



PERGAMON

Evaluation and Program Planning 26 (2003) 355–366

EVALUATION
and PROGRAM PLANNING

www.elsevier.com/locate/evalprogplan

Theory-driven evaluation in action: lessons from a \$20 million statewide Work and Health Initiative

Stewart I. Donaldson^{*,1}, Laura E. Gooler

Department of Psychology, School of Behavioral and Organizational Sciences, Claremont Graduate University, 123 E. Eighth Street, Claremont, CA 91711, USA

Abstract

Despite the burgeoning literature explicating the benefits of theory-driven program development and evaluation, there remains a strong need for practical advice, written insights and experiences, and examples from evaluation practice illustrating how to implement this approach. The purpose of this paper is to move the field closer to a concrete understanding of the strengths, limitations, and challenges of implementing theory-driven program development and evaluation in modern human service organizations. This is accomplished by describing the evaluation process, resulting program theories, and lessons learned from the evaluation a five year, \$20 million statewide Work and Health Initiative.

Published by Elsevier Ltd.

Keywords: Theory-driven evaluation; Program evaluation; Applied psychology; Program development; Organizational development; Work and health

1. Introduction

Theory-driven evaluation is a relatively recent theory of evaluation practice that attempts to build upon knowledge acquired from the practice of program evaluation over the past three decades (Chen, 1990; Chen & Rossi, 1983; Donaldson, 2001a, 2003a). Shadish, Cook, and Leviton (1991) referred to theory-driven evaluation as one of the most advanced forms of evaluation theory—Stage III Evaluation Theory. They carefully described how it integrates and synthesizes previous approaches to develop a contingency approach to evaluation practice, acknowledging that some evaluation approaches and methods work well under some circumstances but fail under others. This approach to evaluation practice has become quite popular in recent years (cf. Donaldson & Scriven, 2003), and now provides a foundation for some of the most widely used textbooks on program evaluation (e.g. Rossi, Freeman, & Lipsey, 1999; Weiss, 1998).

One of the central tasks of the theory-driven evaluator is to fully understand the nature of the program, the true purpose and context of the evaluation, in an effort to design the most rigorous and sensitive evaluation possible within practical constraints. This is accomplished by developing

program theory that is used to identify and prioritize the important evaluation questions, and to tailor the evaluation methodology to answer those particular questions (see Donaldson, 2001a, 2003a). This approach is method neutral and creates a superordinate goal that helps evaluators get past old debates about which methods are superior in evaluation (e.g. the quantitative/qualitative debate, Mark, 2003; Reichhardt & Rallis, 1994). That is, from the contingency point of view, the theory-driven approach argues that quantitative, qualitative, or mixed methods designs are neither superior nor applicable in every evaluation situation (see Chen, 1997; Fitzpatrick, 2002). Instead, methodological choices are informed by program theory, specific evaluation questions ranked in order of priority, and practical constraints (Donaldson, 2003a).

Despite the burgeoning literature on the benefits of theory-driven evaluation, practicing evaluators must still search far and wide to find practical advice about the nuances of implementing this approach. That is, much of the literature on theory-driven program evaluation is written at a stratospheric level of abstraction (Weiss, 1997). Therefore, there appears to be a strong need for practical advice, written insights and experiences, and examples from evaluators who are actually conducting theory-driven program evaluations. Further, practical experiences from the trenches are needed to refine some of the initial theorizing about how best to conduct theory-driven

* Corresponding author. Tel.: +1-909-621-8084; fax: +1-909-621-8905.
E-mail address: stewart.donaldson@cgu.edu (S.I. Donaldson).

¹ www.cgu.edu/sbos/.

evaluations across various settings and program areas. The purpose of this paper is to begin to meet some of those needs. This will be accomplished by describing the evaluation process, resulting program theories, and lessons learned from a multi-program theory-driven evaluation. Because this paper is focused on practical issues related to conducting theory-driven program evaluation, we do not emphasize or report specific evaluation findings of the individual programs discussed. However, these evaluation findings are available elsewhere (Donaldson & Gooler, 2002a,b; Fitzpatrick, 2002).

2. Theory-driven evaluation of a Work and Health Initiative

Lessons learned from using a theory-driven program development and evaluation process with 4 programs under The California Wellness Foundation's \$20 million statewide 'Work and Health Initiative' will be described. Particular attention will be devoted to discussing practical issues involved with using a theory-driven evaluation approach for facilitating continuous program improvement, building program and evaluation capacity, managing stakeholder evaluation anxiety, and providing summative evaluation.

2.1. The Work and Health Initiative

In 1995, The California Wellness Foundation launched a 5-year, \$20 million statewide Work and Health Initiative to promote the health and well-being of California workers and their families through work. The mission of the Work and Health Initiative was to improve the health of Californians by funding interventions that positively influence health through approaches related to employment. Fundamental to this Initiative was the perspective that important relationships between work and health are shaped by an evolving California economy. The goals of the Initiative were (1) to understand the rapidly changing nature of work and its effects on the health of Californians; (2) to increase access to high quality employment for all Californians; (3) to improve conditions of work for employed Californians; and (4) to expand the availability of worksite health programs and benefits. The means to achieving these goals were:

- To better *understand* the links between economic conditions, work and health and their effects on employees, employers, and the community.
- To *demonstrate* the relationship(s) between work and health for employees, potential employees, employers, and the larger community.
- To *influence* corporate and government policy makers to improve working conditions for individuals in California.

- To *establish* a self-sustaining network of researchers and practitioners to advance the work and health agenda.
- To *evaluate* the process and impact of all The California Wellness Foundation (TCWF) funded projects aimed at improving the work and health of Californians.

To accomplish these goals, TCWF funded four inter-related programs comprised of over 40 partner organizations working together to improve the well-being of Californians through approaches related to employment. *The Future of Work and Health (FWH)* and the *Health Insurance Policy Programs (HIPP)* were expansive and comprehensive research programs designed to generate and disseminate knowledge of how the nature of work is being transformed and how that change will affect the health and well-being of Californians. Current statewide trends related to health and health insurance within California were being examined through extensive survey research on an annual basis. In addition, researchers throughout California were systematically analyzing the changing nature of work and health, and searching for ways to improve working conditions and lower employment risks.

