

**THE HISTORICAL DEVELOPMENT OF PROGRAM EVALUATION:
EXPLORING THE PAST AND PRESENT**

R. Lance Hogan
Eastern Illinois University

THE HISTORICAL DEVELOPMENT OF PROGRAM EVALUATION: EXPLORING THE PAST AND FUTURE

Abstract

The purpose of this article is to present the historical development and significant contributions leading to the current status of the program evaluation field. Program evaluation has been defined as “judging the worth or merit of something or the product of the process” (Scriven, 1991, p. 139). Guskey (2000) updated this definition stating that evaluation is a systematic process used to determine the merit or worth of a specific program, curriculum, or strategy in a specific context. The author describes seven significant time periods in the development of program evaluation and identifies five evaluation approaches currently used by practitioners. This article concludes by providing the reader with insight to the future of program evaluation.

Introduction

Organizational decision-makers and stakeholders want to ensure that programs are accomplishing their intended purpose. They are interested in assessing the effects of programs by asking questions like “What changes occurred?” or “Are we satisfied with the results?” (French, Bell, & Zawacki, 2000). Therefore, program evaluation is utilized by organizations to periodically assess their processes, procedures, and outcomes. According to Wholey, Hatry, and Newcomer (2007), the field of program evaluation provides processes and tools that workforce educators and developers can apply to obtain valid, reliable, and credible data to address a variety of questions about the performance of programs. Program evaluation is often defined as “judging the worth or merit of something or the product of the process” (Scriven, 1991, p. 139). Guskey (2000) updated this definition stating that evaluation is a systematic process used to determine the merit or worth of a specific program, curriculum, or strategy in a specific context. Despite its essential function, program evaluation may well be the most widely misunderstood, avoided, and feared activity by practitioners (Shrock, & Geis, 1999).

Purpose of the Study

The scope and purpose of this study was to provide an overview of the historical evolution of program evaluation by describing seven significant time periods. This overview was intended to give students, educators, and practitioners a succinct synopsis of the field of program evaluation and its advancement from the late 1700’s through the 21st Century. The growth and evolution of this field establishes the need for such a study. Further, five program evaluation approaches that are currently used by practitioners were identified. It is the hope of the researcher that a better understanding of evaluation will reduce the fear and misunderstanding identified by Shrock & Geis (1999).

Methodology

The researcher used systematic search methods to collect a broad swathe of relevant literature. The review synthesized the literature on program evaluation as it relates to seven significant time periods in the evolution of program evaluation identified by Madaus, Stufflebeam, and Kellaghan (2000). Primary resources for this study were collected from refereed print-based journals, ERIC documents, and books with an academic focus. A variety of search terms were used, including *evaluation*, *program evaluation* and *history of evaluation*.

Literature Review

Historical Evaluation of Program Evaluation

The historical development of evaluation is difficult, if not impossible, to describe due to its informal utilization by humans for thousands of years. Scriven (1996) noted that “evaluation is a very young discipline - although it is a very old practice” (p. 395). In the past 20 years, the field of evaluation has matured. According to Conner, Altman, and Jackson (1984), evaluation is an established field and is now in its late adolescent years and is currently making the transition to adulthood. Madaus et al., (2000), described seven development periods of program evaluation. First, the period prior to 1900, which the authors call Age of Reform; second, from 1900 until 1930, they call the Age of Efficiency; third, from 1930 to 1945, called the Tylerian Age; fourth, from 1946 to about

1957, called the Age of Innocence; fifth, from 1958 to 1972, the Age of Development; sixth, from 1973 to 1983, the Age of Professionalization; and seventh, from 1983 to 2000 the Age of Expansion and Integration.

Time Period 1: The Age of Reform (1792-1900's)

The first documented formal use of evaluation took place in 1792 when William Farish utilized the quantitative mark to assess students' performance (Hoskins, 1968). The quantitative mark permitted objective ranking of examinees and the averaging and aggregating of scores. Furthermore, the quantitative mark was historically important to the fruition of program evaluation as a discipline for two reasons: (a) it was the initial step in the development in psychometrics; and (b) its questions were designed to measure factual technical competence in subject areas that gradually replaced questions aimed at assessing rhetorical style (Madaus & O'Dyer, 1999).

