AD* = Archival Data (Records of discipline, attendance or academics); *C* = Control sites; *P* = Pre/Post tests.

¹BCPS (Baltimore County School District) actively promotes "character education" efforts in all of its schools.

² Personal Responsibility Education Program (PREP) again is a district wide effort to encourage character education efforts in all schools throughout the district. ³ This is an after-school program that aims to develop students' academic skills and character. ⁴ The Jefferson Center has developed an approach to character education used in various schools throughout the country. ⁵ Similar to the Jefferson Center, this center also develops character education materials used in schools throughout the country. ⁶ Again, Quest develops and used character education materials in schools throughout the country. ⁷ Heartwood has created a literature program that develops character traits and is used in schools throughout the country. ⁸The Child Development Project is part of the Developmental Studies Center which has created an educational approach that works to develop character traits in addition to academic capacities in students. This approach is also used throughout the country.

*The Character Education Institute responded with a compilation of evaluations, some of which used inferential statistics.

Program Implementation Issues

One limitation in the evaluation studies examined in this article is the scarce attention given to program implementation evaluation, also often referred to as process evaluation. Although this issue has received tremendous attention in many areas of education since the publication of Gross, Gianquinta and Bernstein (1971), character education evaluators have not given the level of attention these authors might recommend in the studies examined in this article. Before effects can be measured, evidence that the desired change in educational methods has occurred must be gathered. Several studies, such as those of Quest and Megaskills, describe laudable successes. However, an examination of whether implementation was complete and correct is not mentioned in these evaluation reports.

Others who did explore this issue found, at times, disappointing levels of program implementation. For example, those who evaluated the Character Education Institute (CEI), Personal Responsibility Education Process (PREP) and Heartwood found implementation problematic. This may be the result of several factors such as implementation not having been accurately evaluated, or character education programs being inherently problematic to implement properly.

Speaking to this issue, Farrelly (1993) suggests that "exploratory evaluation," evaluation of program feasibility and implementation, should dominate the initial phases of the evaluation. Marantz (1991), in accord with Farrelly's advice, conducted an action research study on character education program implementation. She found that issues such as the pedagogical soundness of the program, school

reform issues and the role of the teacher in the program require consideration and attention to ensure optimal implementation. Although the pedagogical soundness of the program was reviewed in CEI, Baltimore County Public Schools (BCPS) and Heartwood, these evaluations did not address the latter issues. Therefore, implementation and the wider issues surrounding it require greater attention than received in these studies.

One proposed reason why implementation evaluation of character education programs may be so complicated is that often such programs are integrated into many areas of the curriculum. This integration across curricular areas, recommended by Marantz (1991), BCPS, the Jefferson Center, PREP, and CEI, includes the ephemeral hidden curriculum and "school atmosphere" which is even more challenging to evaluate objectively.

Implementation evaluation, therefore, becomes considerably more complicated than evaluation of more structurally simplistic changes that occur at one specific point in the day or in one isolated curricular area. Although both open-ended (Power, Higgins & Kohlberg, 1989) and objective (Battistich & Hom, 1997) measures of moral atmosphere exist, evaluators typically did not use them, nor do they create their own. For example, BCPS evaluators described a positive environment, however, a structured format or checklist was not used to rate it.

A second reason why implementation evaluation is so challenging is that character education programs are often notoriously difficult to fully or correctly implement. They require rigorous attention to subtle implementation details. Those researchers who did directly observe program implementation in schools, such a PREP and BCPS who did not use structured rating forms, mention no more than one observation of each classroom. The exceptions, Marantz (1991), who explored program implementation, and the Child Development Project, used a clear and structured evaluation format on repeated visits together information about exactly how the program was being used. Although researchers who rely on teacher self-reports of implementation, a method used by Heartwood, or administrator reports of implementation, used by the Jefferson Center evaluators, gain valuable information, they risk not knowing how well or completely the program was actually implemented because sites were not directly observed.

Implementation evaluation serves a number of important functions. First, as programs are beginning there may be external pressures for evaluation reports. Implementation evaluation may satisfy those needs while the program is given time to take hold and produce outcomes to be evaluated later. Second, and more importantly, implementation evaluation can provide data for program improvement by offering feedback on what is and is not working. Third, implementation evaluation adds power to outcome evaluation. When low, moderate, and high implementing schools or classrooms are analyzed together, the effectiveness of a program may be masked. Implementation evaluation allows a researcher or implementer to examine the differential ef-

fectiveness of various levels of implementation, so that schools or classrooms that do a poor job of implementing can be compared to schools or classrooms that do a complete job of implementing. The Child Development Project has found this to be a very useful strategy in studying the effectiveness of their program.