The Initiative also included two demonstration programs in 17 sites throughout the state to assist both youth and adults in building job skills and finding employment. The *Winning New Jobs (WNJ)* program aimed to help workers regain employment lost due to downsizing, reengineering, and other factors driving rather dramatic changes in the California workplace, and thereby put an end to the adverse health consequences that most workers experience as a result of unemployment. Finally, the *Computers In Our Future (CIOF)* program aimed to enable youth and young adults from low-income communities to learn computer skills to improve their education and employment opportunities—thereby improving their own future health as well as the health and well-being of their families and communities (for more details, see Donaldson & Gooler, 2002a,b; Donaldson, Gooler, & Weiss, 1998; Donaldson & Weiss, 1998; Fitzpatrick, 2002).

2.2. Evaluation approach

TCWF was also deeply committed to the science of promoting health and well-being through work (Donaldson et al., 1998). As part of this commitment, systematic evaluation research by an external evaluation team was commissioned to guide the strategic development and management of each program in the Initiative, as well inform the direction of the entire Initiative. The Initiative Evaluator served as an integrating, synthesizing force in evaluating goals, objectives, strategies, and outcomes central to the long-term impact of the Initiative. Cross-cutting goals and synergies were identified, enhanced, and evaluated in an effort to maximize the overall impact of

the Initiative (i.e. the whole was expected to be greater than the sum of the parts). In addition, the Initiative Evaluator developed evaluation systems that provided responsive evaluation data for each program. Those data were used to continually improve program effectiveness as well as evaluate long-term outcomes.

To ensure that the perspectives and problem solving needs of all those with a vested interest in the Initiative programs (e.g. TCWF, grantees, program administrators, staff, and program recipients), collectively known as stakeholders, were understood and addressed, the evaluation team adopted a participatory theory-driven evaluation approach (see [Chen, 1990](#); [Donaldson, 1995, 2001a, 2003a](#)). This approach rested on developing program theories for each program and using empirical feedback to guide program development. Each program theory was based upon the stakeholders' experience with how these types of programs seem to work, prior evaluation research findings, and more general theoretical and empirical work related to the phenomena under investigation. Such frameworks provided a guiding model around which evaluation designs were developed to specifically answer key evaluation questions as rigorously as possible given the practical constraints of the evaluation context. Given the high potential for environmental factors to confound the estimates of program effects, this 'conceptual mapping' approach helped identify and examine sources of variance and to isolate the effects of the factors that each program was attempting to impact. In addition, this approach increased understanding about how programs worked and when they worked.

Finally, data collection efforts were based on the premise that no single data source was bias-free or a completely accurate representation of reality. Evaluation plans were designed to specifically encourage each grantee to use multiple methodological strategies with different strengths and weaknesses to answer evaluation questions ([Chen, 1997](#); [Cook, 1985](#); [Donaldson, 2003a](#); [Shadish, 1993](#)). A special effort was made to understand cross-cultural and language concerns so that the methodologies employed were sensitive enough to detect program effects when they existed. In addition to evaluating program outcomes and impacts, evaluation efforts were both formative (i.e. aimed at developing and improving programs from an early stage) and process-oriented (i.e. geared toward understanding how a program achieves what it does over time) in nature. Evaluation efforts also identified cross-cutting goals and synergies within and across programs to facilitate success of the overall Initiative.

2.3. Program development and evaluation process

The evaluation design for each program was developed through a collaborative process that included the TCWF staff, project coordination teams, community-based organizations (CBOs), and other community stakeholders.

The evaluation team engaged stakeholders in a participatory process that involved constructing models of how their programs work, and then used those models to guide question formation, data gathering, and evaluation. To achieve this, the evaluation team facilitated numerous meetings and discussions of the program models and theories of change, evaluation design, data collection methods, feedback loops, and evaluation reports. Specific attention was given to identifying and measuring realistic outcomes and indicators that would result from specific program efforts. In addition, each evaluation was tailored to answer questions deemed of most importance by each project team. As such, each Initiative program evaluation was designed to include methods and techniques that can be used for understanding lessons learned and providing external feedback to enhance continuous quality improvement and long-term program success.

To support continuous program improvement throughout the life of the Initiative, the evaluation team created three primary vehicles for providing continuous program improvement feedback over time. These included: (1) mid-year evaluation reports, (2) year end evaluation reports, and (3) annual 360 degree feedback from grantees (cf. [Mersman & Donaldson, 2000](#)), that is, grantees were given the opportunity annually to evaluate the evaluation team and foundation staff. In addition, these were supplemented with several interim evaluation reports and frequent telephone and email communications designed to provide timely feedback throughout the year.

3. Resulting program theories

There is often confusion about the nature of program theory ([Donaldson, 2001a, 2003a](#); [Weiss, 1997](#)). For some, the term seems to conjure up images of broad social science theories rather than specific theories of treatments, programs, or interventions ([Donaldson, 2003a](#); [Lipsey, 1993](#)). The following definitions of program theory capture the essence of how we are using the term in this paper:

- The construction of a plausible and sensible model of how a program is suppose to work ([Bickman, 1987](#))
- A set of propositions regarding what goes on in the black box during the transformation of input to output, i.e. how a bad situation is transformed into a better one through treatment inputs ([Lipsey, 1993](#))
- The process through which program components are presumed to affect outcomes and the conditions under which these processes are believed to operate ([Donaldson, 2001a](#)).

It is highly desirable if program theory is rooted in, or at least consistent with, behavioral or social science theory or prior research (see [Donaldson, Street, Sussman, & Tobler, 2001](#)). However, often sound theory and research is not

available for the problem of concern. Other sources of information can also be used to develop program theory, including implicit theories held by those closest to the operation of the program, observations of the program in action, documentation of program operations, and exploratory research to test critical assumptions about the nature of the program (for discussions on how to use these sources, see Donaldson, 2001a; Rossi, Freeman, & Lipsey, 1999). The goal is to develop, in collaboration with key stakeholders, a parsimonious program theory (or competing theories to be tested) that captures the main factors that link a program with its presumed outcomes. This program theory can then be used to generate and prioritize evaluation questions. To illustrate outcomes of this process in practice, we present parsimonious versions of the program theories for each of the programs in the Work and Health Initiative below.