During this period in Great Britain, education was reformed through evaluation. For example, the Powis Commission recommended that students' performance on reading, spelling, writing, and arithmetic would determine teachers' salaries. It was not uncommon to have annual evaluations on pupil attainments (Madaus & Kellaghan, 1982).

The earliest method of formal evaluation in the United States occurred in 1815 when the Army developed a system of policies for "uniformity of manufacturers' ordinance" (Smith, 1987, p.42). These policies set standardized production processes that fostered conformity of materials, production techniques, inspection, and product specification for all suppliers of arms to the military. The first formal education evaluation in the United States took place in Boston, Massachusetts in 1845. Printed tests of various subjects were used to assess student achievement in the Boston education system. Horace Mann, Secretary of the State Board of Education, wanted a comprehensive assessment of student achievement to assess the quality of a large school system. According to Stufflebeam, Madaus, & Kellaghan (2000), this event served to be an important moment in evaluation history because it began a long tradition of using pupil test scores as a principal source to evaluate school or instructional program effectiveness.

From 1887 to 1898, an educational reformer named Joseph Rice conducted a similar assessment by carrying out a comparative study on spelling instruction across a number of school districts. He was concerned about methods of teaching spelling, because U.S. students were not learning to spell. Rice was able to determine that there was no relationship between time devoted to spelling and competence. He reported his findings in *The Forum* in 1897, in an article entitled "The Futility of the Spelling Grind" (Colwell, 1998). Rice's evaluation has been recognized as the first formal educational program evaluation in America (Stufflebeam et al., 2000).

Time Period 2: The Age of Efficiency and Testing (1900-1930)

Fredrick W. Taylor's work on scientific management became influential to administrators in education (Biddle & Ellena, 1964). Taylor's scientific management was based on observation, measurement, analysis, and most importantly, efficiency (Russell & Taylor, 1998). Objective-based tests were critical in determining quality of instruction. Tests were developed by departments set up to improve the efficiency of the

educational district. According to Ballou (1916), tests developed for the Boston public schools were described as being objective referenced. The tests were used to make inferences about the effectiveness of the district. During this period, educators regarded measurement and evaluation as synonyms, with the latter thought of as summarizing student test performance and assigning grades (Worthen, Sanders, & Fitzpatrick, 1997).

Time Period 3: The Tylerian Age (1930-1945)

Ralph Tyler, considered the father of educational evaluation, made considerable contributions to evaluation. Tyler directed an Eight-Year Study (1932-1940) which assessed the outcomes of programs in 15 progressive high schools and 15 traditional high schools. Tyler found that instructional objectives could be clarified by stating them in behavioral terms, and those objectives could serve as the basis for evaluating the effectiveness of instruction (Tyler, 1975). Tyler wrote, "each objective must be defined in terms which clarify the kind of behavior which the course should help to develop" (cited in Walbesser & Eisenberg, 1972). Stufflebeam et al. (2000) concluded that Tylerian evaluation involves internal comparisons of outcomes with objectives; it need not provide for costly and disruptive comparisons between experimental and control groups, as were utilized by comparative studies used by Rice. According to Worthen et al. (1997), Tyler's work formed the basis of criterion-referenced testing.

Time Period 4: The Age of Innocence (1946-1957)

Starting in the mid 1940's, American's moved mentally beyond the war (World War II) and great depression. According to Madaus & Stufflebeam (1984), society experienced a period of great growth; there was an upgrading and expansion of educational offerings, personnel, and facilities. Because of this national optimism, little interest was given to accountability of national funds spent on education; hence the label of this evaluation time period, *The Age of Innocence*.