Defining Variables

Another equally demanding challenge is the process of defining desired outcomes (Berkowitz, 1998). Researchers must have a clear notion of the outcomes they expect, while also not allowing their preconceived expectations to occlude their ability to find unexpected outcomes. PREP evaluators astutely recommend that districts spend time defining student outcomes and determining how they will measure program success. Character education proponents often use alarming youth trends to justify implementation (Rest, 1994). However, if reversing these trends becomes the "desired outcomes," then, if they are not reversed, the public may withdraw support for character education. Therefore, outcomes must be clear, appropriate and, reasonably attainable. Lickona (1994) broadens the realm of outcomes when he suggests that not only improvements in school should be defined, but also out of school and post-graduation outcomes should be targeted. Too narrow of a definition puts "all one's eggs in one basket" and may lead to the rejection of a program because one or two outcomes have not been met. Too broad of a definition requires a burdensomely extensive evaluation effort. Yet, defining these outcomes is confounded by several issues.

First, defining student outcomes can prove exceedingly difficult for some aspects of character education. BCPS and PREP evaluators mention how challenging it is to describe such changes scientifically; i.e., to define, understand, and measure human behavior in this area. Coles (1985/86), in searching for a definition of character, does not discover rigid traits but rather a process, "a quality of mind and heart one struggles for, at times with a bit more success than others" (p. 21). Defining a desired process far exceeds the demands of defining simple product changes such as test score increases or behavioral improvements. Although such changes also may occur with character education, they constitute only one facet of the many and broad desired outcomes. Berkowitz (1997) has described a "moral anatomy" that analyzes the psychological components of moral functioning. This taxonomy defines what is traditionally considered character as moral personality, an enduring tendency to act in certain ways. Enduring tendencies are clearly more challenging to assess than are simple behaviors in limited situations.

A second difficulty is that the intended outcomes will vary dramatically across sites according to the goals of specific schools, characteristics of student populations, the unique nature of the intervention and the individual context or school culture in which it is implemented. PREP, recognizing this, allows each school system individually to de-

fine its own goals concerning which specific values they intend to foster in students. Additionally, once the desired outcomes were defined, CDP reported, they needed to remain open to revision. This makes it difficult to centralize evaluation and produces great inefficiency. For example, an overall evaluation of PREP (now called Character Plus) is complicated by the strategy of allowing each school to design its own character education agenda and program.

Finally, the tendency to over-simplify the nature and definitions of the desired outcomes to make them more easily assessable, should be avoided (Straughan, 1983). Essentially, although he encourages working towards this goal, Thomas (1991) asserts that the nature of character education precludes the possibility of quantifying all the expected outcomes. Otherwise, the true richness of character education effects may be lost in the search for simple conceptions of outcome variables.

Instrumentation Issues

Outcome definition facilitates instrument choice or construction. However, instrument selection or creation poses major challenges, regardless of how clearly defined the outcomes may be. The scarcity of established instruments specific to traditional character education program measurement is probably one reason why many evaluators, such as CDP, PREP, Heartwood and CEI, developed their own, a complex and demanding task.

Most pre-constructed instruments measure only limited aspects of character such as ego-development or self-esteem, while each character education program requires instruments sensitive to its specific individual components and goals. Therefore, it is often necessary for evaluators to construct their own instruments. Detailed suggestions for this process are given in Berkowitz (1998) and Hendrix, Berkowitz, Stone & Asiu (1999). Yet, this could prove a valuable area for future work and research.

The complex nature of the changes sought by character education programs as well as the wide realm of desired outcomes requires a variety of methods and instruments to ensure a comprehensive evaluation. A combination of both quantitative and qualitative research methods are recommended to ensure this (Farrelly, 1993; Thomas, 1991).

Evaluations that rely on only one qualitative method of data collection, such as telephone interviews of administrators – a method used in evaluation of the Jefferson Center, risk gathering incomplete or overly subjective data. While such efforts are valuable, quantitative data, such as direct behavioral observations made by objective researchers can compliment and substantiate such impressions.

The majority of instruments used in the studies examined measure primarily attitudinal variables while several also address cognitive variables. The few evaluators to measure behavioral variables did so through an analysis of archival discipline statistics and potentially subjective teacher or administrator perceptions. None of the researchers, except the CDP, used instruments to directly observe and

record student behavior variables, except the CDP.