3.1. Winning New Jobs program theory

One of the demonstration programs was the Winning New Jobs (WNJ) program. The mission of WNJ was to provide job search training to 5000 unemployed and underemployed Californians over a 4-year funding period. The WNJ program is based on a theory-based intervention, JOBS, which was developed and initially tested via randomized trial in Michigan (Price, Friedland, Choi, & Caplan 1998; Vinokur, van Ryn, Gramlich, & Price, 1991). Using systematic organizational readiness assessments, three organizations in diverse California communities were selected to implement the WNJ program (Donaldson et al., 1998).

The core program theory that was developed using the process described above, and then used to guide

the evaluation of WNJ is shown in Fig. 1. Participants attended a one-week workshop designed to improve job search self-confidence, job search skills, and problem solving strategies including inoculation against setbacks. These skills and psychological factors are presumed to facilitate re-employment and improve mental health. Furthermore, the WNJ program is hypothesized to have impacts at multiple levels: participant (e.g. increased job search self-efficacy and re-employment), organization (e.g. staff skill development, reputation enhancement), community (e.g. increased access to job search services), and the policy environment (e.g. financial support for continuation of program).

This conceptualization of the WNJ program was used to develop and prioritize evaluation questions and to guide data collection. For example, extensive standardized eligibility, demographic, pretest, posttest, and employment follow-up data were collected at each site. Various types of qualitative implementation and outcome data were also collected. Further, databases tracking participants in other parts of the country and world were available for comparison purposes. This collection of databases was used for both formative and summative evaluation of the WNJ program.

3.2. Computers In Our Future program theory

The second demonstration project is called Computers In Our Future (CIOF). The CIOF project created 14 Community Computing Centers (CCCs) in eleven low-income California communities. The 14 CCCs were designed to demonstrate innovative, creative, and culturally sensitive strategies for using computer technology to meet

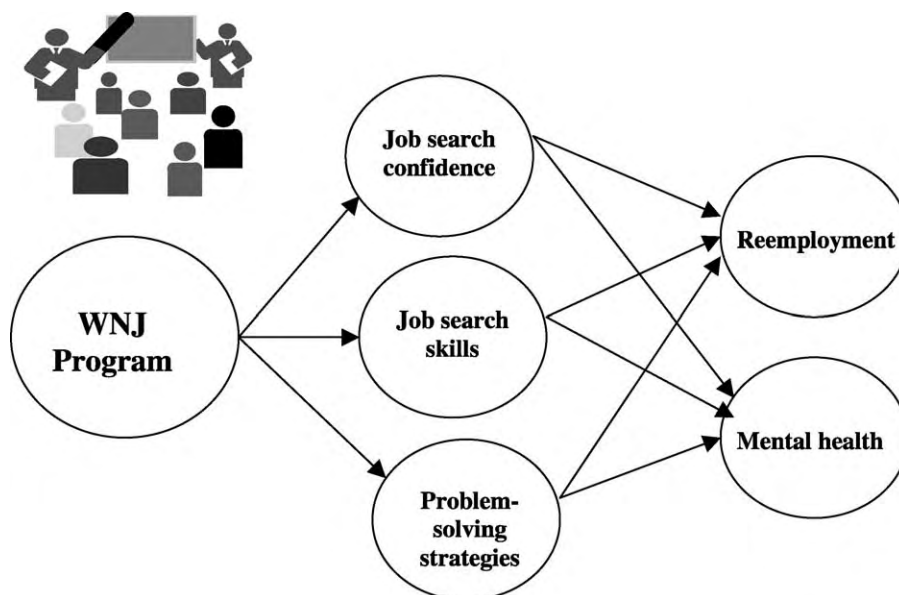


Fig. 1. Winning New Jobs Program Theory.

the economic, educational, and development needs of their local communities. The CIOF program explored and demonstrated ways in which CCCs can prepare youth and young adults ages 14–23 to use computers to improve their educational and employment opportunities, thereby improving the health and well-being of themselves, their families, and their communities.

Organizational readiness criteria were used to select 11 diverse California organizations to accomplish these goals. These organizations are diverse with respect to organizational type, geographical location, and populations served. Collectively, these centers were designed to provide access to more than 200 computer workstations statewide. With respect to open access service goals, each site collectively committed to providing unrestricted open access to 27,705 Californians (approximately 6900 individuals per year, statewide). Similarly, they committed to providing technology training and work experiences to over 4300 youth and young adults over the four year program period. With extensive input from site leaders, the funder, and the CIOF program coordination team, we constructed a guiding program theory of how the CIOF program is presumed to work.

Fig. 2 shows that participation in the CIOF program is believed to lead to improved attitudes toward computers, technology skills, career development knowledge, job search skills, and basic life skills. These acquired skills and knowledge are presumed to facilitate the pursuit of more education, internship opportunities, and better employment options, which in the long-term is expected

to improve participants' health status. This program theory has been used to identify a number of evaluation questions and the types of data needed to answer these questions and verify whether these relationships do indeed exist. Extensive standardized demographic, center utilization, pretest, posttest, and follow-up data were collected from each site. Various types of qualitative implementation and outcome data were also collected. Data were used for both formative and summative evaluation of the CIOF program.

3.3. Health Insurance Policy Program theory

The Health Insurance Policy Program (HIPP) was designed to support the development of state policy to increase access to health insurance for employees and their dependents that is not only affordable, but is also comprehensive and emphasizes the promotion of health and prevention of disease. To this end, the HIPP issued an annual report on the state of health insurance in California based on surveys of: the non-elderly population, HMOs, licensed health insurance carriers, purchasing groups and employers. In addition, HIPP team members developed policy briefs and related health insurance publications for broad dissemination to appropriate policy stakeholders.