In the early 1950's during The Age of Innocence, Tyler's view of evaluation was rapidly adopted. Bloom, Engelhart, Furst, Hill, and Krathwohl (1956) gave objective-based testing advancement when they published the *Taxonomy of Educational Objectives*. The authors indicated that within the cognitive domain there were various types of learning outcomes. Objectives could be classified according to the type of learner behavior described therein, and that there was a hierarchical relationship among the various types of outcomes. Moreover, they indicated that tests should be designed to measure each type of outcome (Reiser, 2001).

Time Period 5: The Age of Development (1958-1972)

In 1957, the Russian's successful launch of Sputnik I sparked a national crisis. As a result, legislation was passed to improve instruction in areas that were considered crucial to the national defense and security. In 1958, Congress enacted the National Defense Education Act (NDEA) which poured millions of dollars into new curriculum development projects and provided for new educational programs in mathematics, sciences, and foreign languages (Stufflebeam, Madaus, & Kellaghan, 2000). Evaluations were funded to measure the success of the new curricula.

In the early 1960's, another important factor in the development of evaluation was the emergence of criterion-referenced testing. Until that time, most tests, called norm-

referenced tests, were designed to discern between the performances of students. In contrast, a criterion-referenced test was intended to measure individual performance in terms of established criteria. It discerns how well an individual can perform a particular behavior or set of behaviors, irrespective of how well others perform (Reiser, 2001).

The passage of the Elementary and Secondary Education Act (ESEA) of 1965 was recognized as the birth of the contemporary program evaluation and included requirements for evaluation. According to Ferguson (2004), the ESEA was intended to supplement academic resources for low-income children who needed extra support in the early grades. Educators were required to evaluate their efforts. Senator Robert Kennedy sponsored the Act because he wanted to authenticate that federal money was not going to support schools' exhausted practices, but rather would help disadvantaged students in new ways (Weiss, 1998).

Time Period 6: The Age of Professionalization (1973-1983)

During the 1970's, evaluation emerged as a profession. A number of journals including *Educational Evaluation and Policy Analysis*, *Studies in Educational Evaluation*, *CEDR Quarterly*, *Evaluation Review*, *New Directions for Program Evaluation*, *Evaluation and Program Planning*, and *Evaluation News* were published (Stufflebeam et al., 2000). Further, universities began to recognize the importance of evaluation by offering courses in evaluation methodology. Among them were the University of Illinois, Stanford University, Boston College, UCLA, University of Minnesota, and Western Michigan University (Stufflebeam et al., 2000).

Time Period 7: The Age of Expansion and Integration (1983-Present)

In the early 1980's, evaluation struggled under the Reagan administration. Cut backs in funding for evaluation took place and emphasis on cost cutting arose. According to Weiss (1998), funding for new social initiatives were drastically cut. By the early 1990's, evaluation had rebounded with the economy. The field expanded and became more integrated. Professional associations were developed along with evaluation standards. In addition, the Joint Committee on Standards for Educational Evaluation developed criteria for personnel evaluation.

Evaluation approaches for the 21st Century

Many evaluation approaches have emerged since the 1930's and range from checklists of suggestions to comprehensive prescriptions. Worthen et al., (1997) classified the different evaluation approaches into the following five categories: (a) objectives-oriented, (b) management-oriented, (c) consumer-oriented, (d) expertise-oriented, (e) adversary-oriented, and (f) participant-oriented evaluation approaches. In addition to these categories, specific evaluation approaches have emerged due to the attention given by researchers and practitioners. These specific evaluation approaches include: (a) CIPP (discussed in management-oriented), (b) CIRO, (c) Kirkpatrick's Evaluation Approach, and (d) Phillip's Evaluation Approach.

Objectives-Oriented Approach

The objectives-oriented evaluation approach focuses on specifying the goals and objectives of a given program and determines the extent to which they have been

attained. Ralph Tyler, who conceptualized the objectives-oriented approach to evaluation in 1932, is recognized as being the pioneer of this approach (Stufflebeam & Shinklefield, 1985). According to Worthen and Sanders (1987), Tyler's early approach to evaluation was "logical, scientifically acceptable, and readily usable by educational evaluators" (p. 63). Tyler hypothesized that, as a pre-requisite to evaluation, goals and objectives must be defined. Evaluation then measured whether these goals and objectives were attained. Tyler used the objectives-oriented approach during his Eight-Year Study.

