Journal of Management Development

Article information:

Users who downloaded this article also downloaded:

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

An outcomes evaluation of a leadership development initiative

Thomas Packard and Loring Jones

*School of Social Work, San Diego State University, San Diego,
California, USA*

Abstract

Purpose – The purpose of this paper is to report the impact of a leadership development initiative in eight organizations and to demonstrate the use of promising evaluation methods.

Design/methodology/approach – This study used multiple methods including surveys with quantitative and qualitative data from participants and their supervisors.

Findings – Program participants and their supervisors reported improved on-the-job performance of participants. There were statistically significant increases in participant self-efficacy over time.

Research limitations/implications – In this time series design, some changes may be due to factors besides the program. Limitations in self-report data, common in studies such as this, were partially mediated by the use of supervisor ratings. Future research could include more objective measures of performance.

Practical implications – Because this evaluation reported on-the-job performance improvements for participants in a leadership development program, other organizations can adapt the program model and expect similar performance improvements. This study also advances leadership development evaluation methods by addressing on-the-job performance to a much greater extent than in past studies.

Social implications – This study of a program training leaders in human service organizations can help similar organizations better train their leaders with expectations of improving the quality of life for clients such as families experiencing poverty, unemployment, and child abuse or neglect.

Originality/value – This evaluation makes a unique contribution in terms of measurement of on-the-job performance of program participants in a time series design which includes ratings of supervisors and self-ratings, benefiting organizations designing such programs and evaluations of them.

Keywords Performance, Leadership development, Outcomes, Organizational effectiveness

Paper type Research paper

Leadership development and, more specifically, leadership development evaluation, have received increasing attention in recent years. The research reported here contributes to this growing literature by demonstrating the use of evaluation methods which are not yet common in leadership development programs and which offer promise for advancing knowledge in this area.

The first purpose of this paper is to address the research question of assessing the results of a leadership development program in terms of changes in self-efficacy and the job performance of program participants. The second purpose is to demonstrate the

use of evaluation methods which show promise in advancing the state of knowledge in the field of leadership development program evaluation.

After a literature review, the Methodology section will present the components of an ongoing leadership development program, its evaluation design, and evaluation results to date, followed by conclusions and implications for research and practice. While this program is in the governmental sector, Russon and Reinelt (2004) recommend sharing across sectors. This program, which serves public human service organizations with annual budgets ranging from $50 million to over $3 billion, and its evaluation should be relevant in a wide range of organizations and industries.

Several study limitations will be noted here, with detail and how the limitations were addressed provided in the Methodology section. There were different respondents for each year; and over the years, there have been occasional minor changes in module content and trainers. Conditions in the agencies may have changed over the years; and other historical factors could have affected results. This study relied partly upon self-report data, and therefore some responses might not always reflect reality. A control group was not used, limiting the ability to establish cause and effect. Attrition was high at the three and six month measures, which raises the issue of how completers might have been different from non-completers.

Leadership development

McCauley (2008) noted a distinction (earlier conceptualized by Day, 2001) between leader development and the more commonly used term, leadership development, with the former focussing on the development of individual leaders and the latter "encompassing the development of collective leadership beliefs and practices in addition to individual development" (McCauley, 2008, p. 6). While the program described below mainly addresses individual leader development, it also is intended to impact the performance and cultures of the eight participating organizations, to enhance leadership development in the larger sense. Therefore, the term leadership development is appropriate here.

Leadership development programs, "structured, off-the-job events that bring individuals together for shared learning and development experiences" (McCauley, 2008, p. 24) are probably the most common explicitly designed development method. Formal training programs – the subject here – are often considered to be only a minor part of leadership development (Aude *et al.*, 2007), with on-the-job experiences, challenges, setbacks, and learning from others seen as more important. Nevertheless, leadership development programs are an important factor in many large organizations: one survey of workforce learning professionals found that "the average amount spent on each employee participating in executive development programs was $12,370" (American Society for Training and Development, 2008, p. 3). The program reported here is a traditional off-site training program, but also has elements which address the work environment, to enhance opportunities for the transfer of new learning.

Formal leadership development programs (e.g. Hernez-Broome and Hughes, 2004; Van Velsor *et al.*, 2010) are available through specialized training organizations, in-house programs for a particular organization, and consortia in which similar organizations pool resources. In the human services field, the setting for the program evaluated here, a large-scale executive development program for county managers has operated in the San Francisco Bay area since 1994 (Austin *et al.*, 2006). A similar program provides leadership development training for managers of nonprofit organizations (Austin *et al.*, 2011).

Leadership development can use several components, often including some combination of off-site training/development programs, 360-degree feedback, the use of instruments filled out by participants on their management styles or characteristics, executive coaching, mentoring, assessment centers, action learning such as real-world problem solving, and plans for applications of new knowledge and skills on the job.

McCauley (2008) examined studies that focussed on the leader's developmental path (including challenging assignments and learning from hardships) and leader development methods such as those just mentioned above. The most common conceptualization of leadership development focusses on skills and abilities, often described as competencies; although a focus on transforming a leader's world views and developing network relationships and sense of identity are also seen as important. Leadership development programs can also focus on improving performance in the work unit or organization.