As part of the evaluation planning process, the evaluation team facilitated the development of a program theory for HIPP. As shown in Fig. 3, the HIPP program seeks to increase target constituents' awareness and understanding of the status of health insurance issues in California, and influence policy development. The program theory shows

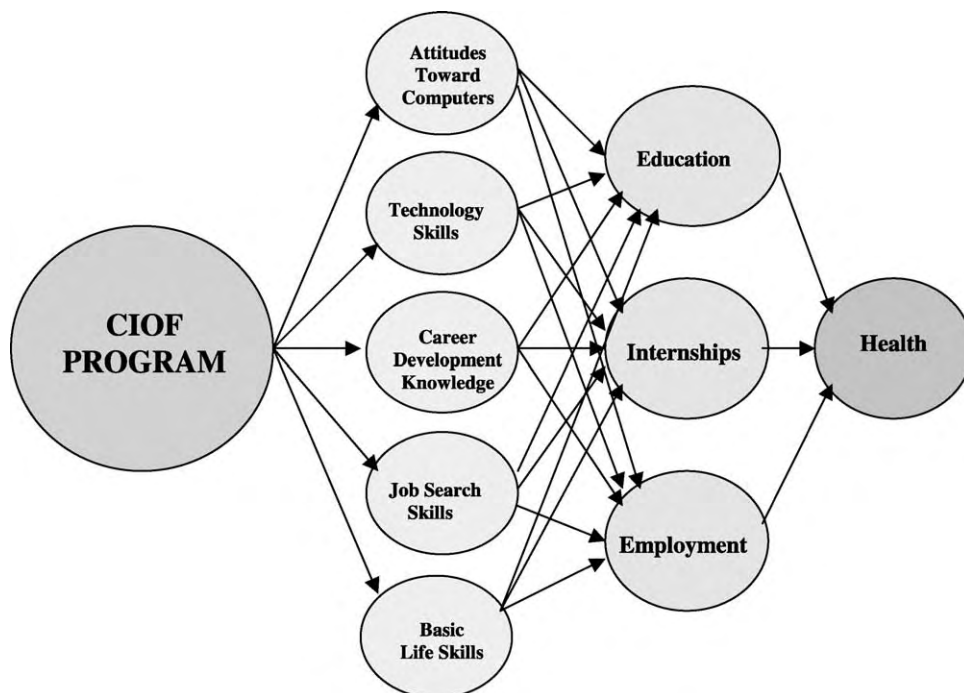


Fig. 2. Computers In Our Future Program Theory.

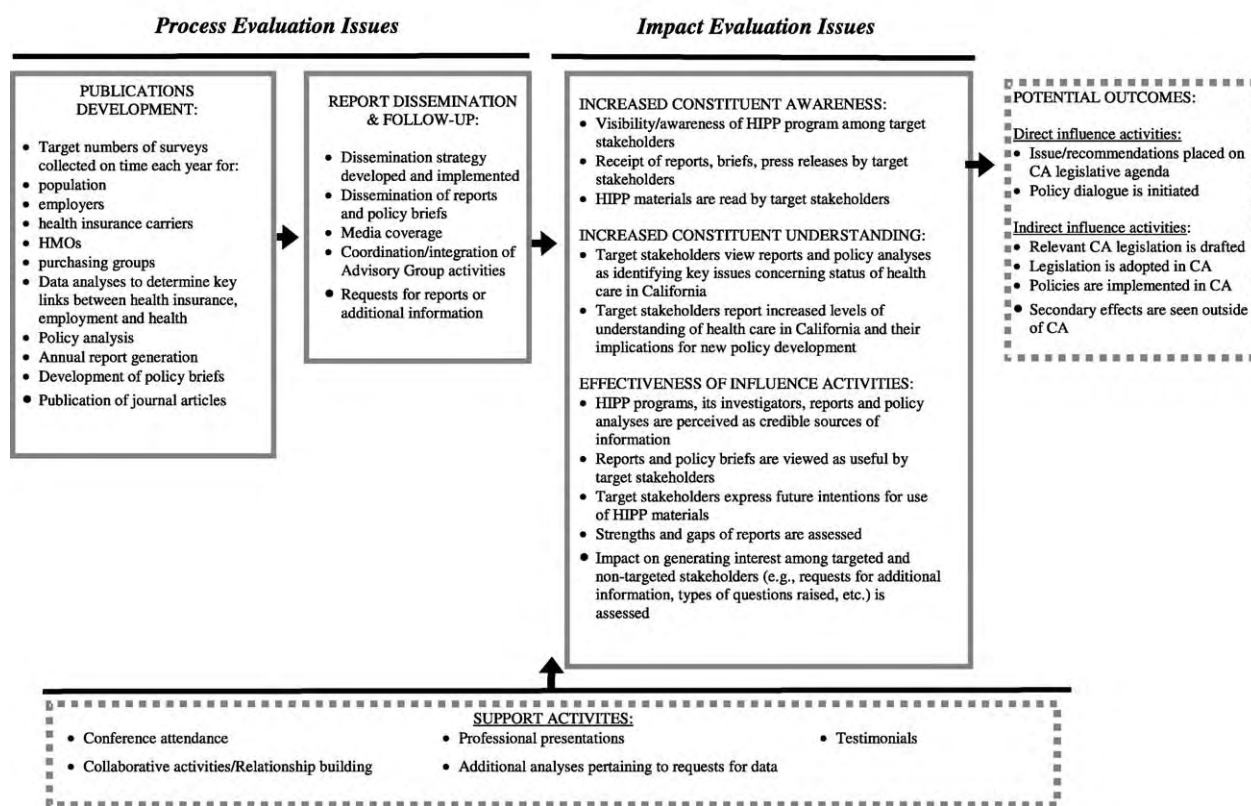


Fig. 3. Health Insurance Policy Program Theory.

that a range of publications development, report dissemination activities, and follow-up activities are being conducted in the effort to reach those desired outcomes. Support activities and potential outcomes are shown in dotted line boxes to indicate that these are expected to occur but are not required by the funding agency. This program theory was used to prioritize evaluation questions and collect a range of data to improve the program as well as examine challenges and identify successes.

3.4. Future of Work and Health program theory

The second research program is called the Future of Work and Health (FWH) program. The aim of the FWH program was to understand the rapidly changing nature of work and its effects on the health of Californians. This program was designed to support a range of research projects and statewide convenings consistent with this aim. After numerous discussions with program stakeholders, the program theory focused on overall program impacts rather than individual project outcomes. Various types of quantitative and qualitative data were collected to determine if the FWH program: (1) identified issues and trends important to the future of work and health of Californians; (2) developed a network of people involved in building

knowledge and improving practice to advance the future of work and health in California; (3) funded projects that illuminated trends important to the future of work and health of Californians; (4) identified policies that can influence work and health trends to improve the health of Californians; and (5) disseminated research findings on California work and health trends from FWH program activities.