In 1930, the Progressive Education Association established the Commission on the Relation of School to College and appointed Ralph W. Tyler as Director of Research for the Evaluation Staff. The purpose of the commission was to conduct long-term research studies to determine the relevance of high school curriculum and its impact on success in college admissions. Tyler's Eight-Year Study determined that student success in college is not predetermined by high-school curriculum requirements. The study determined that students attending more experimental schools performed better than students in less experimental schools. Finally, the study found that integrative curricula approaches produced students that performed better in college than students who did not have integrative curricula.

According to Guba and Lincoln (1981), there were problems associated with the objectives-oriented approach. Critics of this evaluation approach claimed that the selection of appropriate objectives to evaluate was problematic, as not all objectives could be evaluated and the process by which objectives were selected was open to bias (Stufflebeam & Shinklefield, 1985). Also, Worthen and Sanders (1987) cautioned that objectives-oriented evaluation could limit the scope and perception of the evaluation, similar to blinders, causing the evaluator to miss important outcomes not directly related to the goals of the evaluation.

Management-Oriented Approach

The management-oriented evaluation approach was intended to serve organizational leaders by meeting the informational needs of managerial decision makers. The foremost management-oriented evaluation approach was developed by Daniel Stufflebeam. Corresponding to the letters in the acronym, CIPP, are the following core concepts: context, input, process, and product evaluation. According to Mathews and Hudson (2001), context evaluation scrutinizes the program objectives to determine their social acceptability, cultural relativity, and technical adequacy. Input evaluation involves an examination of the intended content of the program. Process evaluation relates to implementation of the program, that is, the degree to which the program was delivered as planned. Finally, product evaluation is the assessment of program outcomes. Stufflebeam et al. (2000) noted:

The model is intended for the use of service providers, such as policy boards, program and project staffs, directors of a variety of services, accreditation officials, school district superintendents, school principals, teachers, college and university administrators, physicians, military leaders, and evaluation specialists. The model is configured for use in internal evaluations conducted by organizations, self-evaluations conducted by individual service providers, and contracted external evaluations. (p. 279)

According to Worthen et al., (1997), potential weaknesses of the management-oriented approach may occur from evaluators giving partiality to top management, from evaluators' occasional inability to respond to questions, from costly evaluation processes, and from the assumption that important decisions can be clearly identified in advance.

Consumer-Oriented Approach

The consumer-oriented evaluation approach is commonly used by government agencies and consumer advocates who compile information to evaluate a product's effectiveness. According to Stufflebeam et al., (2000), a consumer-oriented evaluation requires a highly credible and competent expert with sufficient resources to conduct a thorough evaluation. Scriven (1991) was a pioneer in applying the consumer-oriented approach to program evaluation and was responsible for distinguishing between the formative and summative roles of evaluation. The primary purpose of formative evaluation is to improve the quality of the program being developed so it will be possible to achieve the objectives for which it was designed (Beyer, 1995). Summative evaluation is conducted to provide decision-makers or potential customers with judgments about the worth or merit of a program in relation to important criteria (Brown & Gerhardt, 2002).

Expertise-Oriented Approach

The expertise-oriented evaluation approach is the oldest and most widely used evaluation approach to judge a program, activity, or institution (Worthen, Sanders, & Fitzpatrick, 1997). Evaluators utilizing this approach draw on a panel of experts to judge a program and make recommendations based on their perceptions. The review process can be formal or informal. Worthen et al. (1997) defined a formal review system as, "one having (a) structure or organization established to conduct periodic reviews; (b) published standards; (c) a prespecified review schedule; (d) a combination of several experts to judge overall value; and (e) an impact depending on the outcome of the evaluation" (p. 121). Any other evaluation lacking one of the five components is considered to be an informal review system.