Leskiw and Singh (2007, p. 444), after an extensive review of the literature, concluded that "Six key factors were found to be vital for effective leadership development: a thorough needs assessment, the selection of a suitable audience, the design of an appropriate infrastructure to support the initiative, the design and implementation of an entire learning system, an evaluation system, and corresponding actions to reward success and improve on deficiencies." In a similar vein, McCauley (2008) listed success factors including alignment of leadership development objectives with business strategies, top-level executive support, shared responsibility between line managers and human resources staff, manager accountability for the development of subordinates, competency models, multiple development methods, and evaluation. The importance of contextual factors was also noted by Atwood *et al.* (2010), who noted the importance of trainees sharing learnings in the workplace.

Summarizing a special journal issue on leadership development, Riggio (2008, p. 390) asserted that such programs "need to fit the requirements of both the organizations and the leaders undergoing development. They need to be theory-driven, use proven methods, be integrated into ongoing organizational processes, evaluated for effectiveness and substantial." All of these criteria except involvement of human resources staff were included in the program described below.

Leadership development evaluation
In spite of a growing number of published reports, much remains to be learned about the effectiveness of leadership development programs. Many methods have been used to evaluate leadership development in many contexts, from businesses to social change.

A meta-analysis by Collins and Holton (2004) assessed designs including post-test only with control group, pretest-posttest with control group, correlational analyses, and single group pretest-posttest. Possible outcomes studied included knowledge, expertise (e.g. behavioral or performance change), and system change (e.g. subordinates' job satisfaction or commitment to the organization, group effectiveness, reduced costs, or improved quality or quantity). Grove *et al.* (2007) used an open-systems model with primarily qualitative methods to assess impacts at the individual, organizational, community, or societal levels in fields including public health and community development.

Martineau and Patterson (2010) described methods for evaluating individual, group, team, and organizational outcomes, and emphasized the importance of organizational context and incorporating "evaluative thinking" into the program design. Several of these methods, including pre-post surveys, behavioral observation, climate surveys,

and assessment of organizational systems changes, were used in the study reported here. Houghton and DiLiello (2010) studied the relationships among leadership development, creative self-efficacy (self-efficacy is a variable in the study presented here), organizational support for creativity, and creativity, with one conclusion that "adult leadership development activities appear to magnify the effects of perceived organizational support for creativity on perceptions of opportunities to actually practice creativity" (p. 240).

A handbook on leadership development evaluation (Hannum *et al.*, 2007) includes chapters on subjects ranging from methods (e.g. experimental and theory of change models, return on investment) to contexts for evaluation, such as social change, communities, schools, and neighborhoods.

A special issue of *Leadership Quarterly* (Hannum and Craig, 2010) addressed leadership development evaluation, with subjects including coaching, return on leadership development investment, leader self-development, mentoring, and the use of methods including hierarchical linear modeling, social network analysis, and Q-methodology.

Clarke (2012) and Watkins *et al.* (2011) noted the importance of evaluating programs on multiple levels. In this vein, Nicolaidouab and Petridouabc (2011) evaluated factors including participant satisfaction, learning and application of new knowledge and skills, and organizational support, concluding that such training needs to become more embedded in organizational operations. They also noted methodological challenges in such evaluations.

In their meta-analysis of 83 leadership development program studies, Collins and Holton (2004) found many positive effects from a wide range of programs. They concluded that "practitioners can attain substantial improvements in both knowledge and skills if sufficient front-end analysis is conducted to assure that the right development is offered to the right leaders" (p. 217). After her review of the two major meta-analytic studies of such programs (Burke and Day, 1986; Collins and Holton, 2004), McCauley concluded that "the effectiveness of leader development programs varies widely, although effect sizes were positive across all criteria" (2008, p. 26).

However, a limitation in this research is that organizational outcomes are rarely mentioned. In the Collins and Holton (2004) meta-analysis mentioned above, only 11 of 130 studies measured "system objectives," and only one measured financial outcomes. In another meta-analysis of leader impact research (Avolio *et al.*, 2009) only two studies out of 207 had organizational performance as a dependent variable. In the special issue of *Leadership Quarterly* on leadership development evaluation, only one study (Gentry and Martineau, 2010) attempted to measure performance in any way.

In spite of this growing literature, comprehensive evaluation of leadership development programs is not common, partly due to the staff time and funding commitments required; and to date, there is not much research which fully assesses the results of such programs, particularly regarding outcomes and impacts on the organization (Russon and Reinelt, 2004; Day, 2001).

Another limitation in leadership development evaluations was noted by Ely *et al.* (2010), who recommended more multi-source data collection. The findings of the study reported here did include superiors' evaluations and actually found the opposite effect: supervisors rated program participant improvements more highly than did the participants themselves.