4. Key lessons learned

This section is intended to capture some of the insights, experiences, and practical advice gleaned about the theory-driven approach to program planning and evaluation in the context of this initiative. While some of these lessons may be primarily relevant to conducting theory-driven evaluations (e.g. lessons pertaining to developing program theory), we acknowledge others may extend to evaluation practice more generally (e.g. strategies for managing evaluation anxiety, providing continuous improvement feedback, building program and evaluation capacity, and managing the formative to summative evaluation transition). However, our focus here is to present lessons learned from the evaluation of the Work and Health Initiative that

may be useful for refining some of the initial theorizing about how best to conduct theory-driven evaluations, as well as improving theory-driven evaluation practice.

4.1. Interactive, bi-directional program theory development was most effective

There seems to be a natural tendency to develop program theory moving from the left of a model to the right (i.e. as we were trained to read). This can be very limiting because most of the time with stakeholders is then spent talking about the program, and the etiology of outcomes and potential mediators and moderators are given short shrift. Another main consequence of the standard linear, ‘left to right approach’ is that stakeholders typically generate a long list of desired program outcomes, many of which appear unrealistic and only remotely connected to the program.

We have found that starting with desired outcomes is usually a more productive first step. Once we have identified a relatively short list of the outcomes the program is suppose to achieve, we begin the discussion of how this might occur. This leads us into the details of the how the program is expected to affect intermediate outcomes sometimes known as proximal outcomes or mediators. Once we develop a first draft of a program theory based on discussion with stakeholders, we examine the various links in light of prior theory, research and evaluation findings. This often leads us to suggest substantial revisions to the program theory or program itself because it is highly inconsistent with what is known about the topic. The point here is that this process seems to work best when evaluators use a highly interactive, bi-directional approach to developing program theory (cf. Donaldson, 2001a, 2003a; Fitzpatrick, 2002). This approach involves moving back and forth between stakeholder discussions and the literature several times, moving right to left and left to right and back again when thinking about the program theory and its various links between constructs, and carefully considering the characteristics of the participants, service providers, settings and the like that might affect (moderate) the relationships suggested by the program theory (see Donaldson, 2001a).

4.2. Reaching consensus was not difficult

One concern about developing program theory in a collaborative fashion is that it will be difficult, if not impossible, to reach consensus among stakeholders. It is common to find that stakeholders have very different views and conceptions of their program initially. However, our experience in this initiative, as well as in other evaluations, is that the interactive process described above usually leads to a high level of agreement and common understanding. If there are competing views of a program’s theory, these views can usually be tested in the evaluation. As is the case in basic science, competing theoretical perspectives makes

the empirical confirmation/disconfirmation (evaluation) process even more informative.

4.3. Complexity is the enemy: strive for parsimonious program theories

We have found that there seems to be a natural tendency for stakeholders to represent all possible program components, mediators, moderators, proximal and distal outcomes and the like in their program theories. However, overly complex program theories often confuse stakeholders and hinder discussions of the important evaluation questions that need to be answered (see Fitzpatrick, 2002). The interactive, bi-directional approach to program theory development described above strives to produce parsimonious program theories that can be used to develop a common understanding of the program amongst stakeholders, and to help identify and prioritize evaluation questions (Donaldson, 2001a, 2003a). While we believe it is important to acknowledge some programs and their effects may be very complex (see Donaldson, 2001a), at least a version of a program theory should focus on the main pathways between a program and its presumed outcomes. The resulting program theories presented in this paper represent the most parsimonious versions of each program theory used to describe the programs in the Work and Health Initiative. As can be observed, the specification of program theory tends to vary across different evaluation contexts. That is, the level of complexity and type of model/display vary based on program type and stakeholder preference. It should be noted that the evaluators often discuss the program theory in much more detail when deciding how to design an evaluation and collect data to answer key evaluation questions (Donaldson, 2003a).

4.4. Potential for substantial time and resources savings

A common criticism of theory-driven evaluation is that it is more time consuming and costly than other evaluation approaches (Scriven, 1998). We have found many situations where this is simply not true. For example, the process of developing program theory often reveals that a program is not ready for full-scale evaluation (i.e. evaluability assessment; Wholey, 2003). In this case, substantial time and resources are saved by redirecting efforts toward further program development and/or implementation activities, as opposed to summative evaluation certain to reveal null effects. Secondly, evaluation questions are prioritized in this approach, which helps stakeholders decide how to allocate evaluation resources in a cost-effective manner. We have found that developing program theory usually enables stakeholders and evaluators to make informed decisions about evaluation design and methods, often leading to cost effective evaluation (see Donaldson, 2003a).

4.5. Program designs are often improved prior to evaluation

The interactive, bi-directional approach to developing program theory often reveals that new programs are poorly conceptualized or at least need some fine-tuning. For example, stakeholders may have designed a program that is not consistent with what is known about changing the behaviors of concern. That is, prior research, well-supported theory, or previous evaluations may suggest that more program development is needed before empirical evaluation methods should be employed. In this case, program theory development may result in program improvement before any empirical evaluation work is conducted.

4.6. Program implementation improvement without evaluation

Once program theory is developed, sometimes it becomes apparent that a program is not being implemented well enough to affect key mediators or outcomes conceptualized in the program theory. For example, there may be inadequate resources or activities in place to affect target mediators. Again, before employing costly empirical evaluation methods, stakeholders can be given the opportunity to improve the implementation of the program, and thus prevent the very common occurrence of discovering null effects through summative evaluation.

4.7. Formative evaluation can lead to goal change instead of program improvement

Although certainly not unique to theory-driven evaluation, the evaluation of the Work and Health Initiative had a substantial formative evaluation or continuous improvement focus. Formative evaluation findings were used to help stakeholders understand implementation issues and the short-term effects of their programs.

A common expectation of emphasizing formative evaluation in the theory-driven evaluation approach is that empirical evidence or findings will be used to improve the delivery of the program. However, we also learned that these data can also be used effectively to adjust program goals and expectations. Once stakeholders see evaluation data, they sometimes decide that negative findings suggest initial goals or expectations are ‘unrealistic.’ Rather than continue to fail to meet unrealistic goals, they seek approval to make goals and objectives more reachable given time and resources constraints. As theory-driven evaluators, we view this as a healthy process under some conditions (e.g. when goals are indeed out of line), but it can also be used to as an excuse for poor performance. We believe evaluators should carefully document stakeholder reactions to formative evaluation data, and be an active voice in the discussions about appropriate responses.