In the eyes of critics, the overall limitation to the expertise-oriented evaluation approach is the central role of the expert judge. Critics suggest that the use of expert judges permits evaluators to make judgments that are personally biased, inherently conservative, potentially incestuous, and are not based upon program objectives (Worthen et al., 1997).

Adversary-Oriented Approach

The adversary-oriented evaluation approach utilizes a judicial process in examining a program. Worthen et al., (1997) identified the central focus of adversary-oriented evaluation is to obtain results through the examination of opposing views. The pros and cons of an issue are examined by two separate teams who then publicly debate to defend their positions and mutually agree on a common position. The evaluation process involves a hearing, prosecution, defense, jury, charges and rebuttals. According to Levine (1982), the adversarial approach operates with the assumption that the truth emerges from a hard, but fair, fight in which opposing sides present supporting evidence.

One advantage to this approach is that it illuminates both positive and negative view points. Additionally, the approach is open to participation by stakeholders and

decisions place greater assurance in the conclusion of the trial. This evaluation approach is not commonly adopted because of its determination of guilt. Worthen et al (1997) stated, "Evaluation should aspire to improve programs, not determine their guilt or innocence." (p. 149)

Participant-Oriented Approach

The participant-oriented evaluation approach stresses firsthand experiences with program activities and emphasizes the importance of the participants in the process. As defined by Royse, Thyer, Padgett, and Logan (2006), participative evaluation "centers on enlisting the cooperation of the least powerful stakeholders in the evaluation from start to finish" (p. 93). Stakeholders define the evaluation approach and determine the evaluation parameters. The participant-oriented approach allows for the evaluator to engage with the stakeholder as a partner in solving the problems.

Empowerment evaluation has been considered a sub classification within participative-oriented evaluation (Secret, Jordan, & Ford, 1999). Strober (2005) described empowerment evaluation as a type of formative evaluation in which participants in a project generate goals for a desired change, develop strategies to achieve them, and monitor their progress. Fetterman (2001) identified three steps as part of empowerment evaluation: (a) developing a unifying purpose; (b) determining where the program stands, including strengths and weaknesses; and (c) planning for the future by establishing goals.

The participant-oriented evaluation (including empowerment) approach is not without disadvantages. According to Worthen et al., (1997), because of the reliance on human observation and individual perspective there is a tendency to minimize the importance of instrumentation and group data. Additionally, advocates have been criticized because of the subjectivity of the evaluation process and possibility of conflicts to arise among participants. Finally, participants could manipulate the situation or withdraw at crucial times causing the evaluation to be negated.

CIRO Evaluation Approach

In 1970, the CIRO model for the evaluation of management training was proposed (Warr, Bird, & Rackham, 1970). This model was based on the evaluation of four aspects of training: context, input, reaction, and outcome. Context evaluation focuses on factors such as the correct identification of training needs and the setting of objectives in relation to the organization's culture and climate. Input evaluation is concerned with the design and delivery of the training activity. Reaction evaluation looks at gaining and using information about the quality of trainees' experiences. Outcome evaluation focuses on the achievements gained from the activity and is assessed at three levels: (a) immediate, (b) intermediate, and (c) ultimate evaluation.

Immediate evaluation attempts to measure changes in knowledge, skill, or attitude before a trainee returns to the job. According to Santos and Stuart (2003), "Intermediate evaluation refers to the impact of training on job performance and how learning is transferred back into the workplace." Finally, ultimate evaluation attempts to assess the impact of training on departmental or organizational performance in terms of overall results. According to Tennant, Boonkrong, and Roberts (2002), the CIRO model focuses on measurements both before and after the training has been carried out. The main

strength of the CIRO model is that the objectives (context) and the training equipment (input) are considered.