In conclusion, McCauley suggested that evaluations of leadership development programs should go beyond the common focus on impacts on individuals to assessing

"their impact at the group, organization, and even industry and society levels" (p. 29). The evaluation design reported here builds upon this suggestion and the existing literature by going beyond the individual level to assess, in a preliminary way, organizational performance, and by including multiple methods of evaluation.

Because self-efficacy has been associated with performance in organizations, it was considered an important desired outcome of this program and a variable for study. Therefore, this literature review concludes with a brief discussion of this important concept.

Self-efficacy has been widely studied in the social psychology and leadership fields. It can be defined as "beliefs in one's abilities to mobilize the motivation, cognitive resources, and courses of action needed to meet situational demands" (Hannah *et al.*, 2008). They added that "research has demonstrated strong positive relationships between self-efficacy and various criteria of human performance in organizations" (Hannah *et al.*, 2008, p. 671). Others, including Holden (1992), Multon *et al.* (1991), Paglis and Green (2002), Prussia *et al.* (1998), and Stajkovic and Luthans (1998) reported similar findings. In a meta-analysis of 89 empirical studies, Blume *et al.* (2010) found "moderately strong relationships" between self-efficacy and transfer of learning to the work setting. In a study of the relationship between self-efficacy and leadership effectiveness, Anderson *et al.* (2008) found that "managers' self-evaluations of perceived competence [...] were highly related to raters' descriptions of their effectiveness in a variety of areas" (p. 604).

These streams of literature suggest that, while progress has been made in the evaluation of leadership development programs, much still remains to be learned about variables to study and evaluation methods to use.

Methodology

Common elements of leadership development are formal development programs, multi-source feedback, developmental, executive coaching, action learning, and mentoring (McCauley, 2008). All of these except assessment centers and mentoring have been included in the program described here.

The program was developed to train middle managers in eight county human service agencies in order to create a talent pool for promotion and to enhance organizational performance. The directors of these agencies have been meeting quarterly since 2001 as a consortium in which members discuss and develop strategies for issues facing public human services by engaging in strategy discussions, research, policy development, and succession planning. In 2003, the directors noted that in the coming years, many of their executive managers would retire, creating a pressing need for succession planning and the development of middle managers who could move into executive leadership positions when these senior executives retired.

To respond, staff of the consortium, including a faculty consultant from the school of social work overseeing this project, completed a best practices review of leadership development programs and lists of specific competencies commonly seen as important in human services leadership. In addition to generic management competency models, social work competencies (Wimpfheimer, 2004; Menefee, 2000) were considered. These findings were presented to the directors, who, in collaboration with staff, developed a set of competencies relevant to their agencies which would guide development of the program. Staff also conducted four focus groups with 45 managers of these agencies. Then, with additional consultation from other leadership development experts, staff designed a program which was approved by the directors for implementation.

In a review of 55 leadership development programs, Russon and Reinelt (2004) noted that the use of a "theory of change" can be useful in evaluating a leadership development program. Gutierrez and Tasse (2007) also articulated the value of using such a theory of change framework in leadership development evaluation, in which program designers "articulate the premises, assumptions, and hypotheses that might explain the how, when, and why of the processes of change" (p. 49).

The theory of change in this study forms the theoretical underpinning which guided this research, suggesting a hypothesis about the program's effects: that the program will affect participant self-efficacy and work performance. This model also shows the importance of the program as one aspect of a strategic human resources management process intended to improve individual manger performance and ultimately organizational performance: that individual-level changes for participants should eventually lead to organization-level changes in the learning environment and program outcomes. This program was intended to be augmented by on-the-job development activities such as stretch assignments and mentoring (currently only informal in this model).

Figure 1 represents the theory of change that informed the implementation and evaluation of this program. It shows the point of origin – the need for executive talent to address organizational goals and strategies – and how the program, in the middle column, is seen as contributing to improved manager and organizational performance.

Core program elements include 360-degree feedback, individual development plans, workshop sessions conducted by both professional trainers and the directors of the participating agencies, individual journaling, coaching, a book club, ongoing support of
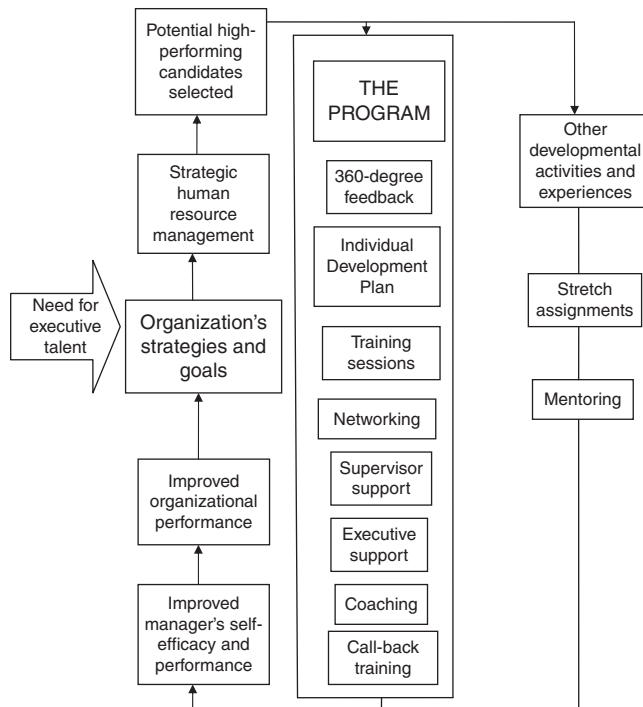
Figure 1.
Leadership development theory of change