4.8. Strategies for managing evaluation anxiety are often required

As we have described, the theory-driven evaluation of the Work and Health Initiative required regular interaction between the evaluation team and stakeholders. The fear of a negative evaluation on a regular basis tended to make some stakeholders very anxious at times (particularly those not confident about their performance). Excessive evaluation anxiety (XEA) can result in problems related to gaining access to critical information, lack of cooperation, compromises the quality of data collected, undermine the validity of findings, lack of data utilization and program improvement, and a general dissatisfaction with program evaluation and evaluators (Donaldson, Gooler, & Scriven, 2002).

One natural response to XEA by evaluators is to emphasize mere description in evaluation or avoiding to provide evaluative conclusions (Scriven, 2003). While the fear of a negative evaluation is greatly reduced and the evaluation team is likely more welcomed and popular with some stakeholders (e.g. program staff), we argue this is an inappropriate response by evaluators to XEA (Donaldson et al., 2002).

Instead, we have developed and implemented a number of strategies designed to manage evaluation anxiety which have been extremely valuable in the theory-driven evaluation of the Work and Health Initiative. We started with the strategy of training our team to expect and accept evaluation anxiety as a natural component of the interaction with stakeholders. Some of the other strategies we have used include (1) legitimating opposition to bad evaluation, (2) determining program psychologic—what stakeholders are hoping the program will do for them personally, (3) discussing purposes of evaluation in detail to avoid fake evaluations, (4) discussing professional standards for program evaluation, (5) discussing why honesty with the evaluator is not disloyalty to the group, (6) providing role clarification on an ongoing basis, and (7) creating opportunities to role model productive ways to respond to and use evaluation findings (e.g. we encouraged the stakeholders to evaluate the evaluation team formally on an annual basis; see Donaldson et al., 2002 for an in depth discussion of strategies for managing evaluation anxiety). Strategies for managing evaluation anxiety are often needed while working with stakeholders to develop program theory and when conducting rigorous, external formative and summative evaluation using the theory-driven approach.

4.9. Continuous improvement feedback: do not forget the positive findings

A central goal of the theory-driven evaluation of the Work and Health Initiative was to facilitate continuous program learning, problem solving, and program improvement throughout the life of the Initiative. The concept of

continuous improvement has been recognized by many organizations as a means to improve customer satisfaction, enhance product quality, lower operating costs, and to improve organizational effectiveness (Donaldson, 2003b). When evaluation is used to improve organizational performance, there are many requirements for it to be effective. Such an evaluation must provide feedback that can be used to (1) improve the design of projects and demonstrations, (2) develop the indicators and performance targets needed to improve effectiveness and responsiveness to challenges, (3) monitor how programs are being implemented for diverse populations across multiple locations, (4) identify and describe key accomplishments, challenges, and lessons learned, and (5) offer recommendations for program improvement.

Core challenges in the evaluation of this Initiative were how to produce evaluation reports and develop information systems that were useful to multiple stakeholders. Key issues for evaluators included meeting the diverse information needs of multiple stakeholders, focusing reports on high level issues and challenges, synthesizing tremendous amounts of data from multiple sources, providing diplomatic feedback, and keeping reports to a readable length. In community-based settings, another key challenge to using evaluative data for program improvement purposes concerns stakeholders' general lack of understanding of how evaluation fits in with their program goals and is useful for achieving other program goals (e.g. dissemination, policy influence). In addition, evaluation anxiety over perceived negative consequences that may result from improvement feedback can pose a tremendous barrier to utilization of findings by intended users.

A common mistake evaluators can make here is to focus exclusively on the areas that need improvement. This is a natural mistake because it is often how we define our job and describe the value we add to the social programming endeavor. Of course, the problem is stakeholders may react very negatively to the process if they do not believe their efforts and successes are fairly acknowledged. Table 1 summarizes some observations about delivering continuous improvement feedback.

4.10. Building program and evaluation capacity

There is a growing recognition of the need for capacity building in organizations that provide social service programs and interventions. In the United States, most organizations providing these types of services are value-driven organizations that are dependent upon external resources for operating and sustaining programs. Because of the complexity of social problems and issues they address, they often face a tremendous scarcity of resources, knowledge, and skills needed to develop and implement effective programs. Those managing and staffing these programs, in turn, have limited time and energy (and experience) to spend on developing and implementing

Table 1

Summary of observations about delivering continuous improvement feedback

-
- Establishing a safe environment that facilitates open communication and a willingness to share both positive and negative information is essential for facilitating utilization of continuous improvement feedback. To achieve this, the evaluator must establish trust, rapport, and credibility with program stakeholders, and oftentimes, engage in long-term relationship management
 - It is not easy overcoming negative perceptions about evaluation and expectations that evaluation plays a punitive role in community-based organizations. Achieving this requires careful listening to stakeholder concerns, values, and needs, and dedicating time for discussing and prioritizing the information needs of multiple stakeholders. Related to this, it is important that evaluators possess awareness and sensitivity to the cultural values of intended users of improvement-oriented feedback
 - Continuous improvement feedback is useful only to the extent that it is timely and meets the core information needs of program stakeholders. This entails providing frequent reports and feedback. Evaluators delivering this type of feedback should allocate sufficient staff and resources toward writing numerous evaluation reports. Because information needs will vary for different program stakeholders, and will change over time, it is important that feedback reports be tailored to different target audiences
 - It is helpful to revisit information needs and adjust data collection practices accordingly as programs evolve, grow, and learn from experience
 - In a continuous-improvement model, evaluators should not underestimate the important role of discussing the evaluation findings with program stakeholders. Evaluation reports cannot substitute for dialogue when it comes to understanding the complex issues facing human service organizations. Feedback reports are critical for examining facts about program achievements and challenges. However, it is only through ongoing dialogue about the meaning and implications of those facts that program stakeholders can identify lessons learned and use evaluation findings for problem solving and decision making
 - Balancing program improvement feedback with feedback pertaining to program accomplishments and strengths is essential for reducing program stakeholders' resistance to discussing and using evaluation findings. Failure to acknowledge the context in which the findings take place, including efforts to address program issues and key accomplishments will result in defensiveness toward findings and a lack of utilization of findings by intended users
-

evaluation practices that can be used to monitor and improve their organizations. Therefore, externally funded evaluations that support capacity-development efforts arguably represent one of the most valuable resources these organizations have for building program capacity.