Kirkpatrick's Evaluation Approach

In 1959, Donald Kirkpatrick presented his evaluation approach. The widely adopted Kirkpatrick (1967) evaluation approach proposes four levels of training outcomes: (a) trainees' reactions to the training curriculum and training process (reactions), (b) knowledge or skill acquisition at the end of training (learning), (c) behavior change in the job (behavior), and (d) improvements in individual or organizational outcomes (results).

According to a survey by the American Society for Training and Development (ASTD), the Kirkpatrick four-level evaluation approach is still the most commonly used evaluation framework among Benchmarking Forum Companies (Bassi & Cheney, 1997). The main strength of the Kirkpatrick evaluation approach is the focus on behavioral outcomes of the learners involved in the training (Mann & Robertson, 1996).

Phillips' Evaluation Approach

In the past decade, training professionals have been challenged to provide evidence of how training financially contributes to businesses. Phillips (1996) suggested adding another level to Kirkpatrick's four-level evaluation approach to calculate the return on investment (ROI) generated by training. According to James and Roffe (2000), Phillips' five-level evaluation approach translates the worth of training into monetary value, which, in effect, addresses ROI. Phillips' framework provides trainers a logical framework to view ROI both from a human performance and business outcome perspective.

Phillips noted (1991):

Evaluation should occur at each of the four levels and a comprehensive evaluation process will focus on all four levels in the same program. The common thread among most evaluation experts is that emphasis should be placed on the ultimate outcome, which results in improved group or organization performance. It is the most difficult to obtain, document and measure. The other three levels will not suffice in an ultimate evaluation. There is evidence in studies to indicate that the fourth level, a results orientation, is a method most desired and receives the most support. (p. 51)

In light of the excitement over the past decade with Phillip's evaluation approach, advantages and disadvantages with this ROI methodology have surfaced. Apparent advantages of this evaluation approach are twofold: (a) gain a better understanding of factors influencing training effectiveness, and (b) determine the monetary value of specific training initiatives. Despite the obvious advantages, the ROI methodology can become overly complex in determining a bottom line organizational value on training, as it is not an inexact science. Specifically, it can be difficult to isolate the effects of training. According to Shelton and Alliger (1993), one way to measure the effectiveness of training is to compare the results of a control group with the results of the experimental group or trainee group which can be burdensome for practitioners.

Current and Future Status of Program Evaluation

Worthen, Sanders, and Fitzpatrick (2004) identified twelve emerging trends that have and will have the greatest influence in shaping the current and future status of evaluation. Following are the twelve trends:

1. Increased priority and legitimacy of internal evaluation.
2. Expanded use of qualitative methods.
3. A strong shift toward combining quantitative and qualitative methods in each program evaluation rather than depending exclusively on either method.
4. Increased acceptance of and preference for multiple-method evaluations.
5. Introduction and development of theory-based evaluation.
6. Increased concern over ethical issues in conducting program evaluations.
7. Increased use of program evaluation within business, industry, foundations, and other agencies in the private and nonprofit sector.
8. Increased use of evaluation to empower a program's stakeholders.
9. Increased options that program evaluators should assume the role of advocates for the programs they evaluate.
10. Advances in technology available to evaluators, and communication and ethical issues such advances will raise.
11. Educators' increased use of alternative assessment methods (as opposed to traditional testing) to assess students' performance, and increased pressure on educational evaluators to use such methods in evaluating school programs.
12. Modifications in evaluation strategies to accommodate increasing trends of government decentralization and delegation of responsibilities to state/provinces and localities. (pp. 49-50)

Conclusion

Most experts would agree that program evaluation has an exciting and dynamic history. Due to its development over the past 200 years, program evaluation has matured significantly into an established field of study. The overarching trend of this field of study has been the transition from more traditional summative evaluation approaches focusing on outcomes toward formative evaluation (Marshall, Crowe, Oades, Deane, & Kavanaugh, 2007). Through this trend, universities have accordingly developed courses and organizations utilize its approaches to understand their processes, procedures, and outcomes.