agency directors, action learning projects, and intensive opportunities for dialogue and networking. Also included are annual follow up sessions to review progress on action plans and "call back" trainings offered to program alumni to meet for a day of additional training based on emerging needs. The first class, with 24 participants, began in 2005. The program has now completed its ninth year. Additional detail on the program is available in Packard *et al.* (2005).

This program includes twelve-and-a-half days of training in five blocks, delivered over a period of five months. Prior to the sessions, each participant completes a 360-degree feedback process and an individual development plan. The 360-degree feedback process is done on line using an instrument for executive managers (Envisia Learning, n.d.), with the training program coordinator conducting individual debriefing sessions to review results. The instrument chosen covered nearly all of the key competencies identified by the directors as described above.

To enhance the transfer of learning to the workplace and build a climate of organizational learning, participants' supervisors are involved before, during, and after the program; and department directors meet with their participants before the program begins and at its completion to discuss action plans for their own agencies. Progress on these action plans is reviewed by directors and participants after program completion.

After participants are selected by their directors, they attend an orientation session prior to beginning the training. This includes a review of the curriculum and a discussion regarding participant and director expectations.

Training sessions currently include the following topics:

- leadership;
- essential critical thinking skills and processes for executives;
- coaching for results;
- organizational change;
- strategic management;
- presentation skills;
- fiscal essentials;
- managing accountability/knowledge management;
- media relations/interview skills;
- multi-generations in the workplace;
- political savvy (including a mock session of a meeting of elected officials: the county board of supervisors); and
- executive critical thinking/writing.

To evaluate this program, an evaluation methodology was designed based on Kirkpatrick's (1996) four-level evaluation schema (participant satisfaction, enhanced skills and knowledge, changes in on-the-job behavior, and changes in organizational outcomes).

Results at Levels 1-3 were addressed in Coloma *et al.* (2012). This paper focusses primarily on Level 4: outcomes.

Data were gathered in the same format and on the same schedule each year. For individual cohorts, data were gathered at several times: before the program; at program

completion; and three, six, and 12 months after program completion. This time series design was used to better assess the impacts of the program than would be a more common post-test only design.

As noted by Solansky (2010) and others, evaluation data based only on self-reports may be subject to bias. To address this limitation, participant reports of performance changes were assessed for corroboration by their supervisors, who were asked for their observations regarding changes in their supervisees' performance.

Changes in job performance since program completion were measured by self-reports and supervisor assessments at the three, six, and 12-month follow-ups in terms of increased responsibilities, improved quality or quantity of work, improved performance as a leader, and better performance as a manager.

Self-efficacy was measured using the ten-item Generalized Self-Efficacy Scale (Schwarzer and Jerusalem, 1995). This scale asked participants to rate themselves on a five-point Likert scale, where 5 is the highest rating, and 1 is the lowest score, on ten different competencies, including items such as "I can always manage to solve difficult problems if I try hard enough" and "I can usually handle whatever comes my way." Items were summed to provide a score. Increases in scores over time were taken as evidence of the effectiveness of the training in increasing the ability of workers to better perform their job. This scale had good reliability ($\alpha = 0.786$).

Both quantitative and qualitative data were gathered for this evaluation. Changes were tested with the Paired t-test. Qualitative data were gathered by asking participants and their supervisors to provide up to two examples for each of the areas of possible program impact (e.g. improved quality or quantity of work, performance as a leader or manager).

This design contains several limitations. There were different respondents for each year; and over the years, there have been occasional minor changes in module content (noted above) in response to evaluation findings and to ensure that content remained relevant.

Trainers remained constant for most modules over the years. One exception is that directors of the participating agencies who presented on some topics, such as leadership, organizational change, and political savvy, varied over the years, largely based on their schedule availability.

Conditions in the agencies may have changed over the seven years; and other historical factors could have affected these scores. Therefore, it is not possible to say definitively that the changes described below are due to any exact extent to the program. Partly for this reason, a multiple methods evaluation process was used. Triangulating several evaluation methods does help to strengthen the overall conclusions.

As is common in many studies, this study relied upon self-report data, and therefore responses might not always reflect reality. This limitation is partially addressed by using supervisor's responses to corroborate workers' reports.

A control group and random assignment was not used, which limits the ability to establish cause and effect. Because of the nature of a leadership development program such as this, control or comparison groups could not be used. The use of a comparison group that had not undergone the training might have helped to more precisely determine the effect of training, but this was not possible because the program was designed to select for participation high performers with future potential.