One additional instrumentation option worthy of further exploration is "alternative forms of assessment," also often referred to as "authentic assessment," or "portfolio assessments." Such forms of assessment have been widely discussed in the educational research literature (Taylor, 1994) and used in many school systems. This refers specifically to non-traditional assessment methods such as portfolios, demonstrations, culminating projects, exhibitions, or oral examinations. Essentially, students are required under this system to provide evidence of what they have learned. Teachers serve as facilitators and give feedback as students create meaningful representations of their work.

A primary benefit of this option is that it can require less additional funding, or better use of funding, though it does place an additional burden on teachers. Teachers, rather than outside researchers could be paid for the additional time needed by this approach. If teachers do not wish to commit the additional time, perhaps time can be reallocated in ways proposed by Darling-Hammond (1997) that would allow teachers greater opportunity to engage in this option.

Potential benefits of this option abound. Teachers, based on their extended observations of students, can provide additional valuable insights into their students' character growth that external observers might miss. The process nature of character education, so difficult to measure quantitatively, may be better measured with such assessment forms that focus primarily on the process nature of education. Additionally, multi-media forms of assessment, such as videotapes or project portfolios, used in alternative forms of assessment, may also better capture the essence of character growth. Finally, evaluation could continue after the formal program evaluators leave, thus ensuring continuity and long-term observation potential. This is merely one alternative option. Future research could work to generate others.

Developmental Considerations

According to Piaget (1965), Kohlberg (1980) and Selman (1980), character education evaluation methods should also consider developmental differences in children. The moral reasoning of children, they posit, evolves in predictable sequences which raises a number of developmental issues for evaluation. Assessment instruments and techniques could measure such developmental differences. A variety of types of assessment have been developed for this purpose; e.g., Colby and Kohlberg (1987), Gibbs and Widaman (1982), and Rest (1986). In addition, different methods and interventions may be appropriate for students at various levels of development. Verbally complex instruments (e.g., Rest, 1986) and unstructured group tasks may not be appropriate for less mature students. It is also important to consider what behaviors are developmentally appropriate at what ages. Lying, for instance, is much more common in early childhood than later. Finally, some outcome and process measures may be inappropriate for children at some developmental levels. For example, a sense of com-

munity requires a mature ability to conceptualize groups. A moral identity requires self-reflection.

Longitudinal Concerns

Development is typically a slow and modest process (Berkowitz & Keller, 1994). To truly effect developmental change, one must intervene and assess over an appropriate time-span. As Thomas (1991) notes, "...immediate effects may be merely temporary responses to external factors and not necessarily indications that values have become ingrained in students' characters" (p. 5). Therefore, the results of programs evaluated after a period of only several months such as Heartwood or CEI, may be inconclusive if true developmental progressions are sought. Immediate post-testing may only assess ephemeral intervention effects, rather than true developmental gains.

Lickona (1991) mentions the discovered "sleepers effect," the fact that effects of a character education program may at times not appear until years after the program has been implemented, further underscoring the need for lengthy interventions and delayed post-testing. In fact, if most character education programs considered this issue fully, it would be recognized that the true outcome goals of character education are years down the road. In other words, character education is supposed to prepare the individual to live a good life. For instance, the character development program at the United States Air Force Academy does not define its goals by *cadet* (student) behavior, but rather by *officer* (alumnus) behavior (Center for Character Development, 1994). In summary, programs must not be labeled ineffective when they may merely require more time for results to appear. Although PREP, CEI, and Quest mention using post-tests, they were not delayed. CDP and Weber have begun to find some encouraging delayed effects of character education efforts with long term delayed post-tests.

An Exemplary Evaluation Model

Among the evaluation studies sampled for this article, one stands out as exemplary and not subject to many of the limitations just described – the Child Development Project (CDP). This project has become one of the best known and most respected in the country for a wide variety of reasons. Yet, it can be speculated that one factor in its success may be its exemplary approach to evaluation.

Although evaluators would ideally assess implementation before program effects, researchers carefully assessed program implementation while assessing program effects. "Blind" to program status of schools, trained observers made eight unannounced visits to experimental and control sites annually to observe classroom and school wide program implementation, thus avoiding potential observer bias. The observers recorded specific teacher and student behaviors in structured checklist format every few minutes during the observation. Global observations and impressions were subsequently recorded in a more open-ended format to gather

both quantitative and qualitative data, ensuring comprehensive and objective research. Principals and parents were interviewed and given questionnaire to further examine implementation of the school-wide and related family activities. Therefore, a complete picture of program implementation was gathered.