References

- Ballou, F.A. (1916). *Work of the Department of Educational Investigation and measurement*. Boston, MA: Chicago: University Press.
- Bassi, L.J. & Cheney, S. (1997, November). Benchmarking the best. *Training & Development*, 51(11), 60-64.
- Beyer, B.K. (1995). *How to conduct a formative evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Biddle, B.J. & Ellana, W.J. (1964). The application of program theory to the evaluation of a managed mental health care system. *Evaluation and Program Planning*, 19(1), 111-119.
- Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals handbook 1*. New York: David McKay.
- Brown, K.G. & Gerhardt, M.W. (2002, Winter). Formative evaluation: An integrative practice model and case study. *Personnel Psychology*, 55(4), 951-984.
- Colwell, R. (1998). A long trail awinding: Issues, interests, and priorities in arts education research. *Arts Education Policy Review*, 99(5), 21-30.
- Conner, R.F., Altman, D.G., & Jackson, C. (1984). *Evaluation studies review annual*. Beverly Hills, CA: Sage.
- Ferguson, R.F. (2004). An unfinished journey: The legacy of Brown and the narrowing of the achievement gap. *Phi Delta Kappan*, 85(9), 656-670.
- Fetterman, D. (2001). *Foundations of empowerment*. Thousand Oaks, CA: Sage
- French, W.L., Bell, C.H. & Zawacki, R.A. (2000). *Organizational development and transformation: Manage effective change*. (5th ed.). New York: Irwin McGraw-Hill.
- Guba, E.G., & Lincoln, Y.S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass
- Gusky, T.R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press, Inc.
- Hoskins, K. (1968). The examination, disciplinary power and rational schooling. *History of Education*, 8(1), 135-146.
- James, C. & Roffe, I. (2000). The evaluation of goal and goal-free training innovation. *Journal of European Industrial Innovations*, 24(1), 12-26.
- Kirkpatrick, D.L. (1967). *Evaluation of training*. New York: McGraw-Hill.
- Kirkpatrick, D.L. (1996). Revisiting Kirkpatrick's four-level model. *Training and Development*, 50(1), 54-57.
- Levine, M. (1982). *Communication strategies in evaluation*. Beverly Hills, CA: Sage.
- Madaus, G.F. & Kellaghan, T. (1982). *Trends in standards in Great Britain and Ireland*. In G. Austin & H. Garber (Eds.). *The rise and fall of national test scores*. New York: Academic Press.
- Madaus, G.F. & O'Dwyer, L.M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan*, 80(9), 688-697.
- Madaus, G.F., Scriven, M., & Stufflebeam, D.L. (1983). *Evaluation models: Viewpoints on educational and human services evaluation*. Hingham, MA: Kluwer Nijhoff.
- Madaus, G.F., Scriven, M., & Stufflebeam, D.L. (1984). Educational evaluation and accountability: A review of quality assurance efforts. *The American*, 27(5), 649-673.