Attrition was high at the three and six month measures (over 50 percent), which raises the issue of how completers might have been different from non-completers.

Perhaps completers perceived more benefit from the training, and thus were more likely to complete follow-ups. If so, then results would skew positive.

On the other hand, this study had a number of strengths, including sufficient sample size for analysis, multiple data collection points, and collection of data from the both participant and supervisor.

Results

Table I provides a description of the sample at the 12-month follow up (age was collected at pre-test only). Over the first seven annual iterations of the program, a total of 191 participants completed the program, for an average of 27 participants per annual program. In Year 1, directors nominated for the program their managers at or near the

	Frequency	Percentage
Gender (n = 124)		
Male	53	43
Female	71	57
Age (n = 123)		
30-39	20	16
40-49	54	44
50-59	44	36
60+	5	4
Ethnicity (n = 125)		
African-American	25	20
Asian Pacific Islander	12	10
Hispanic	26	21
White	62	50
Years of management experience (n = 122)		
0-5	24	20
6-10	41	34
11-20	45	37
21+	12	10
Number of staff responsible for (n = 99)		
0-50	45	45
51-100	18	18
101-250	19	19
250+	17	17
Education (n = 124)		
Less than a BA	17	14
BA	36	29
Some graduate work	9	7
Master's	60	48
Other	2	2
Field of service (n = 152)[a]		
Administrative	25	16
Aging and adult	11	7
Welfare to work	23	15
Child welfare	45	30
Indigent services	11	7
Medicaid	15	10
Other	22	14

Notes: *n*'s vary because of missing data. [a]Several respondents were responsible for more than one field of service; differences from 100 percent are due to rounding errors

Table I.
Demographic description of program participants at 12-month follow up (age at pre-test)

top levels of management; and in subsequent years, they have nominated a larger proportion of managers from lower levels. Nevertheless, the average years of management experience of each cohort has remained stable, clustering around ten years. There have not been notable trends in variations regarding the other demographic characteristics of participants over the years.

Of the 191 participants, 147 completed the post-test (77 percent). Attrition reduced the number of participants with both a pre-test score and a three and six month score to 85 (44 percent). In all, 97 participants completed both a pre-test and a 12-month follow-up (66 percent). Table II provides a summary of findings regarding perceived changes in self-efficacy and participant performance, and the effects of the program.

The Self-Efficacy Scale was used to reflect participants' knowledge/skill acquisition. Participants were asked to self-assess their competency regarding a topic before and after each module using a five-point Likert scale, where 5 is the highest rating. As shown in Table II, respondents reported significant gains from the pre-test measure to the post-test, suggesting that they felt better able to carry out each competency after training. These gains were sustained at each follow-up.

Performance outcomes were assessed by two impact measures, reported in Table II. One measure looked for actual changes on the job which were reported separately by the participant and her or his supervisor (impact: participant view and impact: supervisor view). Measures of performance by supervisors are not common in leadership development evaluations, but were gathered here by asking supervisors for their observations regarding changes in their supervisees' performance. This approach addresses the criticism of self-reports that they may be providing data about self-perceptions and not actual behavior.

The second measure (Effect of Program), which was completed by the participant only, assessed the degree to which they thought the change could be attributed to the training on a four-point scale, from "to a great extent" to "not at all." Table III is summary of individual changes (impacts) up to 12 months after training, and includes both worker and supervisor responses.

Impact: participant view, a count of the changes in participants' responsibilities (see Table III for specific items), shows an increase in responsibilities as reported by both worker and supervisor, but this change only approaches significance. Effect of

Scales	Pre-test M (SD)	Post-test M (SD)	3-month M (SD)	6-month M (SD)	12-month M (SD)
Self-efficacy	21.91 (3.18)	22.78 (3.90)**	22.14 (3.10)**	23.09 (3.85)***	23.16 (3.28)***
Impact: participant view			$n = 64$ 2.64 (1.21)	$n = 64$ 2.67 (1.22)	$n = 46$ 2.83 (1.06)****
Effect of the program			$n = 67$ 7.88 (1.40)	$n = 50$ 7.67 (1.41)*****	$n = 50$ 7.50 (1.34)*
Impact: supervisor view			$n = 56$ 3.02 (1.27)	$n = 56$ 3.27 (1.29)	$n = 47$ 3.29 (3.30)*****

Notes: Pre-test was given prior to the start of training. The post-test was given at the conclusion of training. The follow-ups were given after training at the intervals specified in the table. Supervisors were given measures only at the follow-ups. Paired t-tests used to test significance between pre-test or first measure with subsequent measures: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; ****$p < 0.10$; *****$p < 0.15$

Table II.
Performance on individual scales

Change area	Worker						Supervisor					
	3-month		6-month		12-month		3-month		6-month		12-month	
	n	%	n	%	n	%	n	%	n	%	n	%
Increase in responsibilities	87	39.1	100	51	112	58.3	91	58.2	84	76.2	110	74.5
Quality or quantity of work improved	86	64.0	98	64.3	109	67.9	91	79.2	84	84.5	109	86.2
Performance as a leader improved	86	86.0	98	87.8	107	83.2	89	83.1	66	84.8	109	89.9
Performance as a manager improved	84	76.2	82	42.0	97	70.1	90	80.0	83	85.5	92	87.0

Table III.
Individual impacts:
n of respondents;
percent reporting
changes

Program suggests that respondents thought the training program was a factor in the reported job changes, but this declined significantly over time.