Core outcome variables were carefully defined so that the researchers would have an exact way of knowing whether the program was effective. All assessment instruments and procedures were then developed by the researchers who sought to evaluate changes in all of the following areas; attitudes, values, knowledge, and behavior. Assessment techniques and scoring systems, developed by the researchers, were pilot tested and revised several times to increase their sensitivity and best adapt them to the individual settings. Developmental differences were addressed in the differing interventions given to various age groups. The program was carried out over a period of several years, with much subsequent delayed post-testing, so that effects over long periods of time could also be examined. Recent reports, for example, demonstrate that immediate elementary gains in academic motivation become delayed middle school gains in academic achievement (Schaps, 1999).

Additionally, objectivity was ensured through the careful process of school selection and control groups. Six similar schools were initially chosen to participate in this study, with three then randomly selected as experimental schools and three as control sites. Before the inception of the program, a student cross-section was observed in the experimental schools to assess pre-existing similarities and differences and to serve as a comparison group within the school. For a detailed description of the evaluation methodology see Soloman et al. (1988) or Battistich (1988).

Efforts to be as thorough as the CDP evaluation often depend upon adequate funding. Funding limitations can reduce the number of researchers, shorten the time available for careful instrument development, limit the size of the sample studied and decrease the time length of the study. Therefore, even when researchers are made aware of the limitations mentioned in this article, other restrictions may impede their efforts. Fortunately, creative solutions to compensate for some of these restrictions are available. Programs that do not have adequate funding can use alternative forms of assessment or, perhaps show dramatic increases in specific variables (e.g., test scores) that are widely accepted as indicators of school reform success.

Conclusion

The goal of the current character education movement is not program implementation or evaluation. Ultimately, it is student development. To assist in this endeavor, researchers can assist by exploring which programs or approaches are effective and which are appropriate in which contexts. If evaluative research is to assist in accomplishing this, its methods must be scrutinized, improved and better adapted to the unique and complex needs of character education

evaluation.

The issues raised and explored in this article are only a beginning. Before evaluation of student development is possible, the challenges of current evaluation methods must be more fully explored and attended to by program evaluators. Only then can the movement hope to build a stronger research base, and, hence, secure a lasting, prominent and defensible place in the educational process. Further research could include an exploration of other viable alternatives to overcome the obstacles, particularly those resulting from funding limitations, to optimal evaluation procedures in this area.