- Madaus, G.F., Stufflebeam, D.L., & Kellaghan, T. (2000). *Evaluation models: Viewpoints on educational and human services evaluation*. (2nd ed.). Hingham, MA: Kluwer Academic Publishers.
- Mann, S. & Robertson, L.T. (1996). What should training evaluations evaluate? *Journal of European Industrial Training*, 20(9), 14-20.
- Marshall, S.L., Crowe, T.P., Oades, L.G., Deane, F.F., & Kavanaugh, D.J. (2007). A review of consumer involvement in evaluations of case management: Consistency with a recovery paradigm. *Psychiatric Services*, 58(30), 396-401.
- Mathews, J.M., & Hudson, A.M. (2001). Guidelines for evaluating parent training programs. *Family Relations*, 50(1), 77-87.
- Phillips, J. (1991). *Handbook of Evaluation and Measurement Methods* (2nd eds.). London: Gulf Press.
- Phillips, J. (1996). *Accountability in Human Resource Management*, Butterworth-Heinemann, Oxford.
- Reiser, R.A. (2001). A history of instructional design and technology: Part II a history of instructional design. *Educational Technology, Research and Development*, 49(2), 57-68.
- Royse, D., Thyer, B.A., Padgett, D.K., & Logan, T.K. (2006). *Program evaluation: An introduction* (4th eds.). Belmont, CA: Brooks-Cole.
- Russell, R.S. & Taylor, B.W. III. (1998). *Operations Management: Focusing on quality and competitiveness*. (2nd eds.). Upper Saddle River, NJ: Prentice Hall, Inc.
- Santos, A. & Stuart, M. (2003). Employees perceptions and influence on training effectiveness. *Human Resource Management Journal*, 13(1), 27-45.
- Scriven, M. (1980). *Educational thesaurus* (2nd eds.). CA: Edgepress.
- Scriven, M. (1991). *Evaluation Thesaurus* (4th eds.). Newbury Park, CA: Sage.
- Scriven, M. (1996). The theory behind practical evaluation. *Evaluation*, 2(4), 393-404.
- Secreat, M., Jordan, A., & Ford, J. (1999). Empowerment evaluation as a social work strategy. *Health & Social Work*, 24(2), 120-128.
- Shelton, S. & Alliger, G. (1993). Who's afraid of level 4 evaluation? *Training and Development*, 47(6), 43-46.
- Shrock, S.A. & Geis, G.L. (1999). Evaluation in Handbook of Human Performance Technology. In H. Stolovich & E. Keeps (Eds.) *Handbook of human performance technology: A comprehensive guide for analyzing and solving human performance problems in organizations* (p. 360 – 365). San-Francisco: Jossey-Bass.
- Smith, N.L. (1987). *Army ordnance and American system of manufacturing*. Cambridge, MA: MIT Press.
- Strober, E. (2005). Is power-sharing possible? Using empowerment evaluation with parents and nurses in a pediatric hospital transplantation setting. *Human Organization*, 64(2), 201-211.
- Stufflebeam, D.L., Madaus, G.F., & Kellaghan, T. (2000). *Evaluation models: Viewpoints on educational and human services evaluation*. (2nd eds.). Boston: Kluwer Academic Publishers.
- Stufflebeam, D.L. & Shinkfield, A.J. (1985). *Systematic evaluation*. New York: Kluwer-Nijhoff Publishing.

- Tennant, C., Boonkrong, M., & Roberts, P.A. (2002). The design of a training program measurement model. *Journal of European Industrial Training*, 26(5), 230-241.
- Tyler, R.W. (1942). General statement of evaluation, *Journal of Educational Research*, 35(4), 492-501.
- Tyler, R.W. (1975). Educational benchmarks in retrospect: Educational change since 1915. *Viewpoints*, 510(1), 11-31.
- Walbesser, H.H., & Eisenberg, T.A. (1972). *A review of the research on behavioral objectives and learning hierarchies*. Columbus, OH: Ohio State University, Center for Science and Mathematics Education. (ERIC Document Reproduction Service No. ED 059 900).
- Warr, P., Bird, M. & Rackham, N. (1970). *Evaluation of management training*. London: Gower Press.
- Weiss, C.H. (1972). *Evaluation research*. Englewood Cliffs, NJ: Prentice Hall.
- Weiss, C.H. (1998). *Evaluation: Methods for studying programs and policies*. (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wholey, J.S., Hatry, H.P., & Newcomer, K.E. (Eds.). (2004). *Handbook of practical program evaluation*. (2nd ed.). San Francisco, CA: Jossey-Bass Publishing.
- Worthen, B.R., & Sanders, J.R. (1987). *Educational evaluation*. New York: Longman.
- Worthen, B.R., Sanders, J.R., & Fitzpatrick, J.L. (1997). *Educational evaluation: Alternative approaches and practical guidelines*. (2nd ed.). White Plains, NY: Longman.
- Worthen, B.R., Sanders, J.R., & Fitzpatrick, J.L. (2004). *Educational evaluation: Alternative approaches and practical guidelines*. (3rd ed.). Boston: Allyn & Bacon.