Table III shows a similar pattern as do the scales. Supervisors reported more changes than did workers, seeing changes in all four areas. Participants see change in two areas: receiving more responsibility after the training (39 percent at three months to 58 percent at 12 months) and improved quality or quantity of work (64-67.9 percent). For each area in which participants noted changes in their performance, they were asked to provide up to two examples. Following are some of the examples mentioned:

- I was given an assignment involving workload reductions while maintaining current staffing. My staff and I created workload reduction scenarios and some have been implemented.

- I restructured the delivery of information to our regions. I developed a new strategy in collaboration with the regional managers, initiated the change, and conducted a follow-up meeting to refine the process.

- I am more attentive to looking at all the facts and trying to visualize a bigger picture in all areas of my assignment.

- I feel more confident in my basis for making "political" type decisions and understanding the impact of those decisions on my group as well as other stakeholders.

- I am better able to delegate and coach my staff more effectively in order to spend more time on higher level strategic issues.

- I am more effective at identifying skill deficiencies in staff and developing assignment opportunities that challenge/stretch those staff to improve those skills.

Similar to the participant survey, the supervisor survey asked respondents to list up to two examples of changes in the performance of their subordinates since program participation. Some examples are below. In some cases, supervisors used the same examples that their subordinates did, even though they did not know what examples their subordinates had provided. In some cases, supervisors mentioned promotions as examples of performance improvements:

- Smooth implementation of a new program. Able to manage a very diverse staff operating a variety of benefits and service programs.

- Breadth of her responsibility has increased, and she is more empowered to make decisions. Able to focus staff on meeting performance standards and expectations. Holding staff accountable. Increased political awareness.

- Sees the bigger global picture for the Agency and the importance of planning and measuring outcomes; has facilitated business process improvement meetings.

- Put in charge of restructuring the entire _____ Division.

- More confident, has a more global view of setting vision for her program. Looking at issues more globally has provided more clarity as to how her program fits into the overall Agency.s

- Able to adjust her leadership style to the different managers that report to her; open to learning from them and offers them various opportunities for leadership development.

Finally, ratings of self-efficacy were correlated with perceived changes in work performance (see Table IV). This was true at both the three- and 12-month follow-ups.

Conclusions and implications

This evaluation of a leadership development program over a period of seven years for a diverse group of human service managers showed significant improvements in their job performance and feelings of self-efficacy.

Participants reported positive changes in their performance over time. However, the extent to which they assign as training as a factor in these changes declined at later measures. It may be that a worker's reported changes were over time seen as a result of other factors such as maturation in their position, in which the changes would have occurred as a normal part of the job without the training.

On the other hand, the mean scale score suggests that workers thought the training contributed at least moderately to the positive changes they experienced on the job. For each of these changes, the mean participant rating of the extent to which the program was a factor was between "to a moderate extent" and "to a great extent." Supervisors reported more positive changes in participant performance than their subordinates did.

It is particularly notable that changes in self-efficacy were associated with the extent to which participants saw changes in the four aspects of work performance (increased responsibilities, increased quality or quantity of work, improved performance as a leader, and improved performance as a manager), and also the extent to which the participants saw the program as a factor in these changes. This finding supports earlier research (e.g. Hannah *et al.*, 2008 and others cited above), with further detail that participants noted the extent to which they believed these changes were due to the program.

These positive findings suggest that the program content and the evaluation methods used here are relevant for applications elsewhere. Program elements used here which have also been shown in other studies to have impacts include off-site training, 360-degree feedback, individual development plans, coaching, and networking. The use of agency directors as trainers is not commonly reported in the literature (an exception is the program described by Austin *et al.*, 2006).

The ongoing involvement of the agency directors has included their hands-on role in initial planning, their involvement as trainers, their annual review of evaluation findings, and their continuing support through selecting promising staff for participation each year. This was seen as an essential factor for the program, suggesting that it should be useful in other leadership development initiatives.

This evaluation also highlights the challenges in gathering truly valid data regarding the precise impacts of a program on participant performance back on the job. Measurement of managerial performance, including the performance of programs within managers' areas of responsibility, are still not at an advanced stage; but this

	Self-efficacy
Effect of the program at 3 months	0.422**
Effect of the program at 12 months	0.351**
Impact: participant view at 3 months	0.338**
Impact: participant view at 12 months	0.216*
Notes: *$p < 0.05$; **$p < 0.01$	

Table IV.
Correlations:
self-efficacy and
program impact

study's inclusion of supervisor ratings of participants was seen as a valuable step forward in evaluating program impact.