Unfortunately, a lack of understanding and/or appreciation of evaluation pose key barriers to prioritizing and allocating precious resources toward evaluation. For example, it is not uncommon for organizational members to question the credibility of evaluators, balk at the costs, and highlight potential undesirable consequences using external evaluation (Donaldson, 2001b). This becomes particularly salient when the demands of evaluation compete with resources needed to meet program service goals.

One of the key dilemmas to building capacity in the community-based and human service organizations participating in the Work and Health Initiative was the loss of capacity due to relatively frequent turnover in program personnel. With respect to building program evaluation capacity, the consequences of personnel turnover include the loss of grantee knowledge and understanding of evaluation goals, terminology, and approach. In addition, turnover contributes to a loss of ownership and buy-in to the evaluation design and methodology among newcomers. In some cases, staff turnover may result in evaluation measures, tools and resources being taken by departing

individuals, leaving newcomers with little to no evidence of evaluation progress that was made. Similar losses are experienced with respect to programmatic knowledge and skills that have been learned by staff over time. Unfortunately, there remains a lack of clear guidelines and practices for building and sustaining capacity in human service organizations. We have presented some of the lessons we learned in this evaluation related to building program and evaluation capacity (Table 2).

4.11. Managing the formative to summative evaluation transition

Using evaluation results to improve a program is fundamentally different from rendering judgment about merit, worth, or significance for an external audience. In the formative phase of the theory-driven evaluation of the Work and Health Initiative, questions tended to be more open-ended and focused on identifying program strengths and weaknesses. In the summative phase, in contrast, we were focused on using explicit criteria of effectiveness in forming the basis for evaluative conclusions. A key challenge to these contrasting purposes of evaluation is that they often come into conflict because the information needed for program management and improvement is different from the data needed for accountability. These approaches, in turn, lead to different experiences for program managers and staff. Whereas conducting formative evaluation is experienced as developmental, conducting summative evaluation is sometimes perceived to be judgmental and punitive.

One of the main challenges we experienced in the current evaluation was making the transition from formative to summative evaluation. In short, for obvious reasons, some stakeholders seem to resist the reallocation of resources from program improvement to summative evaluation activities. We learned that it is critically important for stakeholders to be fully informed about the need for this transition, and they should know well in advance about when this fundamental change in the nature of the evaluation is going to occur.

5. Conclusion

It is now well documented that many well-intentioned social programs fail because they are poorly conceptualized or not properly implemented (Donaldson, 2003a). In addition, many program evaluations are often criticized because they are not sensitive enough to detect program effects when they do exist (Lipsey, 1988, 1990). The theory-driven approach to evaluation described in this paper was designed to deal with these major shortcomings of modern program design and evaluation.

Since the initial writings and conceptualization of this evaluation approach, there has been an explosion of

Table 2
Lessons learned related to building program and evaluation capacity

-
- Evaluators must expect and plan for staff turnover in community-based and human service programs. This entails initial and ongoing instruction and training in evaluation terms and procedures for new staff and program directors
 - Program stakeholders may not see the connection between program evaluation and the uses of findings for facilitating program improvements. To be effective, stakeholders must understand how the evaluation fits in with their overall program goals and how they will benefit from incorporating evaluation practice in their organization. Evaluators should expect that users will hold negative stereotypes about evaluation's punitive role and must work to clarify the many uses of evaluation for intended users
 - Evaluators should plan on developing clear and straightforward evaluation tools and resources for program stakeholders. Key resources may include evaluation plans, evaluation procedure manuals, glossary of evaluation terms, evaluation roadmaps, evaluation training, evaluation training manuals, and frequent contact and communications with program stakeholders
 - Evaluators should not assume that technical assistance providers and program management teams understand the role and practice of evaluation. Lack of understanding, commitment and support from these groups can derail evaluation efforts in human service organizations. To minimize role conflict and to prevent miscommunications about evaluation practice, evaluators need to build evaluation knowledge and skills in these groups as well. It is also critical that they buy-into the evaluation and support evaluation efforts
 - Evaluators must be strategic and sensitive in asking for time and involvement from busy people in evaluation efforts, and they must be sure they are interacting with the right people around relevant issues. Evaluation capacity cannot be increased in community-based and human service organizations without addressing the practical implications of conducting evaluations in these types of organizations. Data collection and tracking efforts that are overly cumbersome will not be supported
 - Evaluators will typically have to work to build and sustain evaluation capacity. Building effective relationships with intended users requires an understanding and sensitivity to the populations served, the sociocultural and economic contexts in which the program operates, and the values and needs of program constituents. It also requires ongoing relationship building and training in evaluation procedures
 - Evaluators need strong interpersonal skills in building relationships, facilitating groups, and managing conflict. They must communicate effectively and be keenly sensitive to human and organizational behavior issues
-

literature about its strengths and limitations (e.g. Birckmayer & Weiss, 2000; Crano, 2003; Donaldson, 2001a, 2003a; Fitzpatrick, 2002; Mark, 2003; Scriven, 2003; Weiss, 1997). However, it has been noted often in recent times that practical advice, explicit examples of developing and using program theory in evaluation, and written insights and experiences from the trenches of evaluation practice are sorely needed. This paper has addressed some of the practical issues related to the implementation of theory-driven evaluations in modern organizations and field settings, provided examples of program theories resulting from an interactive, bi-directional program theory development process, and has highlighted lessons learned from using this approach to evaluate a large Work and Health Initiative implemented in the Western United States. It our hope that this paper nudges the field closer to a concrete understanding of the strengths, limitations, and practical challenges of conducting theory-driven evaluations.

Acknowledgements

The authors would like to thank the editor and two reviewers for comments, which lead to notable improvements in this article. This work was supported by a grant from The California Wellness Foundation to Dr. Stewart I. Donaldson, Principal Investigator. The authors thank Work and Health Initiative Evaluation Team members (in alphabetical order) Judith Au, Dale E. Berger, Cindy K. Gilbert, Michelle D. Jones, Anna Malsch, Michael Scriven, and Rachel Weiss for their contributions toward the Initiative Evaluation. Special thanks to Gary Yates, Tom David, Ruth Brousseau and Lucia Corral Pena of The California Wellness Foundation; Sharon Rowser and Rick Price of the Wining New Jobs Program; Helen H. Schaffler and E. Richard Brown of the Health Insurance Policy Program; Linda Fowells, Paul Vandeventer, Wendy Lazarus, and George Gundry of the Computers In Our Future program; and Ed Yelin of the Future of Work and Health Program.