References

- Battistich, V., & Hom, A. (1997). The relationship between students' sense of their school as a community and their involvement in problem behaviors. *American Journal of Public Health, 87*, 1997-2001.
- Battistich, V. (1988). The measurement of classroom processes and children's prosocial characteristics in the Child Development Project. *Moral Education Forum, 13*, (4), 3-9.
- Berkowitz, M. W. (1997). The complete moral person: Anatomy and formation. In J. M. DuBois (Ed.), *Moral issues in psychology: Personalist contributions to selected problems*. Lanham, MD: University Press of America.
- Berkowitz, M. W. (1998). *A primer for evaluating a character education initiative*. Washington, D.C.: Character Education Partnership.
- Berkowitz, M. W. & Keller, M. (1994). Transitional processes in social cognitive development: A longitudinal study. *International Journal of Behavioral Development, 17*, 447-467.
- Buttram, J., Kruse, J., Sidler, J. (1992). *Evaluation of Heartwood Program*. (Available from Research for Better Schools, 444 North 3rd Street, Philadelphia, PA.)
- Center for Character Development (1994). *United States Air Force Academy Character Development Manual*. Colorado Springs: United States Air Force Academy.
- Colby, A. & Kohlberg, L. (1987). *The Measurement of moral judgment (Volumes 1 & 2)*. New York: Cambridge University Press.
- Coles, R. (1985-1986). The moral life of children. *Educational Leadership, 43*, (4), 19-25.
- Costa, A. L. (1986). Re-assessing assessment. *Educational Leadership, 46*, (7), 2.
- Darling-Hammond, L. (1997). *The Right to Learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.
- Farley & Associates. (1989). *National Survey of SFA Parents*. Columbus, Ohio.
- Farrelly, T. M. (1993). A New Approach to Moral Education: The Integrated Character Education Model. *Journal of Correctional Education, 44*, (2), 76-81.
- Gibbs, J. C. & Widaman, K. R. (1982). *Social intelligence: Measuring the development of sociomoral reflections*. Englewood Cliffs, NJ: Prentice Hall.
- Gilman, D. (1991). *A study of the Lions-Quest Program for the Academic Year 1990-91: A Report Prepared for the Analysis of Pretest and Posttest Measures for the North Gibson School Corporation of Princeton, Indiana*. Unpublished manuscript, Indiana University.
- Grady, Michael P. (1993). *Personal Responsibility Education Program: 1992-1993 Evaluation Report*. Unpublished manuscript, St. Louis, Mo.: St. Louis University.
- Gross, N., Giaquinta, J. & Bernstein, M. (1971). *Implementing Organizational Innovations: A Sociological Analysis of Planned Educational Change*. New York: Basic Books.
- Hendrix, W. H., Berkowitz, M. W., Stone, L. J., & Asiu, B. (1999). *Assessing the outcomes of character education: A model program*. Paper presented at the Annual Forum of the Character Education Partnership, Charlotte, NC.
- Home and School Institute. (1992). *Megaskills Students Improve School Performance*. Washington, D.C.: Home and School Institute.
- Kalish, J. D., Anastasi, R. E. & Kenny, R.A. (1991). *An Evaluation of the Baltimore County Public Schools' Values Education Program: A Synopsis*. Washington, D.C.: Ethics Resource Center, Inc.
- Kohlberg, L. (1980). *Essays on Moral Development. Volume 1: The Philosophy of Moral Development*. New York: Harper Row.
- Leming, J. S. (1993). In search of effective character education. *Educational Leadership, 51*, (3), 63-71.
- Lickona, T. (1991). *Educating for Character: How Our Schools Can Teach Respect and Responsibility*. New York: Bantam Books.
- Lickona, T. (1994). *Educating for Character*. Paper presented at the International Conference on Education and Values. Mexico City: Mexico.
- Marantz, R. D. (1991). *An analysis of the implementation of a curriculum innovation for character development: Sathya Sai education in human values*. Unpublished doctoral dissertation, Teachers College, Columbia University, New York.
- McClellan, B. E. (1999). *Schools and the Shaping of Character: Moral Education in America, 1607 - Present*. New York: Teachers College Press.
- Mulkey, Y. J. (1994). *Compilation of pieces from many evaluations*. (Available from Y. J. Mulkey, Character Education Institute, 8918 Tesoro Dr., Suite 575, San Antonio, TX 78217-6253.
- Piaget, J. (1965). *The Moral Judgment of the Child*. New York: Macmillan Publishing Co.
- Piscolith, M. (1991). *Heartwood Program Pilot Year Evaluation: 1990-91 Summary Report*. Pittsburgh, Pa.: Pittsburgh Board of Public Education.
- Power, F.C., Higgins, A., & Kohlberg, L. (1989). *Lawrence Kohlberg's Approach to Moral Education*. New York: Columbia University Press.
- Quest International Research and Evaluation Department

ment (1991). *Analysis of Lions-Quest's Skills for Adolescents Program in Detroit Schools Using School Records, School Attitude Measure and California Achievement Test Data*. Detroit, MI.

Rest, J. R. (1986). *DIT: Manual for the Defining Issues Test* (3rd Edition). Minneapolis: University of Minnesota.

Rest, J. R. (1994). Review of *Reclaiming our schools: a handbook on teaching character, academics, and discipline*. Unpublished manuscript, University of Minnesota: Minneapolis.

Satnick, R. D. (1991). *The Thomas Jefferson Center Values Education Project: A Survey of Administrators in the Los Angeles Unified School District*. Van Nuys, CA: California Survey Research.

Schaps, E. (1999). *Building community in a diverse world*. Paper presented at the Annual Forum of the Character Education Partnership, Charlotte NC.

Selman, R. L. (1980). *The Growth of Interpersonal Understanding: Developmental and Clinical Analyses*. New York: Academic Press.

Solomon, D., Watson, M. S., Delucchi, K. L., Schaps, E. & Battistich, V. (1988). Enhancing Children's Prosocial Behavior in the Classroom. *American Educational Research Journal*, 25, (4), 527-554.

Straughn, R. (1983). Values, Behavior and the Problem of Assessment. *Journal of Moral Education*, 12, (3), 187-191.

Taylor, C. (1994). Assessment for Measurement or Standards: The Peril and Promise of Large-Scale Assessment Reform. *American Educational Research Journal*, 31, (2), 231-262.

Thomas, R. S. (1991). Assessing character education: Paradigms, Problems, and Potentials. *The Clearing House*, 65, 51-55

Leslie Laud is a doctoral student in the Department of Curriculum and Teaching at Teachers College in Columbia University.

Marvin W. Berkowitz, a developmental psychologist and character educator, is Sanford N. McDonnell Endowed Professor of Character Education at the University of Missouri-St. Louis.