The greatest advance in the evaluation of such programs would be the development of methods to gather more powerful follow up outcome data, ideally using other relevant unit or program performance measures; and using more effective methods of connecting such improvements to program activities. As noted in the literature review above, organizational outcomes are rarely examined in leadership development evaluation. Measures of pre-post performance of the programs overseen by participants would be powerful data, but the methodological and logistical challenges with this are substantial. Nevertheless, future research should be able to refine and strengthen the follow-up measurement processes here, providing greater validity to evaluations of programs in the future.

Because evaluation of such programs is not at an advanced stage, this study adds to the knowledge base regarding viable leadership development processes and evaluation methods, and suggests opportunities and challenges regarding improving such programs and their evaluations.

References

American Society for Training and Development (2008), "Executive development: strategic and tactical approaches executive summary", available at: www.astd.org/NR/rdonlyres/A684D741-AD27-4268-BDE1-1C8E35CF34B8/0/ExecDevel_ExecSumm.pdf (accessed September 20, 2010).

Anderson, D., Krajewski, H., Goffin, R. and Jackson, D. (2008), "A leadership self-efficacy taxonomy and its relation to effective leadership", *The Leadership Quarterly*, Vol. 19 No. 5, pp. 595-608.

Atwood, M., Mora, J. and Kaplan, A. (2010), "Learning to lead: evaluating leadership and organizational learning", *Leadership & Organization Development Journal*, Vol. 31 No. 7, pp. 576-595.

Aude, S., Keller-Glaze, H., Riley, R. and Fallsen, J. (2007), "Center for army leadership technical report: commander's handbook for unit leader development", available at: www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA492849andLocation=U2anddoc=GetTRDoc.pdf (accessed September 10, 2010).

Austin, M., Regan, K., Samples, M., Schwartz, S. and Carnochan, S. (2011), "Building managerial and organizational capacity in nonprofit human service organizations through a leadership development program", *Administration in Social Work*, Vol. 35 No. 3, pp. 258-281.

Austin, M., Weisner, S., Schrandt, E., Glezos-Bell, S. and Murtaza, N. (2006), "Exploring the transfer of learning from an executive development program for human services managers", *Administration in Social Work*, Vol. 30 No. 2, pp. 71-90.

Avolio, B.J., Hannah, S., Reichard, R., Chan, A. and Walumbwa, F. (2009), "A meta-analytic review of leadership impact research: experimental and quasi-experimental studies", *The Leadership Quarterly*, Vol. 20, pp. 764-784.

Blume, B., Ford, J., Baldwin, T. and Huang, J. (2010), "Transfer of training: a meta-analytic review", *Journal of Management*, Vol. 36 No. 4, pp. 1065-1105.

Burke, M. and Day, R. (1986), "A cumulative study of the effectiveness of managerial training", *Journal of Applied Psychology*, Vol. 71 No. 2, pp. 232-45.

Clarke, N. (2012), "Evaluating leadership training and development: a levels-of-analysis perspective", *Human Resource Development Quarterly*, Vol. 23 No. 4, pp. 441-460.

Collins, D. and Holton, E. (2004), "The effectiveness of managerial leadership development programs: a meta-analysis of studies from 1982 to 2001", *Human Resource Development Quarterly*, Vol. 15 No. 2, pp. 217-248.

Coloma, J., Gibson, C. and Packard, T. (2012), "Participant outcomes of a leadership development initiative in eight human service organizations", *Administration in Social Work*, Vol. 36 No. 1, pp. 4-22.

Day, D. (2001), "Leadership development: a review in context", *The Leadership Quarterly*, Vol. 11 No. 4, pp. 581-613.

Ely, K., Boyce, L.A., Nelson, J.K., Zaccaro, S.J., Hernez-Broome, G. and Whyman, W. (2010), "Evaluating leadership coaching: a review and integrated framework", *The Leadership Quarterly*, Vol. 21 No. 4, pp. 585-599.

Envisia Learning (n.d.), "Executive view 360 degree feedback assessment for executives", available at: www.envisialearning.com/360_degree_feedback/executive_view (accessed June 12, 2013).

Gentry, W.A. and Martineau, J.W. (2010), "Hierarchical linear modeling as an example for measuring change over time in a leadership development evaluation context", *The Leadership Quarterly*, Vol. 21 No. 4, pp. 645-656.

Grove, J., Kibel, B. and Haas, T. (2007), " 'EvaluLEAD: an open-systems perspective one valuating' leadership development", in Hannum, K., Martineau, J. and Reinelt, C. (Eds), *The Handbook of Leadership Development Evaluation*, Jossey-Bass, San Francisco, CA, pp. 71-110.

Gutierrez, M. and Tasse, T. (2007), "Leading with theory: using a theory of change approach for leadership development evaluations", in Hannum, K., Martineau, J. and Reinelt, C. (Eds), *The Handbook of Leadership Development Evaluation*, Jossey-Bass, San Francisco, CA, pp. 48-70.