References

- Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation*, 33, 5–18.
- Birckmayer, J. D., & Weiss, C. H. (2000). Theory-based evaluation in practice. What do we learn? *Evaluation Review*, 24(4), 407–431.
- Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Chen, H. T. (1997). Applying mixed methods under the framework of theory-driven evaluations. *New Directions for Evaluation*, 74, 61–72.
- Chen, H. T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7, 283–302.
- Cook, T. D. (1985). Post-positivist critical multiplism. In R. L. Shotland, & M. M. Mark (Eds.), *Social science and social policy* (pp. 21–61). Beverly Hills, CA: Sage.
- Crano, W. D. (2003). Theory-driven evaluation and construct validity. In S. I. Donaldson, & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 145–157). Mahwah, NJ: Erlbaum.
- Donaldson, S. I. (1995). Worksite health promotion: A theory-driven, empirically based perspective. In L. R. Murphy, J. J. Hurrell, S. L. Sauter, & G. P. Keita (Eds.), *Job stress interventions* (pp. 73–90). Washington, DC: American Psychological Association.
- Donaldson, S. I. (2001a). Mediator and moderator analysis in program development. In S. Sussman (Ed.), *Handbook of program development for health behavior research and practice* (pp. 470–496). Newbury Park, CA: Sage.
- Donaldson, S. I. (2001b). Overcoming our negative reputation: Evaluation becomes known as a helping profession. *American Journal of Evaluation*, 22(3), 355–361.
- Donaldson, S. I. (2003a). Theory-driven program evaluation in the new millennium. In S. I. Donaldson, & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 109–141). Mahwah, NJ: Erlbaum.
- Donaldson, S. I. (2003b). Using professional evaluation to improve the effectiveness of non-profit organizations. In R. Riggio (Ed.), *Improving leadership in non-profit organizations*. San Francisco, CA: Jossey-Bass.
- Donaldson, S. I. & Gooler, L. E. (2002). *Summary of the evaluation of The California Wellness Foundation's Work and Health Initiative*. Institute for Organizational and Program Evaluation Research, Claremont Graduate University (www.cgu.edu/sbos).
- Donaldson, S. I., & Gooler, L. E. (2002b). Theory-driven evaluation of the Work and Health Initiative: A focus on Winning New Jobs. *American Journal of Evaluation*, 23(3), 341–346.
- Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation. *American Journal of Evaluation*, 23(3), 261–273.
- Donaldson, S. I., Gooler, L. E., & Weiss, R. (1988). Promoting health and well-being through work: Science and practice. In X. B. Arriaga, & S. Oskamp (Eds.), *Addressing community problems: Psychological research and intervention* (pp. 160–194). Newbury Park: Sage.
- Donaldson, S. I., & Scriven, M. (2003). *Evaluating social programs and problems: Visions for the new millennium*. Mahwah, NJ: Erlbaum.
- Donaldson, S. I., Street, G., Sussman, S., & Tobler, N. (2001). Using meta-analyses to improve the design of interventions. In S. Sussman (Ed.), *Handbook of program development for health behavior research and practice* (pp. 449–466). Newbury Park, CA: Sage.
- Donaldson, S. I., & Weiss, R. (1988). Health, well-being, and organizational effectiveness in the virtual workplace. In M. Igbaria, & M. Tan (Eds.), *The virtual workplace* (pp. 24–44). Harrisburg, PA: Idea group Publishing.
- Fitzpatrick, J. (2002). Dialog with Stewart Donaldson. *American Journal of Evaluation*, 23(3), 347–365.
- Lipsey, M. W. (1988). Practice and malpractice in evaluation research. *Evaluation Practice*, 9, 5–24.
- Lipsey, M. W. (1990). *Design sensitivity*. Newbury Park, CA: Sage.
- Lipsey, M. W. (1993). Theory as method: Small theories of treatments. In L. B. Sechrest, & A.G. Scott (Eds.), *Understanding causes and generalizing about them. New Directions for Program Evaluation*, 57, (pp. 5–38). San Francisco, CA: Jossey-Bass.
- Mark, M. M. (2003). Toward a integrative view of the theory and practice of program and policy evaluation. In S. I. Donaldson, & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 183–204). Mahwah, NJ: Erlbaum.
- Mersman, J. L., & Donaldson, S. I. (2002). Factors affecting the convergence of self-peer ratings on contextual and task performance. *Human Performance*, 13, 299–322.
- Price, R. H., Friedland, D. S., Choi, J. N., & Caplan, R. D. (1998). Job-loss and work transitions in a time of global economic change. In X. B. Arriaga, & S. Oskamp (Eds.), *Addressing community problems:*

- Psychological research and intervention* (pp. 195–222). Thousand Oaks, CA: Sage.
- Reichhardt, C., & Rallis, C.S., (Eds.), (1998). The qualitative quantitative debate: New Perspectives. *New Directions for Program Evaluation* 61. San Francisco, CA: Jossey-Bass.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.
- Scriven, M. (1998). Minimalist theory: The least practice requires. *American Journal of Evaluation*, 19, 57–70.
- Scriven, M. (2003). Evaluation in the new millennium: The transdisciplinary vision. In S. I. Donaldson, & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 19–41). Mahwah, NJ: Erlbaum.
- Shadish, W. R. (1993). Critical multiplism: A research strategy and its attendant tactics. *New Directions for Program Evaluation*, 60, 13–57.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review*, 21, 501–524.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wholey, J. S. (2003). Improving performance and accountability: Responding to emerging management challenges. In S. I. Donaldson, & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 43–61). Mahwah, NJ: Erlbaum.
- Vinokur, A. D., van Ryn, M., Gramlich, E., & Price, R. H. (1991). Long-term follow-up and benefit-cost analysis of the JOBS program: A preventive intervention for the unemployed. *Journal of Applied Psychology*, 76, 213–219.