Hannah, S.T., Avolio, B.J., Luthans, F. and Harms, P.D. (2008), "Leadership efficacy: review and future directions", *The Leadership Quarterly*, Vol. 19 No. 6, pp. 669-692.

Hannum, K. and Craig, S.B. (Eds) (2010), *Introduction to Special Issue on Leadership Development Evaluation*, Vol. 21 No. 4, pp. 581-686.

Hannum, K., Martineau, J. and Reinelt, C. (Eds) (2007), *The Handbook of Leadership Development Evaluation*, Jossey-Bass, San Francisco, CA.

Hernez-Broome, G. and Hughes, R. (2004), "Leadership development: past, resent, and future", *Human Resource Planning*, Vol. 27 No. 1, pp. 24-32.

Holden, G. (1992), "The relationship of self-efficacy appraisals to subsequent health related outcomes", *Social Work in Health Care*, Vol. 16 No. 1, pp. 53-93.

Houghton, J. and DiLiello, T. (2010), "Leadership development: the key to unlocking individual creativity in organizations", *Leadership and Organization Development Journal*, Vol. 31 No. 3, pp. 230-245.

Kirkpatrick, D. (1996), *Evaluating Training Programs: The Four Levels*, 2nd ed., Berrett-Koehler, San Francisco, CA.

Leskiw, S. and Singh, P. (2007), "Leadership development: learning from best practices", *Leadership and Organization Development Journal*, Vol. 28 No. 5, pp. 444-464.

McCauley, C. (2008), "Leader development: a review of research", available at: www.shrm.org/about/foundation/research/Documents/McCauley-%20Leader%20Dev%20Lit%20Review.doc (accessed June 12, 2013).

Martineau, J. and Patterson, T. (2010), "Evaluating leader development", in Van Velsor, E., McCauley, C. and Ruderman, M. (Eds), *The Center for Creative Leadership Handbook of Leadership Development*, 3rd ed., Jossey-Bass, San Francisco, CA, pp. 251-281.

Menefee, D. (2000), "What human service workers do and why they do it", in Patti, R. (Ed.), *The Handbook of Social Welfare Management*, Sage Publications, Thousand Oaks, CA. pp. 247-266.

Multon, K.D., Brown, S.D. and Lent, R.W. (1991), "Relation of self-efficacy beliefs to academic outcomes: a meta-analytic investigation", *Journal of Counseling Psychology*, Vol. 38 No. 1, pp. 30-38.

Nicolaidouab, M. and Petridouabc, A. (2011), "Evaluation of CPD programmes: challenges and implications for leader and leadership development", *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, Vol. 22 No. 1, pp. 51-85.

Packard, T., Tucker-Tatlow, J., Waechter, J., Rahiser, P. and Dudley, D. (2005), "University-agency collaboration to design, implement, and evaluate a leadership development system", *Professional Development: The International Journal of Continuing Social Work Education*, Vol. 8 Nos 2/3, pp. 98-107.

Paglis, L. and Green, S. (2002), "Leadership self-efficacy and managers' motivation for leading change", *Journal of Organizational Behavior*, Vol. 23 No. 2, pp. 215-235.

Prussia, G.E., Anderson, J.S. and Manz, C.C. (1998), "Self-leadership and performance outcomes: the mediating influence of self-efficacy", *Journal of Organizational Behavior*, Vol. 19 No. 5, pp. 523-538.

Riggio, R. (2008), "Leadership development: the current state and future expectations", *Consulting Psychology Journal: Practice and Research*, Vol. 60 No. 4, pp. 383-392.

Russon, C. and Reinelt, C. (2004), "The results of an evaluation scan of 55 leadership development programs", *Journal of Leadership and Organizational Studies*, Vol. 10 No. 3, pp. 104-107.

Schwarzer, R. and Jerusalem, M. (1995), "Generalized self-efficacy scale", in Weinman, J., Wright, S. and Johnston, M. (Eds), *Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs*, NFER-NELSON, Windsor, pp. 35-37.

Solansky, S.T. (2010), "The evaluation of two key leadership development program components: leadership skills assessment and leadership mentoring", *The Leadership Quarterly*, Vol. 21 No. 4, pp. 675-681.

Stajkovic, A. and Luthans, F. (1998), "Self-efficacy and work-related performance: a meta-analysis", *Psychological Bulletin*, Vol. 124 No. 2, pp. 240-261.

Van Velsor, E., McCauley, C. and Ruderman, M. (Eds) (2010), *The Center for Creative Leadership Handbook of Leadership Development*, 3rd ed., Jossey-Bass, San Francisco, CA.

Watkins, K., Hybersten, I. and deMarris, K. (2011), "Evaluating executive leadership programs: a theory of change approach", *Advances in Developing Human Resources*, Vol. 13 No. 2, pp. 208-239.

Wimpfheimer, S. (2004), "Leadership and management competencies defined by practicing social work managers: an overview of standards developed by the national network for social work managers", *Administration in Social Work*, Vol. 28 No. 1, pp. 45-56.

Corresponding author

Dr Thomas Packard can be contacted at: tpackard@mail.sdsu